



2014-06-26

For PEAT's Sake: The Influence of Student-Led Evaluations on Small Nonprofits

Rilee A. Buttars

Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Sociology Commons](#)

BYU ScholarsArchive Citation

Buttars, Rilee A., "For PEAT's Sake: The Influence of Student-Led Evaluations on Small Nonprofits" (2014). *All Theses and Dissertations*. 4141.

<https://scholarsarchive.byu.edu/etd/4141>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

For PEAT's Sake: The Influence of Student-Led
Evaluations on Small Nonprofits

Rilee Buttars

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Master of Science

Carol Ward, Chair
Curtis Child
Eric Dahlin

Department of Sociology
Brigham Young University

June 2014

Copyright © 2014 Rilee Buttars

All Rights Reserved

ABSTRACT

For PEAT's Sake: The Influence of Student-Led Evaluations on Small Nonprofits

Rilee Buttars
Department of Sociology, BYU
Master of Science

Development evaluation is a booming business often seen as a development activity in its own right. Yet, not every actor in the field engages with evaluation. This is primarily due to inadequacies in time, data, resources and expertise. These challenges are prevalent among small nonprofits that arguably stand most in need to appraise their extensive efforts. In an accountability-focused environment, these organizations not only fail to assess adequately the impact of their work but also to evaluate reflectively their structure and implementation process. In response, the sociology and international studies departments at Brigham Young University set up an internship that trains students in evaluation techniques and then contracts their services to nonprofits that struggle to systematically evaluate their programs. Data collected from the student-evaluators and nonprofits showed that the evaluations influenced the organizations by helping the personnel learn, reflect on, discuss, and adjust their programs. In addition to finding evidence of the factors, mechanisms, and change processes that influence an evaluation's effect, this study extends the conceptual models within the influence literature by outlining the enabling role of students in contexts with high capacity constraints.

Keywords: evaluation influence, nonprofits, non-governmental organizations, organizational learning, process assessment, developmental evaluation, student-evaluators

ACKNOWLEDGEMENTS

I would like to thank the Sociology Department at Brigham Young University: the people and conversations in its corners and halls have been my safe place. I am grateful for the training I received, friends I have made, and the patience of each faculty member as I navigated this graduate program as a former “ad-girl.” I am especially grateful for Margaret McCabe for not cutting me off after the one-hundredth missed punch. Thank you for guiding the troublesome 2011 cohort. I owe you a cannoli from Little Italy and maybe a couple dozen mini cupcakes from the great New York City.

To my Thesis Committee—thank you for reading and editing this brick of a paper. You stuck with it despite my need to be overly comprehensive, detailed, and obsessive. Thank you for stepping in and providing your insight and support. I respect your expertise, advice, and who you are as academics and colleagues. To Dr. Ward, I owe this accomplishment and my graduate experience to your tireless efforts, patience, and guidance. How can I repay you for introducing me to evaluation? You’re a perfect mentor, but after all those years, I am especially grateful to call you a friend. I always felt most at ease in your office laughing about the news of the day. Thank you for taking me in.

To the unofficial committee member, Dr. Ralph Brown—thank you for believing in me and showing me the way. You’re the reason I can call myself a sociologist and the reason I keep fighting. You taught me how to ask questions, how to find answers, and how to see the world in a different light: in a positive light. Thank you for opening your door. Thank you for teaching me. Ralph, you’ll always be here, and there, and with all of us wherever we go.

TABLE OF CONTENTS

Abstract.....	ii
Acknowledgements.....	iii
Table of Contents.....	iv
List of Tables	ix
List of Figures.....	x
Introduction.....	1
Literature Review.....	4
In the Pursuit of Social Betterment.....	5
The Many Faces of Evaluation: In Definition, Purpose, and Inquiry.....	7
What Is Evaluation?.....	8
Evaluation to What End?	9
Evaluation Inquiry and Design	14
The Accountability Movement.....	18
Problems with the Accountability Movement	21
The Use and Influence of Evaluation.....	24
Evaluation for a Key Development Actor	26
Meet PEAT: The BYU Program Evaluation and Assessment Team.....	28
Research Questions.....	29
Methods.....	30
Sample.....	30
Interview Procedure and Instrument.....	31
Survey Procedure and Instrument.....	33

Document Analysis.....	33
Analysis.....	34
Limitations	35
Findings.....	36
Facts of the Case	37
PEAT’s Clients: Context and Mission.....	39
Organizational Model	40
Stage of Development and Resources.....	43
Learning Culture	46
Approach to Evaluation	47
Experience with Evaluation	47
Barriers to Evaluation	49
Facts of the Evaluation.....	50
Evaluation Inputs	51
Decision-making setting	57
Micro-politics.....	58
Competing information.....	60
Evaluation readiness	61
Accountability requirements.....	62
Evaluation Outputs.....	62
Evaluation planning	62
Evaluation purpose: nonprofit survey responses	64
Evaluation purpose and type: interviews and reports	65

Evaluation methods.....	66
Evaluation analysis	68
Evaluation findings	69
Evaluation recommendations.....	70
Report generation and dissemination.....	71
Evaluation Quality	72
Reliability or consistency.....	73
Validity or authenticity	75
Student perceived quality.....	76
Nonprofit perceived quality	78
Context Contingencies	78
Organization Experience.....	81
Working with PEAT	81
Working with Student-evaluators	83
More Helpful Evaluations.....	85
Satisfaction Rating	85
Influence of the Evaluation.....	86
General Influence Processes	87
Elaboration.....	87
Priming.....	89
Justification	91
Change-agent	92
Persuasion	93

Minority-opinion influence	94
Collective deliberation/discussion	95
Skill acquisition	95
Additional general influence.....	96
Individual/collective learning	96
Cognitive and Affective Processes	97
Salience	97
Opinion/attitude change	98
Behavioral Processes	98
Changes in practice	99
Changes to program	100
Led to more funding and/or evaluation.....	100
Source of Influence	101
Anticipated or New	102
Timeframe	103
Student Factor	104
Summary of Findings.....	108
Lessons Learned.....	112
Recommended Approach for the Organization	116
Recommended Approach in Context	118
PEAT's Role	119
Conclusion	121
References.....	124

Appendix A.....	140
Appendix B.....	141
Appendix C.....	144
Appendix D.....	148

LIST OF TABLES

Table 1. Organization Participants: Pseudonym, Position, Years at Organization.....	157
Table 2. Descriptive Information on PEAT Interns, 2009-2013.....	158
Table 3. Descriptive Information on PEAT’s Clients, 2009-2013.....	159
Table 4. Survey Response on Barriers PEAT Clients Face to Implementing Evaluation.....	160
Table 5. Survey Response: Evaluation Purpose and If Purpose Was Met.....	161
Table 6. Survey Response: Evaluation Purpose, If Met, and If Results Were Used in Decision- making.....	162
Table 7. Type of Evaluation Conducted.....	163
Table 8. Type of Methods Used for Evaluation.....	164
Table 9. Survey Responses: Type of Influence.....	165

LIST OF FIGURES

Figure 1. Mark and Henry (2004) Schematic Theory of Evaluation Influence.....	166
Figure 2. Appleton-Dyer et al. (2012) Influence Model.....	167
Figure 3. A Conceptual Model of the Student-Led Evaluation Change Process.....	168

INTRODUCTION

Although evaluation is a powerful tool for poverty alleviation efforts, for some organizations it could work more effectively. Small nonprofits are important actors in the development arena, but they rarely engage in evaluation due to a lack of resources and organizational maturity. Amidst accountability demands, these organizations struggle to show impact let alone use evaluation to improve their programs. The BYU Project Evaluation and Assessment Team (PEAT) was created to assist organizations that face challenges with evaluation. This thesis presents a case study of the PEAT internship program. It analyzes 23 student-led evaluations with the primary purpose to better understand what type of evaluation works best for these organizations and why. The scholars Mark and Henry (2004) and Appleton-Dyer, Clinton, Carswell and McNeil (2012) outlined various aspects of the organization, evaluation and context that contribute to whether an evaluation will have practical influence or not. This study utilized their conceptual frameworks as a guide for accomplishing this purpose. As part of this process, a new conceptual model was developed that builds on the models of Mark and Henry (2004) and Appleton-Dyer et al. (2012) for the purposes of this study. The new model outlines the ways the student-led evaluations were influential and the factors that matter in bringing about those changes. The model extends that of Mark and Henry and Appleton-Dyer et al. by only including the elements that led to changes in the PEAT organizations, adding the role of the student, and separating the mechanisms of learning about a program from elaborating about it (Henry and Mark 2003). The conceptual model highlights the factors that may lead to specific changes in small nonprofits working in complex environments, and could be used by nonprofits as a guide for planning and conducting assessments of their program strategies and selected outcomes.

The data produced by this study provides empirical evidence of the evaluation process and associated benefits of student-led evaluations for organizations that work in developing contexts. Additionally, this study contributes to practice by recommending that these types of organizations are best able to stand accountable to stakeholders if they first utilize evaluation and systematic reflection to develop and improve their programs. This can be accomplished by emphasizing all types of assessment, mixed-methods and capacity development. To come to these conclusions, this study used the categories of the influence models to formulate the following research questions: 1) what do the student-led evaluations reveal about these organizations and the challenges they face in evaluation; 2) what was the nature of the evaluations; 3) what types of influences resulted from the evaluations; and 4) what does the process uncover about how best to provide evaluation for this specific type of organization? Answering the first three questions is essential for identifying the most appropriate evaluation purpose, type, and methods for these organizations and their environments—the last research objective.

To answer the first question, the analysis showed that these organizations work in complex environments, have limited resources, and are in a perpetual state of development. The directors are passionate about their work and want to succeed, but they struggle to research effectively the needs of their beneficiaries and whether the program can and does meet those needs. This is primarily because they lack the resources and expertise to incorporate evaluation at every phase of the program. In light of increased accountability demands, they focus on gathering the information that is more likely to appease stakeholders and lead to more funding as opposed to the information that might improve their services. They engage in minimal formal

reflection and research, and thus, learning and application of that learning is typically sporadic and shallow.

Second, the student-led evaluations were primarily focused on describing the context, service population, program delivery, and perceived outcomes. While the organization staff initially wanted the evaluation to provide evidence of impact, adjustments were required given the status of the program in the field and context. The majority of the students lived with the beneficiaries while in-country, which proved essential in building rapport among the population. This gave the students access to informal information that would have been difficult to obtain otherwise. The students judged the quality of the evaluations more harshly than the nonprofit personnel, but overall, both parties felt that the quality was good enough to meet the organization's needs.

Third, the types of influences were identified following the conceptual models of Mark and Henry (2004) and Appleton-Dyer et al. (2012). Evidence was found for behavioral changes among the individuals working within the organization and to the structure itself. However, a substantial amount of support was found for the mechanisms that led to these changes. Specifically, the student-evaluators helped the organizations learn and think about programs at the individual and collective level. They also provided data that encouraged individuals to speak up about an issue or idea. Each of these led to more informed discussions and decisions about the organization's operations and programs.

Fourth, the findings concerning the organization and the evaluations helped show that the organizations are not in the phase of development that would allow for evaluating impact. They need information that can help them structure their model, mission, and programs to the needs of the service population and to continually improve them along the way. Additionally, the student-

evaluations helped the organization to understand the purpose, benefits and requirements of evaluation as well as how to prepare to measure short- and medium-outcomes that are essential for showing longer-term impact. Future evaluations of these organizations should emphasize gathering information for learning and development purposes, adjusting programs to that insight, and increasing organizational capacity for further reflection, analysis, and impact assessment.

LITERATURE REVIEW

Utilizing literature from both the development and evaluation disciplines, the following review first provides background into the broad relationship between poverty alleviation efforts and evaluation and the evolution of the evaluation field as it aims to meet the demands of its scientific trade and applied purpose. This leads into a section on the various definitions, purposes and forms of evaluation inquiry—a summary of the range of possible evaluation attributes that might best fit small or developing organizations. While the evaluation field has recently embraced methodological pluralism, the accountability movement within development has emphasized the types of assessments that can provide direct evidence of impact as opposed to promoting overall program learning, development and improvement. The flaws with the movement are discussed, including the constraints faced by development actors—and in particular small nonprofits—to adequately meet their donor demands. The issue of non-use is also discussed as an unfortunate by-product of the movement. The evaluation characteristics and the issues discussed in this literature review provide context for the relevance of PEAT and the need to adjust evaluation practices to improve its *use* and *influence* among developing organizations, constructs widely recognized as evaluation's *raison d'être* (Mark and Henry 2004; Preskill and Caracelli 1997).

IN THE PURSUIT OF SOCIAL BETTERMENT

Poverty is an age-old problem felt by millions in both developed and less developed nations. It is experienced generationally, culturally, and systematically among individuals, families, and communities. As a concept, it has been discussed, debated, measured, and analyzed by academics, practitioners, leaders, policy-makers, and activists alike. No matter how it is classified, its persistence is evident, and it is these same actors that attempt to render this state of deprivation obsolete. Some call it a war, a movement, an initiative with institutionalized goals and task forces.¹ Indeed, fields of professionals stand dedicated to see it through to the “the end of poverty” (Sachs 2005).

In 2000, the 189 countries of the United Nations unanimously declared, “we will spare no effort to free our fellow men, women and children from the abject and dehumanizing conditions of extreme poverty (United Nations Millennium Declaration 2000). Charged with such an agenda, the social sector churns out careers and funds dedicated to the cause. This space consists of big and small players aimed at helping marginalized populations (Grasso 2010). According to Vinod Thomas, “there [are] 280 bilateral aid agencies, 24 development banks, 40 United Nations agencies, and a plethora of private foundations and nongovernmental organizations funding approximately 340,000 development activities around the world” (2010:545). The International Development Association reported these numbers in 2007, and undoubtedly the quantity of organizations has steadily increased over the last six years. Additionally, the amount allocated to meet the objective is unprecedented. The United States alone gave \$31.2 billion in foreign

¹ See United Nations Millennium Development goals: <http://www.un.org/millenniumgoals/>.

assistance² and \$550 billion domestically in 2012 (Wostall 2012). US Today reported that same year that individuals in the U.S. donated over \$200 billion to charities.

While the size, purpose, and outcomes associated with these efforts widely vary, a consistent thread connects them all: the sincere belief that a difference can be made, that inequality can be balanced. In the pursuit of social betterment, the organizational model is fundamentally different from other public and private companies because it is underlined with a moral imperative (Henry and Mark 2003; Mark and Henry 2004). The ethical dimension creates an environment in which the continuation of less effective activities, faulty decision-making, or underinvestment in needed services is not just “bad for business” but potentially detrimental for the intended beneficiaries. In the development enterprise, the stakes are high for whether it works or not, and for many, that is precisely the million-dollar question: Given the money and energy spent, are people better off now than they were before?

The answer to this question is wrought with controversy. At the beginning of last year, the United Nations Development Program reported that the “number of extreme poor had dropped by 650 million in the last three decades, a level of progress humankind had never seen” (UNDP 2013). News articles flash headlines claiming that social programs reduce poverty (Burman 2011), that the world is taking great strides in this fight (“Towards the End of Poverty” 2013) and that “extreme poverty is falling in every region,” (United Nations 2012:4). On the far side of the spectrum, the story moves from accomplishment and hope to shame and broken promises. As William Easterly (2006:11) said, “Sixty years of countless reform schemes to aid agencies and dozens of different plans, and \$2.3 trillion later, the aid industry is still failing to reach the beautiful goal.”

² U.S. Overseas Loans and Grants (Greenbook) 2012.

So, which story is correct? Is one a more accurate portrayal of reality? Have the attempts to improve the human condition worked or have they not (Deaton 2010)? In the pursuit to describe, understand and ameliorate social problems, organized efforts have also been undertaken to study the interventions (Rossi et al. 2004; Patton 1990). The scientific inquiry dedicated to untangling the enigma of poverty is accompanied by research endeavors to not only answer the question above, but also improve these very same services so that it can be answered in the affirmative. To borrow from Rossi, “at various times, policymakers, funding organizations, planners, program managers, taxpayers, or program clientele need to distinguish worthwhile social programs from ineffective ones and launch new programs or revise existing ones so as to achieve certain desirable results” (2004:3). This particular task belongs to the field of evaluation.

THE MANY FACES OF EVALUATION: IN DEFINITION, PURPOSE, AND INQUIRY

Systematic evaluation research was first attached to the implementation of social programs, particularly in education and public health, around World War I. By the 1950s, its usage extended to a range of services, and in the early 1970s, it was officially adopted by the social sciences as a distinct specialty within the discipline (Rossi et al. 2004; Weiss, Murphy-Graham, Petrosino, and Ghandi 2008). As the number and funding of social programs increased so too did the concern over the balance of benefits and costs for these interventions. In response, evaluation increasingly became established as their counterpart set to determine the impact of the programs towards improving people’s standard of living (Mulwa 2008; Patton 1990). Charged with this mandate, academics within the field helped to shape and advance the practice through the construction and testing of theories, methods, and approaches deemed most appropriate for its various applications. Yet, given its inherent applied nature, the consumers of evaluation work

contributed significantly to the field's maturation, at times emphasizing different aspects in order to meet specific needs. Thus, the evolution of evaluation jockeyed between the demands of science and practice; the capricious debate over what it is, its value and how it should be carried out mirrors the balancing act between these two competing forces.

What Is Evaluation?

The term evaluation is generally understood as the appraisal of the worth or value of some object. This is no coincidence given the root of the word and, consequently, adjudicating merit has been a central tenant of evaluation work (Scriven 1980; Weis 1972; Worthen et al. 1997). The Merriam Webster dictionary adds one element to this typical definition: to judge the *condition* of something in a careful and thoughtful way.³ This view extends evaluation beyond a summary judgment on the overall effectiveness of a program and allows for an in-depth look into its inner workings, which coincides with Patton's (2007) version of evaluation as an illuminant. "It is a process of illuminating the project in order to see more clearly what needs more attention, whether for modification, expansion, or replication" (Mulwa 2008:45).

Encompassing both of these characteristics, Rossi, Lipsey and Freeman (2004:16) define program evaluation as the "use of social research methods to systematically investigate the effectiveness of social interventions" through describing the performance of a program and applying a set of criteria for judging the success of that performance. This definition entails both data gathering for description's sake and applying criteria to that data for judgment's sake. In many circumstances these two activities are valuable in their own right, but for some scholars, the mere accumulation and summarization of information is viewed as a means, *not* an evaluative end.

³ <http://www.merriam-webster.com/dictionary/evaluate>

Under this premise, to be considered an evaluation activity, there must be a process of converting facts into evaluative conclusions (Scriven 1991:206) and/or attaching significance to particular results (Patton 1990:375). This is done as the researcher applies criteria for appropriateness, relevance or effectiveness. Thus, whether narrowly defined as judgment or broadened to include illumination, evaluation includes collecting, analyzing, interpreting, and communicating information. This information is used to study and appraise, and help improve social programs—to name a few of the various roles evaluation can play. What constitutes as evaluation largely stems from the purpose evaluation serves for the organizations that engage in the process. Evaluation involves applied science, and as such, is judged by its usefulness in making programs that are trying to improve the world more effective (Patton 1990). How it goes about meeting that expectation stems from the way the field and consumers view evaluation's fundamental purpose and the role of the evaluator in fulfilling it.

Evaluation to What End?

While professional evaluators, academics, and those who utilize evaluations still do not agree upon a uniform definition of evaluation, it is widely regarded as a tool to determine the worth and condition of a program. Evaluator as judge can be seen in the classic and popular distinction between formative and summative evaluations.

Michael Scriven (1967) first drew these classifications in his seminal paper on the methodology of evaluation. Scriven envisioned two roles for evaluation with formative evaluation performing the appraisal of a program's worth or merit while there is still time to modify and improve the program (Popham 2006). In contrast, summative evaluation provides a judgment of worth or merit over a mature program's performance (Worthen et al. 1997:15). This

information could help stakeholders in their decisions to continue, terminate or expand the program.

Rossi et al. (2004) further discussed these two types of evaluations. From their perspective, formative evaluation may gather information on the need for the program, the program's design, its implementation and service delivery, its impact, and/or its efficiency. These various types of assessments cover a range of program needs, and if provided in a timely and concrete manner, can be immediately useful for "planners, administrators, oversight boards, or funders with an interest in optimizing program's effectiveness" (Rossi 2004:34). According to these authors, summative evaluation provides stakeholders with the information necessary to determine if the program has met expectations for using resources effectively and efficiently in order to produce the intended benefits (p. 36).

The summative and formative divide covers four primary and traditional reasons for evaluations. The first is evaluation as a basis for decision-making (Grasso 2010:534; Gopalakrishnan, Preskill, Shijie 2013). "Evaluation, we believe, exists or perhaps only should exist, to facilitate intelligent decision-making" (Struening and Brewer 1983:140). As Preskill and Caracelli (1997) found, 99 percent of the evaluators they surveyed placed decision-making as one of the primary purposes for evaluation. If conducted in an appropriate, systematic and high-quality manner, evaluations can provide relevant information that "helps set priorities, guide the allocation of resources, facilitate the modification and refinement of project structure and activities and signal the need for additional personnel and resources" (Mulwa 2008:51). Indeed, evaluation can provide information for decision-making at every stage of the program.

Therefore, two other common functions for evaluation include enlightening decision makers on the nature of the program and producing evidence of the potential impact of the

project. Evaluation for program improvement answers questions regarding the program's strengths and weaknesses, progress towards achieving the desired outputs and outcomes, how well assumptions are holding up, and what new ideas and solutions are emerging. Evaluation for conclusive purposes provides an overall judgment about the effectiveness of a stabilized program answering questions on whether the program worked, and if it contributed to the stated goals and outcomes.⁴

While these efforts help to make interventions more effective, the data used to inform decision-makers and render a summary judgment ultimately helps organizations stand accountable to those who have a stake in a project, including the beneficiaries, managers, staff, donors and public at large. Arguably, evaluation is best viewed and used as a tool in the process of "making the world a better place." Though, at the end of the day, this tool must help organizations show whether the efforts and funds actually contributed to this aim or not. Thus, accountability is an important purpose for evaluation, and while both formative and summative evaluations have their distinct functions, they are both instrumental in meeting this demand.

In the decades since Scriven proposed the summative and formative divide, debates have flourished over their scope, relationship to programs and each other, as well as the supremacy and preferential use of one over the other (See Chen 1996a; Chen 1996b; Patton 1996; Scriven 1996). The meanings have expanded and contracted, and more liberal approaches have been introduced for how evaluation may better serve as the counterpart to social betterment. Michal Patton and many others proposed expanding the role of evaluators beyond that of distant, independent-judge to facilitator "helping decision makers and information users arrive at their

⁴ Purpose of Evaluation Guide provide by the United Nations Population Fund Division for Oversight Services at www.unfpa.org.

own judgments” (Patton 1996:138). This notion of facilitator is present in the various collaboration-based and learning-oriented approaches introduced and advocated over the last 30 years.

In one view, evaluation is seen as an educational process where reflection and self-evaluation are encouraged among all stakeholders (Mulwa 2008; Fetterman 2001). In this sense, “evaluation is a process of learning from the past with the view to improving on the future” (Mulwa 2008:52). Learning emerges as the organization and its beneficiaries work in tandem with an evaluator as equal partners in the evaluation process. Interactive and collaborative elements are fundamental in participatory approaches to evaluation. In these approaches, key stakeholders hold the power and their fears, interests, experiences, and insights are not only valued, but also seen as *crucial* in dialogue and reflection. Participation is meant to increase the ownership and use of findings, as well as empower and build the capacity of stakeholders (see Cousins and Whitmore 1998).⁵

As participatory and empowerment approaches brought the stakeholders into the evaluation process, Rosalie Torres and Hallie Preskill’s (2001) organizational learning model emplaced evaluation *within* organizations. They envision evaluation as on-going and iterative process of inquiry that is fully integrated into the organization’s infrastructure and fosters a culture of continual learning, growth and improvement (see also Preskill and Torres 1999:184-185) Similarly, Owen and Lambert (1998) present evaluation as a tool at the organizational level that focuses on total system improvement, challenges assumptions and organizational dynamics,

⁵ For an in-depth look into this type of evaluation see the following authors: Brandon1998; Brunner and Guzman 1989, Cousins and Earl 1992, 1995; Fetterman 1996; Greene 1987; 1988; O’Sullivan and O’Sullivan 1998; Patton 1997.)

and assists in strategic thinking and planning (see also March and Olsen 1976). For these scholars, evaluation inquiry nurtures “individual and team learning within an organization” (Preskill and Torres 1999:17) and “addresses systematic factors that directly impact the quality of goods and services provided by the organizations (Owen and Lambert 1998:363; see also Weick 1976).

Furthermore, these approaches challenge the notion that programs grow in a linear fashion in which an improvement-oriented evaluation helps form a well-defined intervention model while preparing for a judgment-oriented summative evaluation (Patton 2011:135). Rather, organizations are increasingly becoming unstable and unpredictable. Thus, program managers need continuous problem solving and rapid responsiveness, characteristics not well suited for the classic distinction. Patton designed an approach specifically geared towards these circumstances: developmental evaluation “support[s] adaptation in complex, uncertain, and dynamic conditions” (Patton 2011:46).

While many things set developmental and these other approaches apart from conventional evaluation, some key features distinguish them as both important in the field and useful to clients. Namely, these approaches redefine the relationship between evaluator and evaluand. “The evaluator’s primary function in the team is to infuse team discussions with evaluative questions, thinking, and data, and to facilitate systematic data-based reflection and decision-making...” (Patton 2011:1-2). These approaches help establish learning and development as key outcomes to evaluation. Instead of bringing order to chaos through uniformity and standardization, these approaches recognize and adapt to complexity, change, and contextual realities. Lastly, they redirect the focus from external accountability for resources to self-assessment and internal liability for assumptions, whether the organization is being true to

its vision, connecting the dots between plans and on-the-ground realities and an overall sense to “what is being developed” and “with what implications” (see Patton 2011:13; Briedenhann and Butts 2005).

While oversimplified, the purpose of evaluation under these liberal alternatives can be understood as 1) Learning by way of participation, reflection, illumination, feedback and dialogue 2) Assistance in planning, developing, improving, decision-making and 3) Accountability to external and internal demands for insight into the workings, progress, and effectiveness of programs. It is important to note that the variations in purpose are best conceptualized as lying along a continuum. Researchers engaged in evaluation at various points along the continuum differ on how they describe what is being done and its purpose, both of which are central to preferences in how best to design a research strategy and gather data (Patton 1990:158).

Evaluation Inquiry and Design

The field of evaluation is multidisciplinary, largely made up of the tools and building blocks of the social sciences. Thus, it has been beholden to the theories and principles guiding their work as well as the debates troubling their community (Scriven 1991:24). Stemming from the longstanding debate within the philosophy of science, theorists and methodologists have argued over how best to conduct research based on two fundamentally competing paradigms for inquiry (1991:37).

The positivist orientation dominated Western social science for much of the 20th century, and emphasizes impartiality, accuracy, causation and generalization through quantitative and experimental methods (Mathison 2005:345; Briedenhann and Butts 2005:226; Patton 1990:37; Jewiss and Clark-Keefe 2007, Preskill and Torres 1999). The other inquiry paradigm, rooted in

phenomenological, interpretive, and constructivist philosophies, invested in “qualitative and naturalist approaches to inductively and holistically understand human experience in context-specific settings” (Patton 1990:37).

The positivist strategies characterize stage-one in the evolution of evaluation theory (Shadish, Cook and Leviton 1991). Consistent with the social environment of the 1960s, theorists in this school advocated evaluation as the tool to rigorously assess program effectiveness at solving social problems. In practice, this was best realized through outcome-based practices and maintaining evaluator independence (Briedenhann and Butts 2005). Preferences toward this approach—in both a literal and epistemological sense—arose out of the desire and attempt to strictly model the social sciences after the natural sciences (Howe 1992). This is evident in the early stages of evaluation as the academy eagerly tried to win credibility by adopting true experiments as the ideal design and settling for quasi-experimental designs as an acceptable substitute (Scriven 1991).

Alternatively, interpretivist approaches and qualitative inquiry focus on studying real-world situations as they unfold naturally, immersion in and description of the details, attention to process and the phenomenon as a whole, as well as flexibility and responsiveness to the needs and interests of stakeholders (see Patton 1990:40-41). Given this emphasis, qualitative traditions resist the tendency to focus solely on outcomes and “insist on [and allow for] understanding the character and quality of the program experience” (Mathison 2005:348). Under this premise, description and interpretation of process as well as outcomes are viewed as essential and valid forms of evaluation. In fact, in this tradition, interpreting the meaning behind information on program outcomes requires an understanding of the character and quality of the program, its implementation and function from within its places and spaces (Mathison 2005). These nuances

and meanings can be captured via qualitative data collection methods such as in-depth interviews, direct observation, and focus groups that take place while in the field (Patton 1990:10).

Beginning in the 1970s, the evaluation field became entangled in the qualitative-quantitative debate, pitting the two methodological traditions against each other in a struggle for legitimacy and dominance in evaluation practice (Greene and Henry 2005:345). Instead of aligning with one paradigm over the other, Patton advocated for a more pragmatic approach that favors choosing a method based on appropriateness rather than orthodoxy. This shift characterizes the second and third generation of evaluation theorists that surfaced in the 1970s and 1980s.⁶ In addition to pragmatism, stage-two theorists focused on how information is used in the design and modification of social programs (Shadish, Cook, Leviton 1991). Use and issues of nonuse were dominant concerns for theorists and practitioners in this camp, such that the success of an evaluation should be viewed by its usefulness to the administrators rather than its scientific rigor (Suchman 1967; Patton 1999).⁷ These values are inherent in the responsive and utilization-focused approaches proposed by Robert Stake (1983) and Patton (2007) that stress a mixed-methods approach to inform data collection strategies. Such strategies are based on stakeholders' concerns, the purpose of the evaluation, intended use of its findings and the resources available to conduct the evaluation. Patton in particular supports using both qualitative and quantitative methods as together they can paint a more holistic picture of the program (Briedenhann and Butts 2005:229).

⁶ See work of Carol Weiss, Joseph Wholey and Robert Stake.

⁷ See the Joint Committee on Standards for Educational Evaluation and JCSEE Program Evaluation Standards, 1994. Sage.

Methodological pluralism is a central tenant of stage-three theorists. As Lee Cronbach and others have stated, no single method should take precedence and no rigid guidelines should direct the process of developing an evaluation (Cronbach 1982; Rossi et al. 2004). However, Rossi et al. (2004) proposed a systematic, comprehensive approach to tailoring evaluations to the diverse needs and circumstances of the program and decision-makers as well as keeping to the high standards of research. In this approach, inquiry decisions are based in the direct link between the stages in the program's development, the questions that arise during that stage and specific evaluative functions that are useful in answering those questions. The stage, question and function all correspond to five recognizable types of evaluation including: needs assessment, assessment of program theory, assessment of program process (or process evaluation), impact assessment (impact evaluation or outcome evaluation) and efficiency assessment (p. 54).⁸

Fourth generation scholars' perspectives and approaches vary but follow a pattern of internalizing evaluation. Many of these techniques are harshly criticized among conventional theorists. However, their work has helped to strengthen the capacity and growth of organizations in every sector, including those in less developed countries (e.g. empowerment evaluation) and in dynamic, unstable, unpredictable environments (Preskill and Torres 1999: xix). The approach offered by Preskill and Torres (1999) places the evaluator in a position of mediator in identifying the needs, questions, and assumptions of each team member. The inquiry design uses this information to define the focus of the evaluation then links it to the data collection methods most likely to produce sufficient, trustworthy, useable information (p. 98). This phase resembles many of the approaches discussed earlier as it supports collaborative efforts while acknowledging the

⁸ Compare this typology to those of Chen (1996) and Scriven (1996) proposed or debated during the *American Journal of Evaluation* forum on the nature and scope of the summative-formative divide.

need for “systematic, planned and rigorous collection and documentation of data [essential for] true evaluative inquiry” (p. 99).

While many frameworks have been posed for how best to design and carryout an evaluation, these different theoretical stages and ideal forms for practice show the dynamic and diverse background of the evaluation field, its orientations and procedures. Controversy persists over what constitutes legitimate and “best” practices, but today—in the literature at least—the lines have blurred toward wider acceptance and propagation of using multiple methods or the most appropriate methods to meet the demands and constraints facing program stakeholders and the context they work within (Greene and Henry 2005:345; Bergman 2008). As the evaluation field and its scholars have tended toward pragmatism and plurality, in practice the demands and expectations from consumers and funders have neither mirrored such sentiments nor adjusted to the “variations, combinations, and adaptations available for creative and practical situational responsiveness” (Patton 1990:39).

THE ACCOUNTABILITY MOVEMENT

Over the last ten years, more organizations are delivering public services than ever before. In the field of development, this includes the large-scale multilateral and bilateral initiatives conducted by the UN and World Bank as well as the small nonprofit and non-governmental organizations standing on the front lines of the public and private sector’s response to major social problems (Smith 2005). Within this environment, development funds are relatively scarce and competition is high. Thus, as the number of social interventions have escalated, so too has the demand to account for these limited resources. Each of these organizations is “under increasing pressure from funders to demonstrate that they are being effective and achieving measurable outcomes” (Carmen 2010:256; Wenar 2006; Newcomer

2004). This environment reflects a survival of the fittest mode where funders' attitudes towards organizations are to "show me the evidence, or else" (Mueller-Hirth 2012). Introducing what Victor Murray (2005:347) and others coined the "accountability movement," which refers to the belief that organizations and their staff "should be more accountable to those they were created to serve and those who provide the money to operate them."

This movement can be seen in all sectors, including school systems involved in mandated state testing and surveying to meet various accountability requirements (Sutherland 2004). Furthermore, development-specific evaluation has become a booming business in response to demands for "charities and foundations to serve as responsible stewards of the public's generosity" (Festen and Philbin 2007:2; Bambherger 2000; Wenar 2006; Clements, Chianca and Sasaki 2008). As Mueller-Hirth (2012) notes, monitoring and evaluation are nothing new to this sector, but the difference is that now evaluation has arguably become a development activity in its own right (Everitt and Le Mare 2012).

Accompanying this shift is an external and internal emphasis on results-based approaches and a reliance on impact assessments (Thomas 2010:543; May 2012). These types of evaluations are concerned with the final results of interventions and focus on long-term outcomes attributable to specific programs (Leeuw and Vaessen 2009). "Whereas some of these accountability efforts have been compliance based, focusing on rules, processes, and procedures, most efforts have been performance based, focusing on outcomes and results" (Carman 2009: 257; Appleton-Dyer et al. 2012; Patton 2011:246). Therefore, organizations like the Network of Networks on Impact Evaluation (NONIE), The World Bank's Development Impact Evaluation (DIME), Massachusetts Institute of Technology's Poverty Action Lab (J-PAL) and the

International Initiative for Impact Evaluation (3ie) have helped to promote and guide efforts for conducting sound impact evaluation (Thomas 2010:544).

Yet as Thomas and other researchers showed, controversy has followed this shift centering once again on the most appropriate methods for assessing impact. Randomized Control Trials (RCTs) have increasingly become popular in recent years advocated by some as a gold standard, superior to all other types of evidence and the “new frontier for the collaboration between science and development policy” (Banerjee and Duflo 2009; Duflo 2012). Others argue that RCTs should hold no special place in the hierarchy of evidence (Deaton 2009; Easterly 2009; Cartwright 2007). Moreover, while these studies have comparative advantages in addressing particular concerns and needs, they also face serious limitations and ethical dilemmas. As reported by Maria May (2012), RCTs are bound by context, can only provide answers to smaller, program-specific questions, cannot easily be combined with other studies, and can only tell you whether or not something works not why. Additionally, the benefits of RCTs are often reduced when applied to real-world field settings as opposed to the ideal conditions (Barrett and Carter 2010; Thomas 2010; Bell and Aggleton 2012).

For these and other reasons, the Networks of Networks on Impact Evaluation (NONIE) guideline on impact assessment suggests that “no single method is best for addressing the variety of questions and aspects that might be part of impact evaluations” and that a “more complete picture of impact” is possible when combined in a way that utilizes each of their respective strengths (Leeuw and Vaessen 2009: x). Such quantitative methods include quasi-experimental techniques (e.g. “pipeline approach,” propensity score matching) and regression-based techniques (p. xiii-xiv). For example, the “pipeline approach” compares groups who have experienced the intervention with those who have been selected but not yet started. Leeuw and

Vaessen recommend this approach for programs that roll out gradually. The accountability movement has accelerated the availability and refinement of methods that can provide better evidence on results and development effectiveness, answer questions on whether development interventions make a difference and how cost-effective they are—each with the goal to “ensure that scarce resources are allocated where they can have the most development impact” (p. ix).

PROBLEMS WITH THE ACCOUNTABILITY MOVEMENT

The accountability movement is fundamentally based on the assumption that if funders require organizations to measure and report on outcomes then organizations will learn from this process, adjust their programs accordingly, and thus, provide more effective and efficient services (see Carmen 2010: 260). Within the accountability movement, many federal and large philanthropic funders have heightened their demand for evidence and subsequently increased funding and guidance of those efforts (Carmen 2008; Development Assistance Committee (DAC) 1991; Hatry, Wholey, and Newcomer 2004; Grasso 536).⁹ Consequently, the number of evaluations reported by organizations seems to fall in line with the movement’s theory of change. According to Carmen’s work, two thirds of nonprofits are doing some type of evaluation, and 90 percent of the organizations surveyed by the Innovation Network evaluated their work (2012). Among that number, 79 percent conducted outcomes evaluation, 65 percent conducted before and after measures, and 48 percent conducted impact evaluation (Morariu, Athanasiades, Emery 2012).

At first glance, the above actions support the logic of the movement. However, an increase in performance-based evaluation doesn’t necessarily mean they are effective. In fact, Carmen argues that the accountability system is not working, which she attributes to fundamental

⁹ See the 2005 Paris Declaration on Aid Effectiveness

flaws in its theory of change and implementation. The reports denote positive trends from the movement's efforts, but a closer look reveals wide variability in the quantity, quality, and relevancy of evaluations funded and conducted across the sectors (Thomas 2010; Carmen 2008; Wenar 2006). While many organizations have embraced evaluation and performance measurement, empirical research has shown that others struggle immensely, some just go through the motions, and others reject it altogether (Carmen 2010:259; Alaimo 2008; Hoole and Patterson 2008). "Although the rhetoric about performance-based accountability might be universal, participation is not" (Carmen 2010:268).

Numerous scholars have documented the reasons behind this shortfall in spite of the accountability-oriented environment. Resource constraints rank highest among organization's struggles. Even though the requirements have increased universally, funding and technical assistance have not (DAC 2010; Newcomer 2004; Carmen 2010). Many organizations within the social sector do not have the budget to allocate additional resources for evaluative activities, and few have the luxury of hiring an external evaluator or full-time evaluation staff. Thus, many rely on team members with insufficient knowledge in evaluation concepts and design. This lack of expertise contributes to the confusion and anxiety associated with meeting the rigorous methodological standards and conflicting requirements from the patchwork of funders that organizations resort to for survival (Carmen 2010; Mueller-Hirth 2012). Additionally, already overworked staff sometimes view evaluation as a resource drain, distraction, "nonessential activity for which they have little time" (Torres and Preskill 2001:389; Carman and Fredericks 2008; Thomas 2010).

While organizational capacity constraints affect the number and quality of evaluations conducted, they also contribute to the evaluation's overall relevancy and utility—starting with

the misconceptions over what an actual evaluation entails. Carmen found that what many organizations consider as evaluation and performance measurement really “amounts to reporting, monitoring basic management practices, and, at best, internal self-evaluation” (2009:264; James 2009). These organizations engaged in all sorts of activities perceived as helping them stand accountable, but not in the actual activities that would help improve the program and its performance.

Even for those who do conduct evaluation, it is assumed that once findings are presented, an organization’s staff would know how to use the information and use it to make appropriate changes (Preskill and Torres 1999:xviii; Thomas 2010:546). However, Murphy and Mitchell (2007) found that nonprofits were more likely to use the evaluation information to report to funders rather than for organizational learning and decision-making. Non-use of recommendations is a problem oftentimes because evaluations are irrelevant to the immediate reality of the program. The findings hide in reports too wordy and long for anyone to read and program personnel are rarely consulted on the design and implementation. Thus, the results are often perceived as incompatible with day-to-day operations and priorities (Mulwa 2008).

These issues relate to another flaw within the movement that Carmen calls “confusion among targets,” which refers to the disconnect between what the movement expects to happen and how they expect it to happen (2009:268). The end goal doesn’t match the means to get there. In a sense, the movement suffers from goal displacement where an emphasis on rules and methodological exactness has become the end itself rather the instrument to greater organizational effectiveness (Scott 1992: 324). In its vigor to scientifically demonstrate evidence of impact, the sector has lost sight of the benefits and power associated with the different types of evaluation and methods for inquiry (Weiss et al. 2008). Instead of using the approaches best

suited to decision-makers and operational needs, the focus shifted to producing information that satisfied external wants. “Much more important to most states than empirical evidence of treatment effects obtained through random assignment evaluations was information about how these programs actually operated in the field” (Greenberg et al. 2000: 379). Focusing on outcomes from the work can often result in a loss of understanding the process that led to them (Appleton-Dyer et al. 2012:541). Moreover, whether an organization is conducting process or outcome evaluations and learning or accountability-focused evaluation, more high-quality information is of little value unless it is used (Thomas 2010).

THE USE AND INFLUENCE OF EVALUATION

Use is a core construct in the field of evaluation because every evaluator has an “interest in the nature and extent of their works impact” (Kirkhart 2000:5). “Many if not most evaluators accept the idea that, at least in part, the merit of their work—the success or failure of their evaluation efforts—can be judged in terms of whether and how an evaluation is used (Henry and Mark 2003:293; see also Torres and Preskill 2001).

Use generally refers to the practical impact an evaluation and the information it produces has on the program being evaluated (Alkin and Taut 2003; Johnson et al. 2009). Under the use paradigm, scholars have identified categories for the different ways evaluations are used.

Instrumental use is the direct action and *conceptual use* is the learning that occurred as a result of the evaluation (Henry and Mark 2003). *Process use* refers to the action or learning that takes place because someone participated in the evaluation procedure. In this sense, the evaluation process not the findings is the source of change. *Symbolic use* is when a person uses the results to persuade or convince others of a position (Lederman 2012). In recent years, scholars have debated over the gaps and shortcomings in the use literature and its ability to account for the

wide spectrum of influence of evaluation findings and processes. Primarily, Kirkhart (2000) argued that use suggested purposeful action, when in reality evaluations have the power to produce effects by “intangible and indirect means” (Lederman 2012:161). A shift was proposed in terminology from *use* to *influence*. The influence framework captures effects that are multidirectional, incremental, unintentional, non-instrumental as well as those represented well under *use* such as unidirectional, episodic, intended, and instrumental consequences.

In Kirkhart’s “integrated theory” for evaluation influence, there are three key dimensions: source of influence (process or results), intention (intended or unintended), and timeframe (immediate, end of cycle, and long term) (Appleton-Dyer et al. 2012). Henry and Mark (2003) expanded the theory to include three levels of analysis for potential consequences (individual, interpersonal, and collective) and the underlying mechanisms that could lead to changes in attitudes, motivations, and behaviors. Mark and Henry (2004) expanded the list of mechanisms in the 2003 article and organized them into four different kinds of processes that may mediate influence and the potential outcomes that signify that influence.

These mechanisms and outcomes relate to other influence models that look into the different characteristics of organizations and the evaluation that help increase or decrease usage or perceived benefits of the evaluation. These attributes often include the perceived quality of the evaluation, the participation and investment of key stakeholders, relevance, timeliness of reporting, as well as contextual factors (see Appleton-Dyer et al. 2012; Lederman 2012; Johnson et al. 2009; and Henry and Mark 2004). Mark and Henry (2004) broke down each of these aspects into a model made up of evaluation inputs, activities, outputs, mechanisms, intermediate and long-term outcomes (p. 46). This conceptual model presents the pathways that link evaluation change processes to outcomes and its ultimate goal: social betterment (Henry and

Mark 2003: 293). The authors suggest applying and modifying this framework to specific contexts. In response, this thesis examines the influence of evaluation for a specific type of organization working in a specific context.

EVALUATION FOR A KEY DEVELOPMENT ACTOR

The literature presented above is applicable to myriad organizations working towards social betterment. In the past, the majority of evaluation research focused on educational or social welfare programs within the global north. In light of the calls for “Developing a World Perspective” by the American Evaluation Association (1995) and the push to better track progress against poverty, more studies have been conducted concerning the effectiveness of programs “dealing with world-wide problems” (Snyder and Doan 1995:142; Ravallion 2013; Thomas 2010) Yet, today international aid evaluation is primarily initiated by funding agencies and donors, and mostly among those programs supported by large development institutions such as the UNDP, USAID, OECD, and World Bank among others (Grasso 2010; Carmen 2008). While these agencies design and implement their own interventions, they frequently partner with already established international non-governmental organizations (INGOs) working in strategic locations and sectors such as Oxfam, Care, and Save the Children (Wenar 2011; Baños-Smith 2006). Monitoring and evaluation is often funded and conducted within these organizations as part of their partnership agreements (Mueller-Hirth 2012).

However, the development field is littered with actors of all shapes and sizes. A diverse set of organizations implement social programs and projects. Among them are the small, community-based NGOs and nonprofits that reside in the private sector. This study focuses on these shoestring, “mom and pop” organizations started by passionate entrepreneurs who see a problem in the world and want to fix it. Often forgotten in development and evaluation

conversations, these organizations have increased in number and reach in the last two decades with thousands working domestically and internationally (Nelson 2007). As dedicated social change agents, they exist in every sector, wiggle their way into remote corners of the world and work with scarce resources and limited skillsets in management, development, and even in their technical service-area. Within these organizations, they heavily rely on voluntary efforts, suffer from high turnover and often lack strategic direction due to their reliance on the whims of grants and fundraising. Additionally, in their quest to adapt and survive the sector's harsh realities, a disparity often exists within these organizations between the stated goals and the "real" goals that govern operations on the ground (Scott 1992:52). Yet, despite the setbacks and struggles, they hold great power for developing new, creative ideas and are important, resilient and extensive actors in poverty alleviation (Weiss 2013; Minear 2002).

Just as the effectiveness of the accountability movement on larger NGOs is debatable, even more questions arise over the progress of the small, community-based nonprofit actors in responding to these demands (Carmen 2007; Henry and Mark 2003). Little is known about how well these actors are doing in meeting their goals or bringing about the intended change (Nelson 2007; Naidoo 2004; Ramadan 2013). However, it is evident within the few studies conducted that small NGOs encounter significant difficulties when attempting to evaluate their work (Swindle 2011; Cox, Blust and Heaton 2013). Similar to the challenges described in Carmen's work, these organizations face capacity constraints to meet the standards of conducting sound, meaningful assessments of their programs. Moreover, evaluation is often low priority for these organizations—seen as a luxury—or impossible given their status in a perpetual phase of development and uncertainty. Even those who have some type of evaluation conducted struggle to incorporate any of the information or recommendations provided through the research

activities (Swindle 2011; Aksel and Baran 2006). As Briedenhann and Butts (2005) recommended for rural tourism projects, an inexpensive, feasible and friendly means for conducting evaluations must be found for these specific development actors given their prevalence within the field.

MEET PEAT: THE BYU PROGRAM EVALUATION AND ASSESSMENT TEAM

The BYU Sociology Department, in conjunction with the Kennedy Center for International Studies, created PEAT in 2009 to meet the need for evaluation among development organizations that would otherwise not conduct them given their resource constraints. PEAT is an internship program that trains students in evaluation techniques, and then contracts their services to nonprofits to perform assessments of their programs. While PEAT's limitations are evident, including the relative inexperience of students and concerns with methodological rigor, the intent is to present evaluation as a learning experience for both student and nonprofit. PEAT offers evaluations to various types of organizations as an option for accountability and program improvement while simultaneously providing students the opportunity to enhance and apply skills learned in the classroom (King 2008; Smith 2011). For the organization, the emphasis is on evaluation research for program review and development (King 2008); for the student, PEAT provides the chance to combine theory with practice and to gain marketable skills in data conceptualization and management.

PEAT interns are selected based on completion of required methodology courses within one of the social sciences as well as other characteristics such as language skills, leadership, and prior research experience. Students are then required to enroll in an applied social research methods class that emphasizes evaluation theory with hands-on activities. They also take an additional preparatory course that focuses on proposal writing rooted in the nonprofit's

knowledge and requests and theoretical and methodological literature. Students are then placed in the field to conduct research for a period of two to three months. While on assignment, they remain in contact with professors when the need arises for feedback and advice. However, the structure of PEAT facilitates a high level of autonomy for the interns and allows them to immerse themselves into the daily life and culture of the country.

Since 2009, PEAT has sent more than 30 students to a variety of regions working with 16 organizations that serve marginalized and impoverished populations. As Cox, Blust and Heaton (2013) have shown, PEAT has been a resource to nonprofits who lack the capacity to conduct quality evaluation despite the challenges faced when attempting a cooperative relationship between university and nonprofit. Yet, in addition to merely providing these actors with a practical option, PEAT directors are also concerned with the *value* PEAT evaluations hold for the organizations. This study explored the perceived and discursive influence the PEAT evaluations may have on the organizations and their staff.

RESEARCH QUESTIONS

Given the issues identified in the literature outlined above and the opportunity to address them in this case study of PEAT, the research questions include:

1. What do the student-led evaluations reveal about these organizations and the challenges they face to evaluation?
2. What was the nature of the evaluations?
3. What type of influences result from the evaluations?
4. What does the evaluation process reveal about how best to provide evaluation for this specific type of organization?

METHODS

This thesis utilized a phenomenological and case study approach to explore the value of student-led research for small nonprofits working in developing contexts. The primary case under study was BYU's Project Evaluation and Assessment Team (PEAT) internship program. The author analyzed the experiences associated with evaluations conducted by PEAT interns for nonprofit clients working in a variety of countries over the course of four years: 2009-2013.

The case study strategy identified the unique benefits of the PEAT program by comparing the details and themes across each assessment. The phenomenological approach sought to understand the individual experience of both student and client in the process and outcomes of conducting these evaluations (Creswell 2007). Utilizing both of these approaches provided a more nuanced understanding of the program, its interns and the nonprofit's relationship with each other and the evaluation. They also provided insight into the evaluation process (e.g. its inputs, activities, outputs, and outcomes) conducted in varying but similar developing contexts, which helped identify the evaluative information best suited for the organizations and their needs. Qualitative and quantitative data was collected through in-depth interviews, surveys, and document analysis.

SAMPLE

Primary data were collected from the staff members of the PEAT's clients and the student interns. The sampling took place at the organizational and participant level with each determined according to purposeful maximal, theoretical and criterion strategies (Creswell 2007:127).

Organization interview and survey respondents were selected among PEAT's clients, who often work with marginalized populations, in developing contexts, and have limited resources to conduct evaluation and/or a variety of types of evaluation (see Table 1 for more

information on the nonprofit respondents). The sample of organizations was also narrowed down according to the type of assessment provided, the general experience (both positive and negative), and relationship with PEAT. Interviews were conducted with the five program personnel who had knowledge about the PEAT evaluation and who worked closely with the PEAT interns (Carmen 2007; Lederman 2011). The author also interviewed 15 past PEAT interns for a total of 20 interviews (see Table 2 for more information on the student participants). The sample of student and nonprofit interviews represented the range of relevant experiences and consequences for the individuals and nonprofits involved (Creswell 2007:75; Henry and Mark 2003; Corbin and Strauss 2008). Interview guides for these interviews are discussed in the next section. Additionally, a survey was administered to at least one representative of the nonprofits that have worked with PEAT since 2009. In total, this amounted to a survey sample of 15 respondents.

[Table 1 about here]

Each participant was informed of the purpose of the study and its confidential and voluntary nature. In the end, the sample only included those participants who consented to participate based on these terms. IRB approval was obtained from the sociology subcommittee. As a former student director of the PEAT program, the author had access to each of the past students, the clients, and deliverables offered to them. Access was further secured because of the reciprocal partnership between PEAT and its clients—many of the organizations are still in contact with PEAT.

INTERVIEW PROCEDURE AND INSTRUMENT

Individual, in-depth interviews were conducted in a semi-structured interview process (see Appendix A and B for nonprofit and student guides). Data were collected from April 2013

to February 2014. Each interviewee gave verbal consent before the interview proceeded. Throughout this paper, pseudonyms are used in place of participants name to ensure confidentiality (see Tables 1 and 2 for a list of pseudonyms for the nonprofit and student respondents). Transcriptions and notes are digital and saved in a secure folder only accessible by the immediate research team. The interviews were captured using an audio recorder and typed notes. Interviews were held in BYU's sociology department conference room, meeting rooms at the NONPROFIT's office, and the private homes of some of the staff members.

The interviews with the NONPROFIT personnel were conducted face-to-face or on the phone since some of the partners do not reside within close proximity of each other or the researcher. The interviews lasted from 45 to 120 minutes. The NONPROFIT interview guide consisted of open-ended questions concerning the organization's past experience with evaluation, expectations for PEAT evaluation, perceived impact or value-added, and their experience working with students. The guide was also adapted as new things were learned, problems arose, or when clarification was needed.

The interviews with past PEAT interns were conducted face-to-face, over the phone, and through Voice over Internet Protocol (VoIP) services such as Skype and Google Video. They lasted between 30 minutes and 120 minutes. The author interviewed 15 past PEAT interns of the 41 total, representing the range of relevant intern experiences that reached a saturation point for major themes (Corbin and Strauss 2008). The topics in this interview covered their evaluation plans, how the plans compared to what they actually conducted, their overall field experience, process of analysis and providing recommendations, preparation, and the perceived benefits of the evaluation for the nonprofit. Often this last topic is more than perceived since many of the students have kept in contact with the organization and observed the changes first-hand.

SURVEY PROCEDURE AND INSTRUMENT

After the majority of interviews were conducted, the author constructed a survey instrument that would help pinpoint the various types of influences and perceived outcomes of the evaluation. Since the research topics were adjusted along the way to address new issues that emerged in the course of the interviews, the survey helped to capture some of the points not emphasized as heavily in the interviews. The survey was created using Qualtrics and consists of about 20 questions. The first part of the survey asked about the organization's background with evaluation, specifically if they had engaged in it before the PEAT evaluation and the barriers they face in implementing evaluation. The next section went into the attributes of the PEAT evaluation, including the purpose and whether they felt it met that purpose.

The last section covered the evaluation's influence based on the models presented by Henry and Mark (2003), Mark and Henry (2004), and Appleton-Dyer et al. (2012). The last section has three subsections. The first concerned learning and attitudinal outcomes, and the second section dealt with action-oriented and behavioral changes that took place in lieu of the evaluation. The last section asked them to rate how beneficial the evaluation was from 1 to 5, with 1 being "highly beneficial" and 5 being "detrimental;" its overall quality from 1 to 5 with 1 being "High quality" and 5 being "Poor quality;" and how satisfied they were with the information provided from 1 "very satisfied" to 5 "very dissatisfied (see Appendix C for the full survey instrument).

DOCUMENT ANALYSIS

The author analyzed the proposals, reports and deliverables created by the student-evaluators prior to and after the evaluation. The purpose for this method was to cross-check the interview and survey information provided by the NONPROFITs and students. The author

analyzed the types of assessments, purpose, questions, methodology, analysis, findings and recommendations. Analyzing the documents helped triangulate and gave a more comprehensive understanding of the evaluations, their characteristics and benefits. It provided a deeper look into the student's experience, and the program's structure, stage, and context (Neuman 2011).

ANALYSIS

The data were analyzed using multiple levels of abstraction moving from the particulars of each subcase to more abstract cross-case themes that represent the value of the program and general experience of both student and NONPROFIT (Creswell 2007). First, the interviews were transcribed and the analysis began by reading through each interview, the survey results and report looking for information that helped describe each separate experience and context—gathering the “facts” of the specific case and evaluation (p. 163).

Next, the author used NVivo to code the data. A coding scheme was created to reflect the following categories: the facts about the case (PEAT and organization), facts about the evaluation, the student and organization experience, the evaluation's influence, and the student characteristics. The categories and their features followed the definitions provided in the conceptual frameworks of Henry and Mark (2003); Mark and Henry (2004); Appleton-Dyer et al. (2012); and Johnson (2009). The relevant themes and quotes from the data were then organized into a spreadsheet according to these categories.

Lastly, the author conducted a cross-case theme analysis looking for similarities and differences across all cases. Each of these phases included a search for significant statements, themes, patterns, and meanings that encapsulate the experience associated with a PEAT evaluation, the mechanisms that produce certain changes within the organization and the

contextual and personal conditions that trigger negative or positive experiential and structural outcomes.

LIMITATIONS

This study faced a few but important limitations. First, some methodologists have suggested that the inquirer be a “stranger” to the group or context one is studying as it may raise issues over power, threaten accuracy, and introduce biases (Agar 1986; Creswell 2007; Neuman 2011). As one of the six who started PEAT, the author was essentially conducting the study in her “own backyard;” She was a student director for three years, interviewed multiple interns for the position, and conducted a PEAT evaluation in 2010. This unique position affected her approach to the study and the interpretation of the words and actions of the interviewees. The author entered the study with certain expectations and assumptions given the intimate knowledge of the program, and its organizational processes, strengths and weaknesses. However, this intimate knowledge was also a strength as it aided in understanding and navigating the subtle variations and complexities of each experience and the program overall from a student, director and researcher perspective. Additionally, because the study is primarily of a qualitative nature, the personal experience and contact proved an “important part of the inquiry and critical to understanding the phenomenon” (Patton 1990:40).

Second, some of the evaluations included were conducted from one to three years ago. Certain measures were taken to reduce the risk that the interviewees could not remember the evaluation. The participants were notified in advance and given time to review reports and meeting notes beforehand. They also used these documents during the interview to help trigger memories of the details and instances. For the interns, this was a unique experience, and many of them have copious personal notes and photographs documenting their time in the field.

Additionally, some of the organizations and students have received or conducted multiple PEAT evaluations. In these instances, they reported on the most recent one and then their thoughts on the experiences and PEAT as a whole.

FINDINGS

The findings are divided into five sections that correspond to the study's research questions: "Facts of the Case" (research question one); "Facts of the Evaluation" (research question two); "Organization Experience" (research question four); "Evaluation Influence" (research question three); and "Student Factor" (research question four). These sections discuss the factors that Mark and Henry (2004) and Appleton-Dyer et al. (2012) outline in their conceptual models, seen in Figures 1 and 2 below. The following information describes the findings from the student-evaluations that correspond to the mechanisms that these scholars suggest mediate the effects of an evaluation.

[Figures 1 and 2 about here]

Each section builds on the other to show that the organizations are continually developing and could most benefit from continual learning, problem solving, and improvement (Preskill and Torres 1999). An important aspect of this is inquiry focused on illuminating the character and quality of the program as opposed to jumping into impact assessment prematurely (Mathison 2005). While PEAT's clients felt the need to focus on results, the students were best prepared for and the program in most need of process evaluation. Despite setbacks in evaluator expertise and quality, the wide array of influences that resulted from the student evaluations indicate that description and interpretation of process are valid forms of evaluation and appropriate for these organizations given their stage of development and contextual realities. Additionally, the evidence from these evaluations demonstrates the power and role of evaluation as a tool for

learning, informing deliberation, and catalyzing change-agents in addition to generating modifications in actions and programs. Lastly, while students brought various skillsets to the evaluation, the organization personnel indicated that practicality took precedence over rigor and that the students were qualified to provide the needed work in unexpected and unique ways. Figure 3 is a modified conceptual model of Mark and Henry (2004) and Appleton-Dyer et al. (2012) that includes the influential factors supported by this data.

[Figure 3 about here]

FACTS OF THE CASE

This section explores the features introduced by Appleton-Dyer et al. (2012)¹⁰ and describes in more detail the characteristics, functioning and evaluation behavior of PEAT's clients. The data summarized here answers the first research question and sets up the context of the evaluation. The primary findings support the literature showing that PEAT is in a unique position to provide quality evaluations to organizations that care about having a positive impact but struggle to show that impact because they are continually developing, work in complex environments, and rarely use evaluation as a developmental tool. Such detailed information about small nonprofits is scarce within the evaluation influence literature and critical in understanding how to align evaluation to their distinct features.

The PEAT program was created in 2009 by three professors in the sociology department at BYU. In addition to their careers in the field of sociology, the three faculty members have extensive experience in applied and theoretical evaluation and international development. Yet, each hold a unique perspective and approach to these fields, which has proven influential in

¹⁰ The Appleton-Dyer et al. (2012) model introduced the role of public sector partnerships and their individual characteristics in an evaluation's influence.

getting the program off the ground and expanding it over the last five years. They provide support as they mentor the students, teach the preparation courses, and represent PEAT in all related administrative, promotional, and procurement opportunities. During the first two years of the program, three student-directors were hired to help bring their vision to life. In succeeding years, this number has been reduced, but student-directors continue to assist in recruitment, training, and logistical support.

The program has two primary targets: student interns and the client organizations that they assist with evaluation. Within the university, it is housed under the international development minor and its corresponding internship program. Qualified students must have prior international and research experience. And while it is geared towards advanced students who plan for a career in development, all majors are encouraged to apply. The internship has become more competitive over time with applications increasing from 15 in the first year to about 80 for 2014. In addition to gaining experience, students can earn credit in related courses that count for the minor and other majors throughout the university. They are also offered a scholarship to fund their travel, tuition and in-country expenses. See Table 2 for descriptive information on the participants and program for the years 2009-2013. PEAT has historically targeted small, shoestring NGOs.¹¹ The data from this study confirm many of the characteristics described in the literature review. The following, however, provides a closer look into PEAT clients' mission, model, stage of development, culture and resources.

[Table 2 about here]

¹¹ There are some exceptions: one school, one corporate social program, one ministry program, and one humanitarian arm of a church.

PEAT's Clients: Context and Mission

The client organizations operate in a variety of international regions such as Eastern and Western Africa, Southeast Asia, Central and South America, Middle East, Eastern Europe, and Oceania. Most often they work in rural areas, but some of them have offices and projects located in both urban and rural areas. The standard of living of the service population varies from program to program in terms of the average education and income level as well as the health status. The common thread running is the intent to work among vulnerable or marginalized populations (e.g., domestic violence victims or Navajo students at-risk of dropping out of school).

[Table 3 about here]

The missions of each organization can be described as either outcomes-focused or values-driven. The former are more problem-centered and aimed at specific outcomes while the latter are focused on process or how the work is to be done (Patton 2011:250). For example, one organization described its focus as helping children who have been orphaned by war and AIDS and improving the lives of women and families. All of the mission statements mention the intent to address poverty and its diverse manifestations (such as high child mortality rates or unemployment). However, the majority of these organizations is guided by the participatory or bottom-up narrative and, thus, emphasizes their value-driven approaches. In their mission statements, descriptors such as sustainable, practical, holistic, village-centered and empowering are common with the idea of helping people help themselves.

The organizations have lofty goals, but they attempt to meet them in complicated, even chaotic environments (Patton 2011:109-110). Such complexity is manifested in the ways these organizations are structured and in the situations they operate—both characterized by

unpredictable and turbulent patterns. According to one student director, the interns walk into “pretty strange and pretty difficult circumstances” (David).

Organizational Model

In poverty-ridden contexts, problems are intertwined and often solutions are unknown or under-developed. While these organizations’ missions set high goals, in some cases, there is no definitive formulation of the problem among staff members, nor is there consensus or consistency among the strategies to address them. One student-evaluator described the organization he worked with this way: “They themselves don’t always know the right answer, and don’t always know what they want” (David).

For many of the organizations, multiple directions are pursued at once, as evidenced by the number of projects running simultaneously with differing goals and tactics. One NGO director described this situation, “So yeah, our projects are kind of all over the place, and we treat them individually” (James). For others, there seems to be no specific direction or plan in particular. Another NGO director described the organization’s approach as needing more clarity:

We do need to be more organized about this. We do need to be more clear about what we’re trying to accomplish. We do need to point all of our energy and resources in this direction instead of a shotgun approach. I think each time we’ve gone there, we’ve looked around and done whatever seems good at the moment. It has been more of an opportunistic approach rather than planned. (Ruth)

Most of the organizations’ models include partnering with numerous local organizations or individuals. Some are large, reputable agencies like the World Health Organization while others are small (e.g., a village pastor). As one NGO executive director explained:

We’ve transitioned away from putting an emphasis on starting our own projects. It’s very much finding those local entrepreneurs, the change-makers, and enhancing what they’re doing. So identifying what’s already working, what’s been started by a local, what’s already culturally and contextually appropriate and expanding that (James).

A collaborative model like this has strengths and weaknesses. The same director pointed out, “I’m trying to [catalyze] efforts that already work...the people on the ground say this is our

mission, this is our plan, and this is how it works. So I'm not too worried about messing things up." For this NGO, sometimes its partners even sit back and "want us to take control and come up with ideas for them." But for those organizations that accomplish their mission directly through the efforts of partners on the ground, aligning agendas and expectations can be difficult.

Despite the differences in partnerships, a common theme found among PEAT's clients is the limited presence of primary stakeholders in the location of interest, which often leaves these relationships fragile and unchecked. Many of the organizations are structured with the majority of staff members and board of directors located at headquarters in the United States who only visit the areas intermittently for relatively short periods of time—if at all.

A lot of the directors of these organizations are here in Utah—they're not there. They go over there once a year for like a couple weeks, take pictures, put it up on the website, and they get reports from the country directors. (Past PEAT student-director, David)

One student also observed:

[The founder] is an expat. She is a [nationality] who lives in the US, and then [name of staff member] is native to the US. So the directors were not located in [country]. So the orphanage was mostly run by [founder's] family. (Austin)

Similar to the organization described above, many of PEAT's clients appoint at least one native staff member to oversee project implementation or office maintenance. Yet, the autonomy and power of these individuals to make programmatic decisions is often restricted or minimal.

As one student evaluator described:

I don't know that I got the feeling that there was a lot of micromanaging. I think the feeling that we got is that [staff person] needs more autonomy in terms of hiring, HR policies, and things like that simply because she's on the ground...I think there were things that we feel it would be good if she had that realm--like, if she had the ability to choose the final say on some things. (Stephanie)

Some organizations lack a year-round presence in the country altogether. Consequently, during their active season, they lose precious time and resources rekindling relationships, getting updates on changes, and backtracking on projects. These conditions expectedly widen the

distance between key stakeholders and the people with whom they aim to collaborate and empower.

Additionally, many of the clients structure their organizational model with a dual purpose. In addition to their social-change missions, these organizations also aim to nurture globally-minded citizens by offering volunteer opportunities in the form of humanitarian missions or internships. They target families, students, young professionals, and special interest groups. For some of the organizations, most of their efforts are implemented via volunteers, and thus helping them is an important and valid facet of their purpose:

When I got hired here, they said they needed someone who doesn't care about money and who doesn't care about volunteers, cause guess what, it's development. But at the end of the day, I care about volunteers. I've managed three teams on the ground, and I recognize that they've invested a lot...I am really invested in them having a really substantive, resume-building experience, and I'm interested in them meeting the goals they've set for themselves. (James)

There are distinct concerns related to this type of model, which many of the PEAT interns observed in their evaluations. For the organization that works primarily through volunteers and placing them with their in-country partners, three main issues came forward: limited effect due to time constraints, limited concrete skills, and differing expectations. These are illustrated in the following evaluator observations:

Those organizations locally are working on long-term projects and it's really hard for the team to pop-in, jump into meaningful projects with these NGO's in which people have skills to contribute, and then leave three months later. (Ann)

Another thing, the culture shock of [country] is pretty insane, and so if interns are only there for six weeks, which most of them were, it takes two weeks to figure it out—it's 115 degrees outside—you're useless for the first two weeks, you've finally learned what's going on by four weeks, and [now] you have two weeks to do something. (David)

When they [the partner organization] think of an international volunteer coming, they think intern: 3 months in our office, adding value, probably doing things the volunteer wasn't thinking of. The volunteer is thinking, "I'm going to play with orphans." (David)

The issues with the various models as described above are further amplified within the organizations that have expanded to multiple countries. What works in one context may not work

in another. As one NGO director put it, “[we] look different in each country,” and for organizations with resource constraints, expansion may look good on paper but often at the expense of quality and effectiveness.

Stage of Development and Resources

For many of these organizations, their approach seems piecemeal and unpredictable. Even though some of them have been around for two or three decades, the progress comes across as disjointed, ever evolving, and even limited. While this is not the case for all of PEAT’s clients, despite years of service, these organizations appear to be in the stages of *exploration* and *exploitation* in the adaptive cycle to organizational change (Westley and Miller 2003). These phases are characterized first by widespread experimentation where creative ideas lead to a number of failures; “uncertainty is high, predictability is low, for things are in flux and it’s not at all certain what will result” (Patton 2011:205). In this phase there is consistent exploring, reorganization and renewal. As one student-evaluator described:

I don’t know where they are now, but I know in 2010 they were really having kind of a “soul-searching” introspective time...and our report really reflected quite strongly to a lot of the board members and even some of the executive directors about what was really going on: Is that really what they want to have happen, and where do they want to take the organization? (Ryan)

For many of the organizations, they are only a few years into it, and as one NGO board member stated, “[We’re] just so new to this.” Similar organizations are asking the fundamental questions such as this NGO board member posed: “What is our real mission? What are our assets, resources? What are our programs? How might we organize, reorganize, whatever, to respond more”? (Ruth)

In the next phase, *exploitation*, programs are taking the more favored ideas and piloting, improving and strengthening them. Other PEAT clients have progressed into the *conservation*

phase in which they are focused on scaling projects and efficiency, yet still constantly evolving in regards to their approach and goals. One program coordinator stated:

I think it definitely, it tends [to] keep morphing. We are really working at a systems approach at the school...that actually allowed students support services to spend more and more time on collaborative services in the community, preservation and really work in our meetings on what support the particular student, or groups of students, need that were beyond the scope what we normally offer to the general population of the school. I would say it's more specific as far as what our role is but it is also expanding more services out into the community and re-coordinating those services." (Carol)

This level of development is not standard for the majority of PEAT's clients, which is frequently related to the resource constraints they face as small organizations. They are considered small for a number of reasons. As nonprofits, their budgets are limited, relying primarily on fundraising, grants, or the generosity of their board members. This example seems consistent with many other organizations' experiences:

Their resources were so scarce. I know that [the founder] was doing it piecemeal. With [staff member], I think it was the same kind of deal...they sent funds when they were available; some months there weren't funds for the organization. So they just had to make it by. (Austin)

These organizations are also considered small in terms of the number of staff members, which is still minimal even for those who work in several countries as opposed to one remote village. The amount ranges from as little as two to fifteen at their headquarters and zero to four acting as in-country directors or staff at various times throughout the year.¹² Given the small numbers, these staff members are often overworked; required to handle multiple tasks and initiatives at once. "They are super busy. The guy I interacted with is in charge of way too many things just because he's good at it 'cause it's a small organization [but] he *can* handle a lot" (student evaluator, Kevin). Staff and volunteer turnover is also a rampant within these organizations due to limited funds.

¹² This excludes the three larger organizations in this study that share very different, more foundational structures and scope.

Some of the staff members *can* manage quite a bit and have skills that are a good fit for their positions. The founders are often considered “social innovators” who are action-oriented, visionaries with a can-do attitude (Patton 2011). They are determined and many are committed to excellence—however they define it. The impetus for these innovators was often a personal connection to the location or target population. For PEAT clients, these experiences ranged from serving a religious mission in the country to being a native originally from the community but who has since left. Given this background, the staff understands many of the small nuances of the population or context, such as their history, their troubles and the “full truth” behind their experiences. Such experiences and knowledge qualified many of the staff for the work they do in these organizations.

As described by Francis Westley and his colleagues (2006), many people get involved with nonprofits because they feel a sense of duty; they are called to this work. From their personal experience, they saw the problems and “felt compelled to act” (Patton 2011:246). Consequently, most decisions are rooted in these value-laden motivations. However, despite their intimate relationship to the issue or target audiences as well as their charitable desires, many are ill equipped to manage nonprofits and/or implement their agendas¹³ in important ways. According to one intern, family members of the founder—who have little knowledge of how to manage finances or the program in general—handled the day-to-day operations of the organization. This reality was further captured by one of the student-interns:

“In my experience, one of the biggest problems with international development is that they are very well intentioned—maybe even very well educated individuals—from the developed world that want to help in a development context but they don’t have necessarily the concrete skills that are needed or wanted or that

¹³ There are important exceptions, however. Many have an applicable college degree(s) or work in the public sector in related positions. Others were hired specifically for their job given their professional experience and skillsets (e.g., the coordinator for the school dropout prevention program).

would really make a difference. In so many NGO's and in so many nonprofits, so much work is going on and I think so much of that really is...it's not as effective as it could be. It's not as responsive to needs of communities or as cost-efficient. (Ann)

Mirroring this sentiment, a board member of one NGO expressed, "We are a small band of idealistic women, and even though [the founder], our president, is extremely knowledgeable about Africa and the economic and social issues faced by African women, she's never done an NGO." A lack of dependable funding also prevents these NGOs from attracting or retaining their highly trained staff members. And those who stay and work for marginal pay are often distracted from their work or forced to let other more pressing or lucrative endeavors take precedence.

Learning Culture

The learning culture varied between organizations as well as between staff members within them. For those who viewed themselves as open and learning-oriented, much of it was just rhetorical with little or lackluster institutionalized approaches to learning or feedback. The mechanisms for learning often amounted to bimonthly meetings between employees and the board to keep a pulse on the organization or sporadic phone calls to the in-country staff to report the latest news. One organization devised an evaluation template, but the country directors showed little interest in implementing it. Other resistance to learning was seen in an overall attachment to the founder's or a particular staff member's initial ideas, goals, and/or plans. For one organization, their culture changed over time from one of resistance to acceptance:

It went from a culture of wanting to only show our best face to the world and making sure that [in our] development work, even if there are issues that we should talk about, we don't want to expose that. It went from that kind of culture to one that I was very adamant about—making sure that all of our flaws in development are exposed. We were very candid with our board members and ourselves about our programs. (Adam)

Each of the features of an organization's learning culture contributes to their general approach to or acceptance of evaluation.

Approach to Evaluation

The views of evaluation ranged from being anxious to intimidated and even disinterested. A lack of investment is highly attributable to misunderstanding its purpose or what it takes to implement one. One evaluator stated: “I don’t know honestly how seriously they take it. I don’t know if they have the mentality to respond to evaluation...with the mindset of employees and organization, I don’t know how much they value M&E” (Henry). While some people clearly do understand its value, one coordinator of a large program added another reason why some are not as receptive to or hesitant about evaluation:

I’m just being honest. I think there are people who see it as, hey, we do this when it is required of us...I mean some people see it as valuable and this is how we are going to tell whether we’re being effective and improving and some people see it as a necessary evil. (Carol)

When asked his opinion on how NGOs view evaluation, this student director provided a similar description:

Evil maybe is the wrong word, maybe a necessary annoyance. We have to do it, and we’re told it’s a good thing, but we don’t really get much value out of it. There were a couple of the partners who loved any type of evaluation cause they were just those pure people that really wanted good feedback. They really wanted to do a better job. Others are just busy and they’ve got a lot to do, have a lot of weight on their shoulders, and they really care about helping people and so the evaluators just kind of get in the way of what really matters to them. (David)

Two members of the board of directors from one organization also indicated in the survey that evaluation is annoying and requires too much effort on the part of the NONPROFIT. These approaches are influenced by people’s past experience with evaluation or overall unfamiliarity with it.

Experience with Evaluation

Six of the organizations reported they had an evaluation conducted in the past, but the majority had not (9). Among those who had past experience with evaluations, half reported that evaluations were conducted by an internal staff member and the other half were conducted by an external evaluator. All six reported having some sort of needs assessment of the service

population and problems they were addressing. Four had an assessment of the program process and implementation, and then half had an assessment of outputs or short-term outcomes. The only nonprofit that had had all types of evaluation but an impact evaluation in the past was the large religious-affiliated humanitarian agency. The only one that reported having an impact evaluation was one of the smaller organizations, and an external evaluator conducted it. An open response in the survey revealed that three of the organizations wish they had gained a better understanding of the status of long-term outcomes, more theory and analysis that reflected the effectiveness of their scope, and more than just background information. They wanted more actionable recommendations.

The qualitative data revealed more about why some of the organizations hesitate to conduct evaluation. Some don't really understand what it entails, its purpose, or how it can benefit them. When asked about the barriers faced with implementing evaluation, one NGO responded that while they understand the point to evaluation, they did not perceive it as a "necessity for an NGO" because they seem less structured (Ruth). Others don't look at it as favorably because they are "surveyed out"—it's commonplace now and sometimes the benefits aren't that apparent. Others avoid evaluation because they are hesitant about receiving negative feedback in general. For many social innovators, they don't want to hear about all the problems they have. What they do is personal; they sacrificed a lot in choosing to work for the nonprofit, and they often view it as their "child:"

The other thing I wonder about, not knowing about evaluation or NGOs in the beginning, there is a sense when you put so much of your heart and soul into it, that you don't want someone else coming in and telling you what's wrong with it. Because then it's attacking you. (Ruth)

One student evaluator made an important observation after conducting two PEAT evaluations:

I feel like NGOs don't get a real sense of what they do to people. Like, for the two NGOs I've worked with, I had a whole different idea of what they actually did in the country. They talk like they do all this, but you get there and it's like [they've] done nothing really...there's these small projects that are probably benefiting lives in this area, but it's not. They convey it in a way that makes it look more massive than it really is. And they don't give a clear picture. (Ellen)

For those who do not conduct evaluation or do them often as they would like, the following are the struggles that these organization's face when trying to change this fact.

Barriers to Evaluation

The survey showed that the primary barriers organizations face in implementing evaluation are a lack of money and time. The third most common struggle is lack of expertise. None of the organizations reported that they did not understand the point of evaluation or face a lack of support from other staff members. Four organizations indicated that they are not able to find granting agencies willing to fund evaluation, and another large organization indicated that one barrier is not being able to find a qualified evaluator (refer to Table 4 to see a list of the barriers).

[Table 4 about here]

The qualitative interviews mirrored these responses regarding funding, time, and expertise as major barriers. Concerning the lack of time, one student observed, "it's not like [the director] is incapable of putting together and managing focus groups, but she is so beleaguered by everything she has to do for her job that it just wouldn't have had happened" (Sarah). Another prevalent issue within these organizations, especially when attempting to assess impact, is the lack of baseline data or records needed for tracking different indicators or other measures of success. In perfect summary of these struggles, one NGO board member stated:

I think that in my view, the major problem has been that we're a volunteer organization, and there's so much anxiety and pressure, to try to raise money to sort of maintain the momentum of [the organization], that evaluation seemed to take a back seat. (Ruth)

FACTS OF THE EVALUATION

This section describes the nature of the student evaluations, answering the second research question. It walks through the evaluation inputs, activities, outputs and quality as outlined by Mark and Henry (2004). Many of these elements correspond to Appleton-Dyer et al.'s (2012) "Evaluation Attributes" component. This section also describes contingencies in the environment (Mark and Henry 2012) and other contextual factors (Appleton-Dyer et al. 2012), since they directly and indirectly affect the nature and quality of an evaluation. Knowing what the students actually did for the organizations, the perceived quality of that work and the context they worked within is essential in extrapolating suggestions for future practice from lessons learned in these situations.

The relationship between the client organizations and PEAT usually developed through previous affiliations with BYU. Some of these organizations were recommended to PEAT, others had worked with a professor of the program in the past, and many of the program personnel were BYU alumni themselves. When asked what they expected to get out of the evaluation, the organizations' responses ranged from basic organizational development to identifying what works and why. One organization specified that they wanted help clarifying their mission, sharpening their focus and identifying their niche. The majority of organizations went beyond basic needs assessment specifying a demand for someone to provide a snapshot of what's happening on the ground and follow-up on projects. Often this resulted in them wanting the students to track outputs such as how many pints of water were coming out of their well, or as this executive director explained:

I love how they're just so willing to scout out anything...Because sometimes that's really all they need to give. Accurate feedback is what I really want. Sometimes it's just if something is being used. For example, we built a school in Tanzania, and it wasn't used for a year and a half. But sometimes I just need them to go see—like showing up at the bakery that we started in Peru, showing up randomly and seeing if it's being used, getting a headcount of students, etc. (James)

One organization director specified that he needed the student-evaluators to identify gaps in their program delivery and then come back and provide recommendations for how they could improve in the future. Other program directors wanted feedback on the projects *and* participants. Two directors wanted this in particular because they felt they are unable to obtain accurate information. One example of this:

The reason why I like having students come down here is because you're a third party – if I go and talk to the people, they're not going to tell me because I'm the source of help...and they're afraid if they complain to me I won't help them anymore. But you're not me and they can open up to you and they did...so I just need you to be my eyes and ears and give us that data and that information. (In-country director talking to Katie)

Four of the organizations specifically mentioned that one perceived value-added was the expertise. Given the fact that they have minimal staff members with the needed skillset, they viewed the students as a resource, bringing “something to the table” (Ryan). Yet, for one organization, they wanted this feedback and perspective, but at the end of the day, were not as open to the possibility that it could be negative. “We wanted somebody to come back and say, ‘You guys are doing a fantastic job, and you're really on the right track.’ You know” (Ruth)?

Evaluation Inputs

The following section discusses the inputs of the evaluation: the resources allocated to the evaluation (expertise, time, communication), the decision-making setting, micro-politics, competing information, and the evaluation readiness and accountability requirements of the organizations. In this study, communication and expertise were the most critical inputs to the evaluation's success; the students portrayed adequate competence to perform quality evaluations according to the organization's needs, but often communication and conflicting information hindered the process and its potential for success.

Expertise in this study refers to the competence of both student and NONPROFIT personnel. Concerning the student, Johnson et al. (2009) found that an evaluator's professional and cultural competence is important to an evaluation's influence (Appleton-Dyer et al. 2012). The specific preparation the student completed as a prerequisite to conducting the evaluation is discussed in the "Student Experience" section located in Appendix D. The following description includes the applicable skillset the students offer outside of the training they received through PEAT.

The students majored in a variety of disciplines, namely sociology, English, communications, political science, international relations, and public health (see Table 2 for the complete list of majors). Many of the students appreciated their background and particularly mentioned it as helpful. For instance, most knew how to write and utilize a literature review and incorporate previously tested measures and concepts into their designs. Another student found their background and experience with writing to be helpful, especially when creating the proposal and reports. One student studied business management, which he perceived as important since he was evaluating a program aimed at helping communities become more self-reliant.

[My] business background I think helped with this because a huge part of the program is about businesses. So I was familiar with all the terms. I knew the end goal and where they were trying to go with a lot of it. And in the report I was able to bring a lot of insights just from the business world as to what might be useful in being able to organize and teach those things, or for other adjustment they might make. (Kevin)

The majority of the students finished a minor in international development and/or language studies (e.g., French, Portuguese, and Spanish). The students mentioned this as helpful, providing them with a critical mind when it came to understanding and assessing processes within a development context. Additionally, about half of them were able to communicate in the native language. One NGO director specifically appreciated the students having a background in

international development. Referencing one student in particular: “She knew more about development than a lot of country directors because it’s hard to find people that have that background. I told her, ‘be prepared to be a CD because you know more than anyone else on that team’ ” (James).

All of them had some prior methodological training either from their social science classes (e.g., statistics, qualitative/quantitative data collection and analysis) or in preparation for PEAT. One student who studied psychology specifically mentioned incorporating his experience into the evaluation:

Let’s put it this way, I came up with constructs, and I then came up with several operationalized questions to measure that construct. And then I came up with scales to somehow define if you more or less agree with those questions. That’s all very psychological. (John)

Almost all of the students had been teaching assistants to professors and many of these were related to methodology.

I felt [PEAT team member] and I were pretty prepared. He was the TA for 405 [Multiple Regression Analysis and Social Science Computing] and was the quantitative genius, and I was the TA for 404 [Qualitative Research Methods] and so it was kind of like the tag team. I don’t think there was a methodological question in our survey tool building that we had difficulty with. (Austin)

Overall, the students referenced their coursework and training in helping them become more analytical. Many felt prepared to identify best approaches, critique validity, know when they are doing things wrong and how to adjust as well as find patterns in the copious amounts of data:

I sold my soul as a student to learn everything I could about microfinance and about [the country]. And it helped me when we did have issues and concerns to be able to say really quickly what can change and what can’t change. Yes, we can change these questions because we understand they are culturally insensitive, but no, we cannot give up on this because we need, like this is a really important, defining aspect. (Samuel)

One student connected this with the ability to work in a different context:

I felt that I excelled at navigating the cross-cultural context. This helped me realize that many of the instruments were too long. It also helped me build a quick rapport with the partners who doubled as my guides. It also helped me to start to realize that many of the stakeholders were not being completely honest with me. (Ryan)

Concerning expertise, the organization personnel provided significant contextual experience along with many technical skills. For example, a few of the in-country staff for one organization came up with applicable research categories and helped design the methodology. One student mentioned how impressed she was with the country director's ability to network and maintain relationships. One student, however, found that the NGO had no prior experience with evaluation and expected the student to guide them throughout the entire process. While many of the staff members had some type of training or experience, they were too busy to really assist in the design and methodological decisions.

In addition to expertise, other human, financial, and logistical support was provided to help the evaluations happen. As mentioned before, PEAT helped to fund the student's flight and tuition and then collaborated with the NGO to help pay for program fees, housing and other materials such as cell-phones and incentives (e.g., one program paid for gift cards and food for the interviews and focus groups). Three of the interns were provided with interpreters and in-country contacts that could check for cultural sensitivity and methodological accuracy. The organizations also provided significant support in accessing the target population. One director described, "I laid a lot of the groundwork for them and the credibility, especially the people they needed access to—certain staff members and teachers. Since I was the go-between, the liaison on that, I think that helped them with their credibility" (Carol). Other students experience less support than they were promised—some more serious than others. For one team, the evaluation completely fell through, but for others, it was merely the undependability of rides, in-country contacts, and interpreters or translators.

Time is another resource of evaluation. The total time dedicated to the evaluation ranged from nine to 18 months, depending on how early the relationship was initiated and how fast the

students were able to complete the report. Most data collection took place over the course of six weeks to four months. While one organization worried that they would not be able to keep the students busy during that length of time, most of the students felt the timeframe was inadequate.

Given the variability in start and end time, the students referred to feeling frustrated over different aspects. Some felt they didn't have enough time to interact with the NONPROFIT beforehand or secure the relationship and evaluation focus. Some needed more time to get to know their context and to adjust to the organization staff members' temperamental desires. For at least ten of the students, significant changes were made to the scope of the evaluation once they arrived in the country, which meant they lost precious data collection time reconfiguring the evaluation. While in country, students mentioned instances of wasted time because projects were not underway, permission to access certain populations was delayed and even the day-to-day waiting for people to show up or rides to come. In general, students felt some dissatisfaction with the timeframe, often because it was rushed due to constraints with the program or PEAT structure. According to one student:

We just couldn't do more. I think the time that we were there that we could be doing something we were doing something. So I think we used our time well when we could have. I just don't know if there was something more we could have done when we had down time. I guess I'm not really satisfied unless I feel that I did the best we could. So part of me wonders if we could have done more...it was just slow.
(Stephanie)

Another influential input on the impact and experience of the evaluation was communication. Between PEAT and the client in general, communication suffered and was not as good as it should be. In many cases there seemed to be a communication breakdown. Students used descriptors such as conflicting, unreliable and inconsistent. Poor communication prior to departure often led to students expecting different things when they got out into the field. Four interns mentioned that the stakeholders were insufficiently made aware of what the students were

able to do and what they could do for them. In the most extreme example where the evaluation ultimately didn't take place, the team described the lack of communication as the main reason:

There was a big disconnect. I think primarily it was a lack of good communication before the fact. I don't think that was entirely our fault. There was just a lot of trouble trying to get on the same page with the people in [the country], because [the in-country contact] just disappeared. I felt like we showed up and no one knew who we were, why we were there, or what we were supposed to be doing. So it was really difficult to get anything going. (Samuel)

However, some students described the communication as positive in that the nonprofit would respond to their questions or provide feedback in a timely manner. Four students felt there was a lot of collaboration between them and the NONPROFIT throughout the entire process. So while the majority of students felt that communication could have been better overall, about nine of the fifteen students who were interviewed stated that the communication was adequate enough for them to complete an evaluation and three of them felt that communication was not an issue at all.

The communication between the student interns and the in-country staff was limited prior to departure, which may explain the in-country's staff initial lack of investment compared to those at headquarters. One student shared this experience while in the field where one staff member would agree to something and then not follow through.

So talking with the in-country director about how to solve these issues, she would often put off and say, "I don't know. Let me talk to [founders] and I'll get back," and that was just a ploy to put it off. So we'd be like, "No, talk to us." So, we had to be direct with her and say, "We need an answer like next week and if we don't then we are just going to move forward." (Austin)

Communication between the student and the service population seemed overwhelmingly positive, especially as the students reassured them they were truly interested in their needs, concerns, and perspectives. One additional finding related to communication was less positive: there was a discrepancy between what the staff and participants in the field said to the evaluators and what they conveyed to the staff or board of directors in the U.S. This was a significant hindrance in the influence of the evaluation overall, as well as in the program's effectiveness.

One student also found that communication was poor between the organizations and the partners they work with on projects.

Decision-making setting. The decision-making setting refers to the cultural, political and informational aspects of the organizations involved in initiating and sponsoring the evaluation (Mark and Henry 2004; Johnson et al. 2009). The first feature covers the organization's level of support or commitment to the evaluation.

For one evaluation, the program coordinator did everything she could to prepare for the students to come out, including getting them permission to work with such a sensitive population. Acceptance, understanding of their purpose and support from other staff members increased over time, especially as they interacted with the interns more. From the interns' perspectives, the staff bent over backwards to help them. Another organization paid for the intern to go with them to install solar panels prior to the evaluation that would take place six months later. Other organizations showed their support in waiving their program fees to accommodate the interns and providing funding for in-country travel, translators, supplies, and housing. Sometimes the in-country staff showed support when they provided cultural insight, helped design instruments, and do what they could to make sure it was completed.

For some organizations, the support for evaluation is high, but they were limited by constraints like funding or time. In some cases, they wouldn't conduct evaluation if it weren't for PEAT. For one organization, although they were open to evaluation, the priorities seemed a little backward. "The fact that they could help with org development, build an effective organization, and needs assessment was not an inconsiderable factor, but they were way down the line from, "They were going to help us get grants" (Jolene).

One student described that the support varied between staff members. One was “happy and willing to work with him” and wanted to do more research, another lost interest as time went on, and another didn’t take it as seriously as he would have hoped (Henry). With another organization, one in-country staff member showed support to the evaluators and was candid throughout, but behind their backs, questioned everything to the board. One of the large programs that PEAT worked with has federal evaluations quite frequently. While the coordinator is very supportive of evaluation, some of her fellow staff members feel annoyed or jaded by it, which influenced their initial response to the PEAT evaluation.

The support also varied over time. At first, many showed high interest and made grand promises to the evaluators about what they could provide, what the student should expect, or how they would help in the process. And in a few cases, they did not follow-through, act upon these promises, or the support and commitment changed as other tasks gained higher priority towards the end. Additionally, the issues discussed with communication seemed indicative of a lower level of support or commitment that five students mentioned.

Micro-politics. According to Johnson et al. (2009:385), the political climate entails the extent to which decision makers are dependent on external sponsors, internal rivalries, budget fights and power struggles. The definition applied here is a little broader, referring to any conflicting or politically driven motivations or actions. This section also only covers those that have not been mentioned above or are different applications to a similar idea.

Some of the issues that came up between PEAT and the NONPROFIT have already been mentioned such as the internal staff wanting or expecting different things out of the evaluation. As one student described, the organizations she worked with the past exaggerated what they did in the country and their impact, which did not match what she observed or experienced on the

ground. Other issues were that the organizations are too close to their work and take it quite personally when they receive less than positive feedback. These members feel the evaluators don't understand what they've put into their work and, thus, are limited in their ability to decipher the effects that have really taken place. For the one evaluation that didn't happen, the team attributed its failure to political undercurrents that blocked the evaluation indefinitely.

Conflict was evident between some stakeholder's assumptions and the evaluator's insights and conclusions drawn from the evaluation. Two student evaluators specifically described a situation in which the NONPROFIT's intended focus did not match the needs of the people as observed by the evaluator. "I think they had in their head ideas of 'this is what we should do for projects' and it wasn't necessarily in response to needs that had been posed" (Ann). Another NGO currently focuses on trauma healing, but as one NGO member realized from the evaluation:

People get their backs up and say, "What do you mean trauma? I don't feel any trauma. I just want to know where my next meal is coming from" and were sort of insulted by this idea of "Let's take care of you poor traumatized." They sort of think it's patronizing and irrelevant. (Ruth)

She learned that they needed to implement more pragmatic programs and approaches as opposed to the one the founder insisted on. Similar discrepancies were also found in what the in-country staff said to the evaluators and what they said to the NGO board members. The in-country staff seemed not as candid when talking to the board for fear of hurting their relationship with them, and instead, told the board what they wanted to hear.

Other issues with open communication were found between the organizations and its target population. Some student evaluators observed that the participants were afraid to speak freely in front of the organization in fear of losing their aid or resources. In one situation, the organization intimidated the population from speaking out. In another situation, when deciding

between evaluation approaches, the student-evaluator found that in “one of the communities there had been a lot of internal strife and past issues with [the organization]” (Tyler).

Competing information. Johnson et al. (2009) define this category as referring to any available information that may contradict the evaluation data. The data presented in this section covers situations when any information given or received was in conflict with each other. In some of the situations, the contradictory information further hindered the quality of the evaluation and its applicability to the organization’s needs.

In terms of conducting the evaluation, some of the nonprofits desired less than or different from what would be considered an adequate evaluation. One nonprofit desired information purely on outputs as opposed to taking a critical look into the program, how it produces those outputs and the potential outcomes that stem from them. As mentioned before, many of the organizations couldn’t make up their mind about what they wanted and changed several times throughout.

It was several months of preparation where...they would say they wanted something so I would prepare something on that and when I would come back with it they would want something completely different. “Let’s do something instead.” So I would reiterate. It wasn’t for improvement; it was a start over iteration because they changed their mind. I don’t think they were completely cognizant that they were changing what they wanted, but it was more like, “Today this sounds great.” (Kevin)

Sometimes the nonprofit would give little direction at all because they wanted the student to do what they needed, when in contrast, the student felt that the evaluation should adhere to the needs of the nonprofit instead. Other students observed further gaps between what the organization needed and what was required of them by the PEAT program.

There’s this difference between an academic research paper, which is what [the professor] wants us to do and an evaluation for the humanitarian agency. We get two different signals from both groups. The humanitarian agency wants just descriptive data like “how much do they eat,” which you can’t really write an academic paper about that and that’s what [the instructor] wants and so, we’re like “Ahhh, we don’t know what to do.” (Katie)

Sometimes conflicting information came from staff within the organization, such as the U.S. director and the in-country director (Henry). Other times it from the partner organizations contradicting each other in what they told the students. As one student described:

I worked with two [organizations] on this evaluation, but they both had these different agendas that were conveyed to us. If we had known they were trying to do these other evaluations in Nigeria with these solar units, if we had known more of what they were doing there, we could have tapped into that and gotten a sense to build from there and improve it. (Ellen)

Evaluation readiness. As mentioned in the “Facts of the Case” section, many of these organizations are unorganized, inefficient and in the beginning phases of organizational development. Some are still trying to figure out how to manage finances and their staff. In some organizations, turnover is constant. This directly impacts their evaluation readiness. About half of the organizations were at one point unsure of what they wanted or needed and, thus, how an evaluation could specifically help them.

They didn’t really know what they wanted at first. They knew that an evaluation would be cool and that it could be helpful, but they didn’t have a particular outcome they wanted to glean from it. That’s fine. Part of the job is really going back and forth and hashing out what we can do and what is really going to help them. (Kevin)

While some were unsure of the direct impact it could have, some didn’t quite understand the range of types and their purposes. Their desires and expectations were limited due to their lack of training and exposure to evaluation. One offered little help in the preparation phase, expecting the student to walk them through it without providing much insight into their prior work, experience and knowledge concerning the program.

So we’ve come up with this idea, but then there was no feedback from the other NGOs and when we’d ask the one, who has been there plenty of times and knows her NGO well, would sit back and say, “Well you’re the experts so you tell me how to do evaluation.” Even though we are supposed to know it, we can’t do it without them. We need to know their mission and objectives and that was never really clear. (Ellen)

Some evaluators were hindered in locating participants and past projects because the organizations kept inadequate records. There was also little baseline data to utilize in an impact evaluation. Some of the projects hadn’t even finished yet when the organization asked to know

its impact. “They were just getting started. It is hard to do a full impact assessment of something that is not completed” (Tyler).

Accountability requirements. Six of the organizations mentioned that one of the purposes for working with PEAT was to meet accountability requirements. In the few comments that came out of the interviews on this subject, only some have funds set aside for evaluation, and these were from the organizations that worked with other partner organizations. Some conducted evaluation regularly, but these were part of the three larger PEAT clients. Those that had large grants from the government or other agencies mentioned evaluation as required, but it often was not thorough (e.g., a call from a consulting group every six months) or did not utilize qualitative methods. One organization stated that the previous evaluations only used quantitative methods, and while necessary to receive further grants, did not provide rich data into certain aspects of the program (e.g., the student’s perspective or what program features were helpful in raising grades).

Evaluation Outputs

The findings in this section describe the evaluation’s activities and are synthesized for all students and by organization. While the previous section provided information on the resources available for the evaluation, evaluation outputs describe what took place. The first subsection addresses the planning phase, followed by sections on the evaluation purpose, methods, analysis, findings, recommendations, and report dissemination.

*Evaluation planning.*¹⁴ The students were selected into the program in September and October. By November and December they had their assignments and had met with the nonprofit staff at least once. There are a few exceptions to this timeframe as some were only able to talk to the staff via phone or Skype. The majority of students (13) went on the internship with a partner,

¹⁴ Most of the planning phase is covered in the student and organization experience.

however, nine of the students went to the country alone even if there were other students working for the same organization in a different country.

Most of the students had the preparation course Soc 301: Applied Social Research Methods, and the latter two cohorts had an additional class (Soc 390R: Special Topics in Contemporary Sociology). In these courses, the students learned approaches and methods for evaluation, gained experience in hands-on assessment activities, had access to the professors, and planned their evaluations.

The experience of planning the evaluation seemed quite different depending on the student and the organization. For some, communication started early and was frequent and open. However, even for these students, the majority reported that the staff struggled with knowing what they wanted the evaluation to accomplish. Thus, the plans went through multiple rounds of reconfigurations and edits.

In ideal circumstances, the student worked with the staff to create the proposal, design the instruments, and structure the research project all prior to departing to the country. Yet, for many of them, once in country, they made substantial changes to the proposal. This was most often due to contextual or funding constraints, misinformation or misrepresentation, or at the request of the in-country director. One scrapped her entire plan and joined with two other PEAT interns on a different evaluation. Another group was asked to change the proposal and then ended up not being able to complete the evaluation. Other students changed their plans minimally such as adjusting questions in their methods for cultural reasons, sampling techniques, and who and when they were interviewing. Often the changes improved the plans to better meet the specific needs and wants of the nonprofit once adjusted to the current reality of the organization on the ground. One student's experience provides an example:

I think the professors wanted more of an academic style of method but when we got into country, we recognized it needed to be just results. Like we need to be able to give the government something they could work with. They didn't want to read through a huge list of everything we had learned socially and academically. They wanted to know what they needed to change this year to have better results next year...I think some of the questions were better and some were worse. I think they were better because they actually addressed the issues the government wanted. They were more tailored, more specific. I think some of [the changes] were worse because they didn't provide academically, researched-backed answers that we wanted. Because some of the questions, the government was like, "You can't ask these," which actually made the evaluation less effective on some issues. (Samuel)

Evaluation purpose: nonprofit survey responses. On the survey, the question was "What was the purpose of the evaluation (select all that apply)?" Thus, the respondent could indicate multiple purposes for the evaluation. The most notable finding from this survey question is that the majority of directors hoped the evaluation would determine impact (14 out of 15) and half of them indicated that the evaluation met this purpose. The other half indicated that it somewhat met this purpose. All but two of the fifteen said they wanted it to provide suggestions for improvement. Among them, eight said the evaluation met this purpose and five said it somewhat did (see Table 5 for a comparison between the indicated purposes and whether the respondent felt they were met).

[Table 5 about here]

Eleven indicated the purpose was to provide information that could enhance project management and decision-making. Six felt that it met this purpose and five felt that it somewhat did. Of those eleven organizations, eight indicated on the survey that they used the evaluation findings directly in decision-making and three did not (see Table 6 for a comparison between the purposes, if the respondent felt they were met, and if the results were used in decision-making). The three who did not are the same organizations who indicated that the evaluation somewhat met this purpose. Each of these organizations is small, relatively new (the oldest was seven years), and seemed the least satisfied with the evaluation based on the interviews and open-ended responses in the survey. Two of the student teams reflect similar dissatisfaction, but for one, this

lack of use and dissatisfaction did not match the student’s perspective. However, the respondent for this last organization was not the primary contact for the evaluation.

[Table 6 about here]

Ten nonprofit respondents said the purpose was to apply lessons and feedback to inform future projects. More than half of the respondents said it met this purpose while four said it did not. Seven said it was to report on the day-to-day operations and the strengths, weaknesses, and obstacle it encounters. Three stated that the evaluation met this purpose, four said that it somewhat met this purpose. These numbers show that the majority of the organizations thought the evaluation could provide evidence of impact, but they also wanted the evaluation to focus on process in order to inform decision-making (11 out of 15) and improve their programs (10 out of 15). Every organization personnel indicated that the evaluations at least somewhat met the range of purposes, and eight of the organizations actually used the evaluation in decision-making.

Evaluation purpose and type: interviews and reports. The interviews and reports also indicate a wide variety of purposes perceived by both the students and organizations. Many of them conducted multiple types of evaluations as well (see Table 7 for the amount of each type). Gathered from the available reports and interviews conducted, three of the evaluations were needs assessment with purposes such as: “Determining the met and unmet service needs of the women, both here in the shelter and in the community. We would like to better understand the current status of each woman, and the conditions in [the city] and the community surrounding the issue of domestic violence,” or “To understand the level and ways stigma is experienced by leprosy colony members” (from the reports of organizations four and eight).

[Table 7 about here]

Twelve of the 23 total evaluations conducted mentioned analyzing some aspect of the program's process (mission, theory or strategies, management, and implementation). The following is an example of one of the evaluation goals that covered both needs assessment and process assessment: "The goal of the assessment was to gain a holistic understanding of each of the 24 villages with which the [organization] works, including the different needs of the villagers as well as a better understanding of the various projects and programs the [organization] has implemented over the years" (report for organization one).

Eight of the evaluations conducted were an outcome assessment. In these, some were identifying indicators for "sustainability" of new projects and another the poverty level after a microfinance program. While many of the evaluation's purposes mentioned trying to identify impact of a project, only four specifically utilized methods that had control groups, before/after or pre- and post-test measures, or statistical analysis of secondary data obtained from past participants of different years of the program. One example from a report states, "In this paper I evaluate what effect [the nonprofit's] neighborhood organization project had on indices of social capital using OLS and logit estimates as well as estimates from propensity score matching" (report for organization six). Three additional types of assessment were conducted, namely cost-benefit analysis, Social Impact Assessment, and utilizing other types (such as process and needs assessment) to set up a baseline or evaluation templates to be used later by the organization.

Evaluation methods. Almost every evaluation utilized mixed methods (see Table 8 for list of types of methods). Sixteen conducted in-depth, formal and informal interviews, and often the interviews contained quantitative and qualitative elements. Eleven studies conducted surveys in the forms of questionnaires, quizzes and scorecards (e.g., scorecard to measure characteristics of partnership viability).

[Table 8 about here]

Six evaluations conducted focus groups, and often they were facilitated utilizing Participatory Rural Appraisal (PRA) activities (e.g., mapping and ranking) and the FAMA technique, which consists of participatory group discussions based in the Literacy for Social Change model promoted by Lynn Curtis. Almost all utilized data from formal observation (e.g., two class instructions per week) or informal observation from general field notes. An estimated 16 mentioned analyzing past organizational records or reports. In all cases, students mentioned utilizing questions, measures, and indicators backed by literature.

When possible, these methods were conducted in all of the participating villages, either all of the ones the organization worked with in that area or those that received a specific resource (e.g., solar panel, square foot garden, or micro-loan). Often in these circumstances, almost all of the community members were interviewed or surveyed (e.g., all of the children living in the orphanage). If there were too many projects in the area or participants, the students implemented some type of sampling technique. One team decided to evaluate three projects that were representative of the 50 projects the organization had running at that time. Another student used a random sampling technique to identify which classrooms he would observe.

To identify respondents, many used structured, purposive and/or theoretical sampling techniques. Some were able to maintain these plans throughout. For example, one team utilized quota sampling to select a homogenous group of participants based on race, gender, and grade-level, which would help facilitate the group discussion. However, for various reasons, many had to rely on convenience or snowball techniques. One student explained that this was all he could obtain since most of the participants were away from the colony during the times he was able to visit. Time constraints also limited three other evaluation team's selection or access to particular

individuals (e.g., mothers are the only ones at home during the day). Other student interns mentioned not having enough control over selecting participants (e.g., community leader directing them only to those who would give positive feedback and lack of adequate record-keeping to select them independently).

Access to these populations was often facilitated via local staff members, past project partners, or other in-country contacts (e.g., a local graduate student). These gatekeepers helped locate and recruit participants, find interpreters, translate documents, and sometimes conduct the interviews. Some of the students lived at the program site (e.g., with program participants or at the orphanage). This helped them gain trust and appreciation among the participants, and oftentimes, the intern's most valuable data came from informal discussions or observations outside official evaluation hours. Many of the students utilized an interpreter or knew enough of the dialect to conduct the activities themselves. In these instances, multiple checks were conducted to adapt questions for fit and understanding. Even when the students did not know the language, they learned some phrases to show respect (e.g., formal and informal greetings, and signs of appreciation or acknowledgement).

Evaluation analysis. For the qualitative data, most interns coded the data in some fashion. Some used software like NVivo while others hand-coded the data. Some used pre-made coding templates for guidance and others utilized grounded theory to develop their coding schemes. Two projects quantified their qualitative data for easy display and comparability with their other quantitative data. All of them looked for patterns or themes across the data and presented them in relevant categories for the organizations.

Most of the quantitative data was analyzed in a basic manner utilizing descriptive statistics, spreadsheets, and/or charts to report before and after numbers (e.g., number of

deliveries at different points of time). Two projects utilized more sophisticated analysis, such as OLS regression, propensity score matching, and comparing the average income totals for participants and nonparticipants by increased expenses and lost income due to being a participant in the solar project.

Evaluation findings. The findings documented in the reports and reported by the students in the interviews ranged from some positive and some negative or neither. The more positive information found insights, such as program participants having a good experience, the inexpensive projects produce better numbers than the more expensive ones, and that attending the program's meetings does help participants feel represented.

The more negative information sometimes reported on the status of projects such as the implementation not going as planned (e.g., the desires of the organization's volunteers didn't match the needs of the partner organizations; the price of their produce was below cost and they were losing money; the amount of light emitted from the solar panel wasn't enough to fill the entire room during a surgery held at night), or that the project was not working, or as one student described, "gone to pot" (John). Those that reported on outcomes sometimes showed that the organization's training was not as effective as hoped (e.g., the cooperatives still didn't know how to market their product; the method required a lot of reading and the people were illiterate).

The neutral data primarily provided information on the aspects of the program or important insights about the participants or context. For example, some provided insight into the leadership characteristics (e.g., communication and conflict resolution) that will enable the two communities to implement their cooperatives in a more sustainable way. Another evaluation documented information on all of the various programs and projects the organization had established in each village. The students provided detailed data on each village's economy,

health, and education plans as well as their wants and needs. One simply described the different types of organizations in the country and their needs to help guide PEAT's client when deciding with whom to work with in the future.

Evaluation recommendations. The recommendations were broad with some suggesting ways the organization could improve services, better target needs, adjust their mission, or improve evaluation techniques. As one student team stated in their report:

Many of these recommendations are easily implemented into the [organization's] current structure. Others require a fundamental rethinking of the organization's structure and logic model. As [the NGO] moves forward out of this period of transition, we hope that this evaluation can be a guide to a successful and productive future. (Ryan)

The majority of the reports provided a presentation of findings on the gaps, issues, status and insights into the participants and partners. Recommendations like this addressed the need for program changes and required minor adjustments, such as including training when implementing a new project; instituting consequences for showing up late to meetings; checking the status of the health center's records before installing a new solar panel; customizing the format on their evaluation surveys so they make sense in the local context; increasing communication; and having their evaluators in the future introduce themselves as researchers who were not connected with the organization.

Other recommendations addressed the need for larger adjustments to the program, such as changing their entire focus from trauma healing to community development or having a yearlong presence in the country. Other examples of recommendations that required substantial adjustments include extending the time volunteers are in the country, treating each country and project as a separate entity with specific missions and strategies, or reducing staff turnover by only hiring one committed director as opposed to four for shorter time-periods.

The recommendations that were more substantial were also less feasible, and as one student-director pointed out, less applicable. Many of the students lack training or experience in organizational change and consulting nonprofits on how to implement that change. Thus, they struggled to translate findings into direct suggestions for action.

I think students by the time they [left] had learned methods but hadn't actually learned how to take data and turn [it] into a recommendation that was implementable. All we are really trained to do is find data, analyze it and report. In terms of recommending stuff, it's like, "Well I don't know anything about organizational change. I don't know anything about how much this would cost if I suggested doing this...it was just like, "Well I saw this problem and I would do this, but I have no idea if you could implement that idea." (David)

One organization mentioned in particular liking the findings but struggling over the recommendations that stemmed from them. This struggle could also be attributed to the organization's strong and emotional attachment to their model and mission as well as the way in which they interact with the in-country staff. However, this insight is applicable given the struggles students face in structuring their recommendations. An NGO executive director also addressed this issue:

Sometimes their recommendations were shooting for the moon, and not always a possible reality. But that's okay. It always gave us something to strive towards, but they didn't always know our organization constraints. [It was more], "This is what we think you should do if you had a million dollars" type of thing. (Adam)

Report generation and dissemination. Many students struggled in attempting to write the report. Primarily this was due to a lack of time e.g., translating 50 plus interviews from Spanish into English. Students mentioned struggling with how to manage the large amounts of data, synthesize it, build a report framework to report findings, translate findings into recommendations and write a comprehensive but digestible report in the limited amount of time they could dedicate to it given their demanding classwork in the following semester. Here are two examples from students who struggled with this phase in particular:

I know we had been given, this is an evaluation report, but for some reason we really struggled with how to format it or how to write it. Cause we felt like the data spoke for itself. It was hard to know how to balance the quote with what we said. We didn't know how long to make it. We didn't know, especially 'cause our findings were all over the map. (Stephanie)

I think the length was something that was very daunting to me. I ended up with a 60-page report. It took a long time to compose that much writing. But I felt like I couldn't legitimately cover it any faster. (Ryan)

These struggles indicate why most of the reports were turned in much later than promised (some up to a year later). However, as reported by the nonprofit staff in the interviews and by the students, most of the students worked with the organizations afterward, telling them of their preliminary findings, insights and recommendations. Some students had multiple conversations in person or over the phone and Skype. One of the NGO executive directors described this process: "I picked their brains for hours. We just talked for hours, and I just wrote everything down, and it was awesome" (James).

In addition to these follow-up meetings, many of the students provided shorter reports or presentations on this preliminary information based on some formal and informal analysis conducted while in the country or immediately upon their return. Some of the students even worked with the staff members to produce the report going through a few rounds of iterations to make sure it addressed all of their concerns and desires. Since most of the students did this, many of them took a significant amount of time to compose a report that both matched the nonprofits' request and the requirements set forth by one of the PEAT directors. Many of the students reported expecting the nonprofit personnel to not even read their final reports. However, from the interviews it was confirmed (and the nonprofit survey data confirmed), that many of the organizations did in fact read them as well as glean information from them.

Evaluation Quality

This section describes the reliability and validity of the data produced by the evaluation. Given the resources and investment that went into the evaluation, it is important to understand whether the activities that took place were actually of high quality or not. In addition to the expertise of the evaluator, the quality of the evaluation design has been shown to increase the use

of evaluation findings or recommendations (Henry 2003). Since the majority of the data was qualitative, a high quality evaluation refers to whether it presented accurate and authentic information about the programs, their effects, and participant experiences.

Reliability or consistency. In particular, this subsection considers whether the data were able to paint a coherent picture across the different sources and time (Neuman 2011). Considering external reliability, most of the studies triangulated their information by using multiple types of methods. A good example of this is that one team checked the number of baby deliveries reported in their records against the numbers that the people stated in-person. In this case, the people reported higher numbers than what the records showed. Another team altered their methods to the specific respondent; if they felt more comfortable in a group or individual setting they would choose between having them join a focus group and conducting a one-on-one interview. Other evaluations showed consistency across sources, as they were able to compare perspectives from almost all participants of a particular project (e.g., those who had received a square-foot garden).

However, a few students acknowledged that there were discrepancies between interviews and the questions asked over time and between different respondents. For one student, they often had to describe the question in further detail to the respondent and each time that description changed according to the needs of the respondent (e.g., didn't understand the concept or what they were asking). Different definitions of a concept (e.g., empowerment) were not uncommon when the students were trying to assess the program's goals (e.g., empower individuals) or the effect (e.g., decrease poverty).

The third thing was operating within the cultural context. I was looking for things like empowerment. Are we actually empowering individuals to maybe achieve more of their life goals, to find resources to overcome poverty? However I defined it, I remembered distinctly that their definition of empowerment and my definition of empowerment were worlds apart. My definition of the term just did not necessarily fit them, and it was a struggle to really even get into that. (John)

Another student conducted the interviews himself in Spanish and adjusted the wording throughout the summer depending on the responses. One student adapted his approach as the summer went on, changing from a structured interview to a more conversational one, so that the respondents felt more at ease. However, these adjustments also helped further validate the authenticity of their responses as it increased understanding of the question being asked. One student stated, “The changes were simply for the sake of practicality. I think they are better because they better reflected the situation on the ground” (Ryan).

In terms of the credibility of responses, a fair number of students realized early on that the respondents had reasons to “lie” or say something other than the full truth. Often these motivations stemmed from having a “stake in making the organization look good” (Ellen), feeling worried that they might lose aid, or telling the researchers what they wanted to hear or what they knew was the best answer for general approval. For example, a student evaluating the sanitation habits of children in an orphanage realized that the children were reporting that they washed their hands with soap and running water, but the researcher knew there was actually no place to wash their hands. Another student noticed that some women living in a slum seemed to exaggerate their plight when he interviewed them as opposed to the graduate student who was native to the country.

On many levels consistency depended on the context. Many of the students lived with community members and spent ample time preparing food, eating at their table, and helping around with the household chores. According to these students, some of their best data came from those informal interactions. On the other hand, one student lived with the in-country facilitators in an urban apartment away from the rural community. He noticed that this had an effect on his ability to gather the most accurate information.

You know, I think I would have gotten more truthful answers if I had hung out with them maybe in the evening as they were just sitting around drinking checha and shooting the breeze and what not. I think if there were more informal opportunities for interaction I could have gotten into the real stuff. But since most of my interactions were formal, their responses were formal. (Kevin)

Validity or authenticity. Validity of these evaluations will refer to whether the data are an accurate, truthful representation of the experience in the field (Neuman 2011). In many cases, the students felt that they were able to gather a true portrayal of the organization and situation at hand. “For that organization, just within that organization, I felt like we gave a pretty accurate snapshot. I really do” (Austin). Similarly, other students stated that despite some setbacks, the evaluation accomplished the goal. “The lack of a scientific sample really worries me. But I still think that we found some ‘truth’ that led to some very good recommendations for the NGO. Although I don’t think that the results could be published, I think we accomplished the goal” (Ryan).

Some students were able to check this supposed authenticity. One made sure to conduct ecological validity checks. He talked with the organizations and the some other women in the area saying, “This is what I’ve been hearing from the people,” and they confirmed that they had observed that as well (David).

Sometimes accurate information was obtained more or less easily depending on the characteristics of the interviewer, or as Neuman (2011) described, competent insider performance. The inherent distant between interviewee and interviewer was sometimes increased because of cultural and language barriers. One student felt that, “A local of similar socio-economic background would have gotten more truthful answers” as opposed to a white woman from the U.S (Ann). In these cases, the distance was inescapable simply because the interns were “different” than the respondents.

I say initially any gap in communication and reluctance on the student part wasn’t because they didn’t try to help them feel comfortable. They are two young white girls, speaking to primarily Navajos, especially males. And all of our staff struggle with that sometimes, especially our Navajo males, where staff are

primarily white females. And so, the guys tend to be less verbal with us. But they [the interns] still got some perspectives, and some quotes from the guys...but I would say it was more, you look and sound different than me, I don't know you. (Carol)

Yet, as this program coordinator pointed out, the students were still able to gather information, and as she explained later, the students were able to adjust in certain ways to help close the gap. In this unique case, the intern evaluators were primarily speaking to students and, thus, could relate on some level.

Because if nothing else, we could say, "I've been in your shoes." I've done what you've done...I really like to relate myself to the interviewees, but I think the most I could say, we are students just like you. We talked about has anyone ever failed a class, [PEAT partner] would say things like, "We all have before." (Stephanie)

Other interns were not interviewing students, but were able to decrease distance by the way they interacted in the context.

One coordinator noticed that the student-evaluators took the necessary time to make people feel comfortable. Living among the participants helped reduce the outsider mentality. Others adjusted how they introduced themselves depending on their audience. Sometimes they used the word "evaluator" and other times it was, "Hey, I'm a student. I just want to learn" (Kevin). Some people took better to the student idea (e.g., rural areas, women living in the slums). Others were more receptive when the student presented himself or herself as an evaluator (e.g., the Indian partner organizations familiar with assessment).

Student perceived quality. According to the students, the general rigor of their evaluation varied for different elements. Sometimes there were issues with quality that they couldn't overcome due to time, money, and contextual and organizational constraints.

I think it was professional in the sense that we had done our research. We had understood the issues about what was going on. We knew the best methods for how to approach it. We had a really good plan; our questionnaire was solid; it was backed in literature, but there were some gaps that we just personally couldn't overcome. (Samuel)

Some students reported bias in the sampling procedures (e.g., there was no sampling frame, they weren't random or representative, or inadequate control groups). Since many had to rely on locals to give them access to participants, bias was detected in who the gatekeepers selected as well as in the projects they would select for the students to observe.

I think they were leading us to people who would give us positive answers. There was even one time where I went to a site where we thought a stove was built and there wasn't a stove. I came back maybe a month later with a project partner and they had rebuilt the stove because they knew that I wanted to go there... We had a pretty serious sampling issue just because we didn't have the universe of individuals. (Ryan)

Other issues were reported in the methods such as inadequate time and budget to conduct a longitudinal or experimental design, no baseline data to compare their findings, no pilot testing of their instruments beforehand, or not being able to listen to interviews while they were happening in order to catch mistakes.

I know at the beginning I really struggled. We didn't listen to our interviews until after we did the first wave because we were doing three a day. So we couldn't transcribe until they were over. And so we went back and listened and were like, "Oh my gosh, we were asking so many leading questions" and that alone taints your data... I wish we could have caught that sooner. I think we did a really good job, especially at the end, cause we listened to the transcriptions and listening to the recordings you can hear us catch ourselves. (Stephanie)

In most cases, the students struggled with their inability to find effective ways to judge or convey impact. "[We were] able to gather anecdotal evidence of effect or what works and why but not show causal effect" (Ellen).

Often the struggles came about in balancing the training received in the classroom to adhere to academic standards with conducting an evaluation in a real-world setting. According to one student, there's a benefit to taking time to understand the literature, but then allowing yourself in the field, "to make those adjustments that are necessary and okay" (Samuel). One student admitted allowing the scientific standards to slip as he "negotiated between these standards and the practicality of conducting the evaluation" (Ryan).

In most cases, it seemed the students felt the evaluation was appropriate and relevant for the scope or purpose of the evaluation (e.g., not generalizable), the stage of the organization (e.g., no record-keeping), and the timeframe they were given to conduct the evaluation. Many rated the quality between a seven or eight on a scale from 1-10, with one being “low quality” and ten being “high quality.” If they gave it a lower rating, such as a four or five, they provided a caveat such as “due to sampling bias” or the other limitations stated above. Some even felt they cared more about the rigor of the evaluation than the organizations. “I don’t think any of them were very concerned about the social science standards. I think they took it as truth, and I think we tried really hard to make it valid” (Ryan).

Nonprofit perceived quality. When asked if she or her staff members had any validity concerns with the data, one program coordinator pointedly responded, “No.” In the survey, five of the organizations rated the overall quality of the evaluation as high and that it met their needs. Nine indicated that the quality was acceptable for their needs. And one responded that it was not good enough for him or her to take it seriously.¹⁵

Context Contingencies

This section discusses the factors that facilitate and inhibit the students conducting the evaluations in their respective settings. This information situates the evaluation inputs and outputs and highlights how the context can shape the nature and quality of an evaluation. This is particularly important for this study because most of the organizations work in a developing

¹⁵ This response is interesting given the nature of the evaluation conducted by the students. Given the state of the organization’s efforts in the country, the student team primarily focused on evaluating the quality of the organization’s program delivery and proposed evaluation techniques.

context among marginalized populations. Concerning the socio-cultural contexts, the students noticed that many people were incentivized to give positive answers. Some people distorted their answers or painted a “rosier” picture of reality. Students felt the participants were reticent to give critical feedback, especially when they viewed the evaluators working in tandem with the organization and, thus, a gateway to resources.

We drove out there and they were all clapping and singing and had this whole program prepared for us and it was a really interesting thing. It gave us some insight into how they viewed the organization. It was like they needed to really create a good relationship with us ‘cause they felt that we were a link to more resources for other groups, more livestock so they could increase their livelihood or agricultural products. So I think they felt it was in their best interest to go with the evaluation as best possible. (Austin)

In this and other similar situations, the students found it hard to ascertain accurate information or to get past the “show” until they had spent more time with the service population. In some areas, the participants were used to receiving aid or “handouts” and would do what they perceived as necessary to preserve it.

I think that the relationships of the people who were asking the questions to the people who were being asked the questions would facilitate answers that were biased, right? Because if you’re viewed as the same people that were giving them money every week or whatever and facilitating these financial opportunities for them, of course they’re going to give responses that are not going to, you know potentially interfere with their ability to get these loans in the future. I think it was really hard to get honest feedback from those people. They also, I think, had a big incentive to give really positive feedback about their interaction with [the NGO] and the effect of our projects because first of all, I don’t think they wanted to look ungrateful. Secondly, other organizations might be less willing to work with them in the future or at least they might think that, right? (Ann)

The service population received the students differently. In some cases, a student’s particular characteristics helped and in others it detracted. One student felt people were nicer because he was different, e.g., white, foreign, and wore different clothes. In other instances being different increased the divide. In some contexts, the students felt that being white or American gave them credibility among the population. In others credibility only came from their background. In those situations, being a student wasn’t as beneficial. In others, it helped the younger populations feel more comfortable and older adults respected their university education. Some evaluators felt the people opened up to them more because they viewed the students as

sincere in their desire to learn about them. In some cases gender helped (e.g., women evaluators interviewing women and male team members interviewing males), and with others it didn't seem to be an issue (e.g., rural women in a middle eastern country agreed to be interviewed by a male intern).

In some cases the characteristics played out differently than expected. For example, one student felt that being American helped balance out his youth, but being half Indian helped them accept him better. And in many circumstances, there were things the students could do to help lessen the distance. As one director explained:

You know when I talked to [the student intern] earlier, she came from this area; she knew what she was walking into. I think there wasn't going to be anything she could do about her gender and her race, but from the standpoint of just being very transparent and honest and trustworthy, she did everything I think she could have to bridge as much of that gap as possible. (Carol)

The service population's familiarity with evaluation helped if they saw it as a way to improve the program, but slowed the evaluation down until they realized it could be beneficial. In a few cases, this referred back to the population's past experience with the organization itself. Sometimes the participants felt that their input or concerns didn't matter or reach the organization anyway. Some worried about losing their job because of the evaluation.

In other cases, the organization's model hindered the evaluation. For example, many of the nonprofits worked in multiple countries and the evaluations couldn't be compared between teams; many had never tracked indicators or done follow-up on projects in the past. Sometimes the organizations were experiencing too many growing pains from expanding too fast and the evaluation they asked for seemed irrelevant. The timing of the evaluation also was a challenge, either it attempted to look at impact too soon after the program was instigated (e.g., only six months after the solar panels were installed) or there wasn't enough time to complete data

collection. Sometimes the evaluation simply took place at an unfortunate time (e.g., trying to evaluate a school program during the summer).

In many instances, however, the fundamental inhibiting factor was trying to adapt a plan to the field; balancing expectations with the conditions at hand. Transporting an evaluator into a context they are unfamiliar with poses challenges that sometimes can't be adequately addressed. However, each of the students acknowledged this fact and adjusted as best they could. A complete account of the students' experience is included in Appendix D.

ORGANIZATION EXPERIENCE

This section discusses the experience of the organizations in working with the student, the PEAT program, conducting the evaluation and their satisfaction overall. These sections on experience provide relevant insight into the evaluation process—challenges and successes—something uniquely accessible through the PEAT program that addresses the fourth research question on how to provide the most appropriate evaluation for PEAT's clients. The following four subsections describe the organization's experience in working with PEAT in general and the student-evaluators in particular, what they feel would have been more helpful given their experience, and their satisfaction rating overall.

Working with PEAT

The organizations were made aware of PEAT in different ways, and variations in their experiences typically originated from there. Those that were closely associated with the BYU sociology department usually knew more about the program, what it could offer, the students' training, and what to expect. Additionally, the organizations that had evaluations following PEAT's first year (some have had interns multiple years in a row) also worked within a more structured program.

Many of the organizations met with the primary PEAT faculty member to initiate the relationship. However, many did not and the narrative slightly differed. The differences include variations in the expectations for an evaluation, in terms of being tailored to their verses the student's needs; the role of the student; level of interaction; level of input from the nonprofit personnel or received input from PEAT; and the level of student expertise. For example, one organization most removed from the department felt there was an inadvertent distortion of what the evaluation could accomplish. In contrast, another coordinator, who was initially approached by students and primarily only interacted with PEAT through them, only had positive things to report: the evaluation met the organization's needs and the students were more than qualified to conduct it.

One organization that predominantly works with the main PEAT director, didn't expect an evaluation geared towards the organization's needs. The in-country director has worked with students multiple times and thought the PEAT intern's main purpose was to work on a student research project. He reported to the students that they could do whatever they needed, and as a result, some benefits would occur for the organization. In fact, at least two of the organizations said that they agreed to be a part of the program because they wanted to help students and support BYU. Alternatively, one organization expressed that the evaluation seemed to produce "more information in the areas that interested the students and not enough that was most pertinent to our priorities" (Ben).

Another organization had only positive things to say about the students, their experience, and the benefits. However, the director wonders if they should continue working with PEAT because he's not sure if the organization's structure provides the best experience for the students

given his view of the internship requirements. For this organization, what they want the PEAT interns to look into doesn't seem to match what the students say they *need* to accomplish:

I for one adore the PEAT interns and would love to have them in each country, but it got a little awkward because I said I wanted them to evaluate 15 [sites]. [The students'] eyes all got big, and they said they wouldn't have time to evaluate that many. I told them that the projects weren't that big, but then [the PEAT director] said no. They could only evaluate one. But in my mind, I was thinking it wouldn't keep them busy, and it wasn't what they need...that was tough for me, because of our model...I've thought about not continuing the relationship. Only because I'm wondering if we are giving them the experience [the PEAT director] wants them to have by giving them a lot of shallower evaluations and not just one big project. I thought about next year talking to [the PEAT director] and saying that maybe it's not the best fit. (James)

A student team for another organization described a similar experience and disconnect.

When writing the report, they wanted to focus on the data that would provide the information the organization staff seemed to request. However, producing that report would not meet the expectations of the PEAT director.

Even with the disparities in experience, most of the nonprofits said they want to work with PEAT again, and they enjoyed their interaction with the PEAT program. According to one executive director:

I had such a good experience with PEAT interns that I don't really have a lot that comes to mind on improvement. Honestly, I think [all three of the PEAT faculty] did an excellent job in promoting the internships, getting us to come and talk to classes about the internship possibilities through PEAT, and I think that was a good deal of the program's success and growth, is how well it was guided by the academics that were in charge of it. (Adam)

Working with Student-evaluators

The data from the organization directors also show an overwhelmingly positive account of working with the students. The organization directors described them as professional, receptive, pleasant, willing, invested and hardworking. For example, one program coordinator stated that she didn't hear any complaints from her staff members about the students:

I didn't hear anything negative...they were great to work with and everyone that met with them. They were doing staff interviews and speaking with people as well, and I mean, I thought everyone felt really comfortable and pleased with their interactions and the work they did and the report they wrote. (Carol)

This same director explained that the students were open to all of the suggestions given by herself and her staff. They were willing to adjust their expectations and plans according to the restraints of the program and to work within the given parameters.

Additionally, the executive director who was questioning whether to work with PEAT in the future, stated that despite the requirements for the internship, the students were willing to do whatever he else asked them to do.

I can ask them to do anything and they'll do it...and they all have such great attitudes about it. They all do. Because even though they all have their assignments, all of them have said in private, "Hey, you have me do whatever you need. Even though I have this assignment," and I love that. (James)

This same director even said he loved the student's insight into what needed to change:

I love how PEAT interns can be so brilliant and easily rip everything apart, but they do [it] so maturely, and they do so to the right people. And they do so with recommendations. It doesn't just say, "This sucks, this sucks, and this is what I'd do differently. (James)

However, some of the organizations had some specific things they wish would have been different. Despite her expression that, "generally the students themselves were wonderful people and I enjoyed working with them," this director wanted more interaction, collaboration, and advice beforehand. In her organization's experience, she felt that the evaluation did not take place at the right time. She would have preferred the evaluation to happen one year after the project had begun instead of six months. She felt that the students' findings on outcomes were misjudged because the use of services would have been skewed due to the rainy season.¹⁶ She also stated that this was something the PEAT faculty should have caught before the evaluation was implemented.

¹⁶ The student's report discusses outcomes, but primarily focuses on the limitations in the data due to a lack of record keeping by the staff, underreporting by the recipients, and biases in the organization's prospective evaluation instruments.

Two other organizations described instances where they felt the students acted a little “too objective.” One organization expected the student to integrate more into their structure and blend in with their other volunteers as opposed to acting like an external observer.¹⁷ Similarly, another program director mentioned in the survey that she would have preferred the students to work more collaboratively with the in-country staff and with those at headquarters. She felt they needed to place more value on their personal experiences and insight.

More Helpful Evaluations

In addition to some of these suggestions, there were a few more responses from the organizations that indicated ways in which the evaluation could have been more helpful. In general, there were instances where the organization wanted the evaluation to reflect the priorities of the organization and not the internship program. Some wanted an equal balance between quantitative and qualitative data, with one wanting more rich, diverse data while at the same time needing a way to compare data from year to year. Many expressed wanting more connection between findings and recommendations, that in some way they diverged. At least three reported wanting more actionable, step-by-step suggestions—and one specifically stated expecting the evaluation to help lead to more funding opportunities.

Satisfaction Rating

Six of the organizations responding to the survey indicated that they were *very* satisfied with the evaluation. Another six indicated that they were satisfied with the information, and three reported that they were neutral, neither satisfied nor dissatisfied.

¹⁷ Comparing the student’s and nonprofit staff responses, this seems to result from a lack of communication beforehand. In subsequent years, the organization made this expectation clear.

The interviews revealed insights such the organizations being very happy to work with PEAT, that it was valuable to have the student's perspectives, it was a new experience for them, and that they want the relationship to continue. For example, one director stated:

Yeah, we were happy—all positive experiences. I just think it was incredibly valuable for us as we grew our organization. It was valuable to have that academic approach in the few countries and teams that we were able to get interns. Had we had the opportunity to have a PEAT intern on every one of our teams, I would have jumped on that opportunity. It was a fantastic experience for our organization to work through PEAT. I hope the partnership continues in the future. (Adam)

According to another in-country program director, “it was a whole new experience working on the survey implementation and then knowing the results. I'm glad I've worked with you” (Paul).

INFLUENCE OF THE EVALUATION

The comprehensive theory of evaluation influence directs the researcher towards the source, timeframe and intentions behind an evaluation's outcomes as well as the underlying mechanisms that mediate the various pathways to change (Kirkhart 2000; Henry and Mark 2003). Mark and Henry (2004) classified the mechanisms and outcomes into four different categories: general influence processes, cognitive and affective (or attitudinal) processes, motivational processes, and behavioral processes (Mark and Henry 2004:40). The following is a description of the types of influences that resulted from the student evaluations categorized into these four sections. This addresses the third research question.

This study analyzed about 25 mechanisms and outcomes identified in the Mark and Henry (2004) and Appleton-Dyer et al. (2012) conceptual models and sufficient evidence was found for fourteen of them (Figure 3 lists the fourteen influences). Interestingly, most of the data corresponded to general mechanisms and cognitive outcomes—particularly, elaboration, priming, justification and salience. One of the main contributions of this paper is that it produced ample evidence of these factors as important elements in the change process.

The most notable finding is that the evaluations led to changes in learning, thinking, and behavior (refer to Table 9 for a list of survey responses for each influence). While it was expected that the evaluation would inform the clients, the data indicated that learning should be considered an important outcome of evaluation and separate from elaborating about a program. Additionally, even though novice evaluators conducted PEAT evaluations, significant changes took place at the individual level and in relation to organizational policy. Thus, despite the challenges faced with using student-evaluators, the students' work led to a wide range of positive outcomes. This finding is supported by the ratings of perceived benefit by the organizational personnel: seven of the survey respondents indicated that they felt the evaluation was highly beneficial. Seven other respondents indicated that they felt it was moderately beneficial, and one program director indicated no effect.

[Table 9 about here]

General Influence Processes

According to Mark and Henry (2004), this level includes the mechanisms that serve as a foundation to changes in the attitudes and actions of key stakeholders. Thus, while many evaluators are interested in the more observable outcomes that lead to social betterment—the “so what” of an evaluation—the general mechanisms are an important and irreplaceable aspect of the entire change process (Mark and Henry 2004).

Elaboration. Engaging in an evaluation process or reading findings in a report may cause an individual to think more systematically about the program, the organization's assumptions about it, and the expected outcomes; he/she may elaborate an aspect of the program because of an evaluation or the information it produces. Often this is a first step to change.

From the perspectives of many organization directors within this study, the evaluations influenced the organizations to think more about their service population's needs and desires and how their program intends to address them. Thus, many stated that they are going back to some of the fundamentals: their mission, strategies and overall approach.

I think it stirred things up a little bit. I do think it's really resulted in all of us taking a step back to say, all right, we do need to be more organized about this. We do need to be more clear about what we're trying to accomplish...I think what we got is a more clear sense that we're not clear on what our goals, our mission, and how we hope to accomplish them. So that became much [clearer] as a consequence of this evaluation (Board Member, Ruth).

In relation to this, some of the nonprofit staff members realize that there might be discrepancies in what the people tell the organization firsthand and reality. While some trusted the accuracy of the reports more than others, the evaluations triggered this realization. One board member described a good example of this:

I'm not sure if I got this more from the report or whether it was from my conversation with [the evaluators] afterward, but I think it is hard to sort out how much people are saying what they think we want to hear. So I don't know what to take of [in-country program director] and the people when you have [the] people saying, "Stop talking about trauma" and [the in-country program director] saying that trauma has to be part of what we do, and how much of [her] saying that trauma has got to be part of what we do is [just the] way it's presented by [the founder] saying, "It's a trauma healing thing." I'm still trying to figure that out. (Jolene)

One executive director felt the evaluation helped guide their organization's thought process as it grew, providing them with additional perspectives that they would not have had otherwise. Two organization members mentioned that the evaluation helped the staff realize that their work requires much more of them than what they were giving, or as one board member stated, "that this is a tougher business that we're in than [expected]" (Jolene). In summary, this same board member stated that the evaluation "made us more analytical."

Gathered from the students' perspectives, the organization directors considered the report's recommendations as they moved forward, especially in the ways they intended to reach the population, what they were trying to accomplish and how they measured success.

Seven organization directors indicated in the survey that they, "...thought more about [their] program and the expected outcomes] because of the evaluation. From the interviews, the elaboration took place at the individual, interpersonal and collective level as some spoke about themselves or in reference to conversations that took place with the entire organization staff present. For example, many of the quotes cited above came from an interview with two people from the same organization in conversation with each other about their individual and the organization's response to the evaluation as whole.

Priming. The evaluation may also bring up an issue and that issue becomes "top of the mind." A primed concept is then accessible to a decision maker in his/her judgment process about people, programs or events. The student evaluations provided insight into issues within three areas for the organizations. First, they brought forward issues with how the organization personnel conceptualize the program. For one organization, the evaluation and the evaluators themselves acted as a resource for the organization as it grew and adjusted its model, especially because they provided a perspective that the staff members lacked.

[Intern 1] was an amazing soundboard, as was [intern 2. [Intern 3] would be, too. Bouncing ideas off of them in regards to our development strategy, again because they have the academic perspective and that's awesome. (James)

Another organization's board realized that their focus might not meet the most basic and prevailing needs of the people. Other organization directors realized that they needed to adjust the way they measure success so it better matches the services they actually deliver.

Second, the evaluations addressed issues with the way the organizations delivered their programs. In reference to an evaluation report, one organization director stated:

It definitely spurred results in the organization insofar as their results were used to identify gaps in delivery of programs, and you know, where there were issues with whatever management was or wasn't doing. Those gaps came to light and were given to the board of directors so they knew or were given a chance to know of the issues. (Adam)

The evaluation reports identified issues with delivery such as the program not reaching all intended beneficiaries (e.g., only women with husbands were accessing the program); that there was no follow-up to ensure their services were rendered properly (e.g., the organization provided eye surgeries without instituting a way to provide checkups); that some members of a co-operative were not attending their weekly meetings to the dismay of their fellow members; that the facilitators of the program's discussions were giving the participants the "right" answers; and that the interest of the loans should be based on different criteria than the ones they were currently using.

One evaluation discussed the issues participants faced when interacting with the administrators of the program. The following compares the evaluator's interview response about the influence of the evaluation and the coordinator's take on the same question. The PEAT evaluator observed:

Basically some counselors were telling people that they weren't smart enough to go to college. And stuff like that. A guidance counselor is there to help students feel like they can do it. It was just ridiculous ...[students were] saying, "My teacher thinks I'm stupid or my favorite teacher believes in me. I wish all my teachers believed in me." And we were like, "Oh my gosh if all the teachers read this, or if the principal read this, how different would their lives be at the high school if people really internalized it?" So that was our hope...I don't think the teachers read it. I don't think the principal read it...I really, really wish everyone had read it. (Stephanie)

In a separate interview, the director of this program stated the following:

I know that the report was read by a number of people in the district and that it was discussed in administrative meetings, and I know that it was referred to. For instance, they've been really trying to change the paradigms and job descriptions of counselors, and [the report] was referred to [say], "Look this isn't the voice of the whole student body but it is a voice of a struggling population in our school that are saying this about your services, about your approachability..." It was mentioned a number of times and I would say that when it first came out it probably extended out a good year from when it would be referred or the principal might say, "Well I recall reading in the BYU report this..." So I think there was *an* impact...I think it was shining a bit more light on the perceptions that certain students have. Everyone can come to work and think, "Oh yeah, we are doing a great job." Or look at test scores and say, "We'll this must prove that we are doing better." But reading some of the pieces of the report, I think definitely made it clear to some staff that there's still a lot of work to be done, especially in the areas of rapport and relationships. (Carol)

Priming helps bring to light issues or weaknesses in the program, or as the same program coordinator described, “I definitely think there was some eyebrows raised when the administration read the report and probably some counselors.”

Third, the reports also brought up issues on the status of the program, such as the recipients were missing one of the most important parts to the project (e.g., they received the cistern but not the tube that connects the cistern to the roof). In an interview, a student described another example of this:

We were able to provide really useful recommendations that they hadn't even thought of such as, if you are going to put all these solar panels in here, you've got to really seriously consider that these people have been in a huge drought and where are you going to get the water to clean them? All these things that should have been really obvious that they hadn't really thought through; things like that that could have gone seriously wrong with the project we were able to warn them of. (Shannon)

Overall, priming as a consequence is illustrated in this response from a student: “From the phone call I had so far, it sounds like yeah, I brought a lot of good things to their attention” (Kevin). In the survey, five organization directors indicated that the evaluation “...brought to light certain issues and [they] thought about them more.” In addition, a different set of five organization directors also selected the response option that stated that the organization learned about [their] service population and the problems [they] intend to address.”

Justification. This type of influence refers to the “use of evaluation findings to support a previously held position” (Henry and Mark 2003:301). Then data from the evaluation may be used to present a case or obtain approval. In the survey, five respondents indicated that, “The findings supported a belief or position [they] had before the evaluation.”

From the interviews, evidence of this outcome includes PEAT's clients and students referring to instances where the organization directors used the evaluation report as a conduit to talk to staff, referred to the cost-benefit analysis in a meeting with a venture capitalist, to support

an already held position about the need for the organization to address addiction, disintegration, the family and the school fees. Discussing benefits of the evaluation, one staff member stated:

I thought one of the most valuable things, even though we're still struggling a little bit...I took it very seriously because it sort of confirmed my suspicions of working with the organization for the past 2.5 years, and that is the idea that non-medical trauma healing is probably not a realistic goal. Or at least it does not inspire programs right now. (Jolene)

One student described another example of justification, "I think it helped to order a lot of things for some of the board members that I talked to or confirmed things that they kind of had some suspicions of but didn't feel like were really there" (Ryan).

Last, some student-evaluators and one program director provided a relevant example.

From the student's perspective:

I know that [the program coordinator] really appreciated it because I think that it vindicated so many of the things she had already seen but that she hadn't necessarily felt that she could substantiate. [The coordinator] was like, tell [the student-evaluators] that I carry a copy of their report with me to meetings. Somebody brings up something, I can be like, "Really cause I have a quote right here on this page that says this is something that the students said." So she had used it in that regard in meetings to defend some things she was trying to push forward in terms of meeting the students needs...I think the superintendent finally had some data that shared [the coordinator's] opinion. So I would say that was the largest impact that our report had. (Sarah)

The same program coordinator stated in her interview:

It was a well-crafted document that supported some of the things I was trying to say but it came from an outside source, and yeah, maybe it wasn't like a \$20,000 consulting firm that came in, but they came in very organized, and it was a very well written report. I don't think a lot of what was in it could be denied or just shooed a way. So I feel like, I'm not sure that I learned something significant that made me go, "Oh wow, we should have been doing this when we were doing that," but I feel like that it definitely bolstered what [I] have been trying to do. (Carol)

Change-agent. This outcome refers to the evaluation findings or process motivating someone to "rise to action" (Henry and Mark 2003:302). Typically, these individuals were already focused on change in organizational policies, programs or operations, and the evaluation was the catalyst to get them to speak up. Four respondents to the survey indicated that, "the evaluation motivated someone to speak up and/or take action."

From one of the organizations who selected this response, two staff members indicated in the interview that the evaluation helped articulate their previous held concerns and, “motivated [one of them] to speak up, sitting at this table” (Ruth). The other member said that the evaluation motivated her to include the in-country staff member in their weekly meetings, to change the name of the center they are building in the country, and to redefine the scope of their organization.

A staff member from the school program stated that the evaluation:

Empowered administrators to keep pushing for a higher level of rapport and relationship [with the participants]. You know, less doing this job by the book and more just really digging in and doing some of the hard work with the most challenging students...So I definitely think that their report was a part of what helped with the momentum of what we eventually moved in to. (Carol)

An executive director from a different organization said that the evaluation motivated him to talk to his other staff members and the country directors about adjusting their entire model. Although the director had been thinking this needed to be done, the evaluation supported it and outlined ways to implement the change. The director confirmed that the process of talking to his staff members and finally adjusting the model “...wouldn’t have gone as quickly without their feedback and research, for sure” (James).

Persuasion. When one party tries to change the attitudes of another using the evaluation findings, they have engaged in the change process of *persuasion*. Six respondents indicated on the survey that, “The evaluation findings helped [them] to persuade other stakeholders in [their] favor.”

As presented in the quotes above, the coordinator of one program utilized the reports to justify her previously held position on a number of issues, and then persuaded her team and other staff members to think more systematically about the quality of their work and ways they can adjust their approach to students. This program coordinator mentioned that the school district has

now implemented a “systems approach.” According to the coordinator, “This was a big paradigm shift for a lot of the staff “...to work in a positive behavior paradigm instead of only noticing and commenting on what’s wrong” (Carol). They have been working with this new approach for the last three years. The program coordinator remembers the principal coming up to her and saying, “It really lines up with a lot of the things you’ve been talking about in these meetings and what’s missing for a lot of our students.” To this coordinator, this shift was a “testament to some of the things that the gals were trying to share in their report.”

Another executive director stated that the evaluators wrote up a report on potential partnerships in the country in an effort to persuade the organization that they needed to expand into other parts of the region. As a result, the director stated, “That was really helpful, because we did actually expand there this year, and so we used that report for the [country directors] to open up that area” (James). Along similar lines, another executive director stated that the student’s observation and report was used to “...promote better procedural systems in regards to how we carry out these development projects” (Adam). One other organization mentioned that the evaluation helped to initiate the persuasion process. She indicated that they were in still in the discussion phase where each member tries to persuade other members of their positions, and the findings were being used to facilitate that process.

Minority-opinion influence. This outcome takes place when the evaluation helps move the opinion of the minority, with respect to attitudes, into more of the mainstream. This type of influence took place in the school program, helping the coordinator push her ideas to a point where the entire district adopted the new systems approach. An executive director used the evaluation to move the opinions of the country directors and then the entire board toward an

internship model as opposed their previous model in which volunteers come up with new development projects each summer.

An evaluation also helped persuade a board member that their scope should widen, and she is in the process of advocating for this previously subdued belief. She specifically said she is working on other board members at this point and that it is too early to say whether it will actually lead to substantial changes. But, she is hopeful. In contrast, two respondents indicated that the evaluation helped move the minority opinion into the majority.

Collective deliberation/discussion. Most of the examples provided above included some type of discussion at the collective level. Typically for a small organization, this is between the staff and the board. Some examples of this is one board member including the in-country staff and her fellow board members in a discussion over what the new center should be called and what activities should take place there. The board had already decided on one name and set of activities, but after the evaluation, this board member thought it necessary to renegotiate the name and purpose.

Another example is when the executive director used the evaluation feedback as a catalyst for changing the entire model of the organization. From the interview, it was evident that discussion about this change took place at all levels: between the executive director and the country directors, between him and his fellow staff members, and between the staff and the board.

Skill acquisition. This refers to the development of new abilities. One respondent in an interview indicated that after working through this evaluation process, she felt she could conduct a simple evaluation on her own. This is the only respondent who indicated on the survey that, “We actually learned enough to conduct an evaluation of our own next time.” Two other

interviewees pointed out ways in which the evaluation taught them more about evaluation in general, what is required to implement one, and that it is somewhat harder than first expected.

Additional general influence. Skill acquisition is the last outcome in the category of general influence as outlined by Mark and Henry (2004). However, in this study, one other type of influence came up that seemed different and frequent enough to warrant making an addition to their model. In the evaluation use framework, there is an outcome called “conceptual” or “enlightenment” use. As Preskill and Caracelli (1997) described, the findings of an evaluation may “become part of a knowledge base that influences thinking over time (p. 217). For this purpose, the influence of an evaluation may simply mean it helped the organization learn something. As Ledermann (2012) stated, conceptual use is the type of outcome when the evaluation results “...lead to a better understanding or a change in the conception of the object of evaluation” (p. 160). While the influence framework covers more potential ways for an evaluation to benefit an organization, the aspect of learning in “conceptual use” is not covered fully under “elaboration” in the influence models; there is a distinct difference between learning something and thinking about something. Below is the evidence from this study of data that became part of a knowledge base (i.e., conceptual use); the data taught the organizations about something (i.e., individual/collective learning) that subsequently became part of an elaboration process. The following is a description of the proposed additional general mechanism.

Individual/collective learning. Many of the organizations received a description of their program in the evaluation. Much of this information was more insight into what’s happening on the ground than what they normally receive. This is often due to a lack of or infrequent communication and face-to-face interaction with their in-country staff members and recipients. The account provided in the evaluation gave them access to many aspects of their programs. For

example, the evaluations taught the organizations about the recipients' perceptions and interaction with the organization; the recipients' real needs versus perceived needs versus what the recipients tell the organization because it's what they want to hear; the interests and desires of the population; the staff members; the process of implementing the program (e.g., the strengths and weaknesses of a microloan program in a leprosy colony); the pros and cons of the volunteer experience; and lastly, the partners they work with and hope to work with in the country, their needs and how the organization can fulfill them. A student-evaluator described an example of this type of information:

I suggested there were two types of organizations in India. In general, there is the small mom-and-pop NGO who's like Mother Theresa: trying to help people in this one small area. They will benefit from man hours and useless students who can't do anything but smile...So go ahead send interns there who just want a fun experience and want to be impacted by learning about what poverty is like. But there are also the big organizations that want interns. So you should send two teams to India. One should be your business, student-type interns who want a professional resume-type experience and the other should be your volunteer. (David)

Cognitive and Affective Processes

This type of process refers to shifts in thoughts and feelings (Mark and Henry 2004). The data collected provides evidence of changes in salience and an opinion or attitude change within this category.

Salience. This outcome takes place when an issue or problem becomes more important in the eyes of an individual or group in light of the evaluation. One of themes that came through is the increase in importance of analysis, systematic thinking and reflection among the PEAT client organizations. When asked the benefit of the evaluation, a student responded that evidence of the evaluation was important in, "...starting the trend, rehashing what we're trying to assess and developing that culture of evaluation" (John).

According to one board member, the evaluation helped them reprioritize from focusing just on fundraising to a mission that meets needs but can also appeal to granting agencies. As she

stated, they realized that they need to keep their mission “broad enough that we’re responding to real needs, and yet narrow enough that we can have a real impact and can explain ourselves clearly enough to get support” (Jolene).

Other examples include when the information that was important to one staff member became just as important to the rest; when an issue with the program noticed by a student-evaluator became important enough for the in-country director to make the changes immediately; and when the suggestions of the student-evaluators became important enough in the eyes of the staff for them to make substantial changes. One instance occurred when organization decided that it was more important to focus on outputs rather than inputs. Seven of the 15 respondents indicated in the survey that the evaluation “...brought to light certain issues and those issues became more important in [their] eyes.”

Opinion/attitude change. Examples of opinions changing as a result of the evaluation include when one organization’s staff felt that running a nonprofit was a harder business than they thought they were getting themselves into; that it was okay for the staff members to think outside the box; and that one job within the organization was actually valued rather than viewed as a waste. One organizational director stated that the evaluation changed some opinions but not all. Four respondents indicated in the survey that, “[their] attitude[s] changed (positively or negatively) about the program and/or its assumptions.”

Behavioral Processes

According to Mark and Henry (2004), the category of behavioral processes refers to the changes in actions and often represents the more long-term outcomes from an evaluation “in a chain of influence processes” (p. 42). The data collected provide evidence of individual and

collective change in practice; program continuation, cessation, or change in program details; collective change in policy; and that the evaluation led to more funding and more evaluation.

Four items in the survey corresponded to the following two types of influence. Ten respondents indicated that they "...adjusted program details based on what [they] learned." Three organizations indicated that, "Because of the evaluation, significant changes took place." Five organizations indicated that, "The evaluation helped to alter thought or behavior at the policy level," and two organizations indicated that, "There were changes (minor or major) to policy due to the evaluation's work." Lastly, the survey included one specific example of a change, which stated that the evaluation helped the organizations narrow their focus or mission. Six organizations indicated that the evaluation had this result. However, it is important to note that one organization director also indicated on the survey that the evaluation did not lead to any of the specified behavioral outcomes.

Changes in practice. Evidence of specific actions taken by an individual or collective because of the evaluation include new staff or board members seeking out past interns for advice; changes in how an organization interacts with its current and potential in-country partners; adding better tracking and record-keeping of recipients; and the entire board of directors now including the in-country staff in their weekly meetings as well as taking the evaluation's information into consideration as they listen to the insight of both in-country staff and the board of directors. For example, this last organization stated in the interview that the staff members are trying to stay in closer contact with their in-country program director. "Well, we have made some adjustments, particularly in terms of working more closely with [program director] and getting more frequent input from [her]" (Jolene).

Changes to program. In the interviews, the organization members specifically stated that the evaluation led to changes in who can and cannot access the program; a change from a punitive to a more positive approach when interacting with students; expanding some programs to new regions of a country; stopping some programs that did not work or not continuing them the next year in the same or a new country; altering the organization's model from volunteer-focused to internship-focused, and then, adjusting how they recruit and interview potential interns. The following quotes support two examples of these types of changes. The first is from an executive director talking about how the organization now offers internships instead of just humanitarian-like mission trips. "As soon as I met with [the PEAT evaluator] and got all his feedback, that's when we started creating these. We've never had them before" (James). Another program director stated that the program evaluation has changed their approach. "Instead of just having a punitive discipline system, we are really now district wide working on this positive approach. I think that what the work that [the PEAT evaluators] did has contributed to this change" (Carol).

Led to more funding and/or evaluation. There was little data in the interviews indicating that the evaluations led to more evaluation or funding. However one student stated that her cost-benefit analysis was used by the organization in their meeting with the venture capitalist who was planning on funding a new project. Incidentally, the organization director indicated in the survey that the evaluation led to more funding. Three other organization directors also indicated this in the survey.

Four organization directors indicated on the survey that the evaluation led to conducting additional evaluations. Two of these organizations in particular have worked with PEAT

multiple times, including summer 2014. For one of the organization, PEAT interns have almost become a regular aspect of their program planning. One student also noticed this change.

I was the first intern to do really evaluation for projects in [country], so it really set the gateway and impetus that they wanted to continue to do evaluation, especially since I get called year after year, and my understanding is they continue to push for that. And they saw my struggles and learned to wrap around that. So, definitely [it was] a trend-setting thing. (John)

Despite these changes, the data collected did not support the following types of mechanisms or outcomes as listed in the Mark and Henry (2004) model: heuristics, standard setting, coalition formation, descriptive norms, agenda-setting, policy-oriented learning, new skill performance, or any of the outcomes listed under the motivational category. The following three sections are the remaining elements of the influence picture (source, intent, timeframe) as outlined by Kirkart (2000), who first suggested changing the focus from use to influence.

Source of Influence

The source of influence refers to the whether the influence stemmed from the process of conducting the evaluation or its findings. The data collected in this study provided examples of both, although the source of influence for the majority of the benefits stemmed from the evaluation's findings. Concerning the evaluation process, there was evidence of the organizations learning what it takes to conduct an evaluation from interacting with the students before, during and after. For example, multiple meetings were held with one organization to discuss the importance of evaluation, what would happen during the evaluation, and what they should expect to learn. In the interviews for this thesis, two members of the board of directors incorporated the language used by the evaluation team in those meetings indicating that they had learned something about evaluation from the participating in the process.

An additional, and probably more important, benefit from the evaluation process—and not the findings—was the impact it had on the service population. One PEAT evaluation team

and the coordinator of the program indicated that the evaluation directly helped the student participants they worked with because it gave them a voice. As one student described:

We were evaluating a school program, but I would say that we not only provided valuable feedback for the program, but I think in terms of information, we provided a really important service for the participants, the students. These students in particular aren't often sought after for their opinions on things. The white administrators and teachers, they try but I think they're in a position of intimidation and a lot of time students don't feel like they're going to listen to them. So those students can be very, very quiet and don't articulate their interests and desires and ideas to higher ups. (Sarah)

Another student indicated that he felt the evaluation provided a great service to the participants because it provided them an opportunity to talk about their troubles and to feel like someone was listening.

Anticipated or New

An important aspect of Kirkhart's (2000) comprehensive theory of evaluation is the dimension of intention. This refers to whether the evaluation influence is "purposefully directed, consciously recognized, and 'planfully' anticipated" (p. 11). The specific data collected in this study captured the last element, in particular, whether the evaluation results and the information provided was anticipated or new to the organization.

In terms of the evaluation producing information that was anticipated or expected, one staff member described in the interview that the evaluation confirmed her suspicions about the potential ineffectiveness of a proposed program. The other board member in this interview stated that there wasn't anything too surprising and that it validated some of the things they had already heard or suspected were issues. One student suggested that the evaluation confirmed things that the board knew, but hadn't really acknowledged or dealt with previously. The coordinator of the student program stated that the evaluation findings didn't necessarily surprise her, but "raised eyebrows" of the other stakeholders. According to one of the student's evaluating this program,

the evaluation presented things that were already suspected by the individual closest to the program and its recipients, but were not as clear to those on the periphery.

In the interview data, there were also examples of organization directors who read things that were unintended or new. The information provided to one organization about their volunteers experience was new. Also new to these program directors was the insights about issues with the program delivery, and they subsequently addressed them. One organization also stated that the evaluation stirred things up for them, especially in regard to the observed ineffectiveness of their current programs. The evaluation suggested that their approach could be shifted in some ways to address this major concern. One student indicated that their evaluation provided information that the organization staff just hadn't thought of before.

In the survey responses, two organization directors specified that they learned what they expected to learn. In contrast, one organization director indicated that they were completely surprised by the information in the evaluation. The remaining respondents indicated that they both, "...learned new things and heard things [they] already knew."

Timeframe

Kirkhart (2000) also included time as an important dimension of an evaluation's influence. In particular, this element refers to the points in time an evaluation's influence emerges, exists, or continues: immediate, end-of-cycle, or long-term.

In the interviews, there were a few examples of the benefit happening immediately. While conducting the evaluation, some students made suggestions to the in-country staff about issues needing immediate attention. Immediate positive influence also took place when the service population had something to benefit from participating in the evaluation. Examples of this include the evaluation providing the recipients an opportunity to speak about their troubles

or provide feedback on the program. There were five evaluations in particular that were meant to teach the in-country staff members about evaluation and how to continue it once the PEAT interns had left. In the survey, seven organization directors indicated that they learned things while the evaluation was taking place and were able to make adjustments accordingly.

For most of PEAT's clients, the evaluation took place from one to four years ago. Many of them made adjustments to their programs at the end of the evaluation cycle, but were not able to identify any long-term benefits that have directly taken place because of those changes. Examples of end-of-cycle changes include utilizing the information in presentations to future partners; utilizing the findings to initiate and inform subsequent discussions for change; and the evaluation altering opinions and approaches in relation to the organization's program, staff, and population. The influence at this stage also includes the insight that was used to teach the organizations about their models, strategies and theories of change; gaps and issues in their program delivery; and things that work and do not work well for the population or in the context.

For one organization that has been working with PEAT since 2010, a few long-term changes can be identified. The interviews and observation data indicate that the organization has increased its acceptance and the importance of evaluation in their work. Additionally, the NGO staff and board have significantly altered their mission statement, model, and the strategies and tactics utilized in each country.

STUDENT FACTOR

This last findings section corresponds to the Appleton-Dyer et al. (2012) "Individual Characteristics" component. While their model suggests that the "transformation of these mechanism into outcomes is mediated by an individual's characteristic," this study adds the element of student-evaluator and his/her pivotal role in providing quality evaluations for this

specific type of organization (p. 540). This study found that the process and potential influence of the evaluations are shaped by the fact that students conducted them. The following is a presentation of the benefits and challenges to using students in the evaluation. Overall, the experiences showed that adjusting the approach, clear communication, and trial-and error could offset many of the challenges and that the students had unique qualities to adapt to harsh circumstances and provide valuable information for these organizations. While using students is not the only option for providing evaluations to these organizations, their experience shed light on the different factors that are useful. For example, a sincere desire to learn helps build trust among the population; methodological rigor is not as important when the greatest need is for eyes and ears on the ground; and proximity to the population provides access to information that cannot be obtained through formal methods. For detailed information on the student experience and how it relates to these findings see Appendix D.

The students' demographics in some circumstances were seen as negative and in others more positive. Regarding their youth and student status, one student team felt that it helped bridge the racial and ethnic divide by making them seem more approachable, genuine and invested. As one student stated, "We felt that our age in particular made the students feel more comfortable talking to us. Because they were like, "Okay these aren't scary white folk; these are just college students" (Sarah). In some cases, the students could adjust whether they acted younger or older depending on their audience. For some evaluations, the students observed that their work ethic or nationality offset the negative perceptions regarding their youth. For others, the biases introduced were because they were predominantly white, American outsiders and being young did not necessarily matter.

In some instances being students allowed the evaluators to adjust how they introduced their purpose. They explained that they were there to learn from the respondents and about their way of life, and the people believed them to be sincere. According to a student-evaluator:

Being young and a student definitely didn't inhibit. Being a student probably helped because you have a reason to be asking questions. You're curious and inquisitive. Being young might also help because they can see your excitement about being in a new place and meeting new people. Rather than it being a job or, "I'm so good at this. I've been doing this for 20 years, and you're just another piece of data." To me, it was super cool to talk to each of those individuals. I don't know how professional evaluators would relate to that sentiment. So, I think it really helped me relate to people and open them up. (David)

As this example explained, most students exude a sense of passion, energy, and eagerness that speaks to the participants. This also relates to a general desire to learn, gain experience and even change the world. From past experience, one intern compared students with the typical development practitioner: "[Students] haven't been hardened by development work yet. They haven't been turned off by it. They are fresh. They are new. They want to experience it. They haven't realized the world can't be fixed overnight yet" (Samuel). This sentiment was mirrored by one of PEAT's clients. The executive director explained:

So that's why I prefer working with students. They are still learning their approaches and methodology. They come out of school very fresh with fresh ideas. They are optimistic. They aren't jaded, and they have a lot of new insights that perhaps a 30-year veteran evaluator on the job wouldn't have." (Adam)

Because they are students, the evaluators seem receptive and teachable or as one NGO director stated, "...able to unlearn everything" and observe what's really going on (Katie). A past PEAT student-director also observed the same benefit:

I think a student is in a learning mindset, which is really good for an evaluator—in my opinion. They don't have blinders on. They don't have past experience to cloud what they see, and I think that's really valuable. There's room for "ah-ha" moments. (David)

Additionally, the students are young and adventurous, which means they often live among the program's beneficiaries. This gives them access to the people and their personal lives as well as enables them to build trust and rapport. One student observed:

We lived in the villages during the week with the people we were talking to and surveying and staying in their houses and I just loved being with them...These people love students, and they love having young

people come; there was not that intimidation factor: “Oh the professional who thinks they know what they are doing coming to ask us all these questions.” It was just three young girls coming to live with them in their houses. The women became our mothers while we were down there, and the people were very comfortable with us. I think that may have been a huge advantage. (Shannon)

Given the contexts the organizations work within, the evaluators are often required to navigate difficult circumstances. As one executive director stated:

“[Students] just throw themselves into situations...I’m having [a PEAT intern] visit this really remote community in the middle of the mountains. It is a really uncomfortable ride on the bus, and then it’s a 45-minute trek. As students, they’re willing to do stuff like that. As a professional, you might say, “Where’s my taxi?” (James)

As students, they are also trained at the university level and, thus, have access to the many resources and expertise it offers. They are immersed into the most “recent or cutting-edge” methods and ideas (JM), and are newly training and eager to keep learning.

However, being a student also has its downsides. Because they are still learning, time is required for adjusting to mistakes, the context, or evaluation outside of the classroom. Because they are enrolled in other classes, their available time for the evaluation is limited before and after implementation. Using students also confines the field experience to the summer months, which negatively influenced at least three evaluations.

Utilizing students allows PEAT to offer small organizations high quality evaluations at a low-cost. Yet, the students have varying levels of expertise and training prior to and during the evaluation. Some of the interns have more applied experience in their respective fields than others. Yet, some cannot offer all the skills needed for the evaluation (e.g., complicated statistical analysis) and others are simply hindered because they are inexperienced with conducting evaluation in the field. The students walk into pretty difficult and unfamiliar circumstances. Despite their coursework training, they are unequipped to adapt as quickly or effectively. As one PEAT student-director explained:

I would say the context is probably responsible for most of the things that go wrong, but it’s probably because the student can’t be held responsible for not really knowing how to keep it right. I think it’s not just

that they are *student* evaluators, but that they are *first-time* student evaluators. So everything to them is completely new, and no amount of coursework is going to make it better. I think that's probably the biggest limitation—not that they are students, but that they are new. (David) [italics added]

Both the benefits and limitations associated with using students must be taken into consideration when evaluating the fit and value of pairing students with potential PEAT clients.

SUMMARY OF FINDINGS

The data collected in this study provides an in-depth look into the attributes of student-led evaluations for the clients of the BYU PEAT internship program. The features of the process include: 1) the attributes of the organization, evaluator (students) and evaluation, 2) the various influences that resulted from the evaluations, and 3) contingencies in the environment. These components form a conceptual model of these specific evaluations that represents a modified form the models of Mark and Henry (2004) and Appleton-Dyer et al. (2012) (see Figure 3 for the adapted conceptual model). Below is a summary of the data as it corresponds to these different elements. The section immediately following this summary, “Lessons Learned,” addresses how these findings relate to the research objectives of this thesis.

The first category in the conceptual model derived from this thesis is the “PEAT Partnership,” which includes the characteristics of the organizations and students. A description of the organizations is summarized first and then the students. From the years 2009-2013, 41 students conducted 23 evaluations for 15 organizations. Eleven of PEAT’s clients are considered small, community-based non-governmental organizations. The remaining four have large budgets dedicated to their cause due to their governmental, religious or corporate affiliations. All of the organizations work among vulnerable populations with the intent to provide services that may alleviate a variety of problems found in the human condition. With these good intentions, they often work in poverty-stricken, unpredictable and complex circumstances. Contextual

constraints are only further complicated by the way many of the organizations are structured and operate.

Typically, social innovators started these organizations and are passionate about the cause and providing effective services. They are creative and qualified individuals, but inadequate resources prevent them from hiring the needed amount of staff members with crucial skillsets. With high turnover and limited expertise, many of them struggle to adequately define and identify the problem or needs of the intended beneficiaries and, subsequently, to create a succinct and targeted strategy to address them. Some of them are newer organizations while others have been around for a few decades; yet, most of them are continually developing and adjusting with little expansion and marginal progress.

Furthermore, their understanding of this progress and how to best adapt their programs as they develop is rarely guided by systematic evaluation. All of the organizations face barriers to conducting evaluation at all levels of their organizational development or to learn from a variety of evaluative approaches. These barriers range from a lack of interest, expertise, funding or time internally to the inability to find people capable and willing to provide the service within their budget. Additionally, many profess a desire to learn from their mistakes but lack the formal means and channels to do so. Given these elements, all of the organization directors decided to collaborate with PEAT because they felt they had something to gain from the experience. They expressed a desire for the evaluations to help with organizational development and identify what works and why. And all of them felt beforehand that the students could adequately provide these services with the assistance of the PEAT program.

The stakeholders' investment in the evaluation was demonstrated in a variety of ways. While some provided ample support, some of the staff members' commitment wavered over time

and others seemed shallow, but in the end, all of the evaluations were completed except one. Even still, communication between PEAT (including the faculty and students) and the nonprofit personnel could have been more frequent and clear. This would have helped the process go smoother and align the evaluation to the direct desires and status of the organization. The organizations were not prepared in many ways to provide the students with the collaboration they needed and some staff members set expectations for the evaluation that could not be met given their stage of development and level of preparation.

The aforementioned characteristics describe the organizational environment in which the students performed the evaluation. The students were selected as interns because they had applicable experience and expertise prior to applying. During the program, the majority of the evaluators were further trained in evaluation techniques in two courses dedicated to the internship. The students utilized their previous experience as well as the training from the PEAT coursework in preparing for and conducting the evaluation. However, they felt that the entire experience would have been more positive for them and the evaluation more effective for the nonprofit had they received more training in what was required to implement an evaluation from beginning to end and adapt it to a real-world setting.

The students planned to conduct a range of evaluation types, however, once in the field, the majority adjusted their plans to focus on program process and outcome assessments. Additionally, the students described needing to adapt their role, to stay open-minded, and do whatever necessary to finish the evaluation and provide what they perceived to be quality information. Despite conducting a variety of qualitative and quantitative methods, many felt the most insightful data came from the informal conversations or observations obtained while living and interacting with the participants and in-country staff members. Inadequate time and training

to adjust to the context was among their greatest challenges, but their ability to connect with the population and critically analyze the situation from within was among their favorite moments and individual strengths.

After they returned, the students felt that they were limited in their ability to spend adequate time on analyzing and synthesizing the data for a report. Many could not find enough time away from their other schoolwork, and when they could focus on it, felt they were unable to identify which data was most relevant to the organization's needs while still meeting the requirements of the PEAT program. While many took months to submit their report to the organization, they were able to present a summary of their findings and recommendations shortly after returning from the field. Both the students and organization representatives expressed that these informal meetings and documents were informative and beneficial.

The second category of the conceptual model is the "Evaluation Attributes," including the inputs, activities, and quality. In summary, the evaluations provided information that described the people and environment the organizations work within, how their missions and goals unfold in reality, the process and details associated with program delivery, and the perceived outcomes of their programs. The students felt the quality of this data was lowered due to biases in the sampling techniques, inherent distance between them and the respondent, and their relative inexperience with the culture, context and evaluation overall. However, they were cognizant of these challenges, made every effort to address them, and despite their perceived methodological setbacks, felt the organizations benefited from the evaluation. The organization staff members, in turn, described evidence of the evaluation influencing them in a variety of ways, and the perceived quality of the evaluation did not seem to hinder its usefulness.

The third category in the model is the “Evaluation Influence,” which is further divided into the general mechanisms and the cognitive and behavioral changes they mediate. The data gathered from the organizations and students showed evidence of the evaluation affecting the organizations primarily through the general influence mechanisms of elaboration, priming, justification, persuasion, and motivating a change-agent. These mechanisms in turn led to the cognitive outcomes of opinion change and salience (Henry and Mark 2003). The mechanisms and outcomes also helped mediate behavioral changes in the organizations’ practices and programs. In addition, the student-evaluations brought about learning at the individual and collective level followed by discussion and deliberation over the issues and needed adjustments. The primary source of influence was the results of the evaluation, the organizations heard both anticipated and new information, and most of the changes took place within one year after the evaluation or are currently underway (Kirkhart 2000).

LESSONS LEARNED

This study had the following research objectives: 1) understand the characteristics, function and evaluation behavior of the organizations; 2) describe the nature, purpose, and outputs of the student-led evaluations; and 3) identify the types of influences or change processes that resulted from the evaluations; in order to 4) understand what the evaluation process reveal about how best to provide evaluations for developing organizations. The overall aim was to clarify the most appropriate type, influential purpose, and viable methods of evaluation inquiry for organizations that work in complex environments and struggle to develop via systematic reflection and analysis. The lessons learned regarding these objectives are discussed below.

The first research objective was to uncover important details about the organizations that PEAT works with in order to gauge how best to provide evaluations for organizations that

struggle to incorporate or learn from evaluation. The data showed that PEAT's clients were no exception to the pressures associated with the accountability movement. This was seen in the number of organization directors who indicated on the survey they initially desired the evaluation to help with accountability and to assess impact. While some had engaged with evaluation beforehand, it was primarily among the larger, more established organizations and/or the experiences were negative, the influence was limited, or the focus inappropriate given the organization's stage of development. Thus, the data also indicated that most of PEAT's clients previously struggled immensely with evaluation, were wary of it, just went through the motions or rejected it altogether (Carmen 2010). In the past, the organizations struggled to effectively utilize any evaluative information that came their way. Whether viewed as irrelevant or incompatible, data gathered via critical analysis was often only used to report on success to donors or board members rather than for organizational learning and improvement.

The findings showed that organizations that are small, unstable or lack the capacity to evaluate are in need of continual inquiry and reflection (Patton 2011). They need detailed information on their beneficiaries and status of their programs, to improve record keeping, and create evidence-based strategies to guide their efforts. While understanding the impact of their work is important, the organizations could benefit from various evaluation efforts before impact can be assessed—much of what the student-evaluations accomplished.

The second research objective was to understand the nature of the students' work. Overall, their work involved illuminating and judging the program and its performance by collecting, analyzing, interpreting and communicating information to PEAT's clients (Rossi et al. 2004). The students provided an in-depth look into the inner workings of the program to illuminate the aspects that needed more attention (Mulwa 2008). Considering the general

influence mechanisms like learning and elaboration that emerged from the evaluations (Mark and Henry 2004), the students' process of accumulating and summarizing data served as a valuable evaluative end in and of itself. Yet, unexpectedly, the utility and scope of the student evaluations extended further. Given the behavioral and attitudinal outcomes, the students' evaluative conclusions drawn from their descriptions of the performance and status of the programs can be viewed as important and valid.

For the most part, the students engaged in formative evaluations that proved immediately useful for the various stakeholders (Rossi et al. 2004). Within the traditional purposes of evaluation, the students' work helped in decision-making, enlightened the organizations on the nature of the program, and produced largely *descriptive* data on the potential impact of the program to help improve and judge the program's performance. They accomplished this by answering questions regarding the program's strengths and weaknesses, progress, assumptions, whether the program worked, and if it contributed to the stated goals and outcomes. By providing this information, the evaluations have helped improve the organizations' accountability to those closest to the organization, namely its beneficiaries, staff and donors.

The student-led evaluations also served nontraditional functions, such as those suggested by Patton (1996), Mulwa (2008) and Torres and Preskill (2001). While the students did provide judgments, they primarily acted as facilitators in producing data that allowed the organizational staff to learn, elaborate, discuss, act and "arrive at their own judgments" informatively (Patton 1996:138). Their positions varied throughout the evaluation and deliberation processes with some more immersed than others; yet, they infused "team discussions with evaluative questions, thinking, and data to facilitate systematic data-based reflection and decision-making (Patton

2011:1-2). These processes helped adjust the focus for these organizations from strictly impact and efficiency to learning and development and the associated implications for neglecting them.

The student-evaluations accomplished both the traditional and nontraditional purposes by following a more interpretivist approach to inquiry. Most of the information was gathered inductively taking into consideration the context-specific setting. This allowed the evaluators to immerse themselves in the program and analyze how it unfolds in the field by focusing on the people, context, process, and phenomenon as a whole (Patton 1990). The quality of the data varied between evaluations depending on the biases introduced and the students' level of preparation and ability to adapt. Students also struggled with the underlying emphasis on rigor, impact, and scholastic questions over what they eventually perceived would better meet operational needs. Thus, the students waged internal battles on how to adjust methods to reality while maintaining standards, to identify the most applicable yet academically focused data, and to write reports that met all stakeholders' expectations, including the PEAT faculty. Even though the students provided their preliminary thoughts and conclusions to the organizations afterward, some reports were still significantly delayed, lengthy and verbose.

The third research objective was to understand the benefits that were derived from the evaluations. While there were weaknesses throughout the entire process, the influence of students' evaluations not only brought about indirect and incremental effects through learning, but also purposeful action. Both the students and organizations had overwhelmingly positive experiences and received multiple and measurable benefits from participating in the PEAT program. Using the data collected on the organizations, the PEAT program, the students' experiences and evaluations as a whole, the following section outlines the recommended

evaluation approaches best suited for a specific type of organization, the environments they work in and the evaluators performing them. This information addresses the fourth research objective.

Recommended Approach for the Organization

In general, the organizations that are most likely to work with PEAT and benefit from the relationship are small, continually developing and face barriers to implementing evaluation.

Because many of the staff members are new to evaluation and emotionally involved with every facet of the program, training through presentation and collaboration must play an important role in the entire process. This would include discussing the purpose and methods of evaluation.

In preparation, the PEAT program (including faculty and student) and the organization should openly discuss the nonprofit's mission and assumptions concerning the problem and their approaches to addressing it. It would be wise to learn the staff members' story—the reason for working with the organization and their vision. This process should include discovering together what the nonprofit wants and needs to learn from the PEAT evaluation. In turn, the organization directors should fully understand the abilities of the students and how the evaluation can or cannot meet those demands. Overall, this phase should entail getting all key players on the same page about the nature of the problem, the organization and its programs, and the evaluation. In the face of so much complexity, the evaluation process should be kept simple and iterative with the evaluators willing and able to react, probe, sense and respond to the organization's needs.

The evaluation function should focus on learning and development. In this way, the organizations themselves can become clearer about their direction, beneficiaries and program so that trust in the organization and its ability to serve can increase. In essence, the evaluations can be learning-centered by focusing on the aspects and details often inaccessible to the organization and catalyzing the critical-thinking process. The intent is to provide an in-depth look into the

program and context in order to form strategies rooted in data, improve delivery, understand needs, track progress, and guide these organizations as they grow and “figure things out.” Additionally, the students are in a unique position to help organizations create good baseline measures and build evaluation capacity for process and outcome assessments.

To meet these purposes, the evaluators will need to play multiple roles (Skolits, Morrow, and Burr 2009; Patton 2007). In the preparation phase, they act as detectives, designers, and negotiators to better understand the organization and design a relevant and feasible evaluation (Skolits et al. 2009). While in the field, some students become observer-participants, but most must find the balance between acting as the eyes and ears of the organization and separating themselves enough to gain trust and maintain a critical viewpoint. This allows them to understand and convey needs and obtain a more accurate snapshot. In this role, the student-evaluators act as information brokers, researchers, advocates, and analysts. The students are also facilitators, judges, and consultants as they provide recommendations and deliberate with organizational staff.

The most beneficial types of assessment are needs-, process- and outcome-focused. Most of the organizations have neither conducted formal needs assessment nor designed their programs from data. As Patton stated, “In our accountability-focused and results-driven culture...process matters. The means to ends matter, not just ends (Patton 2011:246). Most of the organizations are not yet equipped to show impacts or do impact evaluations. While interest in evaluation is increasing and pressure for proof is high, PEAT’s clients are still improving and continually adjusting their programs so that they can have the outcomes and impacts they hope for. Assessing impact is improbable if they first do not understand the program and whether it’s

running properly; it is not beneficial to assess it unless the organization is ready—some may never reach the state of development fit for traditional impact methods (Patton 2011).

The methods employed should be mixed with an emphasis on the needs of the organizations rather than rigor. While some may ask to track outputs with numbers, others may benefit most from descriptive information. The evaluations do not need to be generalizable and should be simple enough to be practical—use is more important than pomp. Utilizing previously tested methods and measures combined with creativity will ensure accuracy and consistency. Limitations in time, record-keeping, sampling, adequate comparisons, and accessing targets should be acknowledged upfront as well as the hindrances associated with political undercurrents, ulterior motives, and miscommunication. The quality can then be rooted in the trustworthiness of the data to accurately portray reality and provide practical advice for organizational development.

Recommended Approach in Context

The evaluation design and process should consider contextual elements. Many of the communities are in rural areas and have had little exposure to evaluation or outsiders; some of the participants may have had less than positive experiences with evaluation or view it as a threat. Some beneficiaries may be more open than others to talking honestly about the organization and their experience; others may need the organization's services even if they are less effective than they could be. Each of these factors will determine how the participants respond to evaluation and the evaluators.

The characteristics of the evaluators—merely being different—may bias their ability to gather accurate data. For instance, respondents may view the evaluator as an outsider who doesn't care, as access to more resources, or as someone who can tell their story to those who

matter. Approaches, definitions, interpretations, and ideas may differ between researcher and respondent. Thus, the evaluators must be able to identify and adjust respectfully to potential barriers. They must be able to balance training and expertise with personality—the ability to connect with people and build rapport.

During the evaluation, the evaluators may be required to alter how they introduce and explain their role and purpose. In most cases, they should be prepared in the relevant theories, concepts, tools, and techniques and then be willing to make adjustments when required. While rigor and comparability are important, they may need to be sacrificed for cultural sensitivity. In many circumstances, collaborative or participatory methods will help encourage interaction and understanding; in this case, the evaluator must be equipped with the resources and ideas to facilitate that process.

Overall, the evaluator should expect for many aspects to change, or in the very least, accept that they cannot control everything. As one executive director suggested, “Always remain open-minded and don’t get boxed in by your own methodology. Always keep an open viewpoint to whatever it is you think you may be seeing” (MC). These contexts are unpredictable and resources are limited. Thus, requirements should be tempered to allow room for adaptation, or as many students expressed, to do the best they can with what they are given.

PEAT’s Role

The BYU PEAT internship program provides a myriad of benefits for its clients. However, it also faces limitations in its ability to address needs from a range of organizational types or implement a variety of tools in a professional evaluator’s toolkit. The program utilizes students to conduct the evaluations and, thus, is limited by their relative inexperience and minimal training. Their target audience is specific though. They are most often small

organizations, unstable, work in complex environments, and lack the time or expertise to conduct formal, in-depth evaluations. In most cases, the organizations would not engage in evaluation if it were not for PEAT. Thus, PEAT plays an important role in providing quality, useful assessments to these clients and is uniquely positioned to offer the recommended approaches.

In particular, PEAT's role is to help their clients learn and develop. First, they provide these organizations with information. As one student stated, "Even for small NGOs [that] would never have the resources otherwise, it doesn't hurt. Information is always helpful and needed" (Ann). Most often the information is formative data with insights on outputs, status, concerns, needs, gaps, and perspectives. This feedback and description is essential for improvement but had previously been neglected or carelessly gathered. With the information, the evaluation initiates the reflection process in which learning can lead to thinking, deliberation, motivation, and then action. Concerning the benefits to the organizations, one student evaluator described:

It helps them be reflexive about what they are actually doing. Whether the actual NGO report is read or not, PEAT evaluators bring in a culture of questioning: "Let's not just assume everything is great. Let's think twice about it." That's probably the biggest contribution we made." (Tyler)

The PEAT evaluations have also helped gather baseline data, prepare the organization for more evaluation, improve their previous evaluation techniques, fix records, and begin data collection that will help the organizations better identify their impact. Upon their return, some students acted as a sounding board for the organizations. This is often because they were viewed as experts on the situation after having spent so much time in the area and with the people. Whether the organizations heeded their recommendations or not, the reports and insights provided by the students initiated discussions that would lead to programmatic changes and organizational adjustments. Additionally, the PEAT interns often trained the in-country staff and the organizations learned more about the process of evaluation by simply participating in the

program. Altogether, the information, training, and PEAT experience helped facilitate the organizations' development and growth.

CONCLUSION

This study set out to understand the benefits to and influence of student-led evaluations for nonprofits that struggle to conduct quality evaluations for organizational learning and development. Cox et al. (2013) recently presented evidence of the potential benefits to partnering students with small nonprofits. With three years of additional data, the current study supports their findings in suggesting that university-trained students, while limited, are a cost-effective solution for nonprofits that work in developing-contexts and face challenges to “implement evaluation methodologies that are both academically sound and feasible” (Cox et al. 2013:2). Students are not only trained in applied evaluation techniques, but embody characteristics that allow them to integrate and connect with the service population as well as eagerly adapt to unpredictable circumstances. Along with Cox et al. (2013), this study contributes to the literature by outlining the advantages and limitations to this partnership and suggesting ways to maximize one while minimizing the other.

However, this research extends their work by utilizing the evaluation influence framework to identify the specific types of outcomes and changes that may result from these evaluations. Evaluation influence scholars, like Karen Kirkhart, Melvin Mark and Gary Henry, created schematic theories of change that outlined the key factors, characteristics and mechanisms that mediate an evaluation's influence. They also expanded the range of potential outcomes by classifying the distal, indirect, and incremental outcomes that link together larger change processes. Mark and Henry (2004) and Appleton-Dyer et al.'s (2012) influence models situate an evaluation's effects in the context of its attributes and various sources of

contingencies. The models served as the primary guides in the analysis and interpretation of data, and, in turn, this study produced demonstrable evidence for the models' key elements. With this evidence, the author created a new conceptual model that shows the relevant factors, including the role of students, that provides the basis for more appropriate and effective assessments for these organizations. Nonprofits and NGOs can also use the model as they consider and develop program strategies that will lead to a range of potential outcomes in the complex environments in which they work.

As a systematic inquiry into the consequences of student-led evaluations, this thesis provides empirical evidence about "...what kinds of evaluation worked in which circumstances to produce which specific outcomes" (Mark and Henry 2004:51). Specifically, formative evaluations that focus on feedback and reflection provide the most useable data for organizations in the exploration and exploitation phases of development (Westley and Miller 2003). The student-evaluators utilized interpretivist approaches to observe and report on program status, provide suggestions for improvement, and prepare these organizations for further learning. For this type of organization and context, professional expertise and academic rigor are moderated by adequate preparation and collaboration, flexibility, and a willingness to do "whatever it takes." These methods and approaches led to interim changes in thinking, attitudes and behavior, shining light on the importance of each stage within a complex chain of influence. Additionally, this helps to temper and clarify the responsibility, purpose, and replication of these student-evaluations for developing organizations.

Limitations in this study include the minimal amount of qualitative data gathered from the nonprofit staff members and reliance on self-report and retrospective memory to gauge each element of the evaluation process. Further research could include a variety of perspectives from

the client side and follow them over time to better observe long-term change and whether they conduct additional evaluation. The author advises using more methodical ways to track the pathways between each mechanism and outcome. Lastly, the author urges application of the conceptual model outlined here as a guide for evaluation research and practice in this sector and for the small nonprofit organizations working towards social betterment.

REFERENCES

- Aksel, Ibrahim and Muhtesem Baran. 2006. "Organizational Problems of Non-Governmental Organizations (NGOs)." *Turkish Weekly*, December.
- Alaimo, Salvatore P. 2008. "Nonprofits and Evaluation: Managing Expectations from the Leader's Perspective." Pp. 73-92 in *Nonprofits and Evaluation: New Directions for Program Evaluation*, edited by J.G. Carman and K.A. Fredericks. San Francisco: Jossey-Bass.
- Alkin, Marvin C. and Sandy M. Taut. 2003. "Unbundling Evaluation Use." *Studies in Educational Evaluation*, 29:1-12. doi: 10.1177/1098214009341660.
- Appleton-Dyer, Sarah, Janet Clinton, Peter Carswell, and Rob McNeill. 2012. "Understanding Evaluation Influence Within Public Sector Partnerships: A Conceptual Model." *American Journal of Evaluation* 33(4):532-46. Retrieved September 23, 2013 (<http://aje.sagepub.com/cgi/doi/10.1177/1098214012447672>).
- Avina, Jeffrey. 1993. "The Evolutionary Life Cycles of Non-Governmental Development Organizations." *Public Administration and Development* 13:453-474. doi: 10.1002/pad.4230130502.
- Bambergher, Michael. 2000. "The Evaluation of International Development Programs: A View From the Front." *American Journal of Evaluation* 21(1):95-102. doi: 10.1177/109821400002100108.
- Bambergher, Michael, Jim Rugh, Mary Church, and Lucia Fort. 2004. "Shoestring Evaluation: Designing Impact Evaluations under Budget, Time and Data Constraints." *American Journal of Evaluation* 25(1):5-37. Retrieved on April 30, 2013 from (<http://aje.sagepub.com/content/25/1/5>).

- Banerjee, Abhijit V. and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics* 1:151-178.
doi:10.1146/annurev.economics.050708.143235.
- Baños Smith, Helen. 2006. "International NGOs and Impact Assessment: Can We Know We are Making a Difference?" *Research in Drama Education: The Journal of Applied Theatre and Performance* 11(2):157–74. Retrieved September 23, 2013
(<http://www.tandfonline.com/doi/abs/10.1080/13569780600671021>).
- Barrett, Christopher B. and Michael R. Carter. 2010. "The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections." *Applied Economic Perspectives and Policy* 32(4):515–48. Retrieved November 14, 2013
(<http://aep.oxfordjournals.org/lookup/doi/10.1093/aep/ppq023>).
- Bell, Stephen A. and Peter Aggleton. 2012. "Integrating Ethnographic Principles in NGO Monitoring and Impact Evaluation." *Journal of International Development* 24:795-807.
- Bergman, Manfred M., ed. 2008. *Advances in Mixed Methods Research*. Thousand Oaks, CA: Sage.
- Black, Maggie. 2009. *The No Nonsense Guide to International Development*. Oxford, UK: New Internationalist.
- Boroch, R. and D. Rindskopf. 1984. "Data Analysis." Pp. 121-158 in *Evaluation Research Methods*. 2nd ed., edited by L. Rutman. Beverly Hills, CA: Sage.
- Briedenhann, Jenny and Steve Butts. 2005. "Utilization-Focused Evaluation Evaluation: A Conceptual Clarification." 22(2):221–44.
- Burman, Len. 2011. "New Flash: Antipoverty Programs Can Reduce Poverty." *Forbes Magazine*, November 7. Retrieved October 15, 2013

[\(http://www.forbes.com/sites/leonardburman/2011/11/07/news-flash-antipoverty-programs-can-reduce-poverty/\)](http://www.forbes.com/sites/leonardburman/2011/11/07/news-flash-antipoverty-programs-can-reduce-poverty/).

Carman, Joanne G. 2007. "Evaluation Practice Among Community-Based Organizations: Research Into the Reality." *American Journal of Evaluation* 28(1):60–75. Retrieved September 23, 2013 (<http://aje.sagepub.com/cgi/doi/10.1177/1098214006296245>).

Carman, J. G. 2008. "Nonprofits, Funders, and Evaluation: Accountability in Action." *The American Review of Public Administration* 39(4):374–90. Retrieved November 14, 2013 (<http://arp.sagepub.com/cgi/doi/10.1177/0275074008320190>).

Carman, J. G. 2009. "The Accountability Movement: What's Wrong with This Theory of Change?" *Nonprofit and Voluntary Sector Quarterly*. Retrieved September 18, 2013 (<http://nvs.sagepub.com/cgi/doi/10.1177/0899764008330622>).

Carman, Joanne G. and K. A. Fredericks. 2009. "Evaluation Capacity and Nonprofit Organizations: Is the Glass Half-Empty or Half-Full?" *American Journal of Evaluation* 31(1):84–104. Retrieved September 23, 2013 (<http://aje.sagepub.com/cgi/doi/10.1177/1098214009352361>).

Cartwright, Nancy. 2007. "Are RCTs the Gold Standard?" *BioSocieties* 2(1):11-20.
doi:10.1017/S1745855207005029.

Chen, Huey-Tsyh. 1996a. "A Comprehensive Typology for Program Evaluation." *American Journal of Evaluation* 17(2):121–30. Retrieved September 23, 2013 (<http://aje.sagepub.com/cgi/doi/10.1177/109821409601700204>).

Chen, Huey-Tsyh. 1996b. "Epilogue: Synthesizing Formative/Summative Evaluation Issues and Beyond." *American Journal of Evaluation* 17(2):163–67. Retrieved September 23, 2013 (<http://aje.sagepub.com/cgi/doi/10.1177/109821409601700208>).

- Chen, Huey-Tsyh. 2005. *Practical Program Evaluation: Assessing and Improving Planning, Implementation, and Effectiveness*. Thousand Oaks, CA: Sage.
- Clark, John. 1990. *Democratizing Development: The Role of Voluntary Organizations*. West Hartford, CT: Kumarian Press.
- Clements, Paul, Thomaz Chianca, and Ryoh Sasaki. 2008. "Reducing World Poverty by Improving Evaluation of Development Aid." *American Journal of Evaluation* 29:195-214.
- Corbin, Juliet and Anselm Strauss. 2008. *Basics of Qualitative Research: 3rd edition*. Thousand Oaks, CA: Sage.
- Cousins, Bradley J. and Elizabeth Whitmore. 1998. "Framing participatory evaluation." *New Directions for Evaluation* 1998(80):5–23. Retrieved April 30, 2014 (<http://doi.wiley.com/10.1002/ev.1114>).
- Cox, Matt, Kendal Blust, and Tim Heaton. 2013. "Finding Solutions: How Small Non-Governmental Organizations (NGOs) Can Partner With Universities to Overcome Evaluation Challenges." Pp. 119-138 in *Developing Countries: Political, Economic and Social Issues*, edited by A. Ramazzotti and W. Gravina. Hauppauge, NY: Nova Science.
- Creswell, John. 2007. *Qualitative Inquiry and Research Design*. Thousand Oaks, CA: Sage.
- Cronbach, Lee. 1982. *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass.
- Deaton, Angus. 2010. "Learning about Development." *Journal of Economic Literature* 48(June):424–55.

- Development Assistance Committee. 1991. "DAC Principles for the Evaluation of Development Assistance." Paris, France: The Organization for Economic Co-operation and Development. Retrieved November 14, 2013 (<http://www.oecd.org/dac/evaluation/2755284.pdf>).
- Development Assistance Committee. 2010. "Development Evaluation Resources and Systems: A Study of Network Members." Paris, France: The Organization for Economic Co-operation and Development. Retrieved November 14, 2013 (www.oecd.org/dac/evaluationnetwork/documents).
- Duflo, Esther. 2012. "Rigorous Evaluation of Human Behavior." *Science* 336(6087):1398. doi: 10.1126/science.1224965.
- Easterly, William. 2006. *The White Man's Burden: Why the West's Efforts to Aid the Rest Have done So Much Ill and So Little Good*. Oxford, NY: Oxford University Press.
- Easterly, William. 2009. "Development Experiments: Ethical? Feasible? Useful?" Aid Watch: Just Asking That Aid Benefit the Poor, July 16, 2009. Retrieved November 14, 2013 (<http://aidwatchers.com/2009/07/development-experiments-ethical-feasible-useful/>).
- Easterly, William, ed. 2008. *Reinvigorating Foreign Aid*. Cambridge, MA: MIT Press.
- "Evaluate". 2013. *Merriam-Webster.com*. Retrieved September 20, 2013 (<http://www.merriam-webster.com/dictionary/evaluate>).
- Everitt, Angela, and Ann Le Mare. 2013. "Evaluating as a Means to Achieve Development Ends." *Journal of International Development* 24:556-570.
- Eyben, R. 2010. "The Political and Ideological Context of Assessing and Reporting on Making a Difference in Development. Retrieved on April 30, 2013 on (<http://www.ids.ac.uk/files/dmfile/tyrannyconceptnote17Mayoc.pdf>).

- Frank, Robert H. 2007. *Falling Behind: How Rising Inequality Harms the Middle Class*. Berkeley, CA: University of California Press.
- Festin, Marcia and Marianne Philbin. 2007. *Level Best: How Small and Grassroots Nonprofits Can Tackle Evaluation and Talk Results*. San Francisco: Jossey-Bass.
- Fetterman, David M. 2001. *Foundations of Empowerment Evaluation*. Thousand Oaks, CA: Sage.
- Gopalakrishnan, Srik, Hallie Preskill, and Lu Shijie. 2013. *Next Generation Evaluation: Embracing Complexity, Connectivity, and Change*. FSG. Retrieved September 30, 2013 (fsg.org).
- Grasso, Patrick G. 2010. "Ethics and Development Evaluation: Introduction." *American Journal of Evaluation* 31(4):533–39. Retrieved September 17, 2013 (<http://aje.sagepub.com/cgi/doi/10.1177/1098214010373647>).
- Greenberg, David, Marvin Mandell, and Matthew Onstott. 2000. The Dissemination and Utilization of Welfare-to-Work Experiments in Pstate policymaking. *Journal of Policy Analysis & Management* 19(3):367–382.
- Greene, Jennifer C. and Gary T. Henry. 2005. "Qualitative-Quantitative Debate in Evaluation." Pp. 345-350 in *Encyclopedia of Evaluation*, edited by S. Mathison. Thousand Oaks, CA: Sage.
- Grudens-Schuck, Nancy. 2003. "The Rigidity and Comfort of Habits: A Cultural and Philosophical Analysis of the Ups and Downs of Mainstreaming Evaluation." *New Directions for Evaluation* 2003(99):23–32. Retrieved November 15, 2013 (<http://doi.wiley.com/10.1002/ev.91>).

- Hatry, Harry P., Joseph S. Wholey, and Kathryn E. Newcomer. 2004. "Other Issues and Trends in Evaluation." Pp. 670-684 in *Handbook of Practical Program Evaluation*, edited by J. S. Wholey, H. P. Hatry and K. E. Newcomer. 2nd ed. San Francisco: Jossey-Bass.
- Henry, Gary T. 2003. "Influential Evaluations." *American Journal of Evaluation* 24:515-525.
- Henry, Gary T. and Melvin M. Mark. 2003. "Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions." *American Journal of Evaluation* 24(3):293-314. Retrieved September 17, 2013 (<http://aje.sagepub.com/cgi/doi/10.1177/109821400302400302>).
- Hoole, Emily and Patterson, Tracey E. 2008. "Voices from the Field: Evaluation as Part of a Learning Culture." Pp. 93-114 in *Nonprofits and Evaluation: New Directions for Program Evaluation*, edited by J. G. Carman and K. A. Fredericks. San Francisco: Jossey-Bass.
- House, Ernest R. and Howe, Kenneth R. 1999. *Values in Evaluation and Social Research*. Thousand Oaks: Sage.
- Howe, Kenneth R. 1992. "Getting Over the Quantitative-Qualitative Debate." *American Journal of Education* 100(2):236-256. Retrieved November 8, 2013 (<http://www.jstor.org/stable/1085569>).
- International Development Association. 2008. "Aid Architecture: An Overview of the Main Trends in Official Development Assistance Flows." Washington, DC: The World Bank. Retrieved November 14, 2013 (http://siteresources.worldbank.org/IDA/Resources/Aid_Architecture-May2008.pdf).
- James, Rick. 2009. "Just Do It: Dealing With the Dilemmas in Monitoring and Evaluation Capacity Building." INTRAC M&E Paper 2. Retrieved December 6, 2013

(<http://www.intrac.org/data/files/resources/659/Dealing-with-the-Dilemmas-in-Monitoring-and-Evaluating-Capacity-Building.pdf>).

Jensen, Leif, Diane K. McLaughlin, and Tim Slack. 2003. "Rural Poverty: The Persisting Challenge." Pp. 118-131 in *Challenges for Rural America in the Twenty-first Century*, edited by D. L. Brown and L. E. Swanson. University Park, PA: Pennsylvania State University Press. Retrieved October 25, 2013 (<http://books.google.com/books>).

Jewiss, Jennifer and Kelly Clark/Keefe. 2007. "On a Personal Note: Practical Pedagogical Activities to Foster the Development of 'Reflective Practitioners'." *American Journal of Evaluation* 28(3):334–47. Retrieved September 23, 2013 (<http://aje.sagepub.com/cgi/doi/10.1177/1098214007304130>).

Johnson, Kelli, Lija O. Greenseid, Satcie A. Toal, Jean A. King, Frances Lawrenz and Boris Volkov. 2009. "Research on Evaluation Use: A Review of the Empirical Literature from 1986 to 2005." *American Journal of Evaluation* 30(3):377–410. Retrieved September 17, 2013 (<http://aje.sagepub.com/cgi/doi/10.1177/1098214009341660>).

King, Jean A. 2008. "Bringing Evaluation Learning to Life." *American Journal of Evaluation* 29(2):151-155.

Kirkhart, Karen E. 2000. "Reconceptualizing Evaluation Use: An Integrated Theory of Influence." *New Directions for Evaluation* 2000(88):5–23. Retrieved May 15, 2013 (<http://doi.wiley.com/10.1002/ev.1188>).

Ledermann, Simone. 2012. "Exploring the Necessary Conditions for Evaluation Use in Program Change." *American Journal of Evaluation* 33(2):159-178. Retrieved February 3, 2013 (<http://aje.sagepub.com/content/33/2/159>).

- Leeuw, Frans and Jos Vaessen. 2009. "Impact Evaluations and Development: NONIE Guidance on Impact Evaluation." Washington, DC: Network of Networks on Impact Evaluation. Retrieved November 14, 2013 (http://siteresources.worldbank.org/EXTOED/Resources/nonie_guidance.pdf).
- Lincoln, Yvonna S. and Egon G. Guba. 1985. *Naturalistic Inquiry*. Newbury Park, CA: Sage.
- March, James G. and Johan P. Olsen. 1976. *Ambiguity and Choice in Organizations*. Bergen, Norway: Universitetsforlaget.
- Mark, Melvin M. and Gary T. Henry. 2004. "The Mechanisms and Outcomes of Evaluation Influence." *Evaluation* 10(1):35–57. Retrieved November 18, 2013 (<http://evi.sagepub.com/cgi/doi/10.1177/1356389004042326>).
- Mathison, Sandra, ed. 2005. *Encyclopedia of Evaluation*. Thousand Oaks, CA: Sage.
- May, Maria. 2012. "RCTs: Not All That Glitters is Gold." *Measuring Social Impact: Innovative Ways to Measure the Impact That an Organization has on Society*, August 28, 2012. Retrieved November 14, 2013 (http://www.ssireview.org/blog/entry/rcts_not_all_that_glitters_is_gold).
- Minear, Larry. 2002. *The Humanitarian Enterprise: Dilemmas and Discoveries*. Bloomfield, CT: Kumarian Press. Retrieved November 12, 2013 (<http://books.google.com/books>).
- Mueller-Hirth, Natascha. 2012. "If You Don't Count, You Don't Count: Monitoring and Evaluation in South African NGOs." *Development and Change* 43(3):649–70. Retrieved November 12, 2013 (<http://doi.wiley.com/10.1111/j.1467-7660.2012.01776.x>).
- Mulwa, Francis. 2008. *Participatory Monitoring and Evaluation of Community Projects*. Nairobi, Kenya: Paulines Publications Africa.

- Murphy, Deena M. and Mitchell, Roger. 2007. Building evaluation capacity in North Carolina's nonprofit sector: A survey report. Raleigh, NC: Institute for Nonprofits.
- Murray, Victor. 2005. "Evaluating the Effectiveness of Nonprofit Organizations." Pp. 345-370 in *The Jossey-Bass Handbook of Nonprofit Leadership and Management*, 2nd ed., edited by R. D. Herman & Associates. San Francisco, CA: Jossey-Bass.
- Naidoo, Kumi. 2004. "The End of Blind Faith? Civil Society and the Challenge of Accountability, Legitimacy and Transparency." *AccountAbility Forum 2: Special Issue on NGO Accountability and Performance* 2:14-25. Retrieved November 5, 2013 (<http://www.greenleaf-publishing.com/content/pdfs/af02naid.pdf>).
- Nelson, Jane. 2007. "The Operation of Non-Governmental Organizations (NGOs) in a World of Corporate and Other Codes of Conduct." Corporate Social Responsibility Working Paper No. 34. Cambridge MA: John F. Kennedy School of Government, Harvard University.
- Neuman, W. Lawrence. 2011. *Social Research Methods: Qualitative and Quantitative Approaches*. Boston, MA: Pearson.
- Newcomer, Kathryn E. 2004. "How Might We Strengthen Evaluation Capacity to Manage Evaluation Contracts?" *American Journal of Evaluation* 25:209-218. doi: 10.1177/109821400402500205.
- Owen, John M. and Faye C. Lambert. 1998. "Evaluation and the Information Needs of Organizational Leaders." *American Journal of Evaluation* 19(3):355-65.
- Patton, Michael Q. 1990. *Qualitative Evaluation and Research Methods*. 2nd ed. Newbury Park, CA: Sage.

- Patton, Michael Q. 1996. "A World Larger than Formative and Summative." *American Journal of Evaluation* 17(2):131–44. Retrieved September 23, 2013
(<http://aje.sagepub.com/cgi/doi/10.1177/109821409601700205>).
- Patton, Michael Q. 2007. *Utilization-Focused Evaluation*. 4th ed. Thousand Oaks, CA: Sage.
- Patton, Michael Q. 2011. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford.
- Popham, James W. 2006. "Defining and Enhancing Formative Assessment." Presented at the Formative Assessment Forum for the Student and Teachers Subgroup, October 10, Austin, Texas. Retrieved October 5, 2013
(http://cpre.org/ccii/images/stories/ccii_pdfs/defining%20formative%20assessment_popham.pdf).
- Preskill, Hallie and Rosalie T. Torres. 1999. *Evaluative Inquiry for Learning in Organizations*. Thousand Oaks, CA: Sage.
- Preskill, Hallie and Valerie Caracelli. 1997. "Current and Developing Conceptions of Use: Evaluation Use TIG Survey Results." *American Journal of Evaluation* 18(1):209–25. Retrieved November 8, 2013
(<http://aje.sagepub.com/cgi/doi/10.1177/109821409701800122>).
- Ramadan, Layla Jamil Abu. 2013. "Assessing The Evaluation of Women Projects Implemented by NGOs in Gaza Strip: Projects Coordinators Perspective." Masters' dissertation, Department of Business Administration, Islamic University-Gaza, Gaza, Palestine.
- Ravallion, Martin. 2013. "How Long Will It Take to Lift One Billion People Out of Poverty." Policy Research Working Paper. The World Bank: Development Economics Vice

- Presidency Partnerships, Capacity Building Unit. Retrieved November 18, 2013
(<http://econ.worldbank.org>).
- Robinson, Mark A. 1992. "Assessing the impact of NGO Rural Poverty Alleviation Programmes: Evidence from South India." *Journal of International Development* 4:397-417. doi: 10.1002/jid.3380040405.
- Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman. 2004. *Evaluation: A Systematic Approach*. Thousand Oaks, CA: Sage.
- Sachs, Jeffrey D. 2005. *The End of Poverty: Economic Possibilities for Our Time*. New York: Penguin Books.
- Scott, Richard W. 1992. *Organizations: Rational, Natural, and Open Systems*. 3rd ed. Englewood Cliffs, NJ: Prentice Hall.
- Scriven, Michael S. 1967. "The Methodology of Evaluation." Pp. 39-83 in *Perspectives of Curriculum Evaluation: AERA Monograph Series on Curriculum Evaluation*, edited by R. Tyler, R. Gagne, and M. Scriven. Chicago, IL: Rand McNally. Retrieved on November 11, 2013 (<http://books.google.com/books>).
- Scriven, Michael. 1980. *The Logic of Evaluation*. Inverness, CA: Edgepress.
- Scriven, Michael. 1991. *Evaluation Thesaurus*. 4th ed. Newbury Park, CA: Sage.
- Scriven, Michael. 1996. "Types of Evaluation and Types of Evaluator." *American Journal of Evaluation* 17(2):151-61. Retrieved September 23, 2013
(<http://aje.sagepub.com/cgi/doi/10.1177/109821409601700207>).
- Shadish, William R., Thomas D. Cook, and Laura C. Leviton. 1991. *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, CA: Sage.

- Shivji, Issa G. 2007. *Silences in NGO Discourse: The Role and Future of NGOs in Africa*. Nairobi, Kenya: Fahamu.
- Skolits, Gary J., Jennifer A. Morrow, and Erin M. Burrow. 2009. "Reconceptualizing Evaluator Roles." *American Journal of Evaluation* 30(3):275-295. doi: 10.1177/1098214009338872.
- Smith, Deborah A. 2011. "Making the Case for the Humanities in Evaluation Training." *New Directions for Evaluation* 131:15-20.
- Smith, Steven R. 2005. "Managing the Challenge of Government Contracts." Pp. 131-152 in *The Jossey-Bass Handbook of Nonprofit Leadership and Management*, 2nd ed., edited by R. Herman Associates. San Francisco: Jossey-Bass.
- Snyder, Monteze M. and Peter L. Doan. 1995. "Who Participates in the Evaluation of International Development Aid." *American Journal of Evaluation* 16(2):141-152. Retrieved September 16, 2013 (<http://aje.sagepub.com/content/16/2/141>).
- Stake, Robert E. 1983. "Program Evaluation, Particularly Responsive Evaluation." Pp. 287-310 in *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, edited by G. F. Madaus, M. S. Scriven, D. L. Stufflebeam. Retrieved November 15, 2013 (<http://link.springer.com/book/10.1007/978-94-009-6669-7>).
- Stiglitz, Joseph E. 2003. *Globalization and Its Discontents*. New York: Norton Paperback.
- Struening, Elmer L. and Marilyn B. Brewer. 1983. *University Edition of the Handbook of Evaluation Research*. Thousand Oaks, CA: Sage.
- Stufflebeam, Daniel. 2005. "Joint Committee on Standards for Educational Evaluation." Pp. 214-215 in *Encyclopedia of Evaluation*, edited by S. Mathison. Thousand Oaks, CA: Sage. doi: <http://dx.doi.org/10.4135/978141290558.n294>.
- Suchman, Edward. 1967. *Evaluative Research*. New York, NY: Russell Sage Foundation.

- Sutherland, Stephanie. 2004. "Creating a Culture of Data Use for Continuous Improvement: A Case Study of an Edison Project School." *American Journal of Evaluation* 25(3):277-93. Retrieved September 23, 2013 (<http://aje.sagepub.com/cgi/doi/10.1177/109821400402500302>).
- Swindle, Jeffrey. 2011. "Tough Realities Small NGOs Face When Evaluating Their Efforts: Observations From Guatemala and Mexico." *Consilience: The Journal of Sustainable Development* 6(1):251-263.
- Thomas, Vinod. 2010. "Evaluation Systems, Ethics, and Development Evaluation." *American Journal of Evaluation* 31(4):540-548. Retrieved February 3, 2013 (<http://aje.sagepub.com/content/31/4/540>).
- Torres, Rosalie and Hallie Preskill. 2001. "Evaluation and Organizational Learning: Past, Present, and Future." *American Journal of Evaluation* 22(3):387-95.
- "Towards the End of Poverty." 2013. *The Economist*, June 1. Retrieved October 15, 2013 (<http://www.economist.com/news/leaders/21578665-nearly-1-billion-people-have-been-taken-out-extreme-poverty-20-years-world-should-aim>).
- Tuan, Melinda. 2004. "Cultivating a Culture of Measurement." Pp. 105-142 in *Funding Effectiveness: Lessons in Building Nonprofit Capacity*, edited by Grantmakers for Effective Organizations. San Francisco: Jossey-Bass.
- United Nations. September 2000. "The United Nations Millennium Declaration." Report 55/2. Retrieved September 25, 2013 (<http://www.un.org/millennium/declaration/ares552e.htm>).
- United Nations. 2012. "The Millennium Development Goals Report." Retrieved October 15, 2013 (<http://mdgs.un.org/unsd/mdg/Resources/Static/Products/Progress2012/English2012.pdf>).

- United Nations Development Program. January 2013. "Fast Facts." Retrieved September 25, 2013 (<http://www.undp.org/content/dam/undp/library/corporate/fast-facts/english/FF-Poverty-Reduction.pdf>).
- VSO. September 29, 2003. *The Role of Volunteers in International Development*. Retrieved September 30, 2013 (<http://forum-ids.org/2003/09/the-role-of-the-volunteer/>).
- Wenar, Leif. 2006. "Accountability in International Development Aid." *Ethics & International Affairs* 20(01):1–23. Retrieved September 5, 2013 (http://www.journals.cambridge.org/abstract_S0892679400003920).
- Weick, Karl E. 1976. "Educational Organizations as Loosely-Coupled Systems." *Administrative Science Quarterly* 21(1):1-19.
- Weiss, Thomas G. 2013. *Humanitarian Intervention: War and Conflict in the Modern World*. 2nd ed. Cambridge, UK & Malden, MA: Polity Press. Retrieved November 18, 2013 (<http://books.google.com/books>).
- Weiss, Carol H. 1972. *Evaluation Research: Methods for Assessing Programme Effectiveness*. Englewood Cliffs, NJ: Prentice Hall.
- Weiss, Carol H., Erin Murphy-Graham, Anthony Petrosino, and Allison G. Gandhi. 2008. "The Fairy Godmother—and Her Warts: Making the Dream of Evidence-Based Policy Come True." *American Journal of Evaluation* 29(1):29–47. Retrieved November 14, 2013 (<http://aje.sagepub.com/cgi/doi/10.1177/1098214007313742>).
- Westley, Francis and Philip S. Miller, eds. 2003. *Experiments in Consilience: Integrating Social and Scientific Responses to Save Endangered Species*. Washington, DC: Island Press.
- Westley, Francis, Brenda Zimmerman, and Michael Q. Patton. 2006. *Getting to Maybe: How the World Is Changed*. Toronto: Random House Canada.

Worstell, Tim. 2012. "If the US Spends \$550 Billion on Poverty How Can There Be Poverty in the US?" *Forbes Magazine*, September, 20. Retrieved October 15, 2013 (<http://www.forbes.com/sites/timworstell/2012/09/13/if-the-us-spends-550-billion-on-poverty-how-can-there-still-be-poverty-in-the-us/2/>).

Worthen, Blaine R., James R. Sanders, and Jody L. Fitzpatrick. 1997. *Program Evaluation: Alternative Approaches and Practical Guidelines*. White Plains, NY: Longman Publishers.

APPENDIX A
Nonprofit Staff Interview Guide

Past experience with evaluation

1. What experience has your organization had with evaluation in the past (other than PEAT)?
2. Have there been any challenges with conducting evaluation(s) in the past?
 - a. If so, what were some of the challenges?
 - b. Do you have any funding dedicated specifically to M&E or evaluation and assessment?
3. Do you perform any evaluation “in-house”? If so, can you describe how that is conducted?
4. What other evaluations do you have conducted by third-party evaluators besides PEAT?
5. What is the value of evaluation to your organization? (What is the role of evaluation?)
6. Can you give some examples of how evaluation feedback (from PEAT or elsewhere) helped your organization improve its services? Make structural changes? What effects did these changes have for your service population?

Goal/expectations for PEAT evaluations

1. How were you introduced/begin working with PEAT originally?
2. Why did you go with PEAT over other evaluation alternatives?
3. What value-added do you hope PEAT will provide for your organization? (What do you hope to get out of a partnership with PEAT?)
4. What were your expectations for your PEAT team?
 - a. Were those expectations met?

Impact or value-added of PEAT

5. What impact has PEAT had on your organization?
6. How helpful do you feel PEAT evaluations have been to your organization?
 - a. In what specific ways have they helped?
7. How could PEAT feedback have been more helpful? (What questions did it leave unanswered? Did you feel enabled to implement the recommendations? Were the recommendations feasible?)
8. What has your organization learned/changed through PEAT?
9. How do you plan on using PEAT’s feedback/research going forward?
10. What resources did you have to expend?

Experience working with PEAT interns

1. What were your experiences with PEAT?
 - a. Were there any difficulties or frustrations working with PEAT?
 - b. If so, were you able to address in a satisfactory way?
 - c. How can these issues be avoided in the future?
2. Please describe your experience in working with a student in particular?
 - a. Was it a positive experience overall?
 - b. What were some of the challenges? Highlights?
3. Would you have preferred working with someone other than a student? Why or why not?
4. What recommendations do you have for PEAT that might improve what the interns provide to you organization?

APPENDIX B
Student Interns Interview Guide

Plans

These questions refer to your evaluation plans, before you left for the evaluation site).

1. Please describe the purpose of the evaluation
2. What were the main elements of your evaluation design/plan?
 - What type(s) of evaluation did you plan to conduct for the organization? (Needs assessment, process, outcome, impact, etc.?)
 - a. What types of stakeholders did you plan to work with, and what types of involvement or role did you expect them to have in the evaluation?
 - b. What were the main evaluation questions that you sought to answer during the evaluation period?
 - c. Briefly describe the planned sampling technique?
 - d. Briefly describe the methods you planned to use to collect data?
 - e. What were your plans for analysis and timeline?
3. Please describe the expectations that you had before leaving (e.g. interaction with the organization, people, logistics, PEAT, the culture, evaluation work, etc.).

Field Data Collection

4. Once out in the field, how many adjustments did you make to your original evaluation plan?
 - No adjustments
 - Minor adjustments
 - Significant adjustments
 - Abandoned proposal entirely and created new evaluation plan in-country
5. In what ways did you adjust your **evaluation plan**?
 - a. Did the purpose change? Did you alter the evaluation questions? Did you find that you needed to conduct a different type of evaluation (e.g. due to the reality of the program stage)? Did you end up conducting more than one type?
 - b. Please explain the effect of these changes on the quality of the evaluation . Do you feel that these changes were better or worse than what you had planned? If so, why do you feel they were better or worse?
6. In what ways did you adjust your **sampling** techniques and **methods**? Why did you adjust them in this way?
 - a. Do you feel that these new techniques were better or worse than the planned sampling technique? If so, in what ways?
 - b. In what ways did you adjust your **methods**?
 - c. Please explain the effects of these changes. Would you describe them as better or worse than the methods that you had planned? Why?
7. How do you think they compare to the scientific standards taught in the social sciences? On a scale from 1 to 10, rate the quality of the data you collected. Explain your rating.
8. Please describe the evaluation methods that you actually used in the field?

Field Experience

9. Did your expectations meet reality? Why or why not?
10. What were some of your favorite moments?
11. As a student evaluator, what do you feel were your greatest strengths while in-country? (What did you particularly excel at?)
12. Please describe the most difficult challenges you faced during the entirety of your time in the field?
13. How was your interaction with the organization?
14. How was your interaction with the people?
15. How often did you receive professor feedback while in country? How useful was this feedback? How else could PEAT support you while in country?
16. If you could change anything about your time in the field, what would it be?

Analysis and Results

17. What did you do to analyze the data?
18. How do your plans for analysis compare to what you actually did to analyze the data?
19. What were some of the most difficult challenges that you faced when it came to analyzing the data and writing an evaluation report?
20. What were your primary findings?
21. What were the suggestions you gave to the organization?
22. In your opinion, what purpose did your evaluation serve for the organization?
23. Rate the usefulness of your evaluation findings and report to the organization. Do you feel that this evaluation was beneficial to the organization? Why or why not?
24. Based on your experience with the organization, do you feel that they will implement your recommendations? Explain.

Preparation

Having discussed your plans and what you actually did in the field, now let's discuss how prepared you feel you were to conduct this evaluation.

25. On a scale from 1 to 10 (1 being not at all prepared, 10 being perfectly prepared), how prepared do you feel you were to perform your evaluation? (1-10)
26. What did you find most useful in preparing for your PEAT internship?
27. What do you wish you had learned before you did your evaluation?
28. What can the PEAT student directors do to better prepare interns to perform quality evaluations?
29. What can the PEAT professors do to better prepare interns to perform quality evaluations? (open)
30. How useful was SOC 301 in preparing you to perform your evaluation? What parts were most useful? How could it be improved? (open)
31. How useful was SOC 390R in preparing you to perform your evaluation? What parts were most useful? How could it be improved? (open) (Note: You most likely already answered these questions in your 390R evaluation at the end of Winter semester. You need not repeat what you've already written – just give any new insight or ideas you've gained after completing your internship.)

Moving Forward

32. On a scale from 1 to 10 (1 being awful, 10 being fantastic), please rate your overall experience with PEAT. Explain your rating. (open response)
33. Would you recommend PEAT to your friends?
 - Yes
 - No
 - Not sure
34. Would you be interested in conducting another PEAT internship or something similar to it?
35. In what ways has your experience with PEAT helped you moving forward?
36. Please describe what you are currently doing? (work, student etc.)

APPENDIX C
Nonprofit Survey

Thank you for choosing to participate in our survey. We value your opinion and honest feedback. We want to assure you that your responses will be kept confidential. No one will have access to them outside of the direct research team. If you have any further concerns or questions, please email the principle investigator (Rilee Buttars: rileebuttars@gmail.com).

1. Had your organization engaged in evaluation *before* your evaluation with PEAT?
 - a. Yes
 - b. No, if no skip to question 5

If yes,

2. Who conducted the majority of those evaluations?
 - a. Internal staff
 - b. External evaluator
3. What type of evaluation(s) was conducted for your organization ? (Select all that apply)
 - a. An assessment that provided information about our service population (e.g. needs assessment, community profiles etc.).
 - b. An assessment of the program's goals, logic, model, and on-the-ground activities
 - c. An assessment of the outcomes of the program
 - d. An evaluation to identify the long-term impacts of our program
 - e. None of the above
4. Is there any information that you wish the evaluation(s) had produced for your organization that it did not? Please explain:
5. Please tell us the barriers (if any) you face to implementing evaluation as often or as well as your would like (select all that apply).
 - a. We can't afford it
 - b. We don't have time to devote to evaluation—our staff is overloaded as is
 - c. None of our staff have the expertise to conduct an evaluation
 - d. We don't know where we'd find a qualified evaluator anyway
 - e. Actually, I'm not exactly sure what the point is
 - f. I may see the importance, but no one else in our organization supports it
 - g. We can't find any granting agency willing to fund evaluation work.
 - h. None of the above

Now we will ask you a few questions about the PEAT evaluation conducted for your organization. If you have had more than one PEAT evaluation, pick the one you are most familiar with or that you feel was the most important.

6. Why did you agree to have the PEAT evaluators come work with you?
 - a. We normally can't afford evaluation, but recognize its importance. This was an affordable alternative.
 - b. Our staff/board members do not have the skills to conduct an evaluation

- c. It was free research
 - d. We wanted to provide a learning environment for students
 - e. Students are a normal part of our program implementation plans
 - f. Our service populations like working with students
 - g. Other (Please explain):
7. What was the purpose of the evaluation (select all that apply)
- a. To determine the impacts of a project
 - b. To provide information that could enhance project management and decision making
 - c. To provide lessons that inform future projects and feedback
 - d. To report on the day-to-day operation of a project, its strengths, its weaknesses, the obstacles it counters and suggestions for improvement
 - e. To meet funding stipulations
 - f. To meet accountability requirements
 - g. Other, please explain
8. Did the PEAT evaluation meet this purpose(s)?
- a. Yes
 - b. No

Evaluation influence

9. Did you learn what you anticipated or did you learn something new?
- a. We learned what we expected to learn
 - b. We were completely surprised with the evaluation findings
 - c. We learned new things and heard things we already new
 - d. We did not learn anything from this evaluation
10. We would like to know what you learned from the evaluation process/results **and/or** the attitudinal changes that took place. Please indicate if any of these scenarios match your experience (select all that apply):
- a. We merely thought more about our program and expected outcomes
 - b. We learned along the way and were able to adjust the program accordingly
 - c. Our attitude changed (positively or negatively) about the program and/or its assumptions
 - d. It brought to light a few issues and we thought about them more.
 - e. It brought to light certain issues and those issues became more important in our eyes.
 - f. We learned about our service population and the problems we intend to address
 - g. It helped us to narrow our focus
 - h. We actually learned enough to conduct an evaluation of our own next time.
 - i. The findings supported a belief or position we had before the evaluation
 - j. It helped move the minority opinion into the majority (minority opinion influence)
 - k. None of the above
 - l. Other (Please specify):

11. Now, we would like to know what direct or indirect actions took place after the evaluation. Please indicate if any of these scenarios match your experience (select all that apply):
- Because of the evaluation, significant changes took place (e.g. the program was continued, discontinued or changed)
 - We adjusted program details based on what we learned
 - The evaluation motivated someone to speak up and/or take action
 - The evaluation findings helped us to persuade other stakeholders in our favor
 - The evaluation helped get one of our initiatives onto an agenda (e.g. media, public or government agenda)
 - There were changes (minor or major) to policy due to the evaluation's work
 - The evaluation helped to alter thought or behavior at the policy level
 - We conducted more evaluations
 - Because we conducted evaluation, we were able to get more funding
 - None of the above:
 - Other (Please specify):
12. Were the evaluation results/findings used *directly* in decision making for your organization?
- Yes
 - No
13. How beneficial do you think the evaluation was for your organization?
- Highly beneficial
 - Moderately beneficial
 - Neutral
 - Not beneficial at all
 - Detrimental
14. How do you rate the overall quality of the evaluation?
- High quality
 - Perfect for my needs
 - Good enough
 - Not good enough for me to take them seriously
 - Poor quality
15. How satisfied were you with the information the student evaluators gave your organization?
- Very Satisfied
 - Satisfied
 - Neutral
 - Dissatisfied
 - Very dissatisfied

16. Is there any information that you wish the PEAT evaluation had produced for your organization that it did not? (Please explain):

17. Do you have any additional comments you want to make about your experience with a PEAT evaluation?

Lastly, we just want to ask a couple questions about you.

18. What is your position at the organization?

19. How long have you worked there?

20. What is the name of the organization you work for?

Note at the end of the electronic survey: Even though we ask you these last few questions, we will not link your name to any of the quotes we use. In the report, we will only give enough information to generally describe the organization and the position of the respondent. The organizations name will also not be used.

APPENDIX D Student Experience

This section describes the students' experience during preparation, in the field, and afterward. Part of each section will include their expectations, and their interactions with PEAT, the nonprofit and the people as well as their strengths and challenges.¹⁸ It also discusses the students' rate of satisfaction, the perceived benefits to the student, and their current status.

Preparation

In general, the students felt prepared in some areas and not in others. At one level, some expressed that they "had no idea what they were doing" (John), and felt that they received little training. On another level, some felt very prepared in their training from coursework, but not enough to work within the field, especially when trying to adapt what you learn in the states to a developing context. As one student described, "I didn't exactly know what I was doing. I had books on it, but it was on a 30,000 foot level" (John). This is often where the students' expectations didn't meet reality: preparing a proposal and evaluation plan that in many cases changed substantially once they began work in-country.

The students felt that the most useful aspects in preparing for the evaluation (excluding the program's two training courses) were the experience they brought from their majors, their additional methods courses, background research they conducted on the topic areas (e.g., microfinance), and the collaboration with other evaluators on their team.

During preparation, students had quite different experiences interacting with the organizations. Most students felt that the organizations' staff was kind, respectful and interested in working with them. If possible, most of the students met with them in person and in the best circumstances, communicated with them often. A theme of infrequent or miscommunication from the organization was present among the students who had to make major changes in their plans once in the field. For the team who wasn't able to finish their evaluation, they described their experience as, "building half a bridge" (JM). They had prepared, designed a rigorous evaluation, but then it all fell apart once in country. Those who were able to collaborate openly with PEAT's clients only had to make minor adjustments. The students were also frustrated because the organizations often didn't know what they wanted or couldn't make up their mind.

Overall, while interest was high in working with the students, understanding their role and purpose was not as clear.

I expected a whole lot more direction from the NGO. I expected to know what they wanted and to tell us what they wanted. But the reality, we got down there and [an in-country staff member] said, "Well you're here to do a project, go do it." And we were like, "Well no, we want this project to be useful to you. That's why we are here." There was kind of a disconnect. He just figured they're student they're just here to do a project. I'll let them go do it and come up with whatever they want to on their own (SM).

One student wished that the organization had been informed on what the students were trained in beforehand.

¹⁸ An evaluation report will be provided to PEAT upon completion of this paper. It will include a more concise description and the student and nonprofit's suggestions.

So it seemed like they were more ready to have a very clueless, let's be honest, I was pretty clueless, but not as clueless as I think they thought I was going to be. If I knew that they weren't going to know much about it, then I could have told them what's going on but because it wasn't what I thought it was. It took extra time to figure out what was going on. (Kevin)

301: Applied Social Research Methods. The students were pleased with most aspects of the class, but overall, it seems, they just wanted more. Many expressed that the class was high quality¹⁹, and that they loved the theoretical training and all hands-on activities, especially the experience working on the 21st Century After-school evaluation: observing, interviewing, coding, writing. They preferred the exercises that focused on practical skills, for example, working through the Social Impact Assessment manual by Rabel Burdge (1995). However, all responses by students show that they wanted it to be more specific with more activities, more explanation, more training, and more application.

I learned a lot in [301], but I don't think that at any point during the semester we had actually completed an evaluation. I think we had done various elements of one and we had talked about the different kinds, but I think for me the most valuable thing in the class would have been to see an evaluation through...even down to how you compile that information and what it looks like at the end and how you present it in the most meaningful way possible. (Ann)

390: Special Topics in Contemporary Sociology. The student's satisfaction varied for this class. In retrospect, the students were able to say that this class helped overall. They appreciated the workload and the high expectation to succeed from the professor.

I freaking sold my soul to write that proposal. I think I pulled more all-nighters in that semester than I've pulled in my life. And now looking back, I loved it; it was fun. I hated it at the time. But I recognize I produced a really high-quality product at the end, and I think that it was required for a student in my situation. I think I'm a fairly self-motivated person, but I didn't know where to channel that. You're taking people from all these different backgrounds, and so to have a professor who demanded high quality was very, very helpful. (Samuel)

However, the majority expressed feeling "lost" throughout its entirety, and during the experience, most stated that they felt discouraged and frustrated. They attributed this to a lack of clear expectations and guidance from the professor.

We needed more deadlines, more firm deadlines and maybe a little more structure and kind of a vision of what we were going to achieve from the beginning...students need deadlines and he's so smart but I feel like we could've gotten so much more out of him if he'd just be more direct about it. But at the same time, that's kind of [the professor's] style. He lets us learn and I felt like PEAT was very much our own process of figuring it out. And so I don't know; there's a balance between structure and student-directed research. I feel it was just a little too much on the side of the students figuring it out. So we were all kind of lost. We were like, "We don't know what we're supposed to be doing!" (Katie)

Yet, in the absence of direction, many rose to the occasion and "figured it out." For some, it worked better than others. One person wrote a high-quality proposal meeting the professor's demands, but had to change everything once in-country because it didn't match the

¹⁹ One NGO director even stated that he wished he could take this class after hearing what they learned from the student interns.

organization's needs. Another worked closely with the organization beforehand and nothing changed in application. This level of collaboration seemed to vary quite considerably from one evaluation to the next no matter the class structure or lack thereof.

Desired preparation. After the evaluation was completed, the students listed what they wish they had learned beforehand: how to write a good literature review that applied to the evaluation—they had a literature review but it wasn't as useful as it could have been; more methodological classes specific to evaluation (e.g., sampling); more concrete skills; how to adjust textbook training to field research; how an evaluation works from beginning to end (e.g., exposure to other evaluations and reports); how to interact with a nonprofit; how to analyze and synthesize data (e.g., code without a pre-made coding scheme; how to write an evaluation report; and how to make recommendations that are actionable.

The students also wished they had been able to make the training courses top priority (e.g., “I was burning a candle at two ends that semester” (John); coordinate better with the staff and had plans that reflected that information; and utilized the professors who had experience in their topic areas. In summary, one student expressed these desires:

I wish I would have been able to go through an entire research process and implement an evaluation beforehand—obviously that would have been ideal. If I could go do another [PEAT internship], it would be so much better. I think of things I would have changed like, how much data that we gathered that didn't end up being useful; learning how to design a really good survey; how to ask really good questions; how to focus in on what you actually really want to know without going off on a lot of extraneous things. I mean as much experience as can be provided, the entire research process, just the plan beforehand would be fantastic (SM).

Level of preparedness. When asked to rate their level of preparedness on a scale from 1-10, some indicated eight: “I felt really prepared.” The majority stated that they were “pretty well-prepared.” For example, “I would say seven because we eventually figured it out but it could have been a lot easier.” For those more in the middle, they indicated a five or six, explaining that they were “prepared enough to be flexible and recreate things from square one in the field.” Two students stated that they were “somewhat prepared,” and this rating was due to not feeling prepared enough to connect methods with questions or maintain social science standards in the field.

Field Experience

Before the students departed for their respective locations, many said they had high expectations. One, they had expected that “everything would work out” or that “some things would change but not everything.” For some, the evaluation did not go as planned and major adjustments were required. Some people had expected it to be a lot more “glamorous,” that the people [organization and service population] would be willing and ready to fully collaborate with them to make sure it went smoothly.

Honestly, I didn't know what to expect. I guess I thought it would be a little more glamorous—no, that's not the word. I thought the program would have been more hyped up about us being there, thought what we were doing was super important, and it would take precedence. But that's stupid. Everyone has their jobs; everyone is busy carrying on. (Stephanie)

They also expected, or at least hoped, it would be a lot easier. Some of them expected the organization's field office to be organized, more professional and structured. Some students had higher expectations for the organization's impact on the community. Some expectations were met; some were not. Yet, the students had a lot of good things to say about their experience in the field—this was the highlight of the program—but they also faced some challenges.

Favorite moments. Across the board, every student expressed that their favorite moments were when they were interacting with the people: interviewing them, talking in focus groups, and hearing about their lives.

Interviewing them about stigma was very interesting. A lot of times, they'd get very emotional and start crying. But a lot of them would thank him after he talked to them, because they don't talk about it. It's very painful to be asked about it, and to talk about it, but it was therapeutic to them. I really enjoyed that. And just getting to know them, being in the colony, getting to know the colony and the people there. (Henry)

I was just so stoked to be going to the focus group because like sitting there in the group and asking questions and digging in and finding out students responses, those were my favorite moments, just there sitting and talking to students by far (Stephanie).

They loved interacting in the different cultural settings, and for those who were able, living with them and experiencing their day-to-day lives. One student from the team who was not able to complete the evaluation explained:

I think the most enjoyable thing was just being with them: going to an Arab wedding; going and having food with them; going and eating with them, seeing how they interact with each other. That was definitely the most exciting part. That's what made it worth it. (Samuel)

Strengths. The good experiences were facilitated by some of the individual strengths these students exhibited. The first was an ability to make people feel comfortable. The student utilized different tactics to accomplish this, such as explaining they were just there to learn or taking the necessary time to get to know the population and build trust.

We spent like literally the first two and a half weeks just living with the people; milking the goats and playing baseball with the kids and soccer and we got into a lot of different things that...they just kind of took us on as their kids. (Katie)

One student team discovered early on that they each had different strengths and used them to create a relaxed environment:

We found that I would normally kick them off because I was better at getting people to warm up. I was good with sitting down with people, making them feel comfortable, and to open up. I think the majority of the students were comfortable. (Stephanie)

The program director confirmed this observation:

Their own approach and professionalism, they were both really great. And they were just very good. And I just felt like [their] awareness is very high of that they were doing something out of the ordinary that they might make people uncomfortable. I felt like they bent over backwards to create as much trust, and to treat people with respect and dignity as absolutely possible. I think they went above and beyond actually. (Carol)

Like this director mentioned, most of the students seemed to demonstrate a high degree of professionalism in the sense that they were serious about their work, thorough, and hardworking. As one student described, “I mean you reap what you sow. We were really serious about getting this done. We worked really hard” (Austin). According to one executive director:

The first one I worked with heavily was [PEAT intern]. [The intern] was just so, for one, he took it really seriously. He was willing to train our country directors. He’s always very prompt in getting back to me. Even though one of his CDs was crazy, he came back so invested with [the NGO] and so willing to help with things. (James)

Much of their hard work was fueled by their passion, energy, and excitement about learning and doing well. When asked about what strengths he brought to the table, one student responded:

Well, I was really passionate about it. I think that that carries more weight, than necessarily a lot of methodological rigor. Because I was willing to like do whatever needed to be done to make it work. I also think that I definitely did have the learning spirit. I really wanted to learn the skills, and like develop the skills as I went, so I tried to do things the right way as much as I possibly could. And, I think also just like being young and kind of energetic, it made it easy to connect with people and relate to people. Like I wasn’t there as a professional, it wasn’t my job, and I think that means something to people, especially when they find out that you’re here on your own dime, “Sure I’ll talk to you.” (David)

Similar to what this student described, the students felt one of their greatest strengths was being able to adjust when required and “learn on [their] feet,” as this student described:

Things just were never what you thought they were going to be and so you had to be very flexible and be willing to change your approach, change your um, your interview guides - change the questions that you’re asking because you have a translator that um, is maybe not understanding you, um, I phrase things too complexly so I had to change that and I had to change how I was going to get at the target population or um, so I think the skills that I ended up really utilizing were ones that um, I had developed from previous international experience. (Ann)

Along this same vein, students were proactive. When things didn’t go as planned, they made knew ones.

I don’t think I shut down whenever we had to make a adjustments. I was willing to make adjustments even up to the day we left. Like on things we needed to do. And I think I was proactive, in that as things didn’t work out, I found other things that I could do in order to like justify why I was there. So our evaluation didn’t happen; instead I did an evaluation on the NGO I was over there with. (Samuel)

Sometimes taking initiative required them to be straightforward and assertive. During preparation, one student explained that in January the organization staff members were still “wishy-washy” so at “that point, I stepped in and said this is where my skills and interests are and this is what I’m going to do” (David). Another student described that during research they had make sure all of their expectations and plans were known to the in-country staff, “So if they weren’t met they would know exactly what would happen. We easily could have gone there and come back without an evaluation if we just kind of went with the flow” (Austin).

In addition to these qualities, the nonprofit personnel considered the student, in general, to be friendly and viewed them favorably. According to one director, “so far the personalities that have come through PEAT have been perfect for our sites. They’ve been very even-tempered,

charismatic, everybody's friend type of presence" (James). He continued to say that the staff and other volunteers in the country were drawn to them because they were able to stay positive, which "is really important with some of the projects they were working on."

Challenges. Despite their strengths, the students also faced challenges while in the field. Some of these have already been discussed, such as methodological setbacks (e.g., sampling and interviewer bias, credibility, inconsistent data, or maintaining confidentiality). As one student described,

I think the most difficult challenge was the manipulation of the partners and some recipients. It led to discouragement. It's hard to go out and knowingly get lied to repeatedly and not feeling equipped to tease out the truth. (Ryan)

Time constraints were also an issue (e.g., not enough time, wasted time, what to do with down time). For one student wasted time was significant: "By the time we got there, there were no projects happening even four or five weeks into the [summer] program; there were no projects happening" (Ann). There were also logistical issues with unreliable rides, recruiting and asking for the participants for their time, and using an interpreter.

One big issue with quality was using a translator. They need to be really fluent in both languages and well educated. That was one huge limitation. In common conversation, he spoke English pretty well, but when he probed, it would be harder. Sometimes he wouldn't translate the whole meaning. (Henry)

Another challenge was building interest in the evaluation. Many in the service population responded to the students' explanation that they sincerely wanted to learn about them, but sometimes the interns had a hard time convincing the other volunteers working for the same organization:

There was a general distance to it—everyone thought it was useless and unnecessary—among all the [organization's] interns. It was a pretty lonely process. That was a lot to carry on my own shoulders, especially when I felt like I got more negative feedback. It was also a trial for me because I didn't have a lot of confidence to push the evaluation. Yes, I didn't have a lot of confidence to push it. That was a difficulty. (John)

For this and other reasons, the students who were not paired with another intern said it was a challenge, especially because they are new to this. They need more guidance and mentorship. They felt that having a partner would help balance this out.

Unfortunately, many of the students stated that they had less support and guidance from the organization and PEAT staff than they expected. They felt there was very little interaction or communication with the professors at home or staff at headquarters. For the PEAT faculty in particular, the students wanted more feedback and advice in the weekly emails. Some students stopped emailing altogether because they didn't know what to ask or if the director would respond well.

[PEAT director] was definitely there if we needed help and we could always go to him with questions. It's hard to get a straight answer out of him, and that's a whole other issue. He was definitely open if we had questions...I mean the difficulty is getting to the point where you have a valid question to ask. (Shannon)

Additionally, a few wished they had more interaction from the PEAT student directors as well to learn from their experience and other past interns.

Field research completion. Most of the students expressed struggling with this phase of the internship. As stated above, some students felt unequipped to manage and analyze their data. However, the majority felt that they could figure it out but didn't have the time. Officially, the internship is not complete until the students turn their report into the organization. However, for many of the students, it had "ended" once they returned to campus. As of now, PEAT does not offer a follow-up course. The students went full swing into another semester of coursework and requirements. They had pages of data and notes to be coded, analyzed, and compiled into a report but had little time and energy to devote to it. Additionally, the students felt that the expectations of the PEAT director for the report did not match the requests of the organization. The timeframe in which to complete it seemed unrealistic. These sentiments were expressed across the board. As this student stated:

There almost needs to be a writing report course at the end. It was insane. It was absolutely insane...I said, [professor] you need to lower your expectations for the report, and that's where the real problem is...his expectation for the report is a near publish-worthy paper, and that's not what PEAT is about. PEAT is all about organization x. Here's three recommendations for you and here's some major points to support it. It should be 4-5 pages long, 'cause they're not going to read a 30-page report, which is what I wrote. Everything else I turned in besides that wasn't good enough. And I was like, [professor] you are wasting my time and [the organizations]. [name omitted]

Satisfaction Rating

When asked to rate their experience, the majority felt it was very positive. The high responses included statements, such as "I loved the experience," "PEAT's great," "very valuable," and "it was a great experience." The three who gave it a lower rating attributed it to feeling like it was harder than expected; and while they appreciated how much it benefited them, they wish it had a larger effect on the organization. However, most students expressed gratitude for the experience. Satisfaction seemed to be high, much like this student expressed:

It was a miracle having that report in our hands to say, "Hey, we did this. We organized this. We were the force behind making this happen." It was really satisfying, not to mention just being in rural [country] for three months and experiencing the culture and the people. It was an incredible experience. (Austin)

Additionally, the majority of students said they would recommend PEAT to a friend, and the remaining stated this, but with a caveat: to the *right* friend.

'Cause PEAT is what it is and it's not an internship at Boeing. Anybody can benefit from having this brand name on their resume, but a lot of my friends are business students, and PEAT would do nothing for them. So it depends on what you are after. I think PEAT benefits students who are interested in international development, in academics. If a person is looking for an internship where they have a great learning experience, where they can learn some academic skills and research and hopefully impact poverty for the better, then I recommend PEAT. If they are looking for something else, then I'm like, "Well, PEAT's hard and you'll probably be pretty upset afterwards, so don't worry about it." (David)

Benefit to Student

The PEAT program is structured to not only help the organizations they evaluate, but also to benefit the students. From the data, it was clear the students had benefited in gaining hands-

on-skills, knowledge, experience, and in other notable ways. The students described areas in which their skillsets improved: interviewing (e.g., how to get people to open up, knowing the difference between a good and a decent interview); how to think about big questions and answer them; organize and connect thoughts; how to solve problems and think critically; how to adapt when things don't go your way; and to think innovatively. A student provided an example of this last skill:

I went to Washington DC and did an internship with Ashoka. And one of the specific things they look for in their interns is entrepreneurial drive, and I feel like PEAT is totally entrepreneurial. I mean you have to go out and do it all on your own. You have to create something from nothing and make it good. (Austin)

The students stated that they gained knowledge about evaluation, field research development, nonprofit management, and themselves. Primarily, many students expressed that it gave them direction.

I know for me, even though things didn't work out like we wanted, it gave me a very valuable experience because it helped me clarify my career goals and understand the value of what I was learning in sociology...before I had no idea what I wanted to do. I found a passion with international development by doing PEAT. I recognize the skills of research and development, which helped me realize I want to be involved...and I saw that it was possible. (Samuel)

In terms of providing them with experience, the students talked about interacting with other cultures, the "community aspect:"

I really liked the exposure to the people, and I think that's one of the most valuable things you get no matter where you are--the human connection...I liked being able to meet with people and connect with people. (Samuel)

Many also appreciated the exposure they had to developing contexts and organizations, and seeing how it all works together.

I want to work in international development in some capacity. I'm not sure what that means yet, but living with the people is huge and being able to have that direct experience with international development work and really see the insides of an NGO, which is totally invaluable in preparation in a future career in international development. Just getting familiar with what the work actually is and starting to grapple with some of those questions myself. (Shannon)

The students also felt the experience was unique, more realistic than other internship or study abroad opportunities, and that "You can't get this experience anywhere else." As one student described:

You're going to gain skills, and I think it's invaluable in the experience that I've had. It gives you experience on a lot of different levels. You are basically creating this thing out of nothing. It's a really unique experience that I don't think other internships give you. When I hear about other internships, you go get the coffee. This is not you. There is a lot of responsibility on you, which is really awesome and beneficial, and it's a little scary. (Ellen)

Combining the skills, knowledge and experience gained, the students explained that they now feel more confident in saying their background is in evaluation, that it's their niche in

development. They also felt more confident in their leadership skills and ability to succeed in graduate school where experience in research would be key:

The other benefits about PEAT for me were that it got me intrigued in conducting more international research. So, it made me want to do more graduate school. It's just made everything in development more complex and that was fun for me. (Tyler)

I wasn't a perfect evaluator, I mean I think I was sufficient in research areas there as well, but I honestly think that so much of my interest in research has come from that experience. (Sarah)

Other material benefits included utilizing the data for other classes and publishing papers from it, more research opportunities, a graduate fellowship, and getting job leads because of their experience. One student even expressed, "I attribute all my success to PEAT. Like I can't, honestly, it's snowballed from there—it rolled [everything] forward" (JH).

Current Status

The student's current statuses vary widely. Some are still finishing their undergraduate degree and are working on similar research projects. For those who have graduated, some are in graduate school (e.g., MS in Conflict Management, Statistics, Sociology, Social Work, and Law School). Others are taking time to gain more professional experience before doing more school (e.g., management consulting at McKinsey, Teach for America, or staff member for the Republican National Committee). Whether graduated or not, some continued working with the organization they evaluated and others decided to serve a mission for their church.

Table 1. Organization Participants: Pseudonym, Position, Years at Organization.

Respondent ID	Respondent position	Years at Organization
Michelle	Program Director	2
Adam	Executive Director	3
James	Executive Director (program director)	3
Paul	Project Manager	3
Christian	Board of Directors	5
Dawn	Board of Directors	6
Claire	Director of Regional Institutional Relations	6
Carol	Program Coordinator	10
Ashley	President and CEO	7
Heather	Director of Operations	7.5
Dennis	Director of Operations	6
Thomas	Program Manager	16
Ben	Director of Humanitarian Operations	16
Susan	Coordinator	1
Jolene	Board Chair	2
Ruth	Board of Directors	2

Table 2. Descriptive Information on PEAT Interns, 2009-2013.

Year	Total Students	Participant Pseudonyms	Total Majors	NGOs	Countries
2009	6		English, Latin American Studies, Sociology	3	Bolivia, Guatemala, Uganda
2010	8	John, Tyler, Ann, Ryan, Ellen	Advertising, Political Science, Public Relations, Psychology, Sociology	3	El Salvador, India, Mexico, Ukraine
2011	7	Sarah, Austin, Stephanie	Economics, English, Political Science, Portuguese Linguistics, Public Relations, Sociology,	4	Argentina, Fiji, Navajo Nation, AZ, Uganda
2012	11	Katie, Henry, Daniel, Samuel, David, Ellen, Shannon	International Relations, Organizational Behavior, Sociology	5	India, Jordan, Mexico, Thailand, Uganda
2013	9	Kevin	Business Administrative Management, Publics Health, Writing and Editing, Sociology	6	Cambodia, Malawi, Paraguay, Peru, Uganda

Table 3. Descriptive Information on PEAT's Clients, 2009-2013.

Org. ID	Years in operation	Countries	Context	Staff Headquarters+ In-country	Board Members	Missions/ Internships
1	25-30	Mali	Rural	6	15	Y
2	15	India, Uganda, Peru, Thailand, Belize, Fiji, Philippines	Mixed	6	7	N
3	32	Kenya, Nepal, Bolivia, Guatemala, Mexico	Rural	15	14	Y
4	6	Ukraine	Urban	16	-	Y
5	7	Uganda	Rural	4	-	Y
6	28	Paraguay	Mixed	11	12	Y
7	5	Uganda	Rural	7	?	Y
8	12	India	Mixed	7	35	Y
9	11	Bolivia, Columbia, Ecuador, Peru	Mixed	3	18	Y
10	29	179	Mixed	-	-	N
11	18	50	Mixed	-	-	N
12	19	Cambodia	Mixed	0	9	N
13	4	Uganda	Rural	4	7	X
14	9	Arizona	Rural	1	-	N
15	45	Jordan	Mixed	-	-	X

Table 4. Survey Response on Barriers PEAT Clients Face to Implementing Evaluation.

	Can't afford it	9	
	Don't have time to devote to it	11	
	Requires too much organizing; it's kind of annoying	2	
	None of the staff have the expertise	5	
Please tell us the barriers (if any) you face to implementing evaluation as often or as well as you would like.	We don't know where to find a qualified evaluator	1	
	Not sure what the point is	0	
	No one in organization supports it	0	
	Can't find any granting agency willing to fund it	4	
	None of the above	0	
	Do not face barriers	0	
	Other	4	
	<hr/> <hr/>		
		Total	15

Table 5. Survey Response: Evaluation Purpose and If Purpose Was Met.

		Did the evaluation meet this purpose?			
		Yes	Somewhat	No	Total
What was the purpose of the evaluation (select all that apply)?	Determine impact	7	7	0	14
	Enhance management and decision-making	6	5	0	11
	Provide lessons and feedback to inform future projects	6	4	0	10
	Report on day-to-day operations, strengths, weaknesses and obstacles	3	4	0	7
	Provide suggestions for improvement	8	5	0	13
	Meet funding stipulations	3	3	0	6
	Meet accountability requirements	2	0	0	2
	Other	0	1	0	1
	Total		8	7	0

Table 6. Survey Response: Evaluation Purpose, If Met, and If Results Were Used in Decision-making.

	Did the evaluation meet this purpose?				Were the results used in decision-making?		
	Yes	Some -what	No	Total	Yes	No	Total
Determine impact	7	7	0	14	11	3	14
Enhance management and decision-making	6	5	0	11	8	3	11
Provide feedback to inform future projects	6	4	0	10	8	2	10
What was the purpose of the evaluation							
Report on day-to-day operations: strengths, weaknesses and obstacles	3	4	0	7	4	3	7
Provide suggestions for improvement	8	5	0	13	10	3	13
Meet funding stipulations	3	3	0	6	5	1	6
Meet accountability requirements	2	0	0	2	2	0	2
Other	0	1	0	1	0	0	0
Total	8	7	0	15	12	3	15

Table 7. Type of Evaluation Conducted.

Type of Evaluation	Count
Needs Assessment	3
Process Evaluation	12
Outcome Evaluation	8
Impact Evaluation	4
Cost-benefit Analysis	1
Social Impact Assessment	1
Set-up Baseline	4

Table 8. Type of Methods Used for Evaluation.

Type of Method	Count
Survey/Quiz/Scorecard	11
Interview	16
Focus Group	6
Observation	23
Secondary Data-analysis (e.g., reports)	16

Table 9: Survey Responses: Type of Influence.

General Influence	
Learning/Enlightenment	5
Elaboration	7
Priming	10
Justification	5
Persuasion	6
Change Agent	4
Minority-Opinion Influence	2
Skill Acquisition	1
Cognitive/Affective	
Opinion/Attitude	4
Saliency	7
Behavioral	
Adjusted Program Details	10
Significant Changes	3
Alter Thought/Behavior at Policy Level	5
Policy Changes	7
Led to more evaluation	1

Figure 1: Mark and Henry (2004) Schematic Theory of Evaluation Influence.

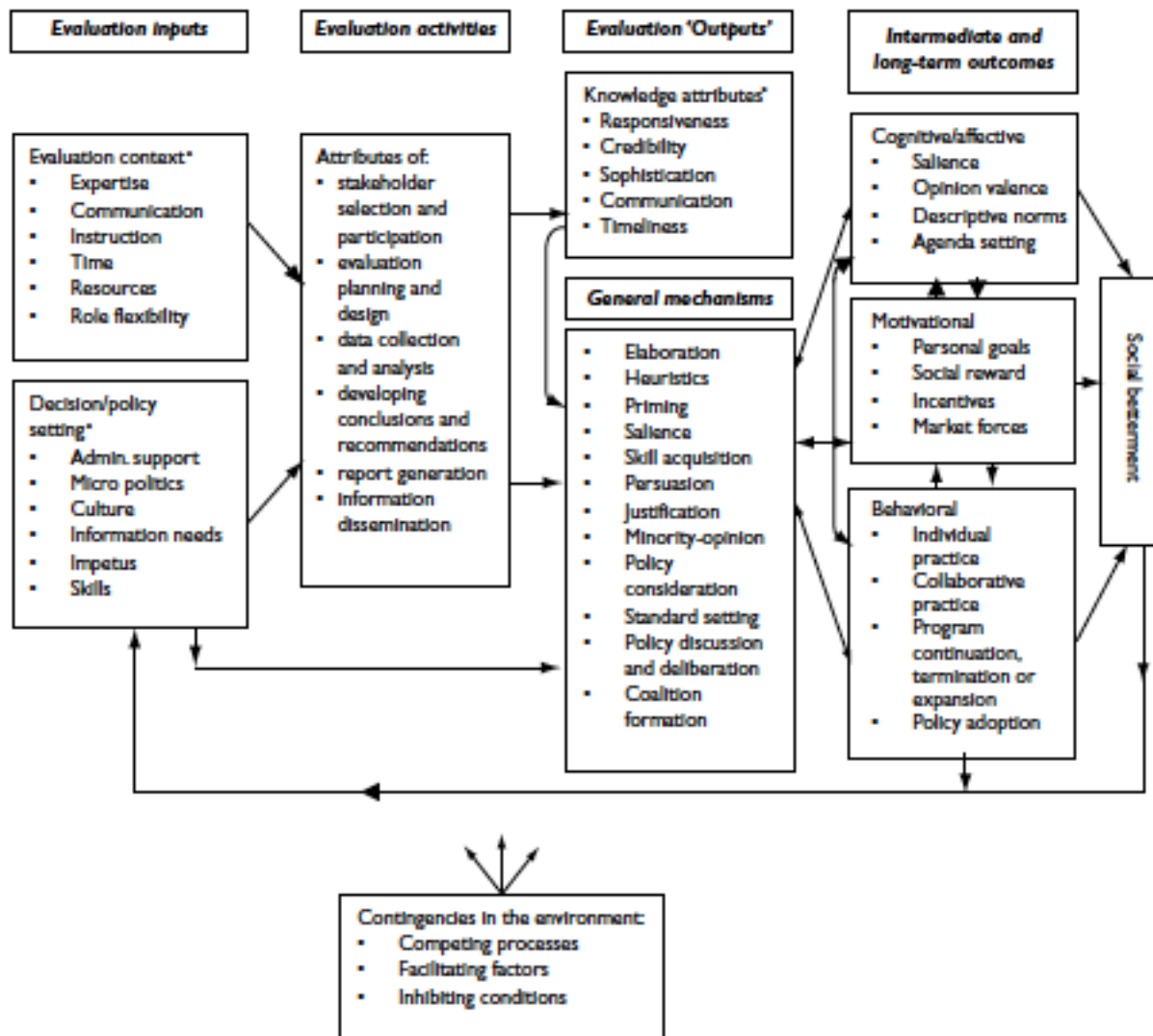


Figure 2: Appleton-Dyer et al. (2012) Influence Model.

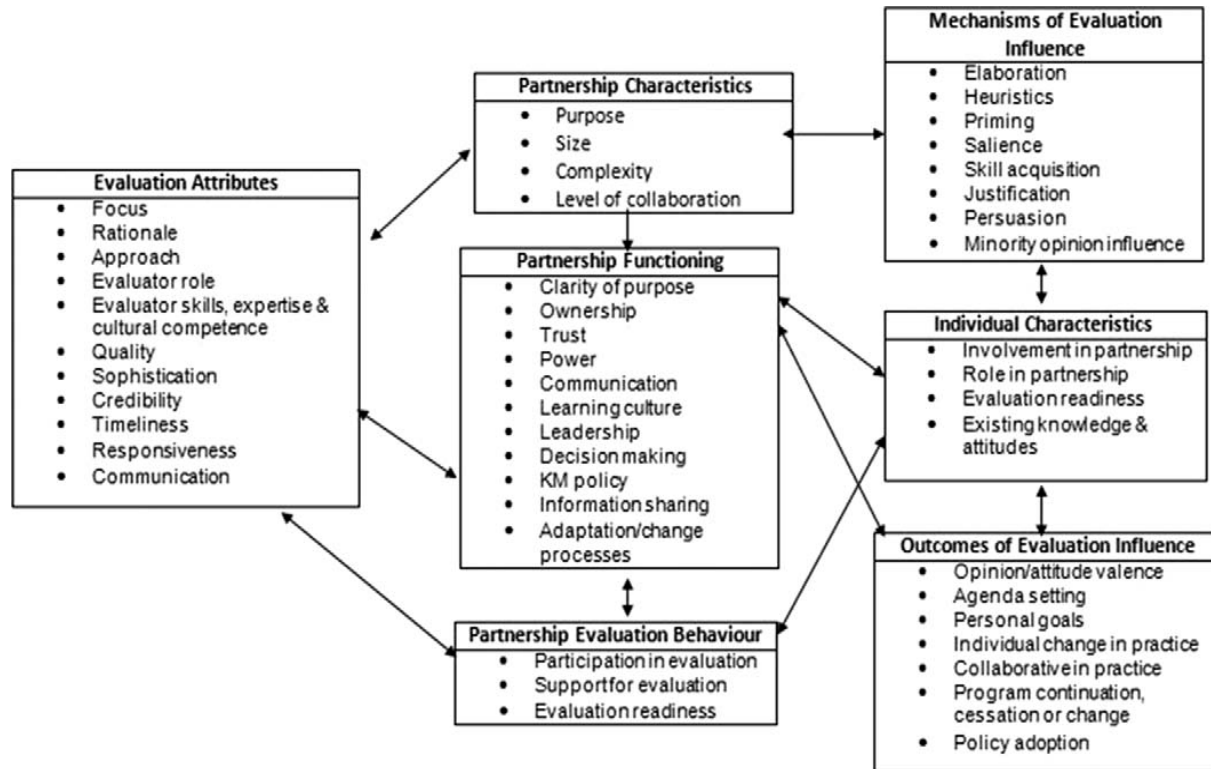
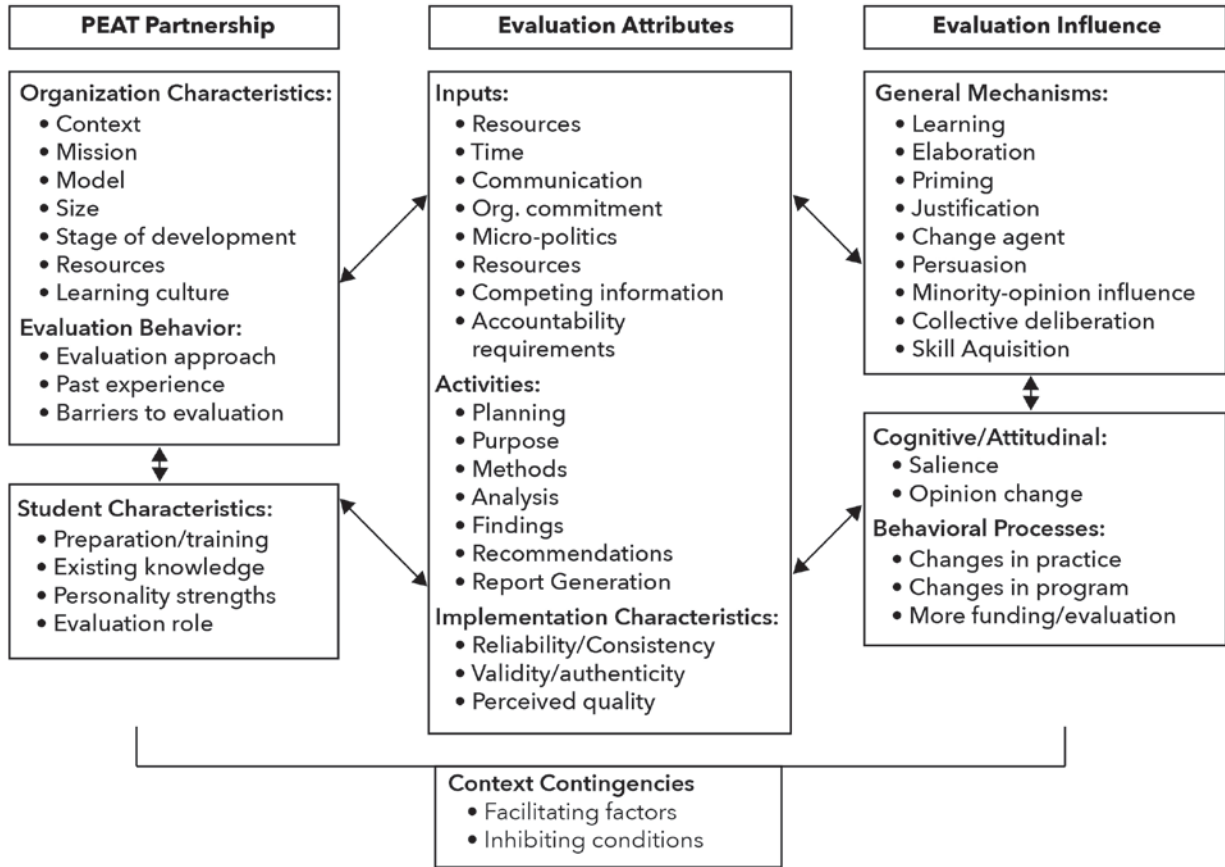


Figure 3: A Conceptual Model of the Student-Led Evaluation Change Process.



*Adapted from Mark and Henry (2004) and Appleton-Dyer et al. (2012)