USING A VALUES- ENGAGED EDUCATIVE EVALUATION APPROACH TO ADDRESS ISSUES OF DIVERSITY AND EQUITY IN A MULTI-SITE SCIENCE TECHNOLOGY ENGINEERING AND MATHEMATICS PROGRAM EVALUATION

BY

AYESHA SHERITA TILLMAN

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Educational Psychology in the Graduate College of the University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Jennifer Greene, Chair
Professor Lizanne DeStefano
Professor Frances Lawrenz, University of Minnesota
Professor Thomas Schwandt
ABSTRACT

As challenges mount in the areas of national defense, climate change, health, energy, economic growth, food safety and accessibility, and environmental protection, so does the demand for skilled scientists, engineers, and health professionals. For decades, the National Science Foundation (NSF) has spearheaded the federal government’s efforts to enhance the quality of Science Technology Engineering and Mathematics (STEM) education and to increase the diversity of STEM communities via a variety of Broadening Participation (BP) initiatives. There is a need for evaluations that better understand BP program characteristics and participant experiences that will lead to success for underrepresented minorities in these STEM programs. Empirical evidence is needed to support the claim that attention to diversity and equity will yield benefits for STEM and society, and provide exemplars of successful BP programs; examples of what such attention looks like in practice is largely absent. Such evaluations would be conducted with some predetermined criteria for judging program quality and with explicit valuing. Evaluators have been discussing values in evaluation for over 40 years (Campbell, 1982; Mabry; 2010; Scriven; 1980). From judging the value of a program based on set criteria, to developing an understanding of stakeholder values stances, explicitly or implicitly all evaluation work involves valuing (Shadish, Cook, & Leviton, 1991; House & Howe, 1999). However, there has yet to be consensus within the field on whose, and what criteria should be used to judge program worth. This dissertation is a case study that examined the potential of the Values-Engaged, Educative (VEE) evaluation approach (Greene, DeStefano, Burgon, & Hall, 2006) to engage the NSF BP agenda through the prescription of the values diversity and equity in a longitudinal multi-site STEM education program evaluation. The VEE evaluation approach seeks to educate stakeholders about their program and emphasizes the inclusion of the interests, concerns, perspectives, and values of those traditionally unheard or least well served in the evaluation
context. With its foundation in STEM educational program evaluation, this approach defines quality at the intersection of the program’s scientific content, pedagogy, and equity. Several findings emerged. First, the VEE approach’s potential to engage with the NSF BP agenda is directly connected to stakeholders’ own values and commitments to BP. Second, the VEE approach impacted program design, especially in relation to participant selection and recruitment. Finally, engaging with stakeholder values, institutional culture, and being educative were aspects of the VEE approach responsible for evaluation success. Implications for the literature and future research endeavors are also discussed.
ACKNOWLEDGMENTS

I must begin by acknowledging God who has strengthened me; through Him all things are possible. Second, to my family: Glenn, Dad, Mom, Quintin, Antonio, grandparents, in-laws, and extended family, thank you for your financial, emotional, and spiritual support. Thank you to my I-STEM family (past and present): Maria Jimenez, Holly Downs, Gabi Garcia, Emily Gates, Lorna Rivera, Michael Culbertson, Susan Kelly, Luisa Rosu, Betsy Innes, Carie Arteaga, Kim Hobby, Linda Morris, Sallie Greenberg, and Saraí Coba-Rodriguez. I am grateful to my friends who have kept me grounded and encouraged: Kelli Center, Raquel Panton, Gerald Blankson, Nora Gannon-Slater, Silvester Mata, Amber Hammons, Tina Davis, Ian Martinez, Kim Brookens, Jackye Peretz, Kim Rice, Joe Trotter, Nikki Hiller, Brandon Meline, Cari Rich, Kate Clancy, and Meredith Olson. Thank you to the Twin City Derby Girls, especially the travel team, for your motivation, support, and love; on and off-skates. Thank you to the NSF Emergent Behaviors for Integrated Cellular System (EBICS) Science and Technology Center for participating in this study. Finally, I would like to sincerely thank my amazing and insightful dissertation committee. Tom, Lizanne, Frances, and Jennifer, you each uniquely contributed to my dissertation completion and I am eternally grateful.
# TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION .....................................................................................................1

CHAPTER 2 LITERATURE ........................................................................................................11

CHAPTER 3 PROGRAM AND EVALUATION CONTEXT: WHO, WHAT, WHEN, AND WHERE .................................................................................................................................29

CHAPTER 4 METHODOLOGY: HOW AND WHY ..................................................................43

CHAPTER 5 FINDINGS: THE CONSCIENCE OF THE NSF BP AGENDA .........................59

CHAPTER 6 DISCUSSION: RESPECTFUL PATIENCE .........................................................103

REFERENCES ............................................................................................................................114

APPENDIX A EXPANDED LITERATURE REVIEWS ...........................................................122

APPENDIX B EBICS REU PROGRAM EVALUATION INSTRUMENTS ............................136

APPENDIX C CASE STUDY INSTRUMENTS .....................................................................159

APPENDIX D EBICS STC EVALUATION DOCUMENTS ....................................................161
CHAPTER 1
INTRODUCTION

The State of STEM Education

Our world increasingly relies on science and technology to solve many of today’s most difficult problems. As challenges mount in the areas of national defense, climate change, health, energy, economic growth, food safety and accessibility, and environmental protection, so does the demand for highly skilled scientists, engineers, and health professionals (National Academies Press, 2005). As a nation, we are currently unable to supply enough competent, well trained scientists for these pressing demands. A critical juncture in the science profession pipeline lies at the undergraduate degree level. Of the four million ninth graders in 2001, only 167,000, or about 4%, were projected to graduate college with a Science, Technology, Engineering, and Mathematics (STEM) degree in 2011 (National Center for Education Statistics, 2008). Despite the previous statistic, the number of students earning science and engineering bachelor’s degrees has risen over the past 15 years (National Science Foundation, 2013), but jobs for competent scientists remain unfilled (National Academies Press, 2010).

Furthermore, science and engineering fields do not have a diversity of people working in them. This limits diversity of thought and, thus, creativity and ingenuity, and thereby threatens America’s scientific leadership in the world. For example, women have earned 58 percent of all STEM bachelor’s degrees awarded since 2002 (National Science Foundation, 2013). By field in 2007, women earned half or more of bachelor's degrees in psychology (77%), biological sciences (60%), social sciences (54%), agricultural sciences (50%), and chemistry (50%). However, large gender discrepancies still exist in bachelor's degrees awarded in the “harder” sciences of engineering, computer sciences, and physics, with men earning about 80 percent of those
degrees. Additionally, women’s share of bachelor’s degrees in computer sciences, mathematics, and engineering has declined in recent years (National Science Foundation, 2013).

Discrepancies in science-related bachelors’ degrees awarded by race and ethnicity are even more pronounced than those by gender. The most recent data available indicate that the current proportion of science\(^1\) and engineering baccalaureate degrees awarded to Asians/Pacific Islander students is 9%; to African American students is 8%; to Latino students is 8%; and to American Indian/Alaska Native students is 0.7% (National Science Foundation, 2013). Scholars, policymakers, and laypersons all agree that women and minorities are an undertapped resource to science innovation and advancement in the United States (American Society of Higher Education, 2011). As such, the National Science Foundation (NSF) has sought to increase understanding of barriers that impede the success of students from groups underrepresented in the science professions—especially engineering, computer science, physics, and mathematics—and is committed to cultivating the academic success of such students in these science and engineering fields (Committee on Equal Opportunities in Science and Engineering, 2004; NSF, 2000, 2002, 2003).

**NSF Commitment to Broadening Participation in STEM: Research Experience for Undergraduates Program**

It is a grand challenge to recruit, train, and encourage undergraduate students from traditionally underrepresented groups\(^2\) (women, minorities [Blacks, Latinos, and American

---

\(^{1}\) Including chemistry, physics, computer science, psychology and biology.

\(^{2}\)“Women, persons with disabilities, and three racial/ethnic groups—Blacks, Hispanics, and American Indians—are considered underrepresented in science and engineering because they constitute smaller percentages of science and engineering degree recipients and of employed scientists and engineers than they do of the population. Asians are not considered underrepresented because they are a larger percentage of science and engineering degree recipients and of employed scientists and engineers than they are of the population” (NSF, 2013).
Indians], and persons with disabilities) to pursue advanced degrees and careers in STEM fields. From improving quality of life by enhanced individual economic opportunity (Baum & Payea, 2005; Kelly, 2005) and increased advances in healthcare technologies, to increasing the vitality and creativity of STEM projects (Burke & Mattis, 2007) and continuing to be a player in the global economy (National Science Board, 2008), these efforts continue to gain traction. The benefits of such endeavors are potentially limitless no matter whose motivations are at stake. For decades, the National Science Foundation (NSF) has spearheaded the federal government’s efforts to enhance the quality of STEM education and to increase the diversity of STEM communities via a variety of “broadening participation” (BP) initiatives. In 2012 NSF spent $910.90 million dollars on broadening participant efforts (Committee on Equal Opportunities in Science and Engineering, 2012). Just over 16% of these funds were spent on the Academic Science and Engineering Careers (ADVANCE), Historically Black Colleges and Universities Undergraduate Program (HBCU-UP), and Louis Stokes Alliances for Minority Participation (LSAMP) programs. Additionally, $57.28 million dollars were spent on the Research Experiences for Undergraduates (REU) program.

The NSF REU program espouses BP commitments to diversity and equity, in theory and in practice. The REU program strives to ensure that underrepresented minorities in STEM have the maximum opportunity to achieve their potential. The REU program is typically an eight- to ten-week summer research opportunity for eight to twenty undergraduate students per participating campus (NSF, 2012). Undergraduate students participating in the REU program are assigned to a research laboratory and faculty advisor on their placement campus or site. They are then mentored by graduate and postdoctoral students as they develop and implement their own research project. During their program, REU participants typically conduct research; attend
professional development workshops (related to Applying to Graduate School, Graduate Records Examination Preparation, How to Make a Professional Poster Workshop, and others); receive ethics training; and prepare written, visual, and oral presentations of their final research projects. Although the demographics of REU students vary from site to site, NSF encourages sites to recruit students who are women, African-American, Latino/a, have a disability, or come from an institution with limited research capacity. While the majority of REU programs are single site, many large NSF funded centers offer multi-site REU programming.

As STEM educational program efforts increase and grow in size, both generally and in service to students underrepresented in the STEM fields, the need for evaluation of these programs also expands. Further, the need for high-quality evaluation within multi-site contexts is ever present (Herrell & Straw, 2002; King & Lawrenz, 2011; Turpin & Sinacore, 1991).

**STEM Education Program Evaluation**

Evaluators have been sought to investigate the impact of NSF programs that aspire to increase enrollment of persons from underrepresented groups in STEM fields. Unfortunately, evaluators qualified to conduct such evaluations and instruments that yield valid assessments of important outcomes of these STEM programs are rare (Katzenmeyer & Lawrenz, 2006). In particular, the need for evaluations that attend meaningfully and respectfully to issues of culture, race, diversity, and equity is not currently met (Mertens & Hopson, 2006). An examination of REU literature and publicly available REU evaluation reports revealed that many of these site-level evaluations have not been very informative or useful, particularly regarding the program’s success at broadening participation (Tillman, 2013). The evaluations display the following shortcomings:

- Mainly tally or count the number of students from underrepresented groups
• Do not assess participant experiences
• Do not attend directly to issues of equity and diversity
• Are often conducted by program staff rather than professional evaluators
• Do not examine which specific program components are successful at broadening participation.

Given these shortcomings, these evaluations are not yielding the kind of evaluative evidence that is necessary to address the lack of minority participation in STEM education programs. Empirical evidence is needed to support the claim that purposeful attention to diversity and equity in STEM programs like the REU will yield benefits for STEM and society and will further provide exemplars of successful BP programs. Examples of what such attention looks like in practice are largely absent. Therefore, evaluations and studies of evaluations of NSF BP initiatives, such as the REU program, are needed that attend to diversity and equity, better understand the program characteristics and participant experiences that will lead to success for underrepresented minorities (URMs) within the STEM field, and critically examine who gets to participate in these programs. Additionally, evaluations can also (1) help program staff better understand their program and how best to improve it, (2) enhance the potential of the program to influence NSF policies, and (3) consider multi-site challenges that multi-site programs can pose.

Evaluations are conducted using predetermined criteria for judging program quality and with explicit valuing. “Values are at the etymological heart of evaluation” (Mabry, 2010 pg. 83), because evaluators are responsible for the determination of merit or value of a program (Scriven, 1991). However, there has yet to be consensus within the field on whose and what criteria should be used to judge program worth. For over 40 years, evaluators have participated in scholarly quarrels about the difference between values and fact (Campbell, 1982; Scriven, 1980), the fact-value continuum (House & Howe, 1999), types of valuing in evaluation (Shadish, Cook, &
Leviton, 1991) and most fundamentally, the role of values in evaluation. Currently, much of the conversation in evaluation pertaining to values is focused on deciding on which stakeholder values are to be included in evaluation and if evaluators should be advocating certain values or not (Hall, Ahn, & Greene, 2011).

The Values Challenge in Evaluation

A brief preview of the extended discussion of values in evaluation follows.

Descriptive and Prescriptive Valuing

Shadish, Cook, and Leviton (1991) define valuing in evaluation in terms of descriptive and prescriptive valuing. Descriptive valuing involves describing the values inherent in any program without advocating one as best or one as better than another. Prescriptive valuing is advocating for certain values, voices, or concerns to be brought to the forefront of evaluation. The Values-Engaged, Educative (VEE) evaluation approach (Greene, Boyce, & Ahn, 2011; Greene et al., 2006; Hall, Greene & Ahn, 2012; Johnson, Hall, Greene, & Ahn, 2013) has recently emerged as a values-explicit approach that espouses both descriptive and prescriptive valuing.

Values-Engaged, Educative Evaluation

Drawing primarily from responsive (Stake 1973, 2004) and democratic (House and Howe, 1999) traditions in evaluation, the Values-Engaged, Educative (VEE) evaluation approach seeks to educate stakeholders about their program while also engaging the perspectives, concerns, and values of all legitimate stakeholders in the evaluation. The approach

---

3 An in-depth description of the literature on values in evaluation will be included in Chapter 2.
4 An in-depth description of the VEE approach will be included in Chapter 2.
emphasizes the inclusion of the interests, perspectives, and values of those traditionally unheard, underrepresented, or least well served in the evaluation context. Additionally, this approach encourages explicit attention to issues of diversity and equity. The VEE approach is well positioned to address the needs of STEM program evaluation. Extensive field tests in STEM education program contexts, commitments aligned with the National Science Foundation (NSF), the desire to use appropriate methodologies, and attention to the social-relational dimensions of evaluation make VEE a good fit for evaluations of STEM educational programs, including engagement with NSF’s BP agenda. The principles and practices of Values-Engaged, Educative evaluation offer one response to the aforementioned state of affairs. However, much of the field testing and implementation of VEE has taken place in single site evaluation contexts, while the proportion and number of large-scale multi-site national STEM initiatives, like nano- or eco-science research centers, continue to increase (Lawrenz & Huffman, 2003).

**Statement of the Research Problem**

Global competitiveness and other pressing national demands underscore the need to increase student interest in and accessibility to STEM degrees and careers. This need is especially salient for women, minorities, and persons with disabilities, as they are the least well represented in STEM fields. Various local, state, and national efforts are being made to support and encourage individuals from underrepresented groups to pursue and stay in science-related degrees and professions. Concomitantly, there is a need in STEM program evaluations to attend to issues of culture, equity, and diversity (the BP agenda) in order to assure that all students are getting the proper support, education, and resources to succeed in such fields. Unfortunately, many evaluations of these efforts to broaden participation in STEM are conducted by program staff rather than evaluation professionals and do not document success at broadening
participation, examine program access, or attend directly to equity and diversity. Furthermore, because these programs are increasingly delivered in a large, multi-site format, thoughtful multi-site evaluation endeavors are also needed. While evaluation has a conflictional history regarding the explicit role of values in evaluation, the principles and practices of the Values-Engaged, Educative evaluation approach offer one response to this state of affairs. Therefore, empirical support is needed for the rationale that evaluation ought to prescribe and attend to the values of diversity and equity in complex STEM educational program contexts.

**Purpose of the Study**

This case study investigated the implementation and results of a multi-year VEE evaluation of a multi-site REU program. First, the case study assessed the ability of the VEE approach to engage the BP agenda, offer information of relevance, and have an impact on program design and implementation. Second, the study explored which strands of the VEE approach account for successes and limitations. Third, the case study examined how the multi-site context presented particular challenges to the VEE approach and how those challenges were addressed.

In order to address these research objectives, the research questions in this study were:

1. How well did the VEE approach to evaluation:
   a. Meaningfully engage the BP agenda of REU programs;
   b. Offer information of value and relevance, especially regarding the program’s connections to URMs (underrepresented minorities), to both local program staff and NSF; and
   c. Offer information of consequence or have an actual impact on program design (notably, stronger BP potential) and program outcomes (effective BP programs)?
2. What components or strands of VEE most importantly accounted for its ‘successes’ in this REU evaluation context?

3. What specific challenges, if any, did the multi-site nature of the REU program raise for the implementation of the VEE approach and how well did the VEE approach respond to these challenges?

Significance of the Study

This study fills multiple gaps in both evaluation and STEM literatures. First, this study contributes to the literature on the role of values in evaluation and provides empirical evidence for the importance of prescriptive approaches to valuing in evaluation. Second, this study also informs the STEM education evaluation community about a viable approach to evaluation that is centered on NSF goals of broadening participation and the accompanying values of diversity and equity. Third, this study contributes to the expanding literature on multi-site evaluation. Specifically, this study investigates three main issues for consideration in multi-site evaluation: evaluation logistics, methodology, and engagement of values. Fourth, this study contributes to the values-engaged, educative evaluation literature. With the formal conception of this approach less than ten years old, this study importantly contributes to the implementation and practice of this approach. Finally this study also contributes to STEM literature by increasing our understanding of the multi-site iteration of NSF’s REU program.

The next chapter provides a review of literature on values in evaluation, descriptive and prescriptive evaluations approaches, and the Values-Engaged, Educative approach. Chapter three provides an overview of the case-study context and briefly introduces multi-site evaluation literature. Chapter four outlines the study methodology, epistemological framework, and data
collection, analysis and quality of the study. Chapter five presents study findings, and chapter six discusses implications for literature, future research, and final thoughts.
CHAPTER 2
LITERATURE

This dissertation examines and documents the advancement of the values of diversity and equity, through the Values-Engaged, Educative evaluation approach. This chapter situates this study within the evaluation literature. The roadmap for this chapter is as follows. The chapter presents (a) an historical overview of the role of values in evaluation; (b) a brief summary of the current conversation on values in evaluation; (c) a description of evaluation approaches that explicitly engage with values; (d) an account of the Values-Engaged, Educative approach; and (e) a snapshot of multi-site evaluation literature.  

5 A brief literature review of STEM and the U.S. workforce and the STEM pipeline is included in Appendix A.1.

Values in Evaluation: Historical Debates

It seems reasonable to begin by defining what I mean by values and valuing. Borrowing from Frankena (1967), values are defined as the worth of a thing, and valuing is defined as an estimate of worth. Evaluation itself is a “values-laden enterprise” (Hall, Ahn, & Greene, 2012, p. 196), with evaluators judging the merit, worth, or value of a program based on a selected set of criteria (Scriven, 1991). However, the history of values in evaluation has played out in a dramatic fashion with key players, opposing views, and hotly contested disputes (House & Howe, 1999). I focus my selection of literature in this section on four major points of contention: (a) objectivity and value neutrality, (b) facts claims versus value claims, (c) prescriptive and descriptive valuing, and (d) a framework of values and ethical aims.
Objectivity and Value Neutrality

In the past, evaluation scholars have argued that objectivity is the regulative ideal (Smith, 1990), and that evaluation claims based on evaluator values have no place in evaluation (Campbell, 1982). It was asserted that objectivity could be obtained by methodological rigor. Much of this argument stems from the epistemology of positivism and post-positivism (Ayer, 1936; Guba & Lincoln, 1989) which argue that both the physical and social world operate via universal laws, thereby excluding a place for values in the domains of science and truth. Positivists argued that observation and all data other collection should be completely value neutral. Further, only data derived from such methods could lead to valid knowledge.

Fortunately, scholars have spent decades pushing back on the idea of value-neutrality in evaluation. Scriven (1980, 1991, 2003) was one of the first to argue that evaluation is about more than gathering facts about things. He argued that, distinctively, evaluation results lead to judgments about the merit, worth, or significance of the evaluand. Valuing is what distinguishes evaluation from forms of other systematic inquiry and is necessary to inform real-world decisions about which programs are good and worthy of the resources required to run them (Julnes, 2012). Scriven’s and others stances on the centrality of values in evaluation have prompted much debate over the years.

Claims: Facts and Values

While Campbell rejected the tenets of positivism, in favor of post-positivism, and believed there was no foundational structure to knowledge, he (1982) asserted that there is a fact-value dichotomy. In this dichotomy, factual claims are made of brute facts, and value claims are base values. He asserted that evaluators must determine the worth of a program based on program goals (and their underlying values), already set by funders, policy-makers, and other
key stakeholders. Further, according to Campbell, these goals and values were beyond critique, because evaluators are unable to argue rationally about values. However, this view constrained evaluators and the acceptable scope of evaluation (House & Howe, 1999). Evaluators who follow Campbell’s framework could potentially miss out on some of the most important aspects of a program because they are unable to critique program goals, values, policies, or politics. Additionally, most evaluators already examine and make available for critique the goals and values of a program in their evaluation practice.

House and Howe’s (1999) presentation of evaluation claims seems more appropriate and flexible than Campbell’s. They argue that all evaluation claims are and should lie on a continuum. At one end are brute facts without any value aspects or individual preference. A brute fact example would be that “Diamonds are harder than steel.” On the other end of the continuum are statements that are purely related to taste or bare values. For example, “Cabernet is better than Chardonnay.” They state, however, that evaluation claims are actually a blend of both values and facts and, therefore, lie somewhere in the middle of this continuum. Examples of evaluation claims that are neither brute facts nor bare values are “This test is valid” and “X is a good program.” These statements blend facts and values together. See Figure 1 below.

<table>
<thead>
<tr>
<th>Brute Facts</th>
<th>Evaluation Claims</th>
<th>Bare Values</th>
</tr>
</thead>
</table>

*Figure 1. Fact-Value Continuum (House & Howe, 1999)*

House and Howe suggest whether a claim counts as a brute fact, bare value, or a blend partially depends on context. For example the statement “John Doe is dead” appears to be a brute fact. However, what if Mr. Doe is hooked up to life support and is in an irreversible coma? The judgment or claim that he is dead is based on the judgment of value as what constitutes quality of
life. This is an example of a brute fact that became value-laden due to medical technology. House and Howe argue that the majority of evaluation claims have values built into them, and this is acceptable as long as valid evidence is presented to support such claims.

Initial conversations of values also included debate about whose values should be advanced in the evaluation.

**Prescriptive and Descriptive Valuing**

Shadish, Cook, and Leviton (1991) argued that when trying to examine stakeholder voice, evaluators will be unable to arrive at value conclusions that will be accepted by all stakeholders. Shadish et al. present valuing in evaluation as taking two forms: descriptive valuing and prescriptive valuing. Descriptive valuing involves describing the values inherent in any program without advocating one as best or one as better than another. Prescriptive valuing is advocating for certain values, voices, or concerns to be brought to the forefront of evaluation. Shadish et al. argue that evaluators should prescriptively value sparingly and only when there is much agreement among stakeholders, which they say it quite rare. Mabry (2010) also argues that prescriptive valuing can give evaluators unwarranted authority.

However, it is impossible to truly separate description from prescription (House & Howe, 1999). While some claims are values free, like “2 + 2 = 4,” for most, context determines the evaluative message of text. House and Howe state that what seems descriptive in one context can become prescriptive or evaluative in another. They give the example: “George Washington was the first president of the United States.” However, if you are discussing this statement with historians, they might point out the racist and patriarchal origins of the country, in which case the statement also becomes evaluative. Howe and Howe further their argument by stating that
prescriptive content is unavailable in many descriptions. They give the multiple examples of two statements:

- “Jones killed Smith with malice,” and “Jones’s actions led to Smith’s death.”
- “The project director stole $50,000 in project funds,” and “The project director deposited $50,000 in project funds into his personal account.”
- “Gay and lesbian youth are oppressed in public schools” and “Gay and lesbian youth experience disproportionate difficulties in public schools.”

The first of each pair is more evaluatively laden, but neither is necessarily inaccurate. Instead, they describe different states of affairs. The first of each pair of statements “better targets problems and remedies than does the second” (House and Howe, 1999, p. 43).

The drawbacks to prescriptive valuing are counterbalanced by ethical and moral commitments to better society, as evaluations more often play a role in political decision making. Just over 15 years ago, evaluators began to agree that evaluands are “social, political, and moral constructions that embody the different (and often conflicting) interests and values of stakeholders (Schwandt, 1997, p. 26).” Gradually evaluators have realized that all evaluations and scientific claims in general are infused with theoretical and value predispositions, especially of the inquirer. Greene (1997, p. 28) notes:

> It is time to unmask them [value stances]. It is time for evaluators to explicitly state the value commitments, programmatic assumptions, and political stances that underlie their chosen methodology. It is time for evaluators to claim and proclaim their advocacy. To do otherwise is to be disingenuous, even deceptive to our audience.

In sum, there was the general view, or rather illusion, that evaluators should be ‘objective’ and not advance any particular values. Decades ago, Scriven, House and Howe, and others pushed back, arguing that values are the centerpiece for evaluators’ work which still must be objective. There has been the gradual realization within the field that all evaluations advance
some values, even ‘value neutral’ evaluations by default advance the values of the main client or funder. Evaluation scholars continue the conversation about values today. Much of the conversation focuses on who is doing the valuing and the need for a valuing framework. Both are discussed below.

**Values in Evaluation: Contemporary Conversations**

Within recent years, there has been a resurgence of literature related to values in evaluation. Challenges and benefits of including a value theory in evaluation were critiqued with a call for evaluators to ensure that societal values are not forsaken in evaluations. Challenges included loss of the profession’s credibility, and benefits included exercising ethical obligations as a citizen. Additionally, the role of valuing in federal programing (Yates, 2012), suggestions for how to explore stakeholder values (Orr, 2012), an examination of critical social theory and values in evaluation (Mabry, 2010), and participant observation and values (Freeman & Hall, 2012) have also emerged as part of the contemporary conversation.

Recently, a typology of evaluator valuing roles has been constructed (Alkin, Vo & Christie, 2012), likening evaluation to teaching. Value judgments have been sorted into three categories: valuing by stakeholders alone, valuing by stakeholders and evaluators, and valuing by evaluators only. In approaches where valuing is done by stakeholders, evaluators provide data but leave valuing to pre-identified stakeholders. In approaches where valuing is done by stakeholders and evaluators, evaluators provide data and help establish a valuing framework, guide others in the valuing process, and participate with others in valuing. In approaches where valuing is done by evaluators, valuing is based on evaluator values, evaluation expertise, program expertise, and scientific appraisal. Alkin et al. list evaluation theorists whose approaches exemplify each classification. Table 1 below presents their taxonomy.
Table 1

**Typology of Evaluator Valuing Roles (Alkin, Vo, & Christie, 2012)**

<table>
<thead>
<tr>
<th>Valuing by</th>
<th>Description</th>
<th>Exemplars</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stakeholders</strong></td>
<td>Stakeholders establish standards a priori</td>
<td>Tyler/Popham</td>
</tr>
<tr>
<td></td>
<td>Evaluator provides data to identified stakeholders</td>
<td>Stake</td>
</tr>
<tr>
<td></td>
<td>Evaluator provides data for use by broad audiences</td>
<td>Cronbach/Weiss</td>
</tr>
<tr>
<td><strong>Stakeholders and Evaluators</strong></td>
<td>Evaluator guides stakeholders by establishing a framework for valuing</td>
<td>Alkin</td>
</tr>
<tr>
<td></td>
<td>Evaluator guides stakeholders by guiding the valuing process iteratively</td>
<td>Fetterman/Patton</td>
</tr>
<tr>
<td></td>
<td>Evaluator participates with stakeholders in valuing</td>
<td>Cousins/King</td>
</tr>
<tr>
<td><strong>Evaluator</strong></td>
<td>Evaluator valuing strongly based on the promotion of evaluator values (e.g., morals)</td>
<td>House</td>
</tr>
<tr>
<td></td>
<td>Evaluator valuing based upon evaluation expertise</td>
<td>Scriven</td>
</tr>
<tr>
<td></td>
<td>Evaluator valuing based primarily on program expertise</td>
<td>Eisner</td>
</tr>
<tr>
<td></td>
<td>Evaluator valuing based upon scientific appraisal</td>
<td>Boruch/Cook</td>
</tr>
</tbody>
</table>

As a field, we agree that evaluators all advance some values, however, there is little consensus on methods of valuing (Julnes, 2012). This contemporary issue is especially problematic because:

1. Evaluation is becoming more central to public-sector decision making
2. The increasing pressure for evidence-based governance is pushing for more evidence-based, and hence systematic, policies on the methods of valuing appropriate for evaluation, often privileging specific approaches to assessing performance and economic impacts (Julnes, 2012, p. 4)

Julnes further argues that the evaluation community needs to articulate our own working consensus of valuation and suggests that warranted valuation is quite possible when yielding
specific judgments about policy or program improvements. Evaluators need to have a voice in governmental policymaking regarding what constitutes evidence and impact. He urges evaluators to develop frameworks for valuation that are more explicit about strengths and limitations of different approaches to valuation in different contexts. This study does just that.

With the recognition that evaluation does advance values, then, has come the development of quite a few approaches to evaluation whose authors have picked what they believe to be the most defensible values that evaluation can advance. Next, I will provide an overview of three evaluation approaches that explicitly engage with values. I will discuss where they fall in the Alkin et al. typology, which values they engage, and with what justification.

**Explicit Valuing Evaluation Approaches**

Responsive evaluation, democratic evaluation and culturally responsive evaluation approaches are presented below. While they are among the most well-known of the values-explicit approaches, others could have been selected, e.g., critical approaches to evaluation.

**Responsive Evaluation**

Responsive evaluation is an approach that was developed to attend to issues of import to program stakeholders and seeks to include multiple stakeholders in dialogue to enhance the understanding of a program from the viewpoint of insiders (Stake, 1975, 2003). In Alkin et al.’s typology, this approach falls within the ‘*valuing being done by stakeholders*’ category because the values being advanced are those of the stakeholders. This approach is not responsive to program theory or stated goals, but to stakeholder concerns (Stake, 2003). In a responsive evaluation, evaluators’ roles include interpreter, educator, facilitator, and Socratic guide (Abma, 2006).
Value pluralism, one of the main tenets of responsive evaluation, is that there is no single true value to anything (Stake, 1975); as such, there can be many defensible interpretations of the same events, activities, or programs. Additionally, there is a “reluctance to push for consensus” (Stake, 2004) among stakeholders to ensure value pluralism. The previous directly challenges the notion associated with positivism that there is a single truth and that truth can be sought only with objectivity and depersonalization. This responsive approach calls for evaluators to be in-depth investigators and recognizes that knowledge is context bound and situational (House, 2002). Therefore, according to responsive evaluation, evaluators need to collect diverse viewpoints from multiple stakeholders, and these viewpoints should be represented in all aspects of the evaluation.

Vicarious experience is a second major tenet of responsive evaluation. As such, case studies are well suited for responsive evaluations because readers can experience the program vicariously through the details of the case. Evaluators are to provide readers with the “experience of personally being there, feeling the activity, the tension, knowing the people, and their values” (Stake, 2003, p. 86). The purpose of this second tenet is for the evaluator to attempt to portray the complexities and contextualities of the program. While this is done best through direct personal experience, vicarious experience is the best substitute. While their view of responsive evaluation is more emancipatory, Guba and Lincoln (1989) contend that responsive evaluation allows stakeholders to be co-owners of the evaluation, and that evaluation is a conversational process between those who have stake in the program.

The previous approach advances and describes the values of multiple stakeholders and not necessarily values of the evaluator. With the acknowledgement that evaluators’ perceptions can be subjective; this approach embraces multiple ways of knowing, personal observation, and attempts to be explicit about the values the evaluation will advocate. Authors proclaim that while
stakeholders’ values and experience are essential to the evaluation and to the development of
criteria for judging quality, responsive evaluation is not the same as participatory evaluation
(Cousins & Whitmore, 1998; King, 1998, 2007). Authors hold that evaluators are trained and
know which methods best suit the evaluation. Justification for whose values are to be advanced
lie in the belief that evaluators must observe and inquire before attempting to understand
program merit.

**Deliberative Democratic Evaluation**

There are three requirements for deliberative democratic evaluation: inclusion, dialogue,
and deliberation (House & Howe, 1999). In Alkin et al.’s typology, this approach falls within the
‘valuing being done by evaluators’ category because the primary values being advanced are
those thought to be important by the evaluator, namely democratic values of equity and justice.
However, here I should note that House and Howe have picked values that they believe
defensive to prescribe, but it is the stakeholders in their evaluation who do the valuing, as such
stakeholder participation is directly related to their democratic commitments. As such, I believe
this approach is more suited for the ‘valuing being done by stakeholders and evaluators’
category.

The first principle in democratic deliberative evaluation is inclusion of all relevant
stakeholder values, interests, and perspectives. The authors argue that one of the biggest threats
to evaluation is power imbalances and that evaluators must design evaluations where a balance
of power among relevant interests is represented. The second principle is dialogue between
evaluators and stakeholders and among stakeholders themselves. This approach asserts that
extensive dialogue with stakeholders increases the opportunity for a greater understanding of
diverse stakeholders’ true interest. Authors caution evaluators not to make assumptions about
interests of stakeholder parties. They must seek out interests and concerns through dialogue. The third principle is deliberation to reach conclusions. House and Howe purport that this deliberation is a cognitive process that must be grounded in evidence and valid arguments. Values cannot be taken as a given, but must be examined through deliberation.

House and Howe (1999) caution that evaluators should recognize the difference between power and authority and should only accept authority. For example, “A has power over B when A can affect B’s behavior contrary to B’s interests. But A has authority over B when B complies because A has influenced B through good reasons attached with B’s own interests” (House and Howe, 1999, p.102). The deliberative democratic evaluator’s role is to determine fact-value claims objectively through inclusion, dialogue, and deliberation. This approach promotes the values of social justice and equity. The justification for such promotion lies in the authors’ belief that because all evaluations will inherently promote some values, voices of those often marginalized are the most defensible to ensure to include because of societal inequalities (House & Howe, 1999).

**Culturally Responsive Evaluation**

Culturally responsive evaluation (CRE) is mostly anchored in and extends the major tenets of responsive evaluation (Stake, 1975, 2003). However, this approach falls within the ‘valuing being done by stakeholders and evaluators’ category because it necessitates that evaluators attend to culture and diversity, in addition to being responsive to stakeholder concerns. Culturally responsive evaluation strategies are intentionally sensitive and responsive to the cultural context of the program and its participants (Frierson, Hood, & Hughes, 2002). Additionally, CRE seeks to honor the cultural content of the program or setting (Ryan, Chandler, & Samuels, 2007). Culturally responsive evaluations attempt to have diverse evaluation team
members, including those who have shared-lived experiences with participants in order to provide sound advice during the evaluation process and in order to better interpret, understand, and disseminate evaluation findings.

The first component of CRE is context. The context of an evaluation includes the “totality of environment in which the project takes place” (Manswell-Butty, Reid, & LaPoint, 2004, p. 38). The second component of CRE is culture. Although many definitions exist, culture is generally thought of as representing shared norms, values, and assumptions of a group (Samuels & Ryan, 2011; SenGupta, Hopson, & Thompson-Robinson, 2004). Culture can refer to shared language, ethnicity, race, social class, religion, gender, sexual orientation, disability, age, and geographic location. The final component of CRE is responsive evaluation (Stake, 1975). In addition to seeking input about stakeholder concerns and values, the CRE evaluators advance values based on the cultural and contextual aspects of the program. Justification for CRE valuing lies in the authors’ belief that attending to the influence of culture, particularly when evaluating projects serving diverse populations, is critical for strengthening the validity and utility of evaluation findings (Frierson, Hood, Hughes, & Thomas, 2010).

In summary, values and valuing are intrinsically and fundamentally embedded in evaluation. All evaluation advances values, and evaluators are called to be thoughtful about whose and what values are implicitly and explicitly advocated. Further, because values will always be promoted, we should be clear about which, whose, and why. Evaluation can be positioned to in service of the public and societal good. As such, it is our responsibility as evaluators to emphasize good practice with mindful attentiveness to theoretical roots. The Values-Engaged, Educative (VEE) approach draws from each of the three approaches listed above. The VEE approach, similar to CRE, attempts to both describe and prescribe values, as
such, valuing is done by stakeholders and evaluators. Below is an overview of the VEE approach.

**Values-Engaged, Educative Evaluation**

Drawing primarily from responsive (Stake 1973, 2003) and democratic (House and Howe, 1999) approaches in evaluation, the Values-Engaged, Educative (VEE) evaluation approach seeks to educate stakeholders about their program while also engaging the perspectives, concerns, and values of all legitimate stakeholders in the evaluation. The approach emphasizes the inclusion of the interests, perspectives, and values of those traditionally unheard, underrepresented, or least well served in the evaluation context. Additionally, this approach encourages explicit attention to issues of diversity and equity.

**Dimensions of Values Engagement**

The purposeful attention to values is the first dimension of values engagement in this approach (Greene, Boyce, & Ahn, 2011). This approach calls for evaluators to engage with the values inherent in the educational program being evaluated (for example, critical thinking skills). Additionally, evaluators are called to attend to and discuss the values of multiple stakeholders and clients. This approach urges evaluators to be inclusive of multiple value stances by promoting stakeholder dialogue and the incorporation of multiple voices in evaluation. The second dimension of values engagement is targeted attention to the values of (a) diversity, defined as both traditional socio-demographic markers such as class, gender, and race, as well as

---

6 Greene, Boyce, & Ahn (2011, pg.4) state: “stakeholders or audiences, comprise four main groups: (a) policy and decision makers, program funders, program developers, and researchers in that field; (b) program administrators, staff, and volunteers; (c) program participants, and their families and communities; and (d) interested advocacy groups, the media, and the general public.”

23
other ways people are different from one another and (b) equity, defined as parity in program access, participation, and accomplishment for all program participants, especially those least well served in the context.

**Being Educative**

Directly influenced by Lee Cronbach’s powerful advocacy for an educative role for evaluation (Cronbach & Associates, 1980) the Values-Engaged, Educative evaluator is further expected to facilitate stakeholder learning about the program itself (Greene et al., 2006). Underlying logic, contextual appropriateness, potential power to effect change, connections to relevant standards and research evidence, and overall quality are all aspects of the program on which the evaluator assists in shedding light. In the VEE approach, evaluators encourage primary stakeholders to critically engage with and dialogue about the program’s theory or underlying logic of activities, outcomes, and their interconnections.

It is often the case that there is more than one program theory underlying a given educational intervention. Different stakeholders may have differing views on the goals and logic of the program. VEE evaluators are encouraged to 1) capture diverse stakeholder understandings of the program theory, 2) bring the voices of those less heard to the forefront, and 3) promote dialogue and reflection. All of the previous are examples of efforts to facilitate learning and enhance the educative aspects of this approach.

**Evaluator Role**

In this approach, the evaluator role calls for him or her to draw attention to the particular histories, contexts, and needs of all types of learners who are underserved by the programs being evaluated. Specifically, the evaluator is called to examine the extent to which an education program is equitable, or has opportunities for participation, meaningful learning, and
accomplishment, particularly for those least well served in that context. Additionally the
evaluator is encouraged to (Greene, Boyce, & Ahn, 2011, p. 14):

- Meaningfully and productively conceptualize, assess, and understand context, especially in relation to the meanings of diversity and equity and their interplay with program quality and effectiveness.

- Use program theory and other strategies to portray and meaningfully engage with varied stakeholder program understandings, values, and interests.

- Respectfully emphasize engagement with values of equity, raising difficult, value-laden questions and working through conflicting viewpoints and, tensions, while being broadly inclusive of and responsive to multiple perspectives and interests.

- Promote and sustain critical reflection and respectful dialogue in order to enhance program understanding and sustain improvement efforts.

- Learn about and become acquainted with multiple and diverse characteristics and rhythms of the program and its context in order to develop an accurate and thoughtful understanding of what is being evaluated.

- Be explicit about the value commitments of the approach upfront and throughout the process, and make the evaluation a visible, open and transparent activity.

- Promote, and also engage in, ongoing communication and critical reflection on practice.

The VEE approach authors suggest that those who endeavor to utilize this approach will be experienced evaluators who have authority, credibility and presence in the evaluation context (Greene, Boyce, & Ahn, 2011). This means that the ideal VEE evaluator will have access to program materials, have adequate opportunities to meet with stakeholders throughout the evaluation, have expertise in the field (or access to it), and have evaluation team members of similar cultural background or lived experiences as program participants.

**Program Quality**

This approach defines educational program quality at the intersection of the program’s content, pedagogy, and diversity. See Figure 2, which follows.
In values-engaged, educative evaluation, the evaluator attends to the contextual quality of the program’s content and pedagogy, and further defines educational quality at the intersection of high-level and current content, appropriate pedagogy, and equity. A good program, that is, not only generates meaningful learning about current content via active and responsive teaching, but also engages, encourages, and respects the perspectives, life experiences, practical wisdom, and understandings that under-represented students bring with them to the teaching-learning context. (Greene, Boyce, & Ahn, 2011, pg. 48)

As such, when evaluating programs, VEE evaluators should be asking questions about content, pedagogy, and diversity as part of or in addition to traditional questions about program design, participant experience, and program outcomes.

Greene (1997), author of the VEE approach, purports that evaluation is always conducted in service to something and that evaluation should be positioned in service to the public good. For example, the VEE approach has attempted to “improve Science, Technology, Engineering, and Mathematics (STEM) education as a means of societal change” (Greene et al., 2006). STEM
education program evaluation is in need of evaluation approaches that attend to and advocate for diversity, equity, and culture (Mertens & Hopson, 2006). The VEE approach is well positioned to address the needs of STEM program evaluation. Extensive field tests in STEM education program contexts, commitments aligned with the National Science Foundation (NSF), the desire to use appropriate methodologies, and attention to the social-relational dimensions of evaluation make VEE a good fit for evaluations of STEM educational programs. Why are STEM education programs and STEM education program evaluations important?

To date, field testing of the VEE evaluation has been in single-site STEM education settings. While these tests have increased our understanding of the VEE approach’s theory to practice, this study takes the logical next step of introducing the VEE approach to a complex, multi-site evaluation context.

**Multi-Site Evaluation**

A mere twenty years ago, literature on program evaluations implemented across multiple locations was scare. Often funded by government agencies or large foundations, multi-site evaluations (MSE) are increasingly showing up in the STEM educational evaluation landscape. Turpin and Sinacore (1991) introduced the field to the term, multisite evaluation (MSE), and provided the first guide for those planning and conducting evaluations across multiple sites in their *New Directions for Evaluation* (NDE) volume. Since then, two additional volumes (Herrell & Straw, 2002; King & Lawrenz, 2011) and a multitude of papers with an MSE focus have been published. The proliferation of knowledge on MSEs suggests the evaluation community’s commitment to better understanding the benefits, challenges, and distinctive issues posed in such contexts.
Such initiatives typically involve multiple higher education institutions and have both scientific and educational objectives. These initiatives usually have a large research component, with secondary goals that address broader impacts where the focus is on STEM training for undergraduates, graduates, and post-doctoral students, each with an emphasis on underrepresented minorities. Further, many large STEM educational efforts are delivered in a multi-site format. For example, the NSF funds multi-site programs such as Science Technology Centers (STC), Chemical Centers for Innovation (CCI), Local Systemic Change, Collaboratives for Excellence in Teacher Preparation Program, the Centers for Learning and Teaching, and Math and Science Partnerships (MSP). The evaluation of these multi-site programs can present distinct challenges and issues for consideration.

Overall, the need for high-quality evaluation within the multi-site context is ever increasing. There are two main factors that differentiate MSEs from other types of evaluation activities: 1) the program evaluation must include multiple sites, two or more, and 2) similar evaluation activities must be conducted across those sites to answer a common set of evaluation questions (Sinacore & Turpin, 1991; Springer, 2000). MSEs present distinctive issues for evaluators to consider carefully (Herrell & Straw, 2002; King & Lawrenz, 2011; Turpin & Sinacore, 1991). I have grouped these items for consideration into three categories: 1) engagement of values, 2) methodology, and 3) logistics (Tillman, 2013). Multi-site evaluation is not the main focus of this study. However, a more complete literature review of MSE can be found in appendix A.2.

The next chapter overviews the evaluation context in which this study took place.
CHAPTER 3

PROGRAM AND EVALUATION CONTEXT: WHO, WHAT, WHEN, AND WHERE

This chapter describes the evaluation context, which includes a description of the program and the evaluation. The program is described in terms of its history, goals, activities, and Center member demographics. In addition, the evaluation description includes the original evaluation background, purpose, approach, key evaluation questions, audience, and criteria for judging program quality, design, data collection methods, and reporting strategies. Finally, the chapter ends with the rationale for choosing this as my dissertation context.

Program Context

NSF STC Program

The National Science Foundation (NSF) Science and Technology Centers (STCs) integrative partnerships program supports innovative, potentially transformative, complex research and education projects that require large-scale, long-term awards. STCs conduct research through partnerships among academic institutions, national laboratories, industrial organizations, and/or other public/private entities, and via international collaborations, as appropriate. They provide a means to undertake significant investigations at the interfaces of disciplines and/or fresh approaches within disciplines. For example, the BEACON STC was “founded with the mission of illuminating and harnessing the power of evolution in action to advance science and technology and benefit society” (BEACON STC, 2014) and spans Michigan State University; North Carolina A&T University; and the Universities of Idaho, Washington, and Texas at Austin.

The National Science Foundation established the Science and Technology Centers Program in 1987. The objective was to mount an innovative, interdisciplinary campaign in

29
important areas of basic research. NSF currently has 17 active STCs (NSF, 2012). NSF has an expectation for STCs that are broader than research alone. Therefore, all funded STCs are required to engage in significant education, diversity, and outreach efforts. As such, many STCs opt to implement a Research Experience for Undergraduates program.

NSF REU Program

As described in chapter one, The National Science Foundation (NSF) Research Experience for Undergraduates (REU) program is an eight- to ten-week summer\(^7\) research opportunity for eight to twenty undergraduate students per participating campus (NSF, 2012). REU sites may be based in a single discipline or academic department or may offer interdisciplinary or multi-department research opportunities with a coherent intellectual theme (NSF, 2012). Large NSF initiatives, including STCs, can apply for additional funding to host an REU program at each of their STC’s participating campuses. Currently, REU sites can be found in the following disciplines: Astronomical Sciences; Atmospheric and Geospace Sciences; Biological Sciences; Chemistry; Computer and Information Science and Engineering; Cyberinfrastructure; Department of Defense; Earth Sciences; Education and Human Resources; Engineering; Ethics and Values Studies; International Science and Engineering; Materials Research; Mathematical Sciences; Ocean Sciences; Physics; Polar Programs; and Social, Behavioral, and Economic Sciences. REU site funds are sought by university STEM departments and other academic entities or national laboratories, such as The National Weather Service.

\(^7\) REUs also offer research opportunities during the academic year as well.
EBICS STC

Funded by NSF in Fall 2010 for over 50 million dollars, Emergent Behaviors of Integrated Cellular Systems (EBICS) is a National Science Foundation (NSF) Science and Technology Center (STC). This STC is an integrative partnership program which spans three universities: University of Illinois at Urbana-Champaign (Illinois), Massachusetts Institute of Technology (MIT), and Georgia Institute of Technology (GT). Illinois, MIT, and GT are the primary partners in the STC. Other participating institutions are Morehouse College; the City College of New York; University of California, Merced; Brigham and Women’s Hospital; Columbia University; Emory University; and University of Georgia.

The mission of EBICS is to create a new scientific discipline for building living, multicellular machines that solve real-world problems in health, security, and the environment. The STC has goals to develop the science and technology needed to engineer clusters of living cells, also called biological machines, which have desired functionalities and can perform prescribed tasks. These machines will consist of sensing, information processing, actuation, protein expression, and transport elements that can be effectively combined to create functional units. It is envisioned that these biological machines will perform tasks such as processing systems that detect toxins in the environment and neutralize them; smart plants that sense and respond to the need for water and nutrients; surrogate organs that are used in place of animals to test new drugs; and biological factories that sequester CO₂ in a continuous flow process.

EBICS Center Member Demographics

Among the 32 total faculty in the Center are 12 women and four underrepresented minorities, two African Americans and two Hispanics. The composition of Center trainees is approximately 44% female and 56% male. Trainees’ educational levels span undergraduates
EBICS trainees’ race and ethnic composition includes Asian (44%); Black, not of Hispanic origin (13%); Hispanic or Latino/a (5%); and White, not of Hispanic origin (34%). Four trainees did not report their race or ethnicity. One Center member self-reported a disability. Ethnic diversity within the Center has remained the same over the course of the past three years. The percentage of women affiliates in EBICS has risen within the past three years. Table 2 below depicts total Center composition by year; Figure 3 presents the race/ethnicity of Center participants.

Table 2

<table>
<thead>
<tr>
<th>Year</th>
<th>N</th>
<th>Gender</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>2011</td>
<td>127</td>
<td>77</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(60.6%)</td>
<td>(37.8%)</td>
</tr>
<tr>
<td>2012</td>
<td>138</td>
<td>84</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(60.9%)</td>
<td>(39.1%)</td>
</tr>
<tr>
<td>2013</td>
<td>128</td>
<td>72</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(56.3%)</td>
<td>(43.8%)</td>
</tr>
</tbody>
</table>
EBICS Goal Areas

EBICS has six goal areas: (a) Education, (b) Ethics, (c) Knowledge Transfer, (d) Human Resources and Diversity, (e) Leadership and Management, and (f) Research. Each goal area mission statement is below.

*Education Mission Statement*

Develop innovative education, outreach, and training programs across the Center that will prepare students to become future researchers and education leaders in the new discipline.

*Ethics Mission Statement*

Implement a cross-disciplinary and multi-institutional ethics program that will inform and guide all members of the Center on ethical and responsible conduct of scientific research. Engage in active dialogue with experts and the public regarding ethically contentious scientific research.

---

8 Goal area descriptions and mission statements are from the EBICS annual reports.
**Knowledge Transfer Mission Statement**

Develop effective mechanisms and pathways to facilitate intellectual exchanges between institutions and partners from different sectors (e.g., industry, K–12, general public, policy makers) that will support the sharing of knowledge, information, and application of new technology.

**Human Resources and Diversity Mission Statement**

The Diversity mission of EBICS is to produce the next generation of researchers, a diverse group of leaders who are equally proficient in biology and engineering, and a supporting community to nucleate and develop the nascent field of engineered biological systems.

**Leadership and Management Mission Statement**

Envision and enable the Center’s mission through inclusive and transparent decision making, as well as effective and responsible implementation; inspire Center participants; and facilitate collaborative efforts within and beyond the Center.

**Integrative Research Mission Statement**

Produce transformative, synergistic research through an inclusive collaborative culture that crosses disciplinary and institutional boundaries and is embedded throughout the Center’s activities.

A faculty member from MIT, GT, or Illinois is assigned as a director to each goal area. Goal area directors work with Center program managers to host meetings about each area and to plan events surrounding one or more area. For example, the Ethics director is responsible for Center wide ethics training activities, including developing ethics modules and hosting ethics lectures. The context of the study took place across the Education and Human Resources and Diversity goal areas.
**Education and Human Resources and Diversity Goal Areas**

The Education and Human Resources and Diversity goal areas are responsible for Center graduate training programming, including the Graduate Teaching Consortium courses, the Student Leadership Council, and professional development activities for graduate and post-doctoral students. These goal areas are also responsible for undergraduate training and high school outreach programming, including the Research Experience for Undergraduates (REU) and high school research experience programs.

**EBICS REU Program**

The EBICS STC hosted an REU program during the summers of 2011, 2012, and 2013, and 2014. Data for this study come from the first three years of the REU program. The goal of the REU program is to provide a 10-week, authentic research experience for undergraduate students and expose them to advanced degree options and careers in the STEM fields, with the ultimate goal of increasing the number of students, especially those from underrepresented groups, that go on to graduate school in EBICS-related fields.

The REU program provided students with a $4,500 stipend in addition to an allowance for travel expenses, on-campus housing, and meals. Some of the REU activities included graduate school preparation, faculty mentoring, professional development, and social engagement with other students on each of the campuses. All participants were required to conduct research and write a scientific paper on their projects. Participants also presented their findings through oral and poster presentations at the end of their REU programs.

**Evaluation Context**

The section below overviews the evaluation context background and evaluation plan.
Background

A longitudinal (2011–2014) evaluation of the EBICS STC and fulfillment of its goals is required by NSF. Dr. Lizanne DeStefano, evaluation team leader, worked with the STC principal investigators (PIs) to develop the Center evaluation plan that was submitted with the proposal. Evaluation is embedded and visible within the EBICS STC. Since the EBICS STC’s inception, the evaluation team has worked to develop a relationship with EBICS leadership, program staff, graduate student trainees, and faculty. Evaluation results are presented at NSF site visits, and annual retreats, during committee meetings, and on the ebics.net website.

Attainment of Center goals has been the focus of the evaluation. Next, I will present evaluation information related specifically to the EBICS REU program.

Purpose and Audience

Evaluation is an integral component of this National Science Foundation (NSF) Research Experience for Undergraduates (REU) program and is conducted for both formative and summative purposes. The REU program evaluation purpose is to provide valid and useful information about STC program implementation, effectiveness, impacts, and institutionalization to NSF, REU program managers, REU program participants, federal policy makers, and ultimately the public. Specific priorities include (1) attending to the quality of the experience for REU participants, faculty advisors and graduate student mentors to guide program improvement; (2) assessing short- and long-term effectiveness and impact; and (3) increasing knowledge on factors that enhance and support undergraduate students to pursue advanced degrees and careers in STEM.

Values and Aspirations

This evaluation was grounded in a commitment to promote an enhanced understanding of the NSF REU program being evaluated, and to engage with issues of diversity and equity in
STEM education. Framed by these value commitments, the major purpose of this evaluation was to assess the quality and importance of the NSF REU Program with regard to its ability to support and enhance STEM education, laboratory experience, mentorship, professional development, and presentation experience for a diversity (gender, disability, ethnicity, college major, research experience opportunities) of student participants. Specifically, the evaluation focused on how well the REU experience supported student interest in and attitudes toward STEM, motivation to further pursue advanced STEM learning, self-concept as STEM students, graduation rates and enrollment in STEM graduate programs, and/or employment in STEM careers. In addition, the evaluation aspired to be responsive to interests and concerns of the REU program managers and the Center diversity director, and yield the kind of information likely to be meaningful and useful.

**Evaluation Approach**

The EBICS REU evaluation utilized a Values-Engaged, Educative approach. This approach has been discussed at length in previous chapters.

**Key Evaluation Questions**

The following are evaluation questions that were used to guide the evaluation. Evaluators used the VEE guidebook (Greene, Boyce, Ahn, 2011) to direct question construction. Questions related to quality of program design, program content and pedagogy, program implementation and effectiveness, outcomes, and institutionalization were used. Questions were developed by the evaluation team and approved by the Center diversity director and program managers.

1. **Implementation**: Is the EBICS STC REU program being implemented on schedule and as planned?
What is the quality of the program implementation and effectiveness at each of the sites, especially as it relates to the needs of the group of students recruited for the REU programs?

2. **Effectiveness**: Is the EBICS STC REU program operating effectively? How might they be improved?
   - What is the quality of the program design, content and pedagogy, as designed for various learners in the context?

3. **Impact**: What outcomes are associated with participation in the EBICS STC REU program? What is the value-added of participation in the EBICS REU Program?
   - What is the quality and magnitude of program outcomes and accomplishments, both intended and unintended, for the REU Participants?

4. **Institutionalization**: How and to what extent are elements of the EBICS STC REU becoming institutionalized at MIT, Illinois, and GT? What opportunities and barriers exist?

**Criteria for Judging Quality**

Again, following the VEE guidebook, evaluators developed criteria for judging program quality. VEE suggests that STEM education programs be judged within three domains: (1) the quality of the program design; (2) the contextual relevance and power of the project design; (3) the advancement of the interests of under-represented and under-served groups. The quality of program implementation and outcomes were also included within the criteria.

**Domain I: Quality of Program Design**

- The professional development curriculum and pedagogy offer relevant, valuable and diverse approaches to learning that are meaningful, appropriate, and adequate for all participants.
• The laboratory experiences include training, mentoring and modeling. Further, the research aspect leads the REU participants from a relatively dependent status to as independent a status as their competence warrants.

**Domain II: Contextual Power of the Program Design**

• Participants in the REU program have meaningful, positive, and consequential learning experiences and demonstrate strong and consistent mastery of valued and relevant laboratory, research, and presentation skills and knowledge. Participants also show increased interest, motivation and self-efficacy for research in STEM areas.

• Program managers, faculty advisors, and graduate student mentors hold high expectations for all participant education, demonstrating care and support for participants and affirming the value of diverse experiences, resources, and the creative minds they bring to the program.

**Domain III: Advancement of the Interests of Underrepresented and Underserved Groups**

• The REU program provides its participants with equitable opportunities for meaningful and high-quality program experiences and accomplishments on par with their peers.

• The program has clear goals and strong rationales that are well aligned with and support the mission of the broader STEM education community and NSF to increase diversity and equity in STEM fields.

**Quality of Program Implementation and Outcomes**

• The REU program was implemented on time and schedule at each site.

• REU participants demonstrate meaningful benefits from their experience with the program initially, including increased technical and professional skills, and ultimately including increased interest in STEM and the pursuit of STEM degrees and careers.
Evaluation Design and Data Collection Methods

The evaluation employed a mixed methods design. In this evaluation, mixed methods were used to (a) generate a comprehensive, in-depth understanding of the program of interest; (b) produce a fair representation of the complexity of the program; and (c) tap into various dimensions of the same phenomena (Greene, 2007). Following VEE approach guidelines, data collection methods were chosen based on knowledge and data needed to answer evaluation questions. Evaluation data collection methods included: observations, program manager interviews, two participant surveys, participant focus groups, a faculty advisor and graduate mentor survey, document analysis, and follow up/longitudinal tracking.

Observations

Observations were conducted throughout the REU programs on all campuses. The purpose of these observations was to generate data to describe activities and the engagement between faculty/trainees with each other and with the activities, to informally interview REU participants, and to collect artifacts distributed to participants. Moreover, evaluators noted information related to the following categories during the observations:

- Physical setting—a rich description of the time and locale of the event, including where participants were situated.
- Social or interpersonal setting—who was clustered with whom and how groups and individuals were arrayed in this context.
- Activities—a systematic description of the activities with timeframes.
- Content—a description of the resources and materials used and discussed.
- Interactions—a description of verbal and nonverbal interactions between the various stakeholders involved with the retreat.
REU Participant Surveys

REU participants were asked to complete two surveys online via the ebics.net website. Participants completed one survey at the beginning of their program (the first or second week). The second survey was completed a week after their REU program ended. Specifically, the surveys were used to obtain information related to participants’ overall experience with the REU program, including items related to campus location, resources given, lab and research assignment and format, lodging accommodations, professional development programs, knowledge learned, and success of the REU program. Additionally, REU participants were tracked longitudinally tracked and followed up each Spring semester. The surveys included both closed- and open-ended items.

REU Participant Focus Groups

Evaluators conducted four semi-structured focus groups. Multi-site focus groups were conducted at Illinois with MIT and GT REU students participating via video conference. Site-specific focus groups were also conducted at each site. Protocols were developed for and followed during the focus groups to prompt interviewees in order to ensure that the relevant questions were covered.

Faculty Advisor and Graduate Student Mentor Survey

After participants completed their REU programs, their faculty advisors and graduate student mentors were asked to complete a brief online survey. The survey was used to obtain information about advisors’ and graduate students’ perceptions about the REU program. Specifically, they were asked about their satisfaction with the REU students’ placement in the labs, their understanding of program goals, and their thoughts about perceived benefits of the program for the participants.
Program Manager Interviews

After the completion of the REU program, all program managers were interviewed via telephone. They were asked about their expectations for the REU program, what they thought went well, and what they thought could be improved for next year.

The next chapter discusses research design and methods.
CHAPTER 4

METHODOLOGY: HOW AND WHY

This chapter outlines the methodology used for this study. I begin by revisiting the study research objectives. Then I describe the following components: research design, case selection rationale, ontological and epistemological framework, data collection methods, data collection timeline, data analysis, and data quality criteria. Finally, I end with a discussion on study limitations.

Research Objectives

The primary objective of this study was to examine and critique the implementation of a values-explicit evaluation approach in a multi-year, multi-site STEM education context, with a particular focus on how well the approach can meaningfully and consequentially engage issues of diversity and equity in STEM education. This study examined the implementation of the Values-Engaged, Educative (VEE) approach from 2011 to 2014 in the evaluation of a summer Research Experience for Undergraduates (REU) program housed within an NSF Science and Technology Center (STC). I sought to contribute to a more comprehensive understanding of the challenges, benefits, and implications of implementing the VEE approach in this kind of context.

In order to achieve the research objectives previously mentioned, the main research questions in this study were:

1. How well did the VEE approach to evaluation:
   a. Meaningfully engage the BP agenda of REU programs;
   b. Offer information of value and relevance, especially regarding the program’s connections to URMs (underrepresented minorities), to both local program staff and NSF; and
   c. Offer information of consequence or have an actual impact on program design (notably, stronger BP potential) and program outcomes (effective BP programs?)
2. What components or strands of VEE most importantly accounted for its ‘successes’ in this REU evaluation context?

3. What specific challenges, if any, did the multi-site nature of the REU program raise for the implementation of the VEE approach and how well did the VEE approach respond to these challenges?

**Research Design**

A case study (Abma & Stake, 2001; Stake, 1995) research design was employed. Case studies are often used to examine the “particularity and complexity of a single case” (Stake, 1995, pg. xi). Case studies are often used to answer questions that begin with ‘why’ and ‘how’ (Yin, 1994). While case study designs are not particularly well-suited for making generalizations, they often provide insights into and richly describe a particular issue. Additionally, case study research is used for particularization and development of an understanding of the uniqueness of a particular case, which ultimately increases knowledge about an overall phenomenon or issue of interest.

Specifically, this was a multi-site, multi-year, instrumental case study, with the case being the evaluation of the EBICS STC REU program. Instrumental case studies are employed when the case is used to understand a larger issue than the case itself (Stake, 1995). This research design was appropriate as it allowed me to collect data to describe the case, which ultimately allowed for the facilitation of knowledge to better understand the larger issue of values engagement in complex evaluation contexts. While this research is framed as a case study it also includes strands of summative meta-evaluation in that this study informs broad questions about the strengths and weaknesses, utility, and feasibility of a particular approach to evaluation.

**Case Selection Rationale**

The evaluation of the EBICS STC REU program was selected as the case for this study for a number of reasons. First, I had direct access to this context through my graduate research
assistantship with the Illinois Science Technology Engineering Mathematics Education Initiative (I-STEM). I have evaluated the EBICS STC with Dr. Lizanne DeStefano, I-STEM director and Educational Psychology faculty, since May of 2011. We have used a VEE evaluation approach for this Center evaluation. My involvement with this evaluation allowed me to conduct this case study for my dissertation. Not only do I have access to the context, but I have access to meetings, discussions, and documents that outsiders would not have. Second, Dr. DeStefano and I have developed a positive working relationship with the EBICS STC leadership, graduate student trainees, Center faculty, and staff. These relationships allowed me to gain permission to use this STC as the context for my dissertation. Additionally, these relationships resulted in various Center members being willing to be interviewed for my dissertation. Third, as described in Chapter 3, the EBICS STC is a complex ecology comprised of multiple sites across the nation; many stakeholders, including university faculty, program staff, and graduate and undergraduate students; and six program goal areas. An NSF STC initiative is one of the most complex evaluation contexts around. Finally, NSF funds a large number of multi-site STEM initiatives. As such, using the evaluation of this program for the case is appropriate because it is a real-world, complex evaluation context.

**Ontological and Epistemological Framework**

The research questions I have chosen are best answered through multiple qualitative data collection methods, which will be discussed in the next section. Qualitative research designs often call for a researcher to make observations; exercise judgment—which may be subjective; and analyze data, all while realizing her own consciousness and paradigm stance (Guba & Lincoln, 1994; Stake, 1995). As such, I will touch briefly on the epistemological and ontological framework that I used to conduct this study.
The framework guiding this study is Constructivism, which asserts that knowledge is socially constructed, realities are apprehendable in multiple forms, and there is no absolute truth (Guba, 1990; Guba & Lincoln, 1994). This framework’s ontological stance is relativism, meaning there is no “objective” truth, and human sense making is an act of construction from a conceptual system and is independent of any foundational reality (Guba & Lincoln, 2001). With an inquiry aim of understanding, this framework purports that values be explicated and explored as part of the research processes and product. The table below displays Constructivism’s position on practical issues (Guba & Lincoln, 1994, pg. 112).

Table 3

<table>
<thead>
<tr>
<th>Constructivism Paradigm Position on Selected Practical Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Issue</strong></td>
</tr>
<tr>
<td>Inquiry aim</td>
</tr>
<tr>
<td>Nature of knowledge</td>
</tr>
<tr>
<td>Knowledge accumulation</td>
</tr>
<tr>
<td>Goodness of quality criteria</td>
</tr>
<tr>
<td>Values</td>
</tr>
<tr>
<td>Ethics</td>
</tr>
<tr>
<td>Voice</td>
</tr>
<tr>
<td>Training</td>
</tr>
<tr>
<td>Hegemony</td>
</tr>
</tbody>
</table>

Data Collection Methods

Multiple qualitative data collection methods were used to answer the study research questions. These methods were: (a) weekly systematic reflection of the 2013 iteration of the
evaluation, (b) review of evaluation artifacts, (c) Center leadership and program manager interviews, and (d) peer review and assessment. All data collection activities were conducted with University of Illinois Institutional Review Board (#14379) approval. The table below explains how each data collection method was connected to the key and subsidiary research questions. A description of each data collection method is included below.

Table 4

Data Collection Method, Research Questions, and Method Collection

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured Reflection of the evaluation</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Review of evaluation artifacts</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Program manager/Center leader interviews</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Peer review and assessment</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

**Structured Reflection of the Evaluation**

The first data collection method was weekly structured reflections of the evaluation that I completed myself, as the researcher. I developed prompts for these reflections with the following purposes in mind: (a) to gather information throughout the evaluation process relevant to how the ________________

---

9 Systematic reflections were conducted throughout the 2013 iteration of the REU program. However, additional reflections about the 2011 and 2012 iterations of the REU program were also constructed.
evaluation was planned, designed, and implemented, with particular attention to logistics, methodology, and engagement of values and engagement of diversity and equity; (b) to capture information related to challenges and triumphs during the evaluation; and (c) to capture the experience of implementing a VEE approach from the evaluator’s perspective. It has been suggested that “a meaningful strategy for self-reflection would be a systematic review of each activity of the evaluation, the demands that resulted, and the effectiveness of the associated role responses” (Skolits, Ann Morrow, & Mehalic Burr, 2009, pg. 291). In order to make the reflections as systematic and useful as possible, I wrote them once a week throughout the duration of the 2013 evaluation. I answered prompts about the implementation and the distinctive challenges of implementing VEE in a MSE context for that particular week. During the summer of 2013, I also reflected on the 2011 and 2012 iterations of the REU program and filled out prompts for what I thought were pivotal moments related to the implementation of the VEE approach in 2011 and 2012. Appendix C.2 presents the reflection prompts.

**Review of Evaluation and REU Program Artifacts**

The second data collection method was a review of all evaluation and program artifacts from the years 2011, 2012, and 2013. These artifacts were reviewed to gather information related to the implementation of the VEE approach. The document review was completed according to Center for Disease Control and Prevention (CDC) evaluation and document review recommendations (Center for Disease Control and Prevention, 2009). Evaluation artifacts were split into two categories—evaluation materials and program materials.

**Evaluation materials**

Evaluation materials that were reviewed and analyzed include: (a) evaluation plans, (b) evaluation instruments, (c) evaluation presentations, (d) evaluation reports (interim and final), (e)
evaluation meeting agendas, (f) emails related to the implementation of the evaluation, and (g) observation notes. The previous were sorted into the following categories based on their relation to: multi-site issues, evidence for engagement of broadening participation, all formal and informal evaluation recommendations, evidence of challenges related to implementation, emails and notes highlighting relationships with staff, and evidence of evaluation success. The data in each category were catalogued by name in Microsoft Excel. For example, some of the data in evidence of BP was listed:

- Email to diversity director 9/13
  - An email to the diversity director that includes a conversation about attempting to recruit more qualified URM the REU program.

- Notes from evaluation meeting 04/12
  - During this meeting we discussed the criteria for REU participant selection. URM and diversity were talked about in this conversation.

- Email to program manager 12/12
  - An email to the program managers about changes to the 2013 REU program.

- Evaluation report 2013 executive summary 10/13
  - Executive summary that included highlights about the participation of URM in the REU program.

Program materials

Program documents, including calendars, agendas, planning agendas, and professional development handouts, were also collected to ensure that an accurate depiction of the program was developed. I completed a descriptive summary excel sheet of the program materials review.
The form included: type of document, types of information included in documents, summary of documents, conclusions pertaining to this study that can be drawn from the document, including:
(a) assess the program values, including diversity and equity and program goals and (b) the context of each site including site-specific history, staffing issues. For example, it was important to note that at Illinois and GT, program managers are half-time employees whose main responsibilities are to implement the EBICS REU program. However, at MIT, the program manager is full time and also has the title of education program manager. As such, she had other EBICS STC responsibilities in addition to the implementation of the REU program. The previous were sorted into the following categories based on their relation to: evidence for program engagement of broadening participation, evidence of program changes based on evaluation recommendations, and evidence of evaluation success/challenges.

**Program Manager and Center Leadership Interviews**

The third data collection method was interviews with key players in the EBICS REU and its evaluation, including the EBICS STC Diversity Director, the lead evaluator, one of the STC associate directors, and two of the program managers. The purposes of these interviews were to:
(a) see how and to what extent the evaluation offered information of value and relevance to interviewees; (b) examine the evaluation’s impact on program design and outcomes; and (c) examine how well the interviewees believed the evaluation approach used for the EBICS REU had engaged with NSF’s broadening participation goal. The full interview protocol is included in Appendix C.1.

**Peer review and assessment**

The final data collection method was a brief assessment by those trained in the VEE approach: Illinois graduate students Jeehae Ahn and Gabriela Garcia. These reviewers were provided with: (a) the evaluation plan and evaluation reports and (b) a summary of the document
review. Then each peer reviewer wrote up their thoughts and met independently with me. The purpose of these peer reviews was to have them help define “meaningful engagement” of BP and to collect their perceptions of strands of the VEE approach that accounted for successes and challenges in this context. Each peer reviewer provided a brief one page written summary of their assessment of the evaluations efforts to engage with BP and evaluation successes and challenges. Then during a one hour conversation each reviewer verbally expanded on her thoughts. I took notes during these meetings, they were not recorded.

**Data Collection Timeline**

The EBICS STC REU program evaluation itself began in fall of 2010. The majority of the data collection for this study took place between June 2013 and February 2014. The figure below displays the case study’s data collection timeline.

![Data Collection Timeline](image)

*Figure 4. Evaluation and Research Study Timeline*

**Data Analysis**

Whether quantitative, qualitative, or mixed, the purpose of data analysis is to:

(a) organize raw data; (b) assess patterns, connections, trends and differences; and (c) support
and validate conclusions (Greene, 2007). Bogdan and Biklen (2003, pg. 159–160) suggest three tips that qualitative researchers should consider before they begin the data analysis process after all data are collected. They are: don’t be afraid to speculate; begin venting about data; and jot down lots of ideas. I will explore each of these below.

**Don’t Be Afraid To Speculate**

Bogdan and Biklen state that speculation is productive for research. They encourage researchers to begin thinking about conclusions with the data that they currently have. Prior to the completion of data collection, I began to write down possible themes that I thought I might find and made a list of possible data sources for each theme.

**Begin Venting About Data**

Bogdan and Biklen encourage researchers to talk about ideas, write notes, and review observer’s comments while the ideas are fresh. ‘Venting’ some of these ideas ahead of time can cut down on data analysis time after all of the data are collected. I began “venting” by chatting informally with my peer reviewers prior to officially sending them the materials they would review.

**Jot Down Lots of Ideas**

They also suggest that while reviewing data during the collection phase, making notes is a good idea. They recommend researchers jot down crazy ideas, speculations, or notes in separate files. I jotted down initial thoughts during interviews, and while I was reviewing the evaluation and program documents.

**Data Analysis**

Prior to analysis, all individual data were reviewed. Interview data were checked for transcription accuracy. Structured reflections and document review artifacts were also checked for accuracy in dates, events, and referenced time frames, etc. This study used analysis
techniques that are consistent with case analysis research. Once all data were gathered, I attempted to make meaning by searching for patterns, constructing themes, and identifying frequencies within the data (Stake, 1995).

First, each data set was analyzed individually. The interviews were transcribed\textsuperscript{10} and then analyzed line by line for themes (Guest, MacQueen, & Namey, 2012). For example, when an interviewee talked about broadening participation, those lines were highlighted in blue. When interviewees talked about multi-site aspects of the program and evaluation, those lines were highlighted in yellow. Initially there were 13 themes found in the interview data: NSF broadening participation, developing relationships with evaluators, multi-site evaluation, evaluation purpose, evaluation informal conversations, evaluation recommendations, evaluation usefulness, value of evaluation, interviewee values, underrepresented minorities and evaluation, evaluation and REU program design, interviewee knowledge about the REU program, and evaluation activities. I then turned to structured reflection and document review data.

Similar to interviews, structured reflections and document review were coded into categorical aggregations (Bogdan & Biklen, 2003; Stake, 1995, 2010). I read through these data to look for certain words, phrases, events, and themes that stood out. This was a two-step process to: (a) search through data for regularities and patterns, as well as topics my data covered, and then (b) write down topics and phases to categorize these patterns/themes. Themes from these data included: NSF broadening participation, developing relationships with stakeholders, multi-site evaluation, evaluation purpose, evaluation recommendations, stakeholder values, challenges

\textsuperscript{10} After interviews were transcribed, per the IRB protocol, the audio files were deleted.
with VEE implementation, URM and selection criteria, evaluation and REU program design, evaluation recommendations, and evaluation activities.

I then went back through all themes and refined each of the main themes by combining themes or deleting themes that had very little data to support them. After meeting with peer reviewers I further refined themes and reordered data into a final set of cohesive themes. All of the final five themes include data from at least two of the four data sources. The final themes are presented in Chapter 5.

**Data Quality**

With any study, there is a need to think about the quality of data collected. This section will discuss the criteria and standards that were employed to gauge the level and nature of confidence (of data quality and of inferences made) within this study. As mentioned before, this study operated under the paradigm of Constructivism. Therefore, criteria for assessing quality related to this paradigm were used. In their checklist for constructivist evaluation, Guba & Lincoln (2001, pg. 6) state that the following criteria for constructivism—credibility, transferability, dependability, and confirmability—“evolved from an effort to produce criteria more or less parallel to those conventionally used, i.e., internal and external validity, reliability, and objectivity.”

**Credibility**

Credibility is generally comparable to internal validity, the approximate truth about inferences regarding cause-effect or causal relationships. Credibility is how well your inferences match various stakeholders’ experiences in the settings at hand. Guba and Lincoln (2001, pg. 6) suggest that credibility can be achieved by:

- prolonged engagement at the site
• persistent observation
• peer debriefing (a kind of external critic)
• negative case analysis (a process of reworking postulated hypotheses)
• progressive subjectivity (continuous checking of developing constructions against records of constructions that were expected prior to data collection)
• member checks
• continuous testing of hypotheses, data, preliminary categories, and interpretations

This study attempted to employ as many of the previous suggestions as possible. Specifically, persistent observation, peer debriefing, member checks, and preliminary data categories and interpretations were used. The peer review process in this study also increased credibility. As previously mentioned, peer reviewers met with me after having reviewed the documents I gave them, and then we discussed the themes unearthed from data analysis. Reviewers were instrumental in restructuring a few themes I was less sure of and in their assessment of the evaluation’s engagement with BP.

**Dependability**

Dependability is generally comparable to reliability. Guba and Lincoln (2001) suggest that this can be achieved by the use of an external auditor who examines all aspects of the social inquiry especially, the methodological decisions made and the reasons for them. Also, overlapping methods, such as focus groups and individual interviews, are suggested. Additionally, efforts should be made to enable readers to develop a thorough understanding of methods and their effectiveness (Shenton, 2004). Authors proclaim the following should be included in reports:

• the research design and its implementation
• the operational detail of data gathering
• reflective appraisal of the project

The previous three suggestions were all built into the data collection of this study. In this study, dependability was importantly established by the inclusion of a dissertation committee (internal reviewers) and external peer reviewers.

**Confirmability**

Confirmability is generally comparable to objectivity, the ability to judge fairly. Constructivism asserts that true objectivity can never be reached. Therefore, this quality criterion suggested that steps must be taken to reduce as much researcher bias as possible and to inform readers of study shortcomings, and researcher’s assumptions. Confirmability is the extent to which the inferences reached are grounded in the data, or can be traced back to the data. Shenton (2004) makes the following suggestions to increase confirmability:

• Triangulation to reduce effect of investigator bias
• Admission of researcher’s beliefs, assumptions, and role
• Recognition of shortcomings in study’s methods and their potential effects
• In-depth methodological description to allow integrity of research results to be scrutinized
• Use of diagrams to demonstrate “audit trail”

This dissertation includes an in-depth, methodological description and information about the researcher’s role in the EBICS REU evaluation. Multiple data collection methods were used to examine the same phenomena. Finally, the employment of a VEE approach clearly established the researcher’s assumptions and beliefs which were detailed in the epistemological framework section.
Research Design Limitations

This study has a few limitations for consideration. First, researcher bias is possible given my role on the evaluation team of the EBICS STC. In this role, under the direction of Dr. DeStefano, I was in charge of data collection, analysis, and reporting. It is possible that I was biased in my collection data, data analysis, and data interpretation. Specifically, using self-reported data could have been a problem as I may have over- or under-estimated, or misremembered evaluation events. Additionally, because I was an evaluator and I also am the person who interviewed program staff and Center leadership, it is possible that they could have answered questions in such a way that they thought would be pleasing to me. Also, is it possible that as the sole researcher I may have been biased in my appraisal of the value of the VEE approach. Rigorous research design, multiple data collection methods, high-quality criteria for data, and internal and external reviewers were all implemented to lower the effects of researcher bias.

Second, the timing of this study was a potential limitation. The EBICS STC evaluation began in 2011, and this study began in 2013. As such, interviewees had to reflect retrospectively to answer some of the interview questions. To ensure interviewees’ responses were as accurate as possible, prior to the interviews, I asked interviewees to review previous evaluation meeting notes, evaluation reports, and evaluation plans.

Third, the longitudinal aspect of the REU evaluation is just beginning to take shape. The EBICS STC evaluation team wrote a comprehensive evaluation report of the EBICS STC in January 2014 that included data from the REU evaluation. An additional summative Center report will be written in 2015. The evaluation team has followed up with REU participants every year; however, it is too early to make long-term outcome evaluation claims about program impact. As such, there are only a few study findings related to the impact the VEE approach has
on summative evaluations. This will be discussed further in Chapter six. The next chapter
discusses study findings.
CHAPTER 5
FINDINGS: THE CONSCIENCE OF THE NSF BP AGENDA

This chapter provides a detailed description of the findings of this study. Findings are presented for each of the five questions. The roadmap for findings are: (a) the research question is presented; (b) where needed, terms are defined and operationalized; (c) data used to answer the question are acknowledged; (d) an overarching finding for the question is offered; (e) themes within the overarching finding are presented; and (f) each of the themes is expanded upon with the presentation of relevant synthesized data. As I shift from theme to theme, I will remind the reader what research question is being answered and which theme is being expanded upon. Finally, I present findings in such a way to ensure interviewee and EBICS site confidentiality.

**Question 1A: How well did the VEE approach meaningfully engage the NSF Broadening Participation (BP) agenda of REU programs?**

**Definitions**

Meaningful engagement within this context is defined as follows:

*Meaningful:* full of significance or purposeful

*Engage:* to occupy the attention and/or efforts of; to involve; to bring into operation

*Meaningfully Engage NSF BP:* to include NSF’s agenda of broadening participation in the evaluation on a significant level. To bring issues related to the recruitment and retention of underrepresented minorities in STEM to the forefront of the evaluation. This could be done by incorporating aforementioned commitments into evaluation questions, data collection activities, reporting, and conversations with stakeholders.\(^{11}\)

---

\(^{11}\) Peer reviewers assisted with the development of this definition
Data Used

Data used to answer this question included: (a) document review items—evaluation and program materials, (b) evaluator structured reflections, (c) stakeholder interviews, and (d) peer review feedback.

Overarching Finding

The EBICS evaluation team used the VEE approach as a framework to meaningfully engage the NSF BP agenda. Their efforts were successful.

The following themes emerged during data analysis and are presented below in support of the overarching finding:

A. Context complexity initially impeded evaluator engagement with BP.
B. Perceptions of the evaluation’s engagement with BP vary by stakeholder.
C. Engagement of BP shows up in conversations.
D. Evaluators were the conscience of the NSF BP Agenda.

Theme A: Context complexity initially impeded evaluator engagement with BP.

- VEE evaluation engagement of the NSF broadening participation agenda was initially challenged by the complexity of the context, specifically by: (a) initial evaluator/client relationship and (b) change in Center management and leadership.

Power dynamics, leadership turnover, development of the evaluator/client relationship and credibility, and a difference of values among stakeholders initially placed constraints on the evaluation team’s ability to meaningfully engage BP. Of the previous, the evaluator/client relationship and change in Center management and leadership had the most significant effects on the level of engagement with BP.
Evaluator/Client Relationship

As in any evaluation context, there is always a period when the evaluators are defining themselves, their values, and the evaluation. It is at this point that it can be difficult to engage deeply with sensitive issues because the relationship between the evaluator and the client is being built. Trust and credibility are paramount for evaluators when broaching such topics. This was the case initially with the overall EBICS STC evaluation and, naturally, the REU evaluation.

The majority of the key stakeholders -- program managers, goal area directors, and Center leadership -- had run NSF REU site programs before. Therefore, even though they were largely unfamiliar with evaluation, they had already developed deep-seated opinions about how the REU program should be run. This was challenging because the evaluation team planned to assess and provide feedback on all aspects of the REU program, including recruitment, criteria for participant selection, participant placement in research laboratories, the actual REU program as it runs in the summer, and longitudinal tracking of participants.

The VEE approach encourages evaluators to begin conversations about diversity and equity at the onset of an evaluation. The VEE approach requests that evaluators make a safe space for these conversations and for stakeholders to be heard. In the first year of the evaluation, these initial conversations caused tension among stakeholders. Differing opinions about BP, lack of developed working relationships, and lack of evaluator proven professional ability and usefulness contributed to this tension. Additionally evaluators were located at one campus. As such, those faculty and staff initially were more familiar with evaluators and their work. Those staff also interacted with evaluators in person and on a more regular basis due to evaluators’ involvement in other on-campus evaluations.
Evaluators were invited to all REU planning meetings. These monthly meetings took place via conference call with five to eight participants, including site-level program managers and administrative staff, the diversity and education directors, the evaluation team, and occasionally Center associate directors. In the first few meetings, evaluation team members’ opinions were not solicited, and information was primarily provided about the REU program plans. During these meetings, evaluators would bring up participant diversity with varying levels of success. For example, some stakeholders agreed with the need to attend to Underrepresented Minority (URM) diversity as a criterion for REU participant selection, while others were less convinced. One stakeholder would state how he or she wanted REU participants to be selected for his or her campus, and another stakeholder would object and state a different process for selection that he or she felt was best. Then a third person conveyed how he or she ran a different REU program on his or her campus and that this EBICS REU should follow similar procedures. Following the guidelines of the VEE approach, evaluators were respectful and attentive to the multiple implicit and explicit values surfaced in these conversations. During the first year, evaluators reminded stakeholders of commitments they made in their grant proposal and pointed them to NSF-stated goals. These reminders were initially gentle, and evaluators spent more time listening to conversations among stakeholders than participating in them. The VEE approach called for evaluators to be respectful of norms, values, and ideals within the Center. As such, the evaluators chose initially to learn about stakeholder values and brought up BP is such a way that they weren’t ‘beating stakeholders over the head' with prescription of particular values.

In subsequent years (2012, 2013, & 2014), conversations about REU participant recruitment, criteria, and selection occurred with less tension for a number of reasons. First, stakeholders and evaluators had developed working relationships among and between each other.
Second, because evaluators called attention to the NSF BP agenda more and more each year, stakeholders knew what type of questions to expect from evaluators. Third, over the course of the Center’s four years in existence, evaluators attended almost all of the REU program meetings. Notes from a structured reflection completed the week of May 13th, 2013 highlight the development of respectful and professional relationships between evaluators and stakeholders.

[This was the] last REU program meeting before REU participants show up to each campus. This meeting went well, and all program managers feel they are ready to begin the summer REU program. At the end of the meeting, one of the managers from Site C thanked me for being flexible about when I would travel to observe their site. At the end of the meeting, I reminded all attendees about [the] REU participant experience and expectations for face-to-face with mentors and advisors. I commented on being sure to help participants acclimate to their lab’s culture and university culture upon arrival so that they have a good experience. The diversity director agreed with that suggestion. (Graduate student evaluator)

This quote also provides an example of a successful attempt by evaluators to respectfully engage the topic of culture with stakeholders. Another factor that affected the engagement of the BP agenda was the changes in Center management and leadership over time.

**Change in Center Management and Leadership**

The EBICS STC is a dynamic and ever-changing coalition of faculty, staff, and students from multiple universities. While overall faculty participation in the Center is consistent, there have been multiple changes in Center leadership and management. Initially, the program management structure was set up such that there were three full-time program managers (one at each site) and one full-time administrative support staff person. After the first year, the program manager who was in charge of the REU program moved on from the Center, and another full-time manager was hired. Then, at the end of the second year, there was a restructuring of Center management, and over the course of eight months, all three program managers left the Center. The current iteration includes two half-time program managers, one full-time manager, and two
full-time administrative staff. Additionally, the positions of Education Director, Diversity Director and Associate Director currently have different faculty members in them than when the Center started.

For the evaluation, this had a number of implications. First, the evaluation team has had to be very attentive to Center structure. These changes affected to whom evaluation reports were submitted, as well as who was in charge of various programs. Second, understanding the subtleties of these changes and power dynamics has also been of the upmost importance. Not all Center management and leadership positions carry the same amount of weight to make decisions. As such, it was important to understand who had the final say, or who could approve program or budget changes. Third, the position of Education director shifted from a faculty member at the same campus as evaluators to faculty at another campus. Therefore, the two main directors evaluators interacted with were at a different site. Finally, relationships with key stakeholders needed to be maintained, while also nurturing relationships with new Center personnel. Overall, ensuring that the right people were at the table for discussions about the REU evaluation and the NSF BP agenda was challenging at times. Per VEE recommendation of engaging all stakeholders, the evaluation team met with all program managers and goal area directors on a semi-regular basis, and more frequently when decisions had to be made about the REU program. Meetings were held more frequently while the REU program was in session over the summer. During times of transition, it was challenging because ownership of the REU program changed hands, and this affected who made decisions. Notes from a structured reflection completed the week of July 14th, 2013, highlight an example of change in Center management:

I just had a focus group with participants. I wrote up major findings and plan to submit them to site-level program managers. Participants have some concerns about professional development activities, and there is still time in the summer to add a few. I am not 100% sure who to send this feedback to because there is
currently an incoming and outgoing program manager. I’ll check with the lead evaluator to confirm, but I plan to submit these results to both managers. (Graduate student evaluator)

Next, I will discuss the difference in key stakeholder opinions (Theme B) of the evaluation’s engagement with the NSP BP agenda (Research question 1A).

**Theme B: Perceptions of the evaluation’s engagement with BP vary by stakeholder.**

- Perception of the level of VEE evaluation engagement of the NSF broadening participation agenda was dependent on stakeholder values and interpretation of the BP agenda.

All things considered, it was difficult to measure level of success and depth of engagement without taking into consideration the perceptions of five key stakeholders. The EBICS diversity director, the lead evaluator, a member of Center executive leadership, and program managers each have their own values and perceptions about the NSF BP goals and have had different types of interaction with the program and its evaluation. As such, their assessment of the evaluation’s level of engagement with the NSP BP agenda varied.

**Center Leadership and Program Managers**

Center leadership and program managers alike reported the evaluation as consistently engaging with NSF’s BP agenda. Their assessment of the level of engagement was high. Each interviewee in this category had a slightly different interpretation of the NSF BP agenda. When asked to define BP in their own words they offered the following statements:

The mission is to broaden the participation of underrepresented groups, women and other minorities within STEM and actually overall a growth in the STEM pipeline per se. I think what we would like to do is to be able to have it integrated from K–12 and connect that to the undergraduate research student and connect that to the graduate and connect that onwards. I think it really needs to be an integrated view from really early on to post graduate. I’m not sure NSF has presented in that holistic way, but I know the goal is certainly to help faculty and
institutions increase the participation in STEM disciplines. (Center Leadership Member)

Yeah, I can’t remember the exact details off the top of my head, but I am familiar and I have read some of the literature with what they consider broadening impact….The way that I interpreted it, is the NSF would like the programs that they fund to easily translate into things that can be shared at multiple institutions and with different types of groups so just basically that if they are contributing to these things people create programs or discoveries or data in a way that can easily be shared. And they make that information available to other people it might be relevant to. (Program Manager)

So the way I understand it is that NSF is interested in increasing diversity from the world and increase diversity in general in among faculty and student and post docs and in the work force. So increasing representation of women, people from minority groups, and people with disabilities. They would like to see kind of a reflection of the general population in, probably, I am assuming in all fields, and in particular in STEM. (Program Manager)

As seen in the interviewees’ own words above, within this group, interpretations of the NSF BP agenda differed. Two of the three interviewees’ see broadening participation as pulling in URMs, while the third interpreted it as sharing data and discoveries. Despite the differing of understandings, all reported the evaluation team as highly engaging the NSF BP agenda. Each of the three interviewees remembered issues related to diversity being brought up during executive committee meetings, the NSF site visits, education, diversity and outreach program meetings, and reporting on REU findings. As such, the evaluation team’s efforts were remembered as successfully engaging with the NSF BP agenda with this particular group.

**EBICS Diversity Director**

The EBICS diversity director was by far the most critical of the evaluation’s engagement with NSF’s Broadening Participation agenda. He did not believe the evaluation engaged with the NSF BP agenda well. This is directly related to his understanding and interpretation of the NSF BP agenda. When asked to define the BP agenda in his own words he stated:
NSF has the mission to broaden participation in science among all individuals in the country, particularly underrepresented minorities, ethnic minorities, women and people living with disabilities and, really, other social and economic groups that may not be exposed to science in the broader scale. As I understand it, it’s because it’s a federally funded agency and everyone pays taxes, so they want to make sure that everyone gets the benefit of the science that is being funded by everyone. (Diversity director)

While he was able to articulate the NSF’s BP goals accurately, he also espouses these commitments in his own professional practice. Long before he was the EBICS STC diversity director, he sought to mentor URM students and help them seek achievement in STEM majors. He has also started a research program that partners the STC with two local minority-serving public schools to raise the awareness of students to the world of biotechnology through real-world, hands-on research projects led by Center scientists. It is these values that led him to determine that the REU evaluation wasn’t doing enough to engage with the BP agenda. When asked to what extent the evaluation puts NSF’s BP goal to the forefront, he stated:

I don’t think it [the evaluation] does. I don’t think it does, because I don’t think it separates the information by demographic, so even when I read the [site A’s] comments and again, sometimes I can guess which students are which. But the year that we had a Latina student from Puerto Rico, White female, and then some Black students, like the comments, they all get lumped together; but I think there are differences in their experiences that get diluted by making it anonymous, and this is what all of them say. (Diversity Director)

Further probing revealed that he wanted to use disaggregated data to improve the program at each site, and within each lab to make it more culturally sensitive. He believed he was unable to do this or justify pushing for certain activities or laboratory placements without the data to back up such actions. He stated that he often receives anecdotal data from REU participants but insisted that the evaluation team seek out and report on experiences of different URM participants. Overall, he felt the evaluation could be doing a better job of engaging the BP
agenda because he has high hopes for EBICS and its potential to increase the number of URMs in STEM or at least increase their access to STEM activities and laboratories.

**Lead Evaluator**

The lead EBICS evaluator described the evaluation team’s engagement with the BP agenda as successful. Further, she believes in many ways during the first year, before the appointment of the current diversity director, the evaluation team was the sole voice and led the conversation about the BP agenda. Early on the lead evaluator was invited to play a role in Center leadership and attended executive committee meetings. Additionally, both evaluation team members were asked to sit in on the diversity and outreach, and education committees. As mentioned previously, during REU recruitment and program planning meetings, which evaluation team members attended regularly, the evaluation team brought up issues of diversity and equity. Even though discussing the NSF BP agenda sometimes brought tension into the evaluation, the VEE approach encourages evaluators to make spaces for conversations among stakeholders about program values and goals. As such, the evaluation often had pointed conversations with stakeholders about the purpose and goals of the REU program. These conversations were met with mixed emotions. The lead evaluator stated:

> So I think we are trying to help them [EBICS leadership] make sure the way they are doing their programs is good, and it is helping to create diversity within the Center. I think there is a little bit of tension or a little bit of a misalignment there. I think that some of the sites do not see diversity as a primary outcome of the REU program. So it seems like every year when we talk about recruitment, when we talk about selection criteria, when we talk about nature of the program for some of the sites we have to revisit the idea that this is a program to bring diverse participants into this pipeline. I think at some sites, they see this as a recruiting mechanism for the most able students and that is, in some case, at odds with this idea of broadening participation. (Lead evaluator)
Even though at times there was a difference of opinion among stakeholders, the evaluation team continued to bring up diversity, equity in access, and the NSP BP agenda. The VEE approach guided evaluation team members in how and when to have conversations and report findings related to BP to stakeholders. Evaluators were committed to describing stakeholder values and prescribing the values of diversity and equity. Similar to the Diversity Director, the previous commitments are held by the evaluation team regardless of the fact that the program is funded by NSF. However, ensuring that funder equity goals are addressed in the evaluation is also an important evaluation practice for the VEE evaluator. When asked to what extent the lead evaluator believed it was the evaluation team’s role to bring up the NSF BP agenda she stated:

I think because NSF has it as an explicit criterion of the program that we would always bring it up. I think if it weren’t in the NSF criteria then we might bring it up, but we have less of a leg to stand on. I think the program staff could say: “We already told you we don’t want to do this. Why do you keep saying this?”, but because it is part of the NSF explicit criteria, I feel like we are within our normal limits to do it, and I am happy to do it. I like doing it. (Lead evaluator)

Next, I will discuss the ways in which the engagement of the NSF BP agenda showed up in conversations between stakeholders and the evaluation team (Theme C). This is a theme within Research Question 1A about the evaluation’s engagement with the NSP BP agenda.

**Theme C: Engagement of BP shows up in conversations.**

- Evidence of engagement with NSF BP agenda was present in evaluation documents. However, the majority of the engagement of the BP agenda took place through dialogue and discussion.

Document review unearthed a number of materials supporting the claim that the evaluation team indeed engaged with the BP agenda throughout the evaluation, mainly during conversations with stakeholders. While evidence of the evaluation team’s efforts to engage with
the NSF BP agenda can be found in the evaluation plan and evaluation finding memos, the majority of BP engagement happened through conversation. Specifically, the evaluation engaged the NSF BP agenda during conversations within:

1) REU Program recruitment meetings and conversations about selection criteria for applicants.

2) Evaluation planning meetings with the evaluation team and with program managers.

3) Meetings with program managers and the Diversity Director about professional development and social activities for REU participants.

4) Meetings for annual site visits and reporting to NSF.

5) Site visit presentations and questions from NSF.

Evidence for these conversations included (a) evaluation meeting agendas and notes; (b) Center diversity, education, and outreach meeting notes; (c) emails between evaluators and stakeholders; and (d) evaluation presentation slides. For example, two meeting notes included the written comments “members asked for diversity numbers” and “Diversity director suggests we change recruitment for REU by attending and advertising at minority-serving STEM conferences.” Additionally, discussion around the NSF BP agenda often arose during NSF site visits. NSF representatives often asked stakeholders and evaluation team members about their efforts to broaden participation and about diversity within the Center. The previous two examples highlight an important aspect of the evaluation team’s engagement with the BP agenda. Most of the engagement with these issues was interpersonal in nature. Less evidence was found pointing to formal reports broaching this subject. For example, evaluation reports contained sections on ‘diversity;’ however there was no ‘broadening participation section’ or recommendation in the report from evaluators to stakeholders that they should change how they

---

12 Annual NSF site visits are held at MIT in Boston. A representative number of faculty and trainees attend and present research, evaluation, and programmatic findings to NSF representatives.
recruit and select REU participants. Again, there were many conversations about the previous topics. The reasoning for this was that VEE evaluators are called to nudge stakeholders regarding such issues, and acknowledge differing stakeholder opinions. Therefore, while evaluators were explicit, often reminding stakeholders about the NSF commitment to BP, they chose not to write inflammatory reports. However, evaluators did support the Diversity Director’s efforts to change program recruitment efforts. These will be discussed within the components for success of VEE approach section.

Next, I will discuss the ways in which evaluators were the conscience of the NSF BP agenda (Theme D). This is a theme within Research Question 1A about the evaluation’s engagement with the NSP BP agenda.

**Theme D: Evaluators were the conscience of the NSF BP Agenda.**

- Evaluators continuously reminded stakeholders about the NSF BP agenda and, in the initial phases of the evaluation, were the leading voice in the BP conversation.

As mentioned in previous sections, evaluators consistently started and participated in conversations with stakeholders about the NSF BP agenda. Furthermore, during the first two years, before the appointment of the current Diversity Director, evaluators often brought up the topics of diversity, access, and equity within the REU program. As such, in many ways, evaluators were the conscience of the NSF BP agenda. Per VEE guidelines, they were the often gentle and respectful reminder that attempted to guide conversations about BP and the importance of URM in STEM to stakeholders.
**Question 1B:** How well did the VEE approach offer information of value and relevance, especially regarding the program’s connections to URMs, to both local program staff and NSF?

**Definitions**

*Valuable and relevant information* within this context is defined as follows:

*Value*: something of worth or usefulness.

*Relevance*: practical and applicable.

*Valuable and relevant information*: to provide stakeholders with information that they found useful, practical, and applicable to their REU program.

**Data Used**

Data used to answer this question included: (a) document review items—evaluation and program materials, (b) evaluator structured reflections, and (c) stakeholder interviews.

**Overarching Finding**

The EBICS evaluation team used the VEE approach as a framework and supplied EBICS stakeholders with valid and useful information to guide program improvement, assess short- and long-term effectiveness, and increase connections to URMs. Relevant information was provided to NSF; however, the value of such information is unknown.

The following themes emerged during data analysis and are presented below in support of the overarching finding:

A. Initial stakeholder expectations and evaluation purpose were misaligned.

B. Information of value and relevance was provided to stakeholders.

C. Evaluators reported on REU program connection to URMs.

D. Information was supplied to NSF as requested.
Theme A: Initial stakeholder expectations and evaluation purpose were misaligned.

- Stakeholder initial understanding about and expectations for the evaluation were low. As a result, the evaluation exceeded their expectations.

Stakeholder expectations and understanding of evaluation purpose directly impacts their assessment of the evaluation’s ability to offer information of value and relevance. This is especially true of the EBICS STC stakeholders (program managers and Center leadership). Initial expectations of the evaluation were relatively low because Center leadership, faculty, graduate students, and staff had little experience with formal program evaluation. Initial conversations revealed that the Center leadership and program staff had minor interactions with professional evaluators in the past and were unsure how to best utilize evaluator skills. As a result, the majority of key stakeholders saw the purpose as mostly related to collecting data on student experience and program impact.

Of all those interviewed, the Center EBICS leadership member was the least involved in the day-to-day planning of the REU program. While he had REU participants in his lab, he was not involved in the social activities or professional development planning of the program. He stated that he saw the purpose of the evaluation to see if the goals set by the Center with respect to the program are being met. He felt the evaluation was able to collect data and provide results that would answer these questions. He also expected the evaluation to determine if REU participants learned about “the science and mission of EBICS and if it was a good experience for program faculty.”

Program managers and the diversity director were entrenched in the planning and execution of the REU program. They had additional expectations and thoughts about the purpose of the evaluation. When asked what the purpose of the evaluation was, they stated:

It seems to me that the evaluation purpose was to look at how the program was going across multiple campuses from an objective standpoint and give feedback
that would allow us to continue to improve it and work on areas of weakness to have a really strong REU with very positive evaluation. (Diversity Director)

The purpose as I understand it is to find out how well our program is actually running, if it’s being effective for the goals that I guess EBICS has decided will be the learning goals of the REU. And I guess the evaluators also do the demographics. But I understand it to be, we can always think that we are doing well, but do you have any metrics to see how well you really are actually doing? (Program Manager)

So the purpose in my interpretation was to get feedback of what the students were experiencing and ways in which we can improve their experience. (Program Manager)

There was a desire from the program managers to know how well the program was doing and if it was running effectively. The diversity director did expect the evaluation to reveal program strengths and weaknesses of each school. He stated:

I like that you keep track of all the data. The things that I like to know about, I mean, I’m at Site C, so I do like to hear about how these programs are being run at the other schools and what the students are feeling at those other schools compared to what they are feeling at my school, because I’m not there, and I might meet those students and I might not. And I’ll say that I also like that the evaluations kind of help reinforce some of the things that I think about what’s right or not right at each of the different schools or at my own school. And I like that it comes out from the students’ mouth. (Diversity Director)

Overall, stakeholders wanted to know how well they were doing, and after the second year, they began to request information on what could be done better and how to improve the REU program. The evaluation team’s purpose and evaluation execution exceeded the expectations of stakeholders. While evaluators did evaluate how well the program was running and describe student experience, additional information was collected and evaluated. Stakeholders were especially pleased with the longitudinal tracking system that evaluators developed to follow REU participants, their interests in STEM, graduation rates, and graduate school pursuits. The lead evaluator saw the purpose of the evaluation as:
I think it kind of goes with our four levels of evaluation. Is it happening on time and is it planned? Are the pieces of the REU program effective? Is the professional development effective? Is the mentoring effective? Is the recruitment effective? Are all those pieces effective? What outcome does it have for the kids? Is it getting them more interested in STEM? Is it encouraging them to go to graduate school? And then finally, is it becoming more institutionalized at the participating schools? Are they thinking about these programs to diversify STEM in the summer to bring more research experience to undergraduates? (Lead Evaluator)

The evaluation team also acknowledged the purpose as to teach stakeholders about their program and its underlying logic and theory. This is important to keep in mind as I shift to what information stakeholders found of value and of relevance to their needs.

Next, I will discuss which information provided by the evaluation was valuable to stakeholders (Theme B). This is a theme within Research Question 1B about the evaluation offering information of value and relevance.

**Theme B: Information of value and relevance was provided to stakeholders**

- Stakeholders found a number of evaluation findings valuable and relevant. These findings were related to: past REU program iterations, program professional/social activities, participant skill development, and long-term impact.

While many evaluation approaches are able to garner data that are useful and relevant to their stakeholders, the development of relationships with stakeholders (including participants), on-site presence, attention to diversity and equity, and evaluation as an educative process are hallmarks of the VEE that set it apart from other evaluation approaches. While other approaches surely would have provided useful data, the way in which VEE evaluators offered data and the type of data were unique. In the following sections, I overview information the evaluation provided that was of value and relevance to key stakeholders.
Past Iterations of the REU Program

The evaluation of the REU program generated a great deal of information that EBICS STC leadership and program managers found relevant and useful. Both of the program managers interviewed for this study were hired during the Center’s third year and were not present for the first two iterations of the REU program. As such, they both found the previous evaluation reports and conversations with evaluators to be useful to guide them as they developed the 2013 REU program. One program manager commented:

I wasn’t coordinating the REU from the beginning of the grant. So it was really useful for me to catch up so in reading over evaluation reports that you’ve sent. I have a sense of how it’s gone as a whole in areas of strength and areas of weakness. I feel like I can look closely at those points both the strengths and weaknesses, and now we can do our absolute best to repeat the good things and correct some of the weaknesses for future summers. I think as a whole, just giving information and kind of guiding me on how I should proceed so I am not just going from my gut, which is also good, but getting an outside perspective from an expert is really nice. (Program Manager)

During the 2013 iteration of the EBICS REU program, evaluation team members not only attended diversity committee and education committee meetings, they also met with each of the three program managers collectively and individually to discuss previous REU evaluation activities, findings, and recommendations.

Professional and Social Activities

As mentioned earlier, VEE evaluators are called to spend a significant amount of time on site. While REU participants received surveys and participated in focus groups throughout their program, evaluators attended professional development and social activities to observe and informally chat with students. Early on, at one of the sites, it began to be evident that REU participants were dissatisfied with the amount and quality of their professional development
offerings. Evaluators met with program managers at that site, and changes were implemented before the end of the program. This is further discussed in findings related to program design.

**Participant Skill Development**

After the 2011 iteration of the REU program, stakeholders were interested in learning more about what type of technical and professional skills participants developed as a result of their REU experience. Evaluators then met with EBICS Center program managers, and with program managers and a diversity director from another NSF STC to examine how they understood skill attainment as a result of REU program participation. During these meetings, evaluators led conversations that explored participant skill attainment expectations versus program inputs. A hallmark of the VEE approach is its attempt to educate stakeholders about their program. This was one of the first times stakeholders were introduced to the logic and theory behind their program. While this process allowed evaluators to better understand what type of skills attainment should be measured, it also provided an opportunity for stakeholders to critically think about their program and its potential impact.

**Long-Term Impact**

Very few REU sites track or follow up with their participants (Tillman, 2013). However, in order to better understand program impact and continue to understand the program’s lasting engagement with BP, the evaluation team employed a longitudinal tracking system. All participants are contacted every spring semester in an attempt to better understand how the program has affected them in their academic and post-academic careers. Questions related to their current aspirations, employment, usage of skills developed during the REU program, and attitudes about STEM and STEM careers are measured. While EBICS stakeholders initially did
not expect a longitudinal evaluation, this is one aspect of the evaluation with which those interviewed were most satisfied.

Next, I will discuss information provided to stakeholders about the REU program’s connection to URM (Theme C). This is a theme within Research Question 1B about the evaluation offering information of value and relevance.

**Theme C: Evaluators reported on REU program connection to URMs.**

- Evaluators provided stakeholders with information about URMs and their experience that was relevant to recruitment and selection criteria.

Initially, a few stakeholders expressed that they wanted to use the REU program as a tool for recruiting future graduate students. Others asserted that they expected to give students who would not normally have access to laboratories a research experience and that REU participants might or might not be students they would recruit for graduate school. Per VEE guidelines of promoting and sustaining critical reflection and respectful dialogue in order to engage with values and sustain program improvement, evaluators along with the diversity director sought to show stakeholders that it is possible to achieve both goals with the REU program.

The first year each site decided to utilize its own selection criteria. One site maintained their minimum 3.75 GPA requirements while the other two sites settled on 3.33 and 3.5. At the end of the summer, graduate mentors and faculty advisors completed assessment surveys of how their REU participant performed in their laboratory. Evaluators provided two pivotal pieces of information to stakeholders. The first showed that the average GPA of Black REU participants was 3.5, with only one participant having over a 3.75. Second, there was no relationship between REU participant GPA and mentor satisfaction. As such, evaluators reminded stakeholders that productive REU participants might not only be those with high GPAs, and that if the whole
Center would have decided upon 3.75 GPA as a cutoff, then they would have had fewer URM participants who worked in the laboratories.

The second year, the evaluation team encouraged stakeholders to solicit and accept participants from STC minority-serving institution partner universities. Evaluators emphasized that these students would likely be URM and would have knowledge of the science required to perform an independent research project in Center laboratories.

One aspect of the program’s connection to URM that the evaluation team failed to formally capture was the URM program experience. No formal individual interviews were conducted, nor were additional survey questions asked of URM participants. Also, data were not disaggregated because there were so few URM at each site that the evaluation team was concerned with confidentiality and the confounding variable of sites. The lead evaluator responded to the diversity director’s request to begin to provide him with survey results disaggregated by ethnicity:

That’s a really good idea and the problem is confounding between institutions. Really, only one institution has been super diverse, so if you disaggregate by African American, the results are going to be also confounded by Site C but it doesn’t mean don’t do it. Just understand that the majority of African American students are at Site C. (Lead evaluator)

She also stated that, in order to uphold confidentiality, the evaluation would not disaggregate data if there was only one participant of a given ethnicity. The evaluation team also commented that the 2014 iteration of REU participants are all URMs, and in the summer of 2014 would make for an interesting case study of participant experience in itself. Both evaluation team members asked all participants, especially URM about their experience in the program. They also informally checked in with URM participants during REU events and chatted with the diversity director and program managers about ensuring success of those participants. These
efforts were not formal evaluation activities, but explicit conversations and recommendations. For example, when evaluators reviewed demographics of future REU participants they found that one participant was Muslim. They contacted program managers and suggested he/she connect with that future participant and give them information on local mosques to ensure that the participant would have a smooth transition to the REU site.

Next, I will discuss information provided to NSF by the evaluation team (Theme D). This is a theme within Research Question 1B about the evaluation offering information of value and relevance.

**Theme D: Information was supplied to NSF as requested.**

- Evaluators supplied REU information to NSF through contributions to the annual report and site visits.

  NSF requires STCs to report progress annually on their research and program goal areas. Each year, evaluators provided information to program managers as they compiled the report. Evaluators provided Center participant (including REUs) demographic data, as well as data on satisfaction, program implementation and effectiveness, and impact. During NSF site visits, evaluators presented evaluation activities, goals, and findings; and assisted faculty goal area directors with their presentation as well. NSF site visit participants received a copy of all evaluation reports, and evaluators were present to answer any questions. Those representing NSF have anecdotally commented that they found the evaluation data and findings useful to understand the “state of the Center.”
**Question 1C: How well did the VEE approach have an actual impact on program design (notably, stronger BP potential) and program outcomes (effective BP programs)?**

**Data Used**

Data used to answer this question included: (a) document review items—evaluation and program materials, (b) evaluator structured reflections, and (c) stakeholder interviews.

**Overarching Finding**

EBICS REU program design was altered by the VEE evaluation findings and recommendations. Data on longitudinal program outcomes are currently being gathered.

The following themes emerged during data analysis, and are presented below in support of the overarching finding:

A. REU program design was impacted by the implementation of a VEE evaluation.

B. Evaluation findings have been primarily formative in nature.

**Theme A: REU program design was impacted by the implementation of a VEE evaluation.**

- REU participant selection criteria, graduate student mentor selection, and program organization were redesigned based on VEE evaluator recommendations.

**REU Participant Selection and Selection Criteria**

Since 2011, the VEE evaluation team has worked to keep diversity on the minds of EBICS leadership and program managers by reminding them about NSF BP commitments during executive committee meetings; in evaluation update emails; during personal conversations; and during education, diversity, and outreach program meetings. As mentioned earlier, until the appointment of the current diversity director, in many ways, the evaluation was
leading this conversation. One example that was carried through 2011 to 2014 is related to the criteria set forth to select REU participants. The lead evaluator commented:

So it was the evaluation that kept bringing up diversity goals. We didn’t always win. Like the first year, the criteria[on] was a GPA of 3.75, and diversity was not at all in the selection. We kept putting it in and the faculty kept taking it out. In the second year, I think it came back a little bit more, but they wanted to put the criteria that the person had already had to have had undergraduate research experience, which excludes that whole group of students who are at institutions where it’s not easy to attain an undergraduate research experience. I think in this REU, more so than the other ones we work in, it’s very hard to keep that diversity criteria out there. (Lead Evaluator)

As mentioned in the quote above, during the first iteration of the REU program in 2011, faculty, especially those at Site B\textsuperscript{13} were interested in using the REU program as a recruitment tool for future graduate students. However, faculty at Site A and Site C were interested in having undergraduate researchers, but not necessarily for the purpose of recruitment. As such, there was a tension between each university’s corresponding values and what they believed should be used for selection criteria of REU participants. In the end in 2011, Site B decided to stick with the high GPA criterion, and the other two universities went with lower GPA requirements. Below is a table that shows 2011 REU participant demographics.

In 2012 and 2013, VEE evaluators continued to stress the importance of attending to the NSF BP agenda. Every year during recruitment conversations, evaluators were present to participate in the conversations about criteria to be used for selection. Additionally, participant demographic data were always included in evaluation reports. In 2012 and 2013, there were small gains with respect to the selection of URM for the REU program. It is in the 2014 iteration of the REU program that the most significant changes can be seen with respect to participant

\textsuperscript{13} Sites have been de-identified.
selection. In the summers of 2012 and 2013, evaluators pursued the VEE guideline of promoting the potential power of the evaluation to be used as a meaningful learning activity with the evaluation of the EBICS high school research experience program. The diversity director was in charge of this program and had complete autonomy to run the program as he saw fit. He relied heavily on evaluators for formative data and recommendations to shape the program and as a result had a true understanding of the potential for evaluation to directly shape a program.

During the 2013-2014 academic year, the diversity director asked that the evaluation team to collect and present him with data on 2011-2013 REU participants demographics. He then used these data to prepare a report for EBICS leadership to argue for additional attention to diversity and equity within the REU program. He also used data the evaluation team provided to him during the January 2014 EBICS NSF site visit. As a result, the 2014 EBICS REU cohort is composed of 100% URM students. URM women participant selection has remained constant across years and in 2014, Site A will host the Center’s first REU participant with a physical disability.
Table 5

REU Participant Ethnicity by Site and Year

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2012</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2013</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2014</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

- **Number of REU participants**
- **REU Site and year**

Legend:
- Asian
- Black
- Latino
- White
- Other
REU Graduate Student Mentor Selection

Per VEE recommendation, each year the VEE evaluation team attempts to spend time ‘hanging out’ with REU participants on each of the campuses. This includes attending social activities, visiting participants in their assigned laboratories, hosting focus groups, and observing professional development. In 2011 and 2012, Sites A and C were the only two sites that were visited. In 2013, a concerted effort was made to visit all EBICS REU sites. As such, all sites and all REU participants interacted with the evaluators in person at least once. As a result of these efforts, evaluators found that one REU participant on each campus had a graduate student mentor who was out of town at some point during the summer. One of the participants noted that he/she didn’t mention it on the survey because he/she didn’t want to get his/her mentor in trouble. However, during the time the mentors were gone, participants were not left with work to do and were unable to collect data in their labs. Evaluators spent time on each campus ‘hanging out’ with REU participants and provided safe spaces for participant dialogue. Without the previous, evaluators would not have been able to collect data to support the recommendation for a change in how graduate student mentors are selected. When asked about how the evaluation has changed the REU program, one program manager referred to graduate student selection. She commented:

That issue with the concerns of across campuses about mentors who are out of town or are unavailable. I think in 2014, we can make that clear when we are recruiting the graduate mentors that at most one week and even two weeks are too much to be gone. If it is going to be anything more than a few days, it’s not going to work out for them to be a mentor. If it is a few days, the mentors would still have to have a backup person who is in close contact and willing to assist the REU participant in case of issues because I think that’s important too. (Program Manager)

Program Organization

As in previous years, but especially in 2013, the evaluation team made efforts to follow VEE guidelines of spending time and hanging out at each REU site to better understand the
program context. Additionally, per VEE recommendation, the evaluation team attempted to purposefully understand and engage participants’ values. Half-way through the summer, evaluators decided to have an impromptu focus group with REU participants at all three campuses to check in with them. Because the participants were familiar with evaluators, they felt comfortable expressing some of their concerns. One was related to feeling as though REU participants were not being provided with enough professional development geared directly towards Biomedical Engineering. The evaluation team decided to immediately contact program staff and notify them of students concerns.\textsuperscript{12} The following email was sent to program managers.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Email from Graduate Student Evaluator to Program Manager}
\end{figure}

The next email follows with the program manager’s response.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure6.png}
\caption{Email from Program Manager to Graduate Student Evaluator}
\end{figure}

\textsuperscript{12} To protect confidentiality, these emails have been de-identified.
As seen above, VEE evaluation efforts resulted in the addition of professional
development activities and additional practice time for participants. In addition to professional
development VEE evaluators also highlighted discrepancies by site in terms of different
pedagogies adopted for different students. As previously mentioned, some sites felt the purpose
of the REU program was to nurture and develop students who were less experienced in research.
Others believed the purpose of the program is to recruit students who have previous research
experience and possibly groom them to be future graduate students. These two views on the
purpose of the program impacted pedagogies within different laboratories. These differences
were emphasized and discussed among REU leadership.

Next, I will discuss VEE evaluation findings (Theme B). This is a theme within Research
Question 1C about the evaluation impacting program design and outcomes.

**Theme B: Evaluation findings have been primarily formative in nature**

- Evaluators have provided primarily formative program feedback to stakeholders.
  Longitudinal outcome data are currently being collected.

Until spring 2014, the main focus of the EBICS STC evaluation has been formative. This
also holds true for the REU evaluation component. Evaluation reports included summative data
about quality of program design, participants’ overall experience, and acquisition of and
potential use for new technical skills developed.

In terms of outcomes data collected thus far, every spring, the evaluation team emails all
REU participants a follow-up survey to see if they have graduated, are currently employed, and
if they are in contact with their EBICS mentors/faculty. REU participants are also asked if they
have used any of the skills they developed in the EBICS REU program during the academic year
at their home institutions and how EBICS has impacted their decision to pursue an advanced
STEM degree. Initial findings indicate that about half of REU participants are still in contact with their advisors and mentors, and that the program has increased interest in STEM research and positively influenced decisions to pursue graduate degrees in STEM.

After the 2014 iteration of the REU program, the evaluation team plans to collect data and write a summative report on the REU program for EBICS leadership. Data collection methods for this report will include document analysis of previous evaluation reports, surveys and interviews with past REU participants, and interviews with EBICS faculty and graduate student mentors. The impact of evaluator’s use of the VEE approach in the evaluation on program outcomes is unclear. Additional data will be collected over the course of the next few years.

**Question 2. What components or strands of VEE most importantly account for its ‘successes’ in this REU evaluation context?**

**Data Used**

Data used to answer this question included: (a) document review items—evaluation and program materials, (b) evaluator structured reflections, and (c) peer review.

**Overarching Finding**

VEE components responsible for evaluation success are primarily related to the engagement of the program/context culture and values and the development of relationships with program stakeholders.

---

13 The majority of the strands of VEE associated with limitations in this context are limitations due to the multi-site context. As such, this question was changed to successes only and limitations will be discussed in the next section.
The following themes emerged during data analysis and are presented below in support of the overarching finding:

A. Deep and sustained engagement with values and culture contributes to evaluation success.
B. Having status and credibility with stakeholders contributes to success.
C. Being educative contributes to evaluation success.
D. Development of respectful professional relationships contributes to evaluation success.

Theme A: Deep and sustained engagement with values and culture contributes to evaluation success.

- VEE approach guidelines of engaging with culture and values accounted for evaluation successes.

At its core, the VEE approach endeavors to purposely engage with the multiple values inherent in all programs while giving special attention to diversity and equity. I have written much about descriptive and prescriptive valuing up to this point. With the guidance of the VEE approach, the evaluation team also sought to understand and engage with the cultures accompanying this context.

The culture of the STEM field itself is very unique with a specific set of values and jargon, which can be challenging to understand. Terms like microfluidics, assay, and biomarker were quite foreign to evaluators. While Center research activities were not evaluated, in order to engage with stakeholders and have a thorough understanding of this evaluation context, it was important for evaluators to engage with STEM culture. This was done by attending as many Center events and activities as possible. These efforts, grounded in a VEE approach, eventually increased the evaluators’ STEM jargon proficiency. These efforts not only increased STEM understanding, but also confidence and comfort level with the evaluation. Notes from a
structured reflection completed the week of July 28th, 2013 highlight the usefulness of engaging with the STEM culture:

I attended Site A’s REU poster session. Surprisingly, I was able engage with two of the five participants about their research in a conversation because I had attended a presentation during the EBICS retreat given by their faculty advisors. I was also able to better judge the quality of the posters/quality of research studies this year compared to last year. I know this is because I have been spending more time talking to EBICS graduate students about their own work and attending EBICS research events. Additionally, my expanded vocabulary meant that I was better able to understand graduate student mentor’s critiques related to in-lab experiences of their REU participants. (Graduate student evaluator)

Per VEE recommendations, the evaluation team purposefully attempted to have deep and sustained engagement with the Center; this meant even though spread across multiple sites, it was necessary to attend as many Center activities as possible. The evaluation team observed courses, attended student talks and seminars, and chatted with REU participants about their research. The annual Center retreat was instrumental in relaying research updates, and allowed Center members to interact and network with each other and most importantly gave insight to Center goals and each member’s role in the Center. Further, every year the evaluation team encouraged site-level labs to allow their REU participants to attend the retreat if it is on their local campus. REU participants who attended the Center retreat had a greater understanding of the concept of the distributed Center EBICS.

Each university had its own culture that included goals, physical space, traditions, etc. As mentioned previously, the evaluation team endeavored to be present and to show up to as many Center events as possible. The VEE approach encourages evaluators to learn about the diverse characteristics of a program and its context. Site A and Site C were visited multiple times a year to evaluate the REU and other Center programs. These visits assisted in developing a deeper understanding of each university’s culture so that the evaluation team could better engage
with stakeholder and program values. Each university’s physical layout, student population, local city size, demographics, and weather all played a role in the university cultures. City size also played a role in university culture as well. Boston and Atlanta are large, urban cities both with populations over 400,000. Champaign-Urbana, on the other hand, is considered a micro-urban city with a population of just over 230,000 (United States Census, 2013). Understanding each of the universities’ and surrounding cities’ culture has helped the evaluation team be more thoughtful about data collection and analysis related to REU participants’ motivations for choosing a certain campus and their acclimation to the program over the summer. Notes from a structured reflection completed the week of June 23rd, 2013, highlight the usefulness of engaging with the STEM culture:

Having the opportunity to visit Site B while the REU program is running has made a big difference in my understanding about the culture of the REU program here, versus other sites. While I knew that faculty utilized the program as a recruitment opportunity, it is now quite clear that all REU participants are also hoping to show their faculty advisors their skills so that they might remember them when they apply for a Ph.D. program. This is completely different than at the other two REU sites. (Graduate student evaluator)

Next, I will discuss evaluator status and credibility with stakeholders (Theme B). This is a theme within Research Question 2 about evaluation success.

**Theme B. Having status and credibility with stakeholders contributes to success.**

- VEE approach guidelines of having status and credibility with stakeholders accounted for evaluation successes.

The VEE approach encourages evaluators to have authority, credibility, and presence in the evaluation context. At the onset of the EBICS REU evaluation, the majority of the key stakeholders were not familiar with program evaluation. Further, because of the multi-site nature of the program, the majority of Center faculty were not familiar with the work of the lead
evaluator. However, her status as the Director of I-STEM, and her reputation among her Illinois colleagues gave her some credibility and authority in terms of the evaluation. For example, she was invited to sit on the executive committee soon after the Center was funded. This gave her access to the Center leadership in a way that would not have been possible otherwise.

At the beginning of the evaluation, few Center faculty were familiar with the graduate student evaluator. Initially, her status as a graduate research assistant excluded her from certain committee meetings, and Center leadership would direct all their questions, suggestions, and concerns exclusively to the lead evaluator. However, as time passed and the evaluation team continued to attend Center events, collect data, produce reports and present results, as such, credibility with the key stakeholders rose. Four years into the evaluation, the Center leadership now directly email the graduate student with questions, suggestions, and requests. This is important for the success of the evaluation because the graduate student evaluator was the primary evaluation data collector, analyzer, and report writer under the direction of the lead evaluator. Below is a string of emails received from the Center Director in January 2014 following the submission of the Center evaluation comprehensive report.

Figure 7. Email from Graduate Student Evaluator to Center Leadership and Program Managers
The previous emails highlight not only the importance of having credibility, but also showcase the need to develop relationships with stakeholders. Increased credibility allowed evaluators to be used to a greater extent than if stakeholders were unsure of their usefulness. Additionally, increased credibility allowed evaluators to continue and increase conversations about the importance of the NSF BP agenda.

Next, I will discuss the VEE guideline of being educative (Theme C). This is a theme within Research Question 2 about evaluation success.

**Theme C. Being educative contributed to evaluation success.**

- Being educative with stakeholders accounted for evaluation successes.

Evaluators attempted to educate stakeholders not only about their program, but also about evaluation itself. The evaluation team explained the evaluation process to stakeholders. They overviewed the multiple evaluation commonplaces including: evaluation context, purpose, audience, questions, criteria for judging quality, data collection activities, and reporting. During evaluation meetings, the evaluation team worked to be educative about the importance of high quality, valid, and useful data. They did this by explaining why they chose to collect certain
types of data (surveys, interviews, etc.) based on the evaluation questions and information needs of the evaluation. This increased evaluator credibility and ultimately stakeholder understanding of evaluation. The Diversity Director commented:

I will say this, I have learned a lot about the importance of evaluation through interacting with the evaluators. It’s been very informative and helpful to me which you know, I’ll contact you and be like, “Oh we need to evaluate this because I don’t know…,” so that has been very eye-opening. When I first started this job as a professor and I was reviewing NSF grants, and NSF were like, “They need to have assessment,” and I was like, “Ok, so it’s just a buzzword, just say it.” Now, I do get the importance of it, so that has come out of working with the evaluators at least at this level as the diversity director. (Diversity Director)

Next, I will discuss developing relationships with stakeholders (Theme D). This is a theme within Research Question 2 about evaluation success.

**Theme D. Developing relationships with stakeholders contributed to evaluation success.**

- VEE approach guidelines of developing relationships with stakeholders accounted for evaluation successes.

Initially, it was difficult to develop relationships with many of the stakeholders due to distance across sites. However, the VEE approach is anchored in the development and nurturing of strong and trusting relationships between the evaluation team and stakeholders. As mentioned earlier, the evaluation team dedicated resources to be able to attend all key Center events, including the annual retreat and NSF site visit. The development of relationships resulted in the evaluation team being utilized in a way that they had not previously, especially pertaining to the engagement of BP.

An example can be found in two emails from the Diversity Director asking for committee members to review a change of REU program and rationale proposal he prepared for the executive committee. He asked the evaluation team to compile the participant diversity
demographics for him as he wrote the report. The first figure shows a snippet of his original email to the graduate student evaluator; the second email is to all diversity committee members.

**Figure 9.** Email from the Diversity Director to the Graduate Student Evaluator

**Figure 10.** Email from the Diversity Director to the Education and Diversity Committee

**Question 3. What specific challenges, if any, did the multi-site nature of the STC REU program raise for the implementation of the VEE approach?**

**Definitions**

Meaningful within this context is defined as follows:

*Meaningful:* full of significance or purposeful.

**Data Used**

Data used to answer this question included: (a) document review items—evaluation and program materials, (b) evaluator structured reflections, and (c) stakeholder interviews.
Overarching Finding

Multiple stakeholders with varying power and values, in multiple locations, program time frame, and data collection were challenges associated with the multi-site nature of the EBICS STC REU program evaluation. Additionally, EBICS Center composition and evaluation priorities raised challenges for VEE approach implementation.

The following themes emerged during data analysis and are presented below in support of the overarching finding:

A. Attending to multiple stakeholders in different locations was a challenge.

B. Distance and program run time challenged on-site data collection.

C. Center complexity influences evaluation priorities.

Theme A: Attending to multiple stakeholders in different locations was a challenge.

- VEE approach guidelines of engaging with all legitimate stakeholders proved to be a challenge in this context due to the multi-site program context.

As previously mentioned, the evaluation team reported to Center leadership, including the Center director, associate directors, program area directors, and program managers. These stakeholders were across multiple campuses, had their own thoughts, opinions, and values related to the REU program and its purpose and goals. The multitude of values was discussed earlier; however, the location difference exacerbated this challenge. The evaluation team had to negotiate evaluation data collection priorities with key stakeholders across three geographically dispersed locations. For example, one stakeholder felt it was necessary for the evaluation team to be present at the final REU culminating poster session. When another program manager also felt the same way, decisions had to be made about how best to split evaluator time and travel resources to ensure that needs were met at both locations. This specific issue could sometimes be solved by having one evaluation team member attend an event at Site A, while the other traveled...
to Site C. The feasibility of this solution varied greatly depending upon evaluation team members’ schedules and availability. Notes from a structured reflection completed the week of May 16th, 2013 highlight these challenges:

I met with the program manager from Site C today. She asked if it would be possible for me to come to one of the REU events whose dates coincides with an event we had planned to observe at Site A. I let her know that it was not possible, and we arranged for me to visit during a different event for the REU participants that she was interested in collecting feedback on. (Graduate student evaluator)

Another challenge was ensuring that all stakeholder voices were all equally heard and considered during evaluation planning. Because evaluators were located at one of the three sites, they naturally had more face-to-face interactions with Center leadership and program managers. As mentioned earlier, evaluators met with program managers and goal area directors located across sites frequently to ensure that they had opportunities to voice opinions and keep up to date with evaluation activities and findings.

Next, I will discuss how distance between sites and program run time affected on-site data collection (Theme B). This is a theme within Research Question 3 about VEE implementation issues due to the multi-site context.

**Theme B: Distance and program run time challenged on-site data collection.**

- The EBICS REU program ran for 10 weeks in the summer. This short time frame made on-site data collection challenging.

While the EBICS REU program evaluation runs year-long, the program itself is implemented during a 10-week period in the summer. This proved to be difficult to implement VEE data collection recommendations with fidelity. The VEE approach calls for evaluators to spend significant time on site; however, because the evaluation team was based at one site and had to travel to the other two sites, it was not possible for evaluators to spend significant amounts
of time at the two remaining sites. On-site observations, interviews, and focus groups were minimal at the two additional sites. During the 2013 iteration of the EBICS REU program, evaluators visited each site at least once during the summer. This was more than in 2011 and 2012. When visiting the REU sites where they were not based, evaluators spent the majority of their time observing professional development activities, visiting students in their laboratories and learning about their research projects, and interviewing participants face to face. While evaluators did collect survey data and conducted multiple focus groups virtually, it was not the same as being present on-site.

Key stakeholders, however, felt that the evaluation spread resources evenly across sites and were pleased with the amount of attention each of their sites received. When asked if the site evaluators are located at receives more evaluation attention, the majority felt that while that site might have received more time, it probably wasn’t much more than their site received. Further, they were satisfied with the evaluation resources dedicated to their site:

I would say I feel like its [the amount of evaluation time] equal because I am actually not sure how much time, I know that you evaluate a lot of different EBICS things, and I didn’t get the impression that you evaluated the REU at Site A or Site B more. (Program manager)

It seems to me, from my understanding, that it’s been basically identical. Whatever you’re doing with Site A, you’re doing with the other two sites, in particular, the students all have the same questionnaires and they participate in the multisite evaluation sessions and so on. So it seems to me to be matching. (Program manager)

Oh, I think that we pay attention to them and that we get evaluated with everyone else. For sure your evaluation team contacts us when you are doing the evaluation for everyone, and we make accommodations for you to meet with the students. So I think we are not left out of any of it. (Diversity director)
Next, I will discuss how the complexity of this context impacted the implementation of this approach (Theme C). This is a theme within Research Question 3 about VEE implementation issues due to the multi-site context.

**Theme C: Center Complexity Influences Evaluation Priorities.**

- Over 10 EBICS STC goal area programs were evaluated. This limited evaluation resources dedicated to the REU program.

Surprisingly, one of the aspects of the EBICS STC evaluation that posed the largest challenge for implementing the VEE approach was the complexity and depth of Center programming. Over the course of the past four years, the evaluation team has evaluated/written evaluation reports on the following aspects of EBICS: graduate teaching consortium, high school research experience, graduate student experience, graduate student curriculum, minority-serving institution partner experience, industrial advisor committee, trainee research symposium, annual retreat, ethics module, graduate teaching consortium, and the research experience for undergraduate program. Similar to the REU program, each of the previous programs/areas of EBICS was distributed across the Center. Additionally, most of the EBICS STC program evaluations took place concurrently.

While the focus of this study has been the EBICS REU program, it is important to note that the evaluation took place concurrently with many other STC goal area program evaluations. As such, evaluators had to split their time and resources not only by site, but also by multiple evaluation activities for multiple EBICS goal areas. For example, every year the EBICS Annual Retreat was held during the summer at Site A or Site C. This four-day retreat was aimed at providing professional development to graduate student trainees, allowing a space for in-person networking, updates on the EBICS research projects, and in-person meetings with the Industrial
Advisory Committee and External Advisory Committee. Even though the Retreat ran while the REU program was also running, all evaluation activities during that time period were focused on the Retreat.

Table 6 below depicts observations conducted on various program activities.  

Table 6

*Observation of the EBICS STC Program Activities by Year (✓ = completed)*

<table>
<thead>
<tr>
<th>EBICS Goal Area</th>
<th>Program Activity</th>
<th>Year 15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2011</td>
</tr>
<tr>
<td>Education</td>
<td>Graduate Teaching Consortium Courses</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Education Committee Meetings</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Trainee Research Symposums</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Journal Club</td>
<td></td>
</tr>
<tr>
<td>Diversity</td>
<td>Diversity Committee Meetings</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Research Experience for Undergraduates</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Project ENGAGE</td>
<td></td>
</tr>
<tr>
<td>Leadership and Management</td>
<td>Annual Retreats</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Executive Committee Meetings</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>NSF Site Visits</td>
<td>✓</td>
</tr>
<tr>
<td>Ethics</td>
<td>Ethics Module Roll-Outs</td>
<td></td>
</tr>
</tbody>
</table>

14 Tables of all EBICS evaluation data collection activities including surveys, interviews, and focus groups, and a list of evaluation reports submitted to Center leadership can be found in Appendix D.

15 Also indicates future administration years.
The complexity of the context challenged the implementation of this approach because at times there was a tension in deciding which STC programs would receive evaluation resources and when. Evaluators had to determine how to split their time to best serve the overall needs of the Center. While the implementation and evaluation of all programs are crucial to the success of the Center, the REU program was not necessarily the most important program of the Center. For example, the evaluation team presented findings to all Center members at the annual Retreat. It was imperative that the End-of-Year survey and Graduate Teaching Consortium data analysis were completed. As previously mentioned, the Retreat took place during the second week of the REU program. Additionally, the evaluation team supplied program managers with a two-week interim report of how the REU program is running and if there are any concerns they should address. Ensuring that all tasks were complete at times meant that some activities were not observed, or it meant that evaluators outside of the evaluation team were pulled in to assist. However, they had less knowledge of the EBICS REU program than evaluation team members. Thus, sometimes conducting a full VEE evaluation was a juggling act for the evaluation team.

In sum, this case study revealed the following findings.

1. The EBICS evaluation team used the VEE approach as a framework to meaningfully engage the NSF BP agenda. Their efforts were successful.

2. The EBICS evaluation team used the VEE approach as a framework and supplied EBICS stakeholders with valid and useful information to guide program improvement, assess short- and long-term effectiveness, and increase connections to URMs. Relevant information was provided to NSF; however, the value of such information is unknown.

3. EBICS REU program design was altered by the VEE evaluation findings and recommendations. Data on longitudinal program outcomes are currently being gathered.
4. VEE components responsible for evaluation success are primarily related to the development of relationships with program stakeholders and the engagement of their culture and values.

5. Multiple stakeholders with varying power and values, in multiple locations, program time frame, and data collection were challenges associated with the multi-site nature of the EBICS STC REU program evaluation. Additionally, EBICS Center composition and evaluation priorities raised challenges for VEE approach implementation.

Chapter six will discuss findings presented in this chapter, implications for the literature, directions for future research, as well as final thoughts.
CHAPTER 6

DISCUSSION: RESPECTFUL PATIENCE

In this chapter I discuss four major findings related to successful engagement of BP including: (a) evaluation team’s entanglement with program planning, (b) the VEE approach and being educative about evaluation, (c) formal data collection on URM experience, and (d) respectful patience and values-engagement. Additional implications for the literature, directions for future research, and final thoughts are also presented.

As presented in chapter five, evaluators attempted to include NSF’s agenda of broadening participation in the evaluation on a significant level by bringing issues related to the recruitment and retention of underrepresented minorities in STEM to the forefront of the evaluation. Their efforts were successful for a number of reasons. Five major takeaways for evaluators about a prescriptive approach to valuing and the VEE approach:

- Deep and sustained engagement with stakeholders and the program, including its culture and values, is essential.
- Teaching stakeholders about the purpose and importance of evaluation is necessary.
- Communicating informally and formally with stakeholders is imperative.
- Evaluators must be respectfully patient when prescribing values.
- The VEE approach may not be able to be fully implemented in a multi-site setting.
Evaluation Team’s Entanglement with Program Planning & Implementation

The VEE approach calls for evaluators to have authority and credibility, show up and spend time at program activities, and develop relationships with stakeholders. Additionally, the VEE approach also suggests that evaluators adopt a membership role by doing small tasks and participating in the life of the sites. VEE guidelines urge evaluators to serve as “teacher, critical friend, and co-learner” (Greene, Boyce, & Ahn, 2011, p. 14). Finally, the VEE guidebooks points to the necessity for the VEE evaluator to have adequate access to program materials, activities, and personal, including key decision makers. These recommendations can lead to blurred lines between evaluator role and participation in program decision-making conversations.

In the EBICS STC REU program evaluation, fuzzy boundaries accounted for much of the implementation of the VEE approach’s success. From the start of the Center, evaluators were present for all major conversations, meetings, emails, and ultimately decisions made about the REU program. While evaluators were not decision makers, they did play an important role in the decision making process. As the Center grew, all stakeholders increasingly relied on evaluator input and evaluation data, reports, and recommendations from evaluation reports and memorandums. This was especially true after the first two site visits from NSF where evaluators played a crucial role in supplying information requested by NSF. For example, every year all Center goal areas, except research, use evaluation data for NSF site visit presentations.

These circumstances reiterate the importance of evaluators having presence, credibility, and relationships with stakeholders, and ultimately decision makers. The evaluation’s success with prescribing the values of equity and diversity and engagement with the NSF BP agenda are directly related to these blurred lines. Even though the evaluators heavily participated in all aspects of the Center, they still produced evaluation claims that were objective and valid but
included both values and facts. The claims were more valid and useful because of evaluators’ contextual and cultural knowledge (Frierson, et al., 2010); and ultimately the evaluators’ recommendations were more valuable and useable by Center leadership and program managers because of the evaluators’ understanding of Center dynamics, power structure, and decision making processes. So while initially it was difficult for evaluators to engage with BP and even though the Center was spread across three sites, this context, while challenging, presented a close to ideal (for a multi-site evaluation) situation for evaluators to practice prescribing and describing values. A major take-away from this study is the importance of deep and sustained engagement with program stakeholders and the program’s culture and values. Essentially, this will play out in the form of blurred lines between program development and evaluation.

**VEE Approach and Being Educative About Evaluation**

The VEE approach advises evaluators to be educative, specifically calling evaluators to assist stakeholders in understanding the logic of connections among program resources, activities, and outcomes. One implicit aspect of the approach is the evaluator’s duty to educate stakeholders about evaluation itself; or rather to cultivate an intelligent belief in evaluation (Schwandt, 2008). While the approach doesn’t actually ask evaluators to do this, in my estimation the next iteration of the VEE guidebook should include such endeavors.

In the EBICS STC REU evaluation the majority of stakeholders, including decision makers were unfamiliar with formal program evaluation. They were even less familiar with evaluators and their role or had only experienced distant external evaluators. A critical piece to the evaluation team’s implementation of the VEE and engagement with BP was in their development of stakeholders understanding of evaluation itself. So while evaluators spent time
educating program managers, goal area directors and Center leadership about logic models and program theory, they also purposefully taught about evaluation. This was done by taking time during evaluation meetings to explain and critique the evaluation purpose, audience, questions, criteria for judging quality, data collection instruments, and results. Again, this education was possible because evaluators had considerable access to stakeholders.

Subsequently, as stakeholders better understood the purpose and power of evaluation they began to utilize the evaluation team more. They asked for evaluations of additional Center programs, and required the evaluation team to present at site level and Center level meetings. Therefore, not just decision makers developed an understanding of evaluation but also other stakeholders including Center faculty and graduate students. While there are often numerous research presentations, time was always made for evaluation presentations. This increase in understanding also assisted evaluators with their prescription of values because stakeholders were more able to see the linkages between evaluation data and program outcomes and potential program outcomes. Therefore, another take away from this study is teaching stakeholders, especially those unfamiliar with evaluation, about the purpose and importance of evaluation is an important strategy to support other educative efforts and assist with endeavors to increase evaluator credibility.

**Formal Data Collection on URM Experience**

The majority of engagement with BP took place through conversations among evaluators and stakeholders. This was the case for two reasons. First, the NSF requires REU programs to report on a number of items, including participant diversity and demographic information. However, there are no specific criteria or targets for what those numbers should look like. As
such, the evaluation team initially had little foundation to support such targets in the criteria for judging quality. Second, all stakeholders were not of the same mind when it came to REU selection criteria. The VEE approach directs evaluators to create opportunities for dialogue, which helps evaluators to describe stakeholder values. However, the prescription of values is not intended to be forced, but rather evaluators should nudge movement on such issues. Thus, as noted earlier evaluators worked to ensure that URMs had access to this program. Without such efforts, there would be little need or possibility to collect information on URM experience within the program.

With respect to formal data collection on URM REU experience, the evaluation team worked with program managers and the diversity director to ensure that all REU participants experienced the program in a safe, meaningfully, positive, and beneficial way. As previously mentioned, in the summer of 2013 evaluators went to all three sites and met with all REU participants face-to-face to better understand their experience and get to know them and their values. Many informal conversations were had with URM participants and subsequently conversations about their experience were had with stakeholders. Evaluators did not include such data in evaluation reports because of the confounding relationship with participant ethnicity and site. Up until the summer of 2014, if evaluators reported on the experience of Latinos or Blacks, it would be easy to guess who quotes and survey results belonged to and reporting such data in a disaggregated form could have ultimately hindered institutional review board confidentiality commitments. Further, these results often would come from one site and the evaluation team attempted to report on results as a Center not three individual, unconnected sites. More thought is needed on how to best formalize such efforts into the evaluation. However, evaluators could have thought better about how to include the experience of STEM URM women in evaluation
reports. Although data revealed women were having a similar experience to male participants, that could have been formally noted in evaluation reports. A third take away is the importance of communicating informally and formally with stakeholders about diversity and equity.

**Respectful Patience, Values-Engagement and Evaluators as the Conscience of the NSP BP Agenda**

The EBICS REU program took place over multiple years with multiple REU program iterations. This study revealed that patience assisted with the evaluation team’s engagement with BP. Additionally, time was a factor the evaluation team had on their side for the prescription of equity and diversity. The VEE approach calls for evaluators to be respectful of norms, values, and ideals within the Center and of individual stakeholders. As such, evaluators gently and patiently engaged with BP year by year with increasingly positive results. As mentioned earlier, evaluators cannot expect to come into a context and ‘beat stakeholders over the head’ with prescription of values. There is a need for evaluators to have patience while cultivating relationships, building credibility, and teaching stakeholder about evaluation.

In the EBICS REU evaluation as relationships and credibility increased so did the evaluation team’s ability to engage with BP. Evaluators were strategic about their engagement with BP and nudged stakeholders every year to think well about these issues. Such engagement would not have been possible or would have been limited if evaluators had not been patient. The lead evaluator commented:

I think with these long term projects you’re in this for the long haul and you just have to time it. You just have to figure out “when am I in a position where I can actually affect change?” You can’t affect change every minute of every day so you have to try to get the moment where you have the biggest opportunity of making an impact and maybe it is now [summer 2014] with that. (Lead evaluator)
The fourth takeaway for evaluators in practice is the need for respectful patience when prescribing values. In addition to takeaways for practice, this study also has implications for evaluation literature.

**Multi-site Evaluation and the VEE Approach**

In the EBICS evaluation, evaluators committed to a high onsite and virtual presence, which in some ways negated the challenges of working across three sites. Some of the best practices related to multi-site evaluation were attending as many virtual and on-site meetings as possible, developing relationships with stakeholder, and relying heavily on video conferencing software. However, as previously mentioned, one of the most difficult aspects of implementing a VEE and MSE evaluation concurrently was during the data collection phase and deciding which sites to visit when. Evaluators worked with program managers at each site to determine data collection priorities. While sites were geographically dispersed, there were only three primary sites. The implementation of the VEE approach of a multi-site evaluation with many more sites would be difficult and ultimately would place stress on implementing the VEE approach with fidelity. Additional multi-site VEE evaluation case studies are needed to further email this potential tension.

**Implications of Findings**

This study is an empirical example of “how evaluators can contribute to the goal of improving STEM outcomes for underrepresented groups” (Mertens & Hopson, 2006, p. 36) by attending to culture, equity, and diversity. It is still unknown how the prescription of these values in this study ultimately yielded benefits for STEM and society. However, it is clear that the
EBICS REU evaluation directly influenced who had access to the REU program. As mentioned in chapter five, the summer 2014 iteration of the EBICS REU program is comprised of all URM in STEM. This is in part due to evaluators continuing to nudge stakeholders to think about the purpose and goals of the REU program and having conversations about the NSF BP agenda. Additionally, educating the diversity director about the importance of evaluation led to a pseudo-partnership where he requested information and support from the evaluation team for a proposal to EBICS leadership to argue for additional attendance to diversity and equity within the REU program, especially in recruitment and selection.

While not fully discussed in this paper this study has implications for the STEM evaluation literature as well. STEM educational programming continues to be an important focus for the 21st century (Katzenmeyer & Lawrenz, 2006). As such, reflections on STEM evaluation best practices are necessary. Introducing EBICS STEM researchers to evaluation, the multi-site context, and differing values made the evaluation context complex. Sharing lessons learned from the first four years of evaluating the EBICS STC could be useful for many evaluators. Some of the best practices that were unearthed in this study that could be covered include formative rapid feedback, summative longitudinal tracking, respectful patience, engaging with the STEM culture and stakeholder values, embedding the evaluation deeply into the STC, and building relationships with Center PIs, researchers, and trainees.

**Future Research**

The evaluation in this study was primarily formative in nature. As EBICS moves into its fifth year there are a number of opportunities for further study in the engagement of values vein. The VEE evaluation team itself has shifted to a summative focus. This presents the opportunity
to study how and in what ways the VEE approach impacts summative evaluation in this context. A research question would be: What difference does the VEE approach make in summative evaluation claims about program value in multiple contexts?

In the EBICS evaluation some attention was given to developing stakeholders’ understanding of the logic, assumptions and theory associated with their program. However, more focus on this strategy is needed. Because of the complex nature of the Center, an interesting study would examine how to develop program theories for each of the EBICS education, diversity, and outreach programs, with overarching theories for each goal area. Developing an understanding of what that looks like might be worthwhile to answer the question: What is the potential for the VEE approach to shape and impact program theory in multiple evaluation contexts?

While the NSF recommends sites to recruit REU participants who are women, African-American, Latino/a, have a disability, or come from an institution with limited research capacity, target numbers for these groups are not explicit. As such, it would be fascinating to explore what type of study would be needed to influence NSF policies. The main research question for that study would be: How and in what ways can the VEE approach impact NSF policies on evaluation?

Finally, this and other studies have helped shape thinking about the VEE approach. A critical next step important to the evolution of the approach is to continue to refine the approach based on empirical evidence and theoretical thinking. This refinement should be reflected in a second iteration of the VEE guidebook.
Concluding thoughts

Some might argue that the VEE approach is simply ordinary good evaluation. However, this study has highlighted four aspects of VEE that make it distinctive and especially relevant for STEM BP evaluations. First, the VEE approach is committed to the prescription of equity and diversity. While the NSF BP context is well suited for these commitments, the VEE evaluator is called to attend to issues of equity and diversity in all evaluation contexts. Second, the VEE approach foregrounds relationships with stakeholders. Evaluators are called to ensure that they have access and time to develop these relationships. Third, The VEE approach defines quality at the intersection of content, pedagogy, and equity. Many STEM evaluations will attend to two out of three of those criteria. Fourth, VEE evaluators are committed to being educative. Not only attempting to educate stakeholders about their program, but about evaluation as well. Finally, in order to enact the VEE approach with fidelity, evaluators must recognize that time and resources are needed. A full scale VEE evaluation is not necessarily well suited for short evaluations or evaluations with limited access to stakeholders.

Julnes (2012) has called for evaluators to develop frameworks for valuation that are more explicit about strengths and limitations of different approaches in different contexts. This study responds to that call by filling the gap in the literature about the VEE approach in complex STEM settings. It also provides an idea of what an evaluation with the prescription of the values diversity and equity looks like. Challenges and successes of the VEE approach have been recorded and reported in this case study. The EBICS REU program evaluation serves as an exemplar of a values-explicit evaluation where both stakeholders and evaluators are doing the valuing.
In the end, this study importantly continues a conversation started decades ago and highlights the importance for thoughtful and reflective evaluations, and studies on evaluation. Evaluators have come to realize that there are value-laden implications of our work. Therefore, evaluators must attend to the social, ethical, political, cultural, and value dimensions of the evaluation context, stakeholders, and the evaluand. Whether prescriptive or descriptive, it is my hope that this study challenges and encourages evaluators to be explicit about the values their work advances.
REFERENCES


APPENDIX A: EXPANDED LITERATURE REVIEWS

A.1: STEM, the U.S. Workforce and the STEM Pipeline Literature Review

Science Technology Engineering Mathematics and the United States

Politicians, researchers, and economists alike assert that the United States should continue to focus on scientific innovation. The world increasingly relies on science and technology to solve some of today’s most difficult problems and for economic growth (Boskin, M.J. and Lau, L. J., 1992.). As challenges mount in the areas of national defense, climate change, health, energy, economic growth, food safety and accessibility, and environmental protection, so does the demand for highly able scientists, engineers, and health professionals (National Academies Press, 2005, 2010). A focus of the U.S. government has been on global competitiveness in each of the aforementioned areas.

Seven years ago a report entitled *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future* (2007) was written by the National Academics Press as response to a request from U.S. Senators and Members of Congress. These elected officials wanted to know what are the top 10 actions federal policymakers could take to enhance science and technology, so that the United States could successfully be more secure in the 21st century global community. In 2010 in a follow up to the first report, committee members recommended the following (National Academy of Science, 2010, pg. 19):

- Move the United States’ K-12 education system in science and mathematics to a leading position by global standards.
Double the real federal investment in basic research in mathematics, the physical sciences, and engineering over the next seven years (while, at a minimum, maintaining the recently-doubled real spending levels in the biosciences).

Encourage more United States citizens to pursue careers in mathematics, science, and engineering.

Rebuild the competitive ecosystem by introducing reforms in the nation’s tax, patent, immigration and litigation policies.

Overall the 2010 (?) report concluded “that America was in substantial danger of losing its economic leadership position and suffering a concomitant decline of the standard of living of its citizens because of a looming inability to compete for jobs in the global marketplace” (pg. 24). This committee felt it was necessary to increase the STEM talent pool and overall number of persons able to work in science and technology fields. While in this report K-12 education was declared the highest priority, graduate and undergraduate education remain important junctures in the STEM pipeline. Some of the report’s action items focused on higher education included providing competitively-awarded scholarships in engineering, science, and mathematics; graduate fellowships; tax credits for STEM employers who support continuing education; and changed immigration laws with respect to those who seek to and earn STEM Ph.Ds.

**STEM Pipeline**

**A Leaky Pipe**

In 2009, 20.7 million students were enrolled in U.S. institutions of higher education, which is 6.2 million higher than in 1994 (NSF, 2012). This increase in enrollment has been
attributed to increased enrollment of minority students, especially Latinos. Additionally, the
number of undergraduate degrees awarded by U.S. institutions has increased in the past 20 years
and is expected to continue this trend until at least 2019. Twenty-three percent of the 3.1 million
Associate’s, Bachelor’s, Master’s, and Doctoral degrees were in science and engineering (NSF, 2012).
While these trends are encouraging, the fact remains that over twenty percent of
undergraduate students who enroll in science and engineering majors never complete their
degree. And this is just one junction in the leaky STEM pipeline. In terms of the large picture, of
the four million ninth graders in 2001, only 167,000 or about 4 percent, were projected to
graduate college with a Science Technology Engineering and Mathematics (STEM) degree in
2011 (NCES Digest of Education Statistics, 2008). See Figure 11.

![Figure 11. STEM Pipeline](image)

**Underrepresented Minority Students: A Leakier Pipe**

The proportion of science and engineering higher education degrees awarded to Latinos
and Asians has risen between 2000 and 2009 (NSF, 2012), and the shares of degrees award to
Blacks and American Indians has remained the same. Despite some progress, there still remains
a large educational attainment gap among the major US ethnic and racial groups. For example,
in 2009 the percentage of 25–29 year olds with bachelor’s or higher degrees was nineteen percent or blacks, twelve percent for Hispanics, and thirty seven for whites (NSF, 2012). Overall, Blacks, Latinos, and American Indians are less likely than Whites to attend college or to graduate. However, for those who do graduate, the degree patterns are similar to those of Whites.

While the latest reports show overall undergraduate degree attainment for men and women has shifted with women earning 58% of bachelor’s degrees in 2000, the gender gap in STEM degree attainment remains. This is especially true of engineering, physical and computer science degrees. Finally, persons with disabilities are also seriously underrepresented in the science and engineering workforce compared to the population as a whole (NCSES, 2013).

Why Focus on Underrepresented Minority Students and STEM?

The U.S. Census Bureau (2008) projects racial and ethnic minorities will make up more than half the national population by 2050 with: the proportions of Latinos and Asian Americans increasing; those of African American and American Indian populations staying the same; and those of Americans of European descent decreasing. Additionally, underrepresented minorities have a higher rate of departure from and failure in STEM education than White and Asian Americans (ASHE, 2011). So, with increasing diversification of the potential STEM student body and continued high rates of departure from the STEM education pipeline by students from groups underrepresented in STEM, we urgently need to better understand how to increase success among all students in general and minority students in STEM in particular. Considering the previous, efforts are being made to increase the success of underrepresented minorities, as well as women and persons with disabilities, in the STEM fields because such success can positively impact their financial well-being (Baum & Payea, 2005; Choy & Li, 2005; Kelly, 2005) and will also assist with keeping the U.S. competitive in the global marketplace.
Finally, the Association for the Study of Higher Education (ASHE) (2011, pg. 5) asserts that the success of underrepresented students in STEM is “a moral and ethical imperative and … by virtually every indicator and at every level of education, racial and ethnic minority students suffer from persisting systemic inequities in the STEM circuit.”

National Science Foundation’s Broadening Participation Agenda

For decades, the National Science Foundation (NSF) has spearheaded the federal government’s efforts to enhance the quality of STEM education and to increase the diversity of STEM communities via a variety of “broadening participation” (BP) initiatives. In 2012 NSF spent $910.90 million dollars on broadening participant efforts (CEOSE, 2012). Just over 16% of these funds were spent on the Academic Science and Engineering Careers (ADVANCE), Historically Black Colleges and Universities Undergraduate Program (HBCU-UP), and Louis Stokes Alliances for Minority Participation (LSAMP) programs. Additionally, $57.28 million dollars was spent on Research Experiences for Undergraduates (REU) sites and supplements.

Broadening participation efforts are advised by The Committee on Equal Opportunities in Science and Engineering (CEOSE). Established in 1980, the CEOSE provides advice to NSF on policies and activities to encourage participation from women, underrepresented minorities and persons with disabilities. The CEOSE is composed of 15 individuals from STEM disciplines, drawn from institutions in higher education, industry, government, and the non-profit sectors who typically serve a three-year term.

The committee provides a report to the NSF director every two years. In the 2012 report, committee members expressed that NSF needed to enhance their BP efforts in science and
engineering education and employment. The main recommendation in this biennial report was to encourage:

NSF [to] implement a bold new initiative, focused on broadening participation of underrepresented groups in STEM, similar in concept and scale to NSF’s centers, that emphasizes institutional transformation and system change; collects and makes accessible longitudinal data; defines clear benchmarks for success; supports the translation, replication and expansion of successful broadening participation efforts; and provides significant financial support to individuals who represent the very broadened participation that we seek (CEOSE, 2012 pg.V).

The committee asserts that the ‘boldness’ in their recommended initiative is a recommendation that NSF make a long term commitment to sufficient resources, including new resources, to improve STEM employment. The committee also noted that while NSF has been successful at recruiting members of underrepresented groups to STEM the “specifics and degree of success remain unclear” because of the lack of evaluation and analysis of systematically collected and longitudinal data (CEOSE, 2012 pg.18).

**NSF BP Program Evaluation**

The Education and Human Resources Directorate (EHR) is primarily responsible for conducting program evaluation at NSF. Evaluation at NSF began in the 1960’s with the launch of Sputnik, the first artificial Earth satellite, and was significantly amplified in 1993 with the passing of the Governmental Results and Performance Act, which emphasized evaluation for federal agencies (Katzenmeyer & Lawrenz, 2006). Evaluation at NSF currently takes one of two forms: status studies (state of the world surveys) or program evaluation. STEM program
evaluation at NSF exists at the individual site or ‘project’ level and at the national or ‘program’ level. Outside educational research organizations and evaluation firms such as Abt Associates often conduct the program level evaluations. However, evaluation at the individual site level varies greatly. One example, the REU program, will be discussed below.
APPENDIX A.2: MULTI-SITE EVALUATION LITERATURE

A mere twenty years ago, literature on program evaluations implemented across multiple locations was scarce. However, these types of evaluations, often funded by government agencies or large foundations, are increasingly showing up in the evaluation field landscape. Turpin and Sinacore (1991) introduced the field to the term, multisite evaluation (MSE), and provided the first guide for those planning and conducting evaluations across multiple sites in their New Directions for Evaluation (NDE) volume. Since then, two additional volumes (Herrell & Straw, 2002; King and Lawrenz, 2011) and a multitude of papers with an MSE focus have been published. The proliferation of knowledge on MSEs suggests the evaluation community’s commitment to better understanding the benefits, challenges, and distinctive issues posed in such contexts. Aside from MSE findings, much of MSE literature has focused on identifying challenges to and suggestions for conducting high-quality, valid, and useful MSEs.

There are two main factors that differentiate MSEs from other types of evaluation activities: (a) the program evaluation must include multiple sites, two or more, and (b) evaluation activities must be conducted across those sites to answer a set of evaluation questions (Sinacore & Turpin, 1991; Springer, 2000).

Multi-Site Evaluation Sites

MSE is the evaluation of a program that is situated in more than one location. The number of evaluation sites can range from two to 20, or even more, and the locations of MSE sites can vary geographically. For example, the sites of an MSE of a local after-school program could be relatively close to one another. Perhaps all sites are in the same district and, therefore, would likely be in the same city. In an MSE of a national teacher professional development initiative, each site could be hundreds of miles away from one another. The main point is that each site is a part of the program being evaluated. It should be noted that because there are
multiple sites, the implementation of the program may vary slightly or greatly from site to site (Sinacore & Turpin, 1991). Differences in staff, participant population, local culture and norms, and a host of other factors contribute to variation in program implementation.

**Multi-Site Evaluation Activities**

In MSE, evaluation activities (collection of data, reporting, etc.) are being conducted across multiple sites. It may not be feasible to conduct all evaluation activities at all sites; however, each site included within the evaluation receives or participates in some activities. For example, in the MSE of a graduate student training program, some of the sites may be observed, while others, due to resource constraints, are not observed. However, it would be likely that all graduate student participants across sites would receive surveys or content knowledge assessments.

**Types of Multi-Site Evaluation**

As with any evaluation, the purpose of an MSE could be formative (provide information for program improvement), summative (examining effectiveness or impact), or both. Additionally, the evaluation activities can be retrospective in nature, where evaluators analyze data already collected, or prospective, in which the evaluator is involved in planning, data collection, analyses, and reporting. Turpin and Sinacore (1991) note that retrospective MSE is less common and warn that additional care should be taken when conducting such an evaluation because data on a similar topic are brought together from different evaluations for analysis.

MSEs are often associated with programs that are funded across multiple sites. The data for these types of evaluations are sometimes collected by each individual site or can be collected by one evaluator or an evaluation team. The purpose of this type of MSE varies. In a summative MSE, evaluators are examining the effectiveness of the program and variations of outcomes.
across the sites. In a formative MSE, evaluators seek to provide feedback to each site and the program overall about implementation or initial program effectiveness. An example of this type of evaluation is the MSE of the NSF REU program sponsored by Science Technology Centers.¹

**Distinctive Issues for Evaluator Consideration**

MSEs present distinctive issues for evaluators to consider carefully (Herrell & Straw, 2002; King & Lawrenz, 2011; Turpin & Sinacore, 1991). I have grouped these items for consideration into three categories: (a) engagement of values, (b) methodology, and (c) logistics (Tillman, 2013). I will provide a brief overview of each of these below.

**MSE Engagement of Values**

Disparate values and cultures at each site and the challenge of including multiple legitimate values in the evaluation are MSE issues for consideration related to the engagement of values. Each site is a unique context with its own culture. Further, each program manager or site coordinator has his or her own views, beliefs, and values that are related to the implementation and logistics of the program. Thus, including and attending to multiple values from various stakeholders can be a challenge. This is especially true when these multiple values are conflicting. Evaluators will have to decide whose interests are primary and whose are not, or provide safe spaces for the articulation of varied value commitments.

King et al. (2011) have suggested that multisite evaluators should attempt to involve program staff in the design of the evaluation. Parallel to the logic of Patton’s (2008) utilization focused evaluation, they suggest that soliciting input on evaluation purpose, design, and instruments can increase the program staff’s commitment to the evaluation, which can lead to increased buy-in from stakeholders and thereby enhanced evaluative validity and use. However,

¹ This context will be discussed further in Chapter 3.
they urge the evaluator to have control over the final design and instrumentation of the evaluation.

However, attempts to be inclusive of multiple value stances by increasing stakeholder involvement in MSE is a challenge because the evaluator is often not in the same location as all of the principal investigators (PIs) or program staff. Further, the evaluator will be dealing with multiple people across sites and developing working relationships with each, which takes time. If key staff are not genuinely involved, a number of problems can arise. For example, during data collection, the evaluator may want program staff to remind participants to take a survey. Yet, if staff do not believe the survey is important, of high quality, or worth the participants’ time, or if they do not value quantitative data, they may disregard the evaluator’s request. Additionally, if staff have minimal involvement in the evaluation, there is a chance they will not find the final interpretations and recommendations useful or valid because they might not reflect their values or what they think is important (King & Lawrenz, 2011).

**MSE Methodology**

It is necessary for the methodology of MSE to be different from that of single-site evaluations, as MSE will have distinct evaluation questions that need to be answered. One such question would be: What is the quality of the program implementation and effectiveness at *each site and across* the sites, especially as it relates to the needs of the group of students recruited for this program? The development of evaluation questions, attention to the program’s implementation across sites, data quality, and data analysis strategies are all methodological issues for consideration in MSE (Turpin & Sinacore, 1991). These will each be addressed in turn.

First, the evaluation questions used in MSEs should be such that they could not otherwise be answered without the inclusion of multiple sites in the evaluation (Herrell & Straw, 2002).
This is important for MSE because evaluators should be attending to both site-specific and universal outcomes for the multi-site program. If conducting a formative evaluation, evaluators will also need to decide if they will have site-specific evaluation questions or questions related to the similarities and differences across sites in addition to the universal questions.

Second, when attempting to assess impact and effectiveness of a program across sites, MSE evaluators need to factor in the extent and quality of the program's implementation at each site. Program implementation can vary from site to site, and if conducting a formative evaluation, site-specific feedback can be given to PIs and program managers to assist with improvement. Having an evaluation question about program implementation can guide data collection efforts and help evaluators meaningfully engage with these issues, especially because they will be examining implementation within and across multiple sites. Fidelity and extent of program implementation at each site should be considered to understand the overall program’s quality. However, evaluators must attend to the tension between fidelity of implementation and contextual appropriateness of implementation and find ways to respect both.

Third, ensuring that data are of high quality can be a concern for MSEs because evaluators are not always on site while data are collected. Evaluators often enlist the help of on-site staff or local evaluators to administer surveys, observe events, and/or conduct interviews. Sinacore and Turpin (1991) recommend that multisite evaluators attend to efficiency of data collection, standardization of data collection, standardization of data organization, and data verification to increase MSE data quality. While the previous are necessary, the development of relationships with program staff is absent from this list. This, as I have stated, is important for many aspects of evaluation, both single-site and multi-site evaluations.
Finally, when conducting an MSE, evaluators should develop strategies for data analysis. When analyzing MSE data for formative reports, evaluators should turn around the data as soon as possible, and report on specific items that indicate the program is doing well and items that indicate places for improvement (DeStefano, 2012). Interim reports should provide targeted feedback to each and across sites. When reporting on summative findings, evaluators cannot simply aggregate data, but must report by site and the overall program. Depending on the population size, evaluators will have to be thoughtful about nesting or weighting data (Angus Council, 2012; Nisbet, Elder, & Miner, 2009). Turpin and Sinacore (1991) suggest looking for program-by-site interactions by viewing the site and the program as factors in an analysis of variance framework. More recently, in very large MSEs, evaluators have employed regression discontinuity designs (Lawrenz, 2013). Evaluators are encouraged to review MSE site data both site by site and universally, across sites. Therefore, in the MSE, smaller reports or short memos of findings related to each individual site and comprehensive larger reports are recommended.

**MSE Logistics**

While all evaluations should be thoughtfully and carefully planned, the presence of multiple evaluation sites warrants additional consideration for the logistical planning and implementation of evaluations. Cross-site communication and evaluation coordination and implementation are logistical items in need of supplementary attention in MSE (King & Lawrenz, 2011; Turpin & Sinacore, 1991). For example, it is necessary to continuously communicate with each site’s local program staff and/or principal investigators (PIs) to ensure that all of the planned evaluation activities can be coordinated. This degree of coordination can add additional time, staffing, and financial costs to an evaluation (Straw & Herrell, 2002). The increased number of sites will often mean additional observations, assessments, interviews,
and/or focus groups. Further, due to resource restraints, evaluators may not be able to implement all evaluation activities in each site equally. For example, distance, travel budget, and time constraints may influence the frequency and magnitude of observations each site receives.
APPENDIX B: EBICS REU PROGRAM EVALUATION INSTRUMENTS

B.1: First/Second Week Survey

Which EBICS campus are you at?

<table>
<thead>
<tr>
<th>Campus</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Illinois</td>
<td></td>
</tr>
<tr>
<td>MIT</td>
<td></td>
</tr>
<tr>
<td>GA Tech</td>
<td></td>
</tr>
</tbody>
</table>

Top 3 reasons for participating in the EBICS REU

<table>
<thead>
<tr>
<th>Reason</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Network with students at other institutions</td>
<td></td>
</tr>
<tr>
<td>Network with faculty at other institutions</td>
<td></td>
</tr>
<tr>
<td>Participate in an interdisciplinary research experience</td>
<td></td>
</tr>
<tr>
<td>Participate in outreach and diversity experiences</td>
<td></td>
</tr>
<tr>
<td>Participate in a paid research experience</td>
<td></td>
</tr>
<tr>
<td>Learn new research methods</td>
<td></td>
</tr>
<tr>
<td>Learn about the graduate programs at U of I</td>
<td></td>
</tr>
<tr>
<td>Participate in research focused on Chemistry</td>
<td></td>
</tr>
<tr>
<td>Other*</td>
<td></td>
</tr>
</tbody>
</table>
Other?

- Location, Graduate School Information

**Why did you choose to participate in an REU at this university?**

<table>
<thead>
<tr>
<th>Reason</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reputation of the school/department</td>
<td></td>
</tr>
<tr>
<td>Reputation of my research mentor</td>
<td></td>
</tr>
<tr>
<td>Diversity of the campus</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>This university is one of my top three choices to go to graduate school</td>
<td></td>
</tr>
<tr>
<td>My undergraduate mentor/professor encouraged me to go</td>
<td></td>
</tr>
<tr>
<td>Other*</td>
<td></td>
</tr>
</tbody>
</table>

Other?

- Wanting to be involved with the Biobot project

**What do you plan to do after you graduate?**

<table>
<thead>
<tr>
<th>Plan</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Full time</td>
<td></td>
</tr>
<tr>
<td>Graduate school (MA or PhD)</td>
<td></td>
</tr>
<tr>
<td>Professional School (medical, law, etc.)</td>
<td></td>
</tr>
<tr>
<td>Study/work abroad</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Serve in the Armed Forces</td>
<td></td>
</tr>
<tr>
<td>I am not sure</td>
<td></td>
</tr>
</tbody>
</table>

**Have you been in contact with your mentor or faculty advisor prior to arriving? If so, in what manner?**

<table>
<thead>
<tr>
<th>Not in contact</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Email</td>
<td></td>
</tr>
<tr>
<td>Phone</td>
<td></td>
</tr>
<tr>
<td>Skype</td>
<td></td>
</tr>
<tr>
<td>Conference</td>
<td></td>
</tr>
<tr>
<td>In person</td>
<td></td>
</tr>
</tbody>
</table>
Please rate each of your abilities:

<table>
<thead>
<tr>
<th>Skill</th>
<th>Very Weak</th>
<th>Weak</th>
<th>Average</th>
<th>Strong</th>
<th>Very Strong</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read, interpret and use research information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop a professional network</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicate across disciplines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicate within my discipline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work independently</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Write a concise abstract</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Give an oral presentation using PowerPoint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Give an oral presentation without slides</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill</td>
<td>Very Weak</td>
<td>Weak</td>
<td>Average</td>
<td>Strong</td>
<td>Very Strong</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>-----------</td>
<td>------</td>
<td>---------</td>
<td>--------</td>
<td>-------------</td>
<td>------</td>
<td>----</td>
</tr>
<tr>
<td>Create a research poster</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Applying for graduate school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leadership skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpersonal (social) skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work effectively with others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manage my time effectively</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpret research data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**What tasks or activities do you expect to do as part of your summer research experience?**

<table>
<thead>
<tr>
<th>Activity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Defining a research problem</td>
<td></td>
</tr>
<tr>
<td>Developing a research hypothesis</td>
<td></td>
</tr>
<tr>
<td>Designing an experiment</td>
<td></td>
</tr>
<tr>
<td>Gathering data</td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Analyzing data</td>
<td></td>
</tr>
<tr>
<td>Modeling data</td>
<td></td>
</tr>
<tr>
<td>Writing about research process/results</td>
<td></td>
</tr>
<tr>
<td>Presenting research process/results</td>
<td></td>
</tr>
<tr>
<td>Working in a research group or team</td>
<td></td>
</tr>
<tr>
<td>Work in a laboratory setting</td>
<td></td>
</tr>
<tr>
<td>Work in an office setting</td>
<td></td>
</tr>
<tr>
<td>Field work outside of a laboratory or office setting</td>
<td></td>
</tr>
<tr>
<td>Developing/using databases</td>
<td></td>
</tr>
<tr>
<td>Developing/using spreadsheets</td>
<td></td>
</tr>
<tr>
<td>Developing computer programs</td>
<td></td>
</tr>
<tr>
<td>Computational analysis</td>
<td></td>
</tr>
<tr>
<td>Statistical analysis</td>
<td></td>
</tr>
</tbody>
</table>
Reflecting upon your first 1-2 weeks during this REU experience, please rate your satisfaction with:

<table>
<thead>
<tr>
<th>Skill</th>
<th>Very Satisfied</th>
<th>Very Dissatisfied</th>
<th>Neutral</th>
<th>Satisfied</th>
<th>Very Satisfied</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lodging accommodations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stipend</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EBICS Campus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction with other REU participants on your campus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction with faculty advisor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction with graduate student/postdoctoral mentor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research project/lab assignment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional development/supplemental programs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Networking Opportunities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Outings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication with EBICS REU Staff</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Additional comments**
**On average how often do you expect to communicate with your Faculty Advisor?**

<table>
<thead>
<tr>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hourly</td>
</tr>
<tr>
<td>Daily</td>
</tr>
<tr>
<td>Weekly</td>
</tr>
<tr>
<td>Not at all</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

**On average how often do you expect to communicate with your Graduate Student / Post Doc Mentor?**

<table>
<thead>
<tr>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hourly</td>
</tr>
<tr>
<td>Daily</td>
</tr>
<tr>
<td>Weekly</td>
</tr>
<tr>
<td>Not at all</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

**What ways do you expect to communicate with your Faculty Advisor?**

<table>
<thead>
<tr>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>In person</td>
</tr>
<tr>
<td>Email</td>
</tr>
<tr>
<td>Phone</td>
</tr>
</tbody>
</table>
What ways do you expect to communicate with your Graduate Student / Post-Doc Mentor?

<table>
<thead>
<tr>
<th>Method</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>In person</td>
<td></td>
</tr>
<tr>
<td>Email</td>
<td></td>
</tr>
<tr>
<td>Phone</td>
<td></td>
</tr>
<tr>
<td>Skype</td>
<td></td>
</tr>
</tbody>
</table>

In what ways do you think your relationship with your graduate mentor and faculty advisor will be beneficial to you?

What knowledge or skills do you hope to gain from your summer research experience with this REU?

If you were not participating in this research experience, what would you be doing this summer instead?

What goals have you set for yourself this summer in relationship to EBICS REU?

What upcoming EBICS REU activities are your most excited about?

Tell us a little bit about your overall experience with EBICS REU thus far.

How and in what ways is your EBICS REU project a match for your future research or career interests?

Additional Comments?
Final REU Survey All Campuses

Please rate each of your abilities.

<table>
<thead>
<tr>
<th></th>
<th>Very Weak</th>
<th>Weak</th>
<th>Neutral</th>
<th>Strong</th>
<th>Very Strong</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read, interpret and use research information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop a professional network</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicate across disciplines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicate within my discipline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work independently</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Write a concise abstract

Give an oral presentation using PowerPoint

Give an oral presentation without slides

Create a research poster

Applying for graduate school

Leadership skills

Interpersonal (social) skills

Work effectively with others

Manage my time effectively

Interpret research data

<table>
<thead>
<tr>
<th>What tasks or activities did you actually do as part of your summer research experience?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defining a research problem</td>
</tr>
<tr>
<td>Developing a research hypothesis</td>
</tr>
<tr>
<td>Designing an experiment</td>
</tr>
<tr>
<td>Gathering data</td>
</tr>
<tr>
<td>Analyzing data</td>
</tr>
<tr>
<td>Modeling data</td>
</tr>
<tr>
<td>Activity</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Writing about research process/results</td>
</tr>
<tr>
<td>Presenting research process/results</td>
</tr>
<tr>
<td>Working in a research group or team</td>
</tr>
<tr>
<td>Work in a laboratory setting</td>
</tr>
<tr>
<td>Work in an office setting</td>
</tr>
<tr>
<td>Field work outside of a laboratory or office setting</td>
</tr>
<tr>
<td>Developing/using databases</td>
</tr>
<tr>
<td>Developing/using spreadsheets</td>
</tr>
<tr>
<td>Developing computer programs</td>
</tr>
<tr>
<td>Computational analysis</td>
</tr>
<tr>
<td>Statistical analysis</td>
</tr>
</tbody>
</table>
Please rate your satisfaction with:

<table>
<thead>
<tr>
<th></th>
<th>Very Dissatisfied</th>
<th>Dissatisfied</th>
<th>Neutral</th>
<th>Satisfied</th>
<th>Very Satisfied</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lodging accommodations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stipend</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Your Campus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction with other REU participants on your campus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction with faculty advisor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction with graduate student/postdoctoral mentor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research project/lab assignment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional development/supplemental programs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Networking Opportunities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Outings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
If you chose very dissatisfied or dissatisfied, please explain:

- 

<table>
<thead>
<tr>
<th>On average how often did you communicate with your REU Faculty Advisor?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hourly</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

- 

<table>
<thead>
<tr>
<th>On average how often did you communicate with your REU Graduate Student Mentor?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hourly</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

What ways did you communicate with your REU Faculty Advisor?

<table>
<thead>
<tr>
<th>In person</th>
<th>Email</th>
<th>Phone</th>
<th>Skype</th>
<th>Other*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What ways did you communicate with your REU Graduate Student Mentor?

<table>
<thead>
<tr>
<th>Method</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>In person</td>
<td></td>
</tr>
<tr>
<td>Email</td>
<td></td>
</tr>
<tr>
<td>Phone</td>
<td></td>
</tr>
<tr>
<td>Skype</td>
<td></td>
</tr>
<tr>
<td>Other*</td>
<td></td>
</tr>
</tbody>
</table>

How has this experience affected you as a researcher?

What was the most important thing you learned from this REU experience?

Which professional development activities were the most useful? Why?

Which professional development activities were the least useful? Why?

What were some of the strengths of this REU program?

What were some things that can be improved for next year?

Additional Comments
### B.2. REU Faculty Advisor/Graduate Student Mentor Survey

<table>
<thead>
<tr>
<th>Extent to which you agree</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel that my REU student understood the research project</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am satisfied with my REU student's poster</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am satisfied with my REU student's final paper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am satisfied with my REU student's final presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working with undergraduate researchers was a rewarding experience.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statement</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>----------</td>
<td>---------</td>
<td>-------</td>
<td>----------------</td>
</tr>
<tr>
<td>Working with undergraduate researchers positively contributed to my research agenda/process.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I developed a mentoring relationship with the student(s) that will last beyond the summer experience.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate your satisfaction with</td>
<td>Very Dissatisfied</td>
<td>Dissatisfied</td>
<td>Neutral</td>
<td>Satisfied</td>
<td>Very Satisfied</td>
</tr>
<tr>
<td>----------------------------------------------------------------</td>
<td>-------------------</td>
<td>--------------</td>
<td>---------</td>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td>Selection and placement of student in your lab</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall REU program structure and organization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amount and type of communication between you and your Program Managers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Your understanding of the program goals and expectations for the faculty, student and mentor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The quality of your student's final presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
On average how often did you communicate with your REU Student?

<table>
<thead>
<tr>
<th>Once a month</th>
<th>Once every 2 weeks</th>
<th>One a week</th>
<th>2-4 times a week</th>
<th>Other</th>
</tr>
</thead>
</table>

What ways did you communicate with your REU Participant?

<table>
<thead>
<tr>
<th>Email</th>
<th>In person</th>
<th>Phone</th>
<th>Other</th>
</tr>
</thead>
</table>

Rate the level of your REU student’s engagement in your lab

<table>
<thead>
<tr>
<th>Not Engaged</th>
<th>Slightly Engaged</th>
<th>Engaged</th>
<th>High Engaged</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
</table>

I would like to participate as a REU research mentor next year.
What are some of the content knowledge and skills your REU student learned this summer.

What are some of the strengths of this REU program?

What are some things to improve for next summer’s REU program?

Any other comments?
B.3: EBICS REU Focus Group Questions

Introduction

- Introduce self and thank them for their participation
- Restate the interview purpose, context, and intended uses of the information to be gathered

Hello my name is______________________. I am a member of the evaluation team from the University of Illinois that is conducting an evaluation of the EBICS REU Program Evaluation. (reiterate group interview confidentiality, permission to audio record) Thank you for agreeing to be interviewed today. The purpose of this group interview is to gather information about your experiences.

Ice breaker

1) To begin, how is your summer going thus far? Has anything interesting happened this week?

REU participant experience

2) Tell us a little bit about your experience with the EBICS REU.
   a. Professional development activities
   b. Social activities
   c. Laboratory experience
   d. On campus housing
   e. New techniques learned

3) What does a typical day related to EBICS look like?

4) Tell us about your relationship with your EBICS Grad Student/ Post Doc mentor.
   a. How often meet
   b. Laboratory setting
   c. Other students in the lab
5) What aspects of EBICS are you most pleased with? Give an example.

6) What aspects of EBICS are you least pleased with? Give an example.

7) What are some of the strengths of EBICS? What could be improved?

**Closing**

8) Please feel free to share anything else important to you that we have not covered yet.

9) Before we end today’s interview, in one or two sentences please describe your overall experience with the EBICS REU program?
APPENDIX C: CASE STUDY INSTRUMENTS

C.1: Key Stakeholder Interview Protocol

1) Are you familiar with NSF’s BP agenda? (Chatted about that)

2) Let’s begin by talking in general about the evaluation. Tell me a little bit about what you know about the REU evaluation.
   
   a. What would you say is the purpose of the evaluation?
   
   b. What kind of activities occurred?

3) What information of value has the REU evaluation generated?
   
   a. What needs did you expect to have met by the evaluation?
      
      i. To what extent were they met?
      
      ii. What were some short comings?
   
   b. What about information of value regarding the program’s connection to Underrepresented minorities?

4) In your recollection, how has the evaluation affected the REU program design?

5) What information has the evaluation generated related to program impact?

6) This program is located is multiple sites. Let’s chat about that briefly.
   
   a. How do you think the MS design has affected the program?
   
   b. What about the impact on the evaluation?
      
      i. How much attention did your site receive versus other sites?

7) Additional comments.
C.2: Structured Reflection Prompts

• List all evaluation activities including planning, meetings, and data collection that have or should have occurred in the past week.
  
  o For each activity rate your perception of how well each activity has been implemented:
    
     1: did not get implemented
     2: implemented but with difficulty
     3: implemented without difficulty
     4: implementation went very well
    
     After each rating write why you rated the implementation of the evaluation activity the way you did and what factors contributed to these ratings?

• Thinking about the implementation of this evaluation this past week, across multiple sites, explain if and how you have dealt with: logistics, methodology, and engagement of values.

• How and in what ways did you engage with issues of diversity and equity this past week?

• Thinking about the evaluation activities that went particularly well this week, what are your thoughts about why they went well?

• Thinking about the evaluation activities that did not go well this week, what are your thoughts about why they did not go well? What could have been done differently?

• How did the multiple sites influence the VEE evaluation activities for this week?
D.1: EBICS STC Evaluation Data Sources

Data Sources

Observations

Observations were conducted throughout the first three years of the EBICS STC. Program activities observed are listed in Table 1 below.

The purpose of these observations was to generate data to describe sessions, engagement between faculty/trainees with each other and with the sessions, to informally interview all EBICS affiliates, and to collect artifacts distributed. Moreover, evaluators noted information related to the following categories during the observations:

- **Physical setting**—a rich description of the time and locale of the event, including where participants were situated.

- **Social or interpersonal setting**—who was clustered with whom and how groups and individuals were arrayed in this context.

- **Activities**—a systematic description of the activities with timeframes.

- **Content**—a description of the resources and materials used and discussed.

- **Interactions**—a description of verbal and nonverbal interactions.
<table>
<thead>
<tr>
<th>EBICS Goal Area</th>
<th>Program Activity</th>
<th>Year¹</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2011 2012 2013 2014 2015</td>
</tr>
<tr>
<td>Education</td>
<td>Graduate Teaching Consortium Courses</td>
<td>✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td></td>
<td>Education Committee Meetings</td>
<td>✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td></td>
<td>Trainee Research Symposiums</td>
<td>✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td></td>
<td>Journal Club</td>
<td></td>
</tr>
<tr>
<td>Diversity</td>
<td>Diversity Committee Meetings</td>
<td>✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td></td>
<td>Research Experience for Undergraduates</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td></td>
<td>Project ENGAGE</td>
<td>✓ ✓ ✓</td>
</tr>
<tr>
<td>Leadership and Management</td>
<td>Annual Retreats</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td></td>
<td>Executive Committee Meetings</td>
<td>✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td></td>
<td>NSF Site Visits</td>
<td>✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Ethics</td>
<td>Ethics Module Roll-Outs</td>
<td></td>
</tr>
</tbody>
</table>

¹ Also indicates future administration years.
Surveys

Multiple surveys were conducted throughout the first three years of the EBICS STC. Faculty, staff, trainees, industrial advisory committee members, and Research Experience for Undergraduates (REU) participants were given surveys. Surveys included open- and close-end items. Appendix A includes all surveys and aggregated responses. Evaluation surveys and year administered are listed in Table 2.

End-of-Year Trainee Survey

Trainees are given an online survey at the end of the Spring semester with questions related to program components (co-advisement, lab rotations, introductory course, international research experience, summer institute, Student Leadership Council, and seminar series/annual symposium) and trainees’ experience over the course of the past year. The survey includes 55 questions: eight demographic, 25 open-ended responses, and 22 close-ended Likert response questions.

Graduate Teaching Consortium Surveys

All students enrolled in the GTC courses were given evaluation forms. The survey was used to obtain information related to students’ final thoughts on course organization, objectives, access to materials, etc. The 34-question survey included four demographic questions, 21 multiple choice/Likert satisfaction questions, and nine open-ended questions.

2 A PDF file of Appendix A is available online at: http://www.istem.illinois.edu/resources/ebics.html.
Table 8. Evaluation Surveys Administration by Semester and Year (√ = completed)

<table>
<thead>
<tr>
<th>Survey</th>
<th>Semester Administered</th>
<th>Year Administered⁴</th>
<th>Recipient</th>
</tr>
</thead>
<tbody>
<tr>
<td>End of Year Surveys</td>
<td>Spring</td>
<td>✓ 2012</td>
<td>Faculty and trainees</td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2013</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2014</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2015</td>
<td></td>
</tr>
<tr>
<td>Graduate Teaching Consortium Surveys</td>
<td>Fall, Spring</td>
<td>✓ 2011</td>
<td>All students enrolled in courses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2012</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2013</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2014</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2015</td>
<td></td>
</tr>
<tr>
<td>Research Experience for Undergraduates Follow-Up Survey</td>
<td>Early Spring</td>
<td>✓ 2012</td>
<td>All REU participants</td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2013</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2014</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2015</td>
<td></td>
</tr>
<tr>
<td>Research Experience for Undergraduates Program Surveys</td>
<td>Summer</td>
<td>✓ 2011</td>
<td>REU participants, graduate student mentors</td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2012</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2013</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2014</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓ 2015</td>
<td></td>
</tr>
<tr>
<td>Trainee Exit Survey</td>
<td>Always live</td>
<td></td>
<td>Trainees who have left/graduated from their home institution</td>
</tr>
</tbody>
</table>

⁴ Also indicates future administration years.
<table>
<thead>
<tr>
<th>Survey</th>
<th>Semester Administered</th>
<th>Year Administered</th>
<th>Recipient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Retreat Surveys</td>
<td>Summer</td>
<td>✔️ 2011</td>
<td>All retreat participants</td>
</tr>
<tr>
<td></td>
<td></td>
<td>✔️ 2012</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✔️ 2013</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015</td>
<td></td>
</tr>
<tr>
<td>Ethics Training Surveys</td>
<td>Spring Summer</td>
<td>✔️ 2011</td>
<td>All trainees</td>
</tr>
<tr>
<td></td>
<td></td>
<td>✔️ 2012</td>
<td></td>
</tr>
<tr>
<td>Ethics Module Survey</td>
<td>Summer</td>
<td>✔️ 2013</td>
<td>All module participants</td>
</tr>
</tbody>
</table>

**Research Experience for Undergraduates Follow-Up Survey**

REU participants are sent a link to an online survey every year in January to further determine REU impact on their productivity at school, graduate/professional school choices, and current research activities. Questions are also asked relation to keeping in contact with their faculty advisor and graduate mentor and current school status.

**Research Experience for Undergraduates Program Surveys**

REU participants and graduate mentors were asked to complete surveys online. Participants completed one survey at the beginning of their program (first or second week). The second survey was completed by graduate student mentors and participants after the REU program was completed. Specifically, surveys were used to obtain information related to participants ‘overall experience with the REU program, including items related to campus location, resources given, lab and research assignment and format, lodging accommodations, professional development programs, knowledge learned, and successfulness of the REU program. Mentors were asked about their satisfaction with the REU students’ placement in the
labs, their understanding of program goals, and their thoughts about perceived benefits of the program for the participants. The surveys included both closed- and open-ended items.

**Trainee Exit Survey**

Trainees who leave or graduate from their home institution complete an online exit survey. The survey has 37 questions: ten demographic, 14 open-ended questions, and 13 closed-ended, Likert response questions. The survey asks questions related to trainee satisfaction with each program component, their overall experience, their current employment, and the value the EBICS STC added to their experience as a graduate student.

**Annual Retreat Surveys**

Retreat attendees were asked to complete an online post-retreat evaluation survey. Surveys links are sent to all attendees via email. While one survey was administered, questions probed feedback for each day and each presentation, as well as overall feedback. Specifically, the surveys were used to obtain information related to IAC/EAC, faculty, and trainees’ overall experience with the retreat, including items related to presentation usefulness, resources given, retreat organization and format, hotel accommodations, food adequacy, knowledge learned, and overall experience and successfulness of the retreat. The surveys included both closed- and open-ended items.

**Responsible Conduct of Research Surveys**

Trainees were given surveys to see if Responsible Conduct of Research requirements were met. Surveys consisted of open-ended questions about the number of hours and activities trainees had completed related to ethics training.

**Ethics Module Surveys**

All ethics module participants were given surveys about their thoughts and experience with the ethics module. The survey included nine close-ended Likert questions about module
usefulness, organization, and expectations. Three open-ended questions related to suggestions for improvement and future topics were also included.

**Focus Groups and Interviews**

Semi-structured focus groups and interviews have been conducted throughout the evaluation. Faculty, staff, trainees, industrial advisory committee members, project ENGAGE scholars and mentors, and research experience for undergraduate participants have all participated in focus groups or were interviewed. Table 3, which follows, displays interviews and focus groups conducted to date.

**Trainee STC Experience Focus Group**

A focus group was held with 14 EBICS trainees (four Illinois, five MIT, and five GA Tech trainees). This focus group took place at the beginning of the EBICS Retreat in Champaign, IL on June 24th, 2012. The focus group lasted approximately 50 minutes, and participants were asked about their experiences as EBICS trainees. Specifically, questions were asked about: interactions with each other, suggestions for improvements, and benefits of their traineeship.

**Project ENGAGE Focus Groups**

Evaluators conducted a semi-structured focus group with Project ENGAGE scholars and graduate mentors to explore their experiences, reactions, and opinions. A protocol was developed for and followed during the focus group to prompt interviewees in order to ensure that the relevant questions were covered.

**REU Focus Groups**

Evaluators conduct a semi-structured focus groups to explore REU participants’ experiences, reactions, and opinions. A protocol was developed for and followed during the focus group to prompt interviewees in order to ensure that the relevant questions were covered. Evaluators usually were in the room with Illinois students, while GA Tech and MIT REU
students participated via video conferencing. In 2013, evaluators conducted focus groups with MIT and GT participants on their home campuses.

**Project ENGAGE Interviews**

Evaluators conducted interviews with Project ENGAGE teachers and program staff experiences, reactions, and opinions. A protocol was developed for and followed during the focus group to prompt interviewees to ensure that the relevant topics were covered.

Table 9. Evaluation Focus Groups and Interviews Administration by Semester and Year ($\checkmark$ = completed)

<table>
<thead>
<tr>
<th>Focus Group/Interview</th>
<th>Semester Administered</th>
<th>Year Administered$^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trainee STC Experience Focus Group</td>
<td>Summer</td>
<td>✓ 2011 ✓ 2012</td>
</tr>
<tr>
<td>Project ENGAGE Focus Groups</td>
<td>Spring, Summer</td>
<td>✓ 2012 ✓ 2013 ✓ 2014 ✓ 2015</td>
</tr>
<tr>
<td>REU Focus Groups</td>
<td>Summer</td>
<td>✓ 2011 ✓ 2012 ✓ 2013 ✓ 2014 ✓ 2015</td>
</tr>
<tr>
<td>Project ENGAGE Interviews</td>
<td>Summer</td>
<td>✓ 2012 ✓ 2013 ✓ 2014 ✓ 2015</td>
</tr>
<tr>
<td>Industrial Advisory Committee Interviews</td>
<td>Spring</td>
<td>✓ 2013</td>
</tr>
<tr>
<td>Minority Serving Institution PIs Interview</td>
<td>Spring</td>
<td>✓ 2012</td>
</tr>
</tbody>
</table>

$^5$ Also indicates future administration dates.

**Industrial Advisory Committee Interviews**

Interviews were conducted with four of the five IAC members on March 23, March 25, and April 17, 2013. Questions related to how they envision their role with EBICS, their interest
in the Center, how they would like to be interacted with, and how much time they would like to spend each year on the IAC.

**Minority-Serving Institution PIs Interview**

Key faculty from EBICS partner Minority-Serving Institutions (MSI) were asked to participate in interviews. Participants included: Kara McCloskey from UC Merced, Alexandra Peister from Morehouse College, and Maribel Vazquez from CCNY. Interviews were approximately 20–25 minutes, and participants were asked about their experiences within the STC thus far. Specifically, the interview questions we asked were regarding their satisfaction with their role in the center, thoughts about the EBICS Graduate Teaching Consortium and Research Experience for Undergraduates, and suggestions for improvement.
D.2: EBICS Evaluation Reports

DeStefano, L., & Tillman, A.S. (2014). *Emergent Behaviors of Integrated Cellular Systems Fall 2013 Graduate Teaching Consortium Evaluation*. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., Tillman, A.S., & Innes, B. (2014). *Emergent Behaviors of Integrated Cellular Systems Year 1 – 3 Comprehensive Evaluation Report*. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., & Tillman, A.S. (2013). *Emergent Behaviors of Integrated Cellular Systems Spring 2013 Graduate Teaching Consortium Evaluation*. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., & Tillman, A.S. (2013). *Emergent Behaviors of Integrated Cellular Systems 2013 Research Experience for Undergraduates Evaluation*. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., & Tillman, A.S. (2013). *Emergent Behaviors of Integrated Cellular Systems 2013 Annual Retreat Evaluation*. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., & Tillman, A.S. (2013). *Emergent Behaviors of Integrated Cellular Systems Project ENGAGE Interim Evaluation Report*. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., & Tillman, A.S. (2013). *Emergent Behaviors of Integrated Cellular Systems Graduate Teaching Consortium Fall 2012 Evaluation Report*. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., & Tillman, A.S. (2012). *Emergent Behaviors of Integrated Cellular Systems 2012 Research Experience for Undergraduates Evaluation*. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., & Tillman, A.S. (2012). *Emergent Behaviors of Integrated Cellular Systems Trainee Curriculum Evaluation Report*. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.
DeStefano, L., & Tillman, A.S. (2012). Emergent Behaviors of Integrated Cellular Systems 2012 Retreat Evaluation Report. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., & Tillman, A.S. (2012). Emergent Behaviors of Integrated Cellular Systems Graduate Teaching Consortium Spring 2012 Evaluation Report. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., & Tillman, A.S. (2012). Emergent Behaviors of Integrated Cellular Systems 2012 Trainee Experience Report. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., & Boyce, A.S. (2012). Emergent Behaviors of Integrated Cellular Systems Minority Serving Institution Interview Report. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., & Boyce, A.S. (2012). Emergent Behaviors of Integrated Cellular Systems Graduate Teaching Consortium Fall 2011 Evaluation Report. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., & Boyce, A.S. (2011). Emergent Behaviors of Integrated Cellular Systems Research Experience for Undergraduates 2011 Evaluation Report. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.

DeStefano, L., & Boyce, A.S. (2011). Emergent Behaviors of Integrated Cellular Systems 2011 Retreat Evaluation Report. EBICS though the University of Illinois at Urbana-Champaign; Massachusetts Institute of Technology; Georgia Institute of Technology; STC award through the National Science Foundation.