

---

Wayne State University Dissertations

---

January 2020

## Advancing Objectives-Oriented Evaluation With Participatory Evaluation Methodology – A Mixed Methods Study

Dustin R. Saalman  
*Wayne State University*

Follow this and additional works at: [https://digitalcommons.wayne.edu/oa\\_dissertations](https://digitalcommons.wayne.edu/oa_dissertations)

 Part of the [Educational Assessment, Evaluation, and Research Commons](#)

---

### Recommended Citation

Saalman, Dustin R., "Advancing Objectives-Oriented Evaluation With Participatory Evaluation Methodology – A Mixed Methods Study" (2020). *Wayne State University Dissertations*. 2370.  
[https://digitalcommons.wayne.edu/oa\\_dissertations/2370](https://digitalcommons.wayne.edu/oa_dissertations/2370)

This Open Access Dissertation is brought to you for free and open access by DigitalCommons@WayneState. It has been accepted for inclusion in Wayne State University Dissertations by an authorized administrator of DigitalCommons@WayneState.

**ADVANCING OBJECTIVES-ORIENTED EVALUATION WITH  
PARTICIPATORY EVALUATION METHODOLOGY – A MIXED METHODS STUDY**

by

**DUSTIN SAALMAN**

**DISSERTATION**

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

**DOCTOR OF PHILOSOPHY**

2020

MAJOR: EDUCATION EVALUATION

AND RESEARCH (Quantitative)

Approved By

\_\_\_\_\_  
Advisor

Date

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

## **DEDICATION**

I dedicate this manuscript to my family and friends. During this journey you have provided unwavering support.

To my husband Jared, who has been waiting patiently for me to complete this degree. Who is looking forward to the return of space in our lives for quality time and fun.

To my parents, Kevin and Lorraine, for providing me with the work ethic and tenacity necessary to successfully get through this process.

To my sister and her family, especially my nephew Zander and niece Amryn. I look forward to having the time to be a more active part of your lives.

To all of my friends, this manuscript is evidence of where I have been over the last two years when I have declined invitations and neglected to make time for enjoying your company.

Finally, to Dr. Langen, who has always served as a mentor. Since my return to Baker College she has provided a constant source of pressure to complete my dissertation, as only a good mentor would.

## **ACKNOWLEDGMENTS**

There are several individuals who deserve to be acknowledged. This has been a journey through the proverbial wild, and these individuals served as essential guideposts along the way. First, my major advisor, Dr. Sawilowsky, for his responsiveness and guidance during the development of this manuscript. Also, for giving me the opportunity to add non-parametric statistics to my repertoire through his teaching.

Second, Dr. Tracey, who was my major advisor when I began my Ph. D. in Learning Design and Technology and then encouraged me to consider the Educational Evaluation and Research program. Sometimes the role of an advisor is to help a student find spaces that will help challenge themselves in new and exciting ways.

Third, Dr. Pogodzinski, for his willingness to serve on my committee and his continuous support and guidance throughout my program.

Fourth, Dr. Kromrei, for her willingness to serve on my committee and support along the way. Early in the process, the opportunities to meet for coffee and go over my early writing and discuss program evaluation were essential to completing this manuscript.

Finally, Dr. Ulmer, who exposed me to qualitative research methods. This was an eye-opening experience and gave me the knowledge necessary to be able to pursue my interest in mixed-methods research.

## TABLE OF CONTENTS

DEDICATION .....	ii
ACKNOWLEDGMENTS .....	iii
CHAPTER 1 INTRODUCTION .....	1
Evaluation of Educational Programming.....	1
Evaluation Models .....	2
Institutionalization of Evaluation and Evaluative Thinking .....	4
Statement of the Problem .....	5
Promoting Evaluative Thinking and Training Opportunities for Stakeholders.....	6
Research Questions .....	6
Assumptions .....	7
Limitations.....	7
Definition Terms.....	8
CHAPTER 2 LITERATURE REVIEW .....	10
Evaluation Theory and Paradigms.....	11
Dialectical Pluralism.....	17
Merit and Worth .....	20
Credible Evidence.....	20
Participatory Evaluation .....	34
Defining and Measuring Participation .....	39
Stakeholder.....	44
Professional Development.....	50
Evaluation of Professional Development .....	56
Evaluative Thinking.....	60
CHAPTER 3 METHODS .....	66

Design.....	66
Setting Description.....	68
Mixed-Methods Design .....	69
<i>Phase One: Quantitative</i> .....	69
Sampling Plan .....	69
Instrument.....	69
Data Collection .....	71
Data Analysis.....	72
Participatory Evaluation Measurement Instrument (PEMI).....	72
Evaluation Involvement Scale (EIS) .....	72
Evaluative Thinking Questionnaire .....	73
<i>Phase II: Qualitative</i> .....	73
Transcript Analysis .....	73
Data Collection .....	74
Chapter 4 Results .....	79
Phase I: Quantitative.....	79
<i>Evaluation Involvement Scale</i> .....	81
<i>Evaluative Thinking Survey</i> .....	85
Phase II: Qualitative .....	89
<i>Control</i> .....	90
<i>Participant Involvement</i> .....	91
<i>Team Diversity</i> .....	91
<i>Prior Experience</i> .....	92
CHAPTER 5 DISCUSSION AND CONCLUSION .....	98
Discussion .....	100
Conclusions .....	105
APPENDIX .....	109
Appendix A: Evaluation Framework Development .....	109
Appendix B: Evaluative Thinking Survey .....	114
Appendix C: Semi-Structured Interview Protocol.....	117
Appendix D: Participatory Evaluation Measurement Instrument (PEMI) Items ..	118

Appendix E: Evaluation Involvement Scale Items .....	119
Appendix F: Qualitative Code Book.....	120
REFERENCES .....	122
ABSTRACT .....	134
AUTOBIOGRAPHICAL STATEMENT .....	136

## LIST OF TABLES

Table 1: Philosophical Assumptions Associated with Major Paradigms.....	15
Table 2: Paradigms and Branch Alignment .....	17
Table 3: Philosophical and Methodological Assumptions of Dialectical Pluralism .....	18
Table 4: Thirteen Evaluation Questions .....	23
Table 5: Example of DEM Evaluation Framework.....	32
Table 6: Coding Scheme for Extent of Involvement .....	41
Table 7: Coding Scheme for Diversity of Participants .....	42
Table 8: Coding Scheme for Control of the Evaluation Process.....	42
Table 9: Evaluation Involvement Scale Items.....	43
Table 10: Groups of Impactees .....	46
Table 11: Evaluation Audience Checklist .....	48
Table 12: Typologies of Nonevaluative Stakeholders .....	50
Table 13: Indicators of Evaluation Thinking .....	63
Table 14: Traditional and Nontraditional Stakeholder Types.....	69
Table 15: General Inductive Analysis Approach .....	75
Table 16: Tools to Address Validity and Reliability in Naturalistic Inquiry .....	76
Table 17: Summary of Research Questions, Variables, Collection, and Analysis .....	77
Table 18: PEMI Descriptive Statistics and Mean Ranks .....	80
Table 19: Kruskal-Wallis Test Statistics .....	80
Table 20: EIS Descriptive Statistics and Levels of Significance.....	82
Table 21: EIS Items and Sum of No or Yes, a little Ratings .....	83
Table 22: Coding Scheme Applied to EIS Results .....	85

Table 23: Evaluative Thinking Survey Descriptive Statistics and Mean Ranks .....87

Table 24: Friedman’s Test Statistics .....87

Table 25: Traditional and Non-Traditional Evaluation Team Members .....95

## LIST OF FIGURES

Figure 1: Evaluation Theory Tree .....	14
Figure 2: Dimensions of Form in Collaborative Inquiry .....	40
Figure 3: A Model of Teacher Change .....	51
Figure 4: Explanatory Sequential Design .....	67
Figure 5: Major Themes and Subthemes Represented in Qualitative Data .....	90

## **CHAPTER 1 INTRODUCTION**

### **Evaluation of Educational Programming**

Fitzpatrick, Sanders, and Worthen (2011) opined the purpose of program evaluation is “the identification, clarification, and application of defensible criteria to determine an evaluation object's value (worth or merit) in relationship to those criteria” (pg. 13). In the context of educational programming there exists a wide variety of forms and contexts, such as traditional classroom instruction, tutoring programs, test preparation programs, and competency-based self-study programs. Components of the program that represent subsets require their own resources, objectives, and programming. For example, a traditional course delivered in a hybrid face-to-face and online method will be interdependent in their design but require separate planning and resources to facilitate the learning in both environments. Education programs can be classified in a variety of ways such as primary, supplementary, and ancillary. This illustrates a high degree of complexity in the design and organization of educational programming when trying to define what is meant by educational programming and what the variables could potentially represent the evaluand.

The outcomes of these evaluations are important to stakeholders such as educational administrators, boards, and policy makers who are accountable for funding and organizational effectiveness. Program directors and others involved in the implementation of a program need information to help them improve the program (Fitzpatrick, Sanders, & Worthen, 2011). The outcomes are important to program recipients so they might understand the impact of their participation on their lives and the lives of the other people around them. When there are multiple stakeholder groups

interested in the outcomes of the evaluation, how defensible criteria are defined and by whom is important to consider and also inherently political given the power dynamics which exist.

Typically, when completing program evaluations, criteria are determined by high level stakeholders, such as program directors and administrators who have knowledge of the program as well as the resources required to operate it and are ultimately responsible for those resources (Fitzpatrick, Sanders, & Worthen, 2011). However, limiting the decision making in an evaluation to these categories of stakeholders comes with limitations such as too narrowly defining the success of the program, failing to capture secondary and tertiary effects, and failure to capture significant details related to the context of the program, all of which can limit the use of the evaluation results (Fitzpatrick, Sanders, & Worthen, 2011; Stufflebeam, 2001; Sorinola & Thistlethwaite, 2013).

### **Evaluation Models**

Objectives oriented evaluation approaches are the most prevalent when evaluating programs due to a high degree of simplicity and clarity when reporting findings to stakeholders. These forms of evaluation are applicable when there are focused programs with supportable objectives (Stufflebeam, 2001). This describes most educational programs as goals and objectives are a fundamental building block of most program design (Richlin, 2006) and often a required element when applying for funding sources (Karsh & Fox, 2014). In objectives-oriented evaluations the evaluation criteria are defined by the program objectives (Fitzpatrick, Sanders, & Worthen, 2011; Stufflebeam, 2001). Due to limitations in time and resources, these sorts of evaluations

have limited stakeholder involvement and operate within the limited scope of the program objectives (Fitzpatrick, Sanders, & Worthen, 2011).

Different evaluation models were developed to provide opportunity to define broader stakeholder groups and allow for more expanded opportunities for stakeholder involvement in the actual evaluation. These participant-oriented evaluation models were developed in some cases to create opportunities to better capture the context of the program and better promote use of evaluation results, such as practical-participant evaluation and stakeholder evaluations (Stufflebeam, & Coryn, 2014). Whereas others were developed to promote social justice values, such as transformative-participatory evaluation and empowerment evaluation (Stufflebeam, & Coryn, 2014). Both have the value of promoting plurality of perspectives in the development and implementation of the evaluation.

However, Stufflebeam (2001) described these approaches (referred to as constructivist evaluation) as being “heavily philosophical, service oriented, and paradigm driven” (pg. 71). They are opposed to the positivist, deterministic, and reductionist structure and are an attempt to make sense of a variety of constructions emerging from the variety of stakeholders, avoiding the idea one of the constructions is “true” (Stufflebeam, 2001, p. 73). They also present barriers when stakeholders at higher levels within the organization are uncomfortable with the processes and outcomes of these sorts of evaluation models (Stufflebeam, 2001). Daigneault and Jacob (2009) described other shortcomings of these approaches as lacking satisfactory conceptualization and operationalization.

## **Institutionalization of Evaluation and Evaluative Thinking**

The beginning of most program evaluation textbooks is a description of the importance of program evaluation and the need for all programs to be evaluated for the purposes of accountability and to demonstrate the merit and worth of the program (Fitzpatrick, Sanders, & Worthen, 2011; Stufflebeam & Shinkfield, 2007; Stufflebeam & Coryn, 2014). However, whether a program is evaluated is often dependent on the availability of limited resources due to evaluators typically being external to the organization. As mentioned above, educational programming is a broad category of activities, which take place in and around the educational environment, and not all programs come with a funding source. Fitzpatrick, Sanders, and Worthen (2011) posited evaluation as a field and as a practice is predicted to increase, because internal evaluation will become more important due to its benefits. This does not inherently provide program evaluation resources for all programs depending on how an internal evaluator is situated within the organization. Stufflebeam and Coryn (2014) provided a checklist to provide guidance for organizations to institute evaluation systems; however, depending on the level of expertise and the resources available to the user(s) this may be of little assistance.

If educational programs must be evaluated to demonstrate their merit and worth and to improve the program, then appropriate personnel in those organizations must find ways to develop capacity to perform more evaluations. If evaluation is to be institutionalized, then strategies need to be identified to promote evaluative thinking amongst those within the organization.

## **Statement of the Problem**

An evaluation method is needed which is eclectic in both its design and methodology. The evaluation findings need to serve the needs of a diverse set of stakeholders; providing evidence to program directors, administrators, board members, and policy makers as to whether the program is successful, provides value to the organization, and provides necessary information to those responsible for implementing and improving the program. The evaluation needs to capture the level of value offered by a program in the context in which it operates and takes into account the day-to-day reality of the program. The value of the program needs to be judged based on different types of data to be able to triangulate findings as an effort to increase the validity and accuracy of the claims made about the program's value.

An evaluation design which borrows the use of program objectives from the objective-oriented evaluation methods as an organizational framework to guide the development of the evaluation framework while using a participatory-oriented model has the potential to draw on the benefits of both models and increase the richness and utility of the evaluation findings compared to using only an objectives-oriented approach. It also has the added opportunity to provide the benefits of promoting knowledge of evaluation and evaluative thinking in program development to a wider audience within the organization, as is often a desired benefit of utilizing participatory-oriented evaluation.

## **Promoting Evaluative Thinking and Training Opportunities for Stakeholders**

Although there are participatory-oriented evaluation models designed to promote evaluative thinking, the models themselves are resource intense and both conceptually and operationally ill-defined (Stufflebeam, 2001; Daigneault & Jacob, 2009). This limits the transferability of evaluative thinking to non-evaluative stakeholders acting independent of a trained evaluator. However, objectives-oriented evaluation models are recognized as being straightforward, less resource intense to implement, and appropriate for focused programs, which could represent many of the programs included in the broad category of educational programs. Philosophically participatory evaluations have the benefit of training non-evaluative stakeholders in evaluation through their inclusion in the evaluation activities; however, the evaluation model needs to be one that allows the stakeholder to implement independently following the end of their experience with the trained evaluator.

### **Research Questions**

Therefore, the questions addressed in this study are:

1. How does the participation of a diverse set of stakeholders influence the development of the evaluation framework?
2. When limiting the scope of a participatory evaluation using an objectives-oriented evaluation model as an advanced organizer, does the evaluation meet the criteria provided by Daigneault and Jacob (2009) using the Participatory Evaluation Measurement Instrument (PEMI) and Evaluation Involvement Scale (EIS) provided by Toal (2009)?
3. Does the use of a participatory evaluation model promote evaluative thinking with

the non-evaluative stakeholders involved in the evaluation?

### **Assumptions**

As a desired outcome of engaging in a participatory evaluation model is to promote evaluative thinking within the organization, it is assumed that non-evaluative staff have no prior training or experience in program evaluation.

When examining the unique contributions of the evaluation team members in the evaluation design, the members of the team will be categorized as traditionally included in evaluation design and not traditionally included in evaluation design. Examples of traditionally included would be program leadership and staff responsible for implementation of the program. Members not traditionally included would be program consumers and downstream Impactees. It is assumed that had the team member not been present the contribution would not have been made.

### **Limitations**

The researcher has experience implementing ad hoc pseudo-evaluations (Stufflebeam, 2001), or participatory evaluation only as a novice.

Given professional development represents a broad concept, the review of literature on this topic includes literature on professional development in both K-12 and post-secondary contexts and represents a variety of contexts including discipline specific professional development, professional development provided by centers, and faculty learning communities.

The generalizability of this study may be limited due to the context in which it took place, and the limited number of participants represented in the data set.

## Definition Terms

**Dialectical:** According to Johnson (2017) and Fang (2016) dialectical refers to a method of dialog and reasoning that tolerates contradiction and accepts the co-occurrence of opposing elements.

**Evaluation:** Fitzpatrick, Sanders, and Worthen (2011) defined evaluation as “the identification, clarification, and application of defensible criteria to determine an evaluation object's value (worth or merit) in relationship to those criteria (p. 13).

**Evaluand:** Scriven (1991) described an evaluand as a generic term for whatever is being evaluated; in the context of program evaluation the evaluand would be the program itself.

**Evaluative Thinking:** Buckley, Archibald, Hargraves, and Trochim (2015) defined evaluative thinking as “critical thinking applied in the context of evaluation, motivated by an attitude of inquisitiveness and a belief in the value of evidence, that involves identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection and perspective taking, and informing decisions in preparation for action” (p. 378).

**Nonevaluative Stakeholders:** Nonevaluative stakeholders are stakeholders included in an evaluation but who have no formal training as evaluators (Daigneault & Jacob, 2009).

**Objective Evaluation:** Objectives oriented evaluation methods are evaluations in which the purposes for activities are specified and the evaluation focuses on the extent to which those purposes, or objectives, are achieved (Fitzpatrick, Sanders, and Worthen, 2011,).

Participatory-Oriented Evaluation: Evaluation methods that use stakeholders- people with an interest or stake in the program- to assist in conducting the evaluation.

Professional Development: Guskey (2000) described professional development as activities designed to enhance the professional knowledge and skills of educators, which in turn might improve the learning outcomes of students.

## CHAPTER 2 LITERATURE REVIEW

Program evaluation formalized as a field of research and investigation in the United States and Canada in the 1960s as the United States government began investing in large educational programs that required evaluations as part of government accountability. There are accounts that program evaluation was being practiced as far back as the 1800s (Scriven, 1991; Madaus, Scriven, & Stufflebeam, 1983). As program evaluation grew and became a more common practice in education, healthcare, and other non-profit organizations the definition of program evaluation has been redefined. Among professional evaluators there is no agreed-on definition of program evaluation. Some definitions have emerged from various theorists in the field as well as professional organizations focused on program evaluation, such as Scriven (1991) and Stufflebeam (2001). Stufflebeam and Coryn (2014) defined evaluation as “the systematic assessment of an object’s merit, worth, probity, feasibility, safety, significance, and/or equity” (p. 7), which is an extension of the Joint Committee’s 1994 definition of evaluation. Fitzpatrick, Sanders, and Worthen (2011) defined evaluation as, “The identification, clarification, and application of defensible criteria to determine an evaluation object’s value (worth or merit) in relation to those criteria” (p. 7). The common element which emerges is determining merit, worth, and accountability.

The practice of program evaluation, like the definition, is broad and varied. There are twenty-three different approaches to program evaluation (Stufflebeam, 2001; Stufflebeam & Coryn, 2014) which were categorized by Stufflebeam and Coryn (2014) as pseudo-evaluation approaches, improvement and accountability-oriented

approaches, social agenda and advocacy approaches, and eclectic approaches. Although the approaches to program evaluation are varied in primary goals, advanced organizers, sources of the questions addressed, and methods typically employed, there are general functions and steps in evaluation design which are common regardless of the approach, impetus, or context for the evaluation (Stufflebeam & Coryn, 2014). Stufflebeam and Shinkfield (2007) proposed four functions common to conducting any evaluation: 1) information collection, 2) organization, 3) analysis, and 4) reporting.

Although the stated functions are common in the design of the evaluation there are a multitude of ways that the activities in the evaluation can be carried out depending on the specific evaluation methodology. This is determined in large part by the political context in which the evaluation is being carried out and the philosophical assumptions and theories on which the evaluation is based (Mertens, 2018).

### **Evaluation Theory and Paradigms**

In a presidential address to the American Evaluation Association (AEA), Shadish (1998) stated evaluators needed to understand evaluation theory because it is “who we are.” Shadish then provided six statements about why evaluation theory was so important to the field and those practicing evaluation:

- 1) Evaluation theory provides the language evaluators use to talk about themselves and about evaluations;
- 2) Evaluation theory encompasses many of the things in the field evaluators care most deeply about;
- 3) Evaluation theory defined the themes of the major AEA conferences;

- 4) As evaluation is a trans-discipline, evaluation theory provides an identity which is different from other professionals;
- 5) Evaluation theory provides the face presented by evaluators to the outside world; and
- 6) Evaluation theory is the knowledge base that defines the profession.

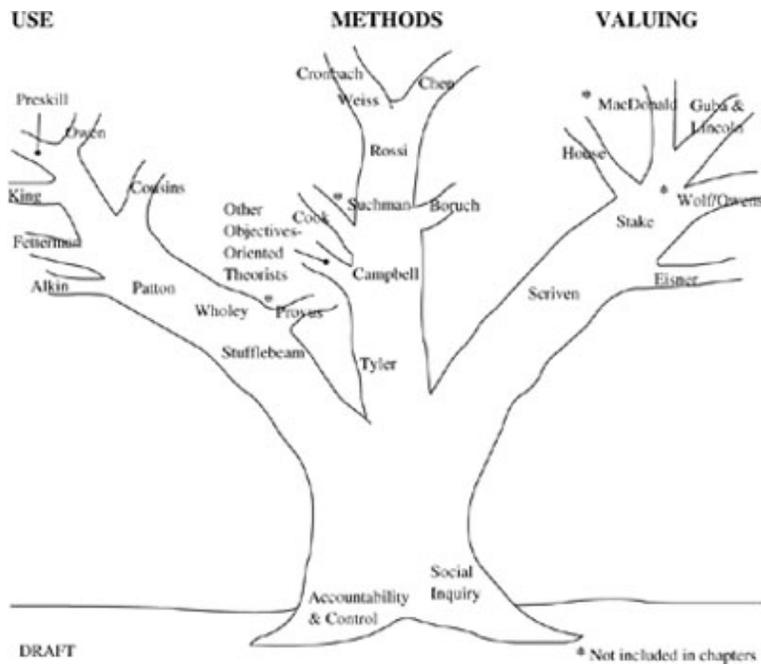
Although theory is important, understanding the major paradigms that the theories are situated in is equally as important. It informs characteristics frequently associated with the particular evaluation theory, such as methods, stakeholders, and overarching purpose of the evaluation.

Alkin and Christie (2004) presented the evaluation theory tree; a graphic representation of evaluation which represented the roots and trunk as accountability and social inquiry to demonstrate the purpose and discipline from which evaluation emanates as shown in Figure 1. The three branches of the tree represent the primary foci of evaluation: methods, use, and valuing. On each branch are represented the major evaluation theorists whose work has contributed to the growth of the branch. The focus of the methods branch of evaluation is generalizability or knowledge construction. The focus of the Valuing branch is placing value on data as the most essential component of the evaluator's role. The focus of the Use branch is the ways in which evaluation information will be used and who will use the information. Mertens and Wilson (2012) added a fourth branch to the tree referred to as Social Justice. The branch "represents the voices of marginalized groups in society and their advocates, the need to explicitly address issues of power, and the design of evaluations to support social transformation in the interest of supporting human rights" (Mertens, 2018, p. 16).

The evaluation theories represented on the Evaluation Theory Tree connect to philosophical paradigms which are composed of four assumptions representing the worldview of the evaluator and the evaluation theorist (Guba, Lincoln, Lynham 2011).

These frameworks are composed of the following four elements:

- Axiology – the nature of ethics and values
- Ontology – the nature of reality
- Epistemology – the nature of knowledge and the relationship between the evaluator and stakeholders
- Methodology – the nature of systematic inquiry

**Figure 1***Evaluation Theory Tree*

Note. From *Evaluation Roots*, by M. Alkin and C. Christie, 2004, p.

13 (<https://doi.org/10.4135/9781412984157>).

In the evaluation world there are four philosophical paradigms represented: positivism/post positivism, constructivism, pragmatism, and transformative (Guba, Lincoln, & Lynham, 2011; Mertens & Wilson, 2012). Detailed in Table 1 are the assumptions for each of these paradigms.

**Table 1***Philosophical Assumptions Associated with Major Paradigms*

Basic Beliefs	Postpositivism	Constructivism	Pragmatic	Transformative
Axiology (nature of ethical behavior)	Respect privacy; informed consent; minimize harm (beneficence); justice/equal opportunity	Balanced representation of views; raise participants' awareness; community rapport	Gain knowledge in pursuit of desired ends as influenced by the researcher's values and politics	Respect for cultural norms; beneficence is defined in terms of the promotion of human rights and increase in social justice; reciprocity
Ontology (nature of reality)	One reality, knowable within a specified level of probability	Multiple, socially constructed realities	Asserts there is a single reality and all individuals have their own unique interpretation of reality	Rejects cultural relativism; recognizes various versions of reality are based on social positioning conscious recognition of consequences of privileging versions of reality
Epistemology (nature of knowledge; relation between knower and would-be known)	Objectivity is important; the researcher manipulates and observes in a dispassionate, objective manner	Interactive link between researcher and participants; values are made explicit; created findings	Relationships in research are determined by the researcher deems as appropriate to the particular study	Interactive link between researcher and participants; knowledge is socially and historically situated; need to address issues of power and trust
Methodology (approach to systematic inquiry)	Quantitative (primarily); interventionist; decontextualized	Qualitative (primarily); hermeneutical; dialectical;	Match methods to specific questions	Qualitative (dialogic), but quantitative and mixed

contextual factors are described	and purposes of research; mixed methods can be used as researcher works back and forth between various approaches	methods can be used; contextual and historical factors are described especially as they relate to oppression
--	--	--

---

*Note.* From *Mixed Methods Design in Evaluation*, D. M. Mertens, 2018, p. 15. Copyright 2018 by Sage Publications.

Mertens (2018) connected the four paradigms represented in the evaluation world to the branches of the evaluation tree, shown in Table 2. Building this bridge between the four paradigms and the branches of the evaluation theory tree demonstrated which paradigm the evaluation theories were rooted in. Knowledge of which paradigms particular evaluation models and theorists were grounded in informs how the evaluator will approach various functions of the evaluation, such as how to define the stakeholders, frame the evaluation questions, frame criteria for evaluating the credibility of evidence, and identifying the methods for collecting and analyzing evidence. It also informs the broader philosophical goals of the evaluation (Donaldson & Lavelle, 2018).

**Table 2***Paradigm and Branch Alignment*

<b>Paradigms</b>	<b>Branches</b>
Positivism/Post positivism	Methods
Constructivist	Values
Pragmatism	Use
Transformative	Social Justice

*Note.* From *Mixed Methods Design in Evaluation*, D. M.

Mertens, 2018, p. 16. Copyright 2018 by Sage Publications.

In addition to the four paradigms in evaluation, evaluators can work from a stance of dialectical pluralism (DP) (Johnson & Schoonenboom, 2016). According to Johnson and Schoonenboom (2016) DP is a metaparadigm; a process philosophy and theory, which uses both/and logic to produce new creative syntheses.

### **Dialectical Pluralism**

Johnson and Stefurak (2013) introduced a metaparadigm referred to as dialectical pluralism (DP) as a process philosophy for engaging with multiple paradigms rather than remaining situated in a monism. This meta-paradigm grew out of the methodological and philosophical movement referred to as “mixed methods research” (MMR), which attempts to mix methods, data, methodologies, disciplines, values, modes of inquiry, paradigms, perspectives, philosophies, and levels of analysis (Johnson, 2017). Table 3 outlines the philosophical and methodological assumptions that underlie DP and the use of the both/and logic which allows for the combining of

principles of two or more branches/paradigms, while maintaining the integrity of each, and then putting the results of the different strategies into conversation with each other (Mertens, 2018).

**Table 3**

*Philosophical and Methodological Assumptions of Dialectical Pluralism*

Axiology	DP takes a pluralist and multidimensional view of ethics; users should explicitly dialogue and thrive on multiple social, economic, and political values.
Ontology	Reality is multiple - there are multiple true statements that can be made about reality. Unless one is stating a simple fact or definitional truth, there is no single way to “carve nature at its joints.”
Epistemology	Users of DP acknowledge the fallibility of knowledge and have the goal of producing somewhat heterogeneous and somewhat homogeneous wholes that respect multiple standpoints and place weight on solutions that work in theory and contextualized practice by enabling dialogue with two or more epistemologies.
Methods	Mixed methods

*Note.* From “Dialectical Pluralism: A Metaparadigm Whose Time Has Come,” by R.

Johnson, 2017, *Journal of Mixed Methods Research*, 11(2), p. 162.

The process of DP is conducted dialectically, dialogically, and hermeneutically to engage in back-and-forth disputation and examination to address the assumption much of reality is plural and dynamic rather than singular and static (Johnson, 2017; Johnson & Stefurak, 2013). It provided a process to consider multiple issues and perspectives interactively (Johnson, 2017; Johnson & Stefurak, 2013). It represented both an intellectual process which one can engage in on their own and a group process, working

with a heterogeneous group (Johnson, 2017). The characteristics of engaging in DP are (Johnson, 2017, p. 163):

- a) Dialectically listening, carefully and thoughtfully, to different disciplines, paradigms, theories, and stakeholder/citizen perspectives;
- b) Combining important ideas from competing values and into a new workable whole for each research study/evaluation
- c) Explicitly stating and “packing” the approach with researchers’ and stakeholders’ epistemological and social/political values and constructing standards to guide and judge the research;
- d) Conducting the research ethically;
- e) Facilitating dissemination and utilization of research findings (locally and more broadly); and
- f) Continually evaluating the outcomes of the research/utilization process

In the context of evaluation, DP requires evaluators and stakeholders to a) dialectically listen to different paradigms/worldviews, disciplines, theories, and stakeholder and citizen perspectives; b) ground each evaluation instantiation in stakeholders’ epistemological and social-political values; and c) recognize that multiple and very different views of “the good” (quotations original) are reasonable, but work together to build on differences and strengths for justice as fairness and democratic equality (Johnson & Stefurak, 2013; Johnson, 2017). The role of the evaluator in DP is to create a forum in which discussions regarding data collection, analysis, interpretation, and use phases can be discussed respectfully and in a way in which different worldviews can be shared (Mertens, 2018).

## **Merit and Worth**

Scriven (1991) defined merit as the “intrinsic value of an evaluand” (p. 227) and compared it to its extrinsic value, which is its worth. Stufflebeam and Coryn (2014) further explained merit as being related to the standards of the particular discipline in which the evaluand resides and can be thought in terms of levels of excellence. An example of merit may be an afterschool math program, which is aligned to a set of standards, put forth by the National Council for the Teachers of Mathematics (NCTM) for remediation programs and has been recognized for its success. The merit of the program would be demonstrated through these achievements.

Worth, Stufflebeam and Coryn (2014) explained, is “the evaluand’s combination of excellence and service in an area of clear need within a specific context and considering the costs involved. Worth is quality under consideration of context and costs” (p. 9). Using the same example of the afterschool math program, even though it had merit, if there were no students who needed remediation it would be deemed to have little to no worth as there is no need for the program in the given context.

## **Credible Evidence**

Donaldson (2017) and Greene (2017) stated the demand for rigorous and influential evaluations is at an all-time high across the globe as there is a need to produce evidence to support important decisions about policies, practices and concomitant resource allocation. Because of the ways in which evaluation results are used it is important to ensure the evaluations produce credible evidence. However, credibility of evidence is defined and judged in different ways and to some degree is dependent on the recipient of the evidence. Miller (2017) defined credibility using the

definition provided by information and communication scientists, as the “extent to which information is perceived as believable or plausible” (p. 40; Tseng & Fogg, 1999; Wathen & Burkell, 2002). The use of the term ‘perceived’ implies there are factors, which influence the perception of the evidence. Three factors, which influence the perception of credibility of evidence are the design of the evaluation or experiment which produces the evidence, the receiver of the evidence, and the characteristics of the evidence itself (Henry, 2017; Greene, 2017; Miller, 2017).

The design of an evaluation is one the first activities completed once an evaluation is initiated with the development of the evaluation framework. The evaluation framework addresses the particular design and methods of the evaluation once the evaluation questions are determined (Greene, 2017; Scriven, 2015; Fitzpatrick, Sanders, & Worthen, 2011). Credibility in the design of an evaluation is addressed through the technical soundness and methodologies selected for use in the evaluation (Greene, 2017). This is addressed by ensuring the methods selected are appropriate for the evaluation questions and qualitative and quantitative approaches adhere to recommended practices and standards for ensuring reliability and construct validity (Chelimsky, 2017; Julnes & Rog, 2017).

There are members of the evaluation community who argued the only evaluation design that is capable of producing credible evidence is experimental design, which utilizes randomization and a control group, often referred to as the gold standard (Donaldson, 2017). This perspective was reinforced by funding agencies, such as the Institute for Educational Science (IES) prioritizing the rewarding of grants to evaluation

designs which utilize experimental designs, followed by high-quality quasi-experimental designs, and then all other designs (Henry, 2017).

Chelimsky (2007) argued the selection of a method prior to determining the evaluation questions does not follow the process evaluators should adhere to in the design of an evaluation. Chelimsky (2007), highlighted this point, “First, the social inquirer identifies the inquiry purposes and questions and only then selects a methodology that fits these purposes and questions” (p. 206). Julnes and Rog (2017) posited evaluation questions can seek different levels of conclusions drawn from an evaluation and selecting appropriate methods needs to be considerate of the level of conclusion.

The different levels of conclusions generalized by Julnes and Rog (2017), built on the list of evaluation questions introduced by Weiss (1998): 1) overall aggregate questions, which describe the overall impact of a program; 2) questions about moderated relationships, which describe the size of the impact of a program; and 3) those focused on inferences about some underlying reality, which describes what are believed to be underlying mechanisms for the observed outcomes. Although experimental designs are appropriate for responding to aggregate questions, they provide limited insight into disaggregate questions and questions of moderating factors; whereas qualitative and mixed methods can provide a great deal of insight into such questions (Julnes & Rog, 2017). This demonstrates that different levels of conclusion will require different methods of inquiry.

In addition to the level of conclusions drawn about a program, evaluations are composed of evaluation tasks that require different evaluation questions in order to

obtain the information necessary to complete the task. The five traditional evaluation tasks include: 1) implementation assessment, 2) outcome assessment, 3) impact evaluation, 4) valuation, and 5) critical review (Julnes & Rog, 2017). Julnes and Rog (2017) organized thirteen of Weiss (1998) evaluation questions with the appropriate evaluation task and provided additional commentary to demonstrate the different evaluation questions which are associated with each evaluation task, shown in Table 4.

**Table 4**

*Thirteen Questions Organized by Evaluation Activity with Comments in Brackets*

Evaluation Task	Evaluation question with Elaboration
A. Implementation Assessment	Describing – What went on in the program over time? [e.g., What has been the reach of the program?] Comparing. How closely did the program follow its original plan? [e.g., Did it reach the expected target population in the numbers expected? Did the population receive the “dose” of the program intended?] Comparing. Did recipients improve? [e.g., pre-post scores and rate of change] Comparing. Did recipients do better than non-recipients] [e.g. differences in change or rate of change]
B. Outcome Assessment	Comparing. Did recipients improve? [e.g., pre-post scores and rate of change] Comparing. Did recipients do better than non-recipients] [e.g. differences in change or rate of change]
C. Impact Evaluation	Aggregate Program Impacts Ruling out rival explanations. Is the observed change due to the program? [Is attribution possible? Is an assessment of contribution possible?] Disaggregated Impacts Disaggregating. What characteristics are associated with success? [interpreted as causal factors moderating impact] Profiling. What combinations of actors, services, and conditions are associated with success and failure? [interpreted as combinations moderating impact] Causal Explanation (Assessment of Causal Mechanisms) Modeling. Through what processes did change take place over time? [What are the key mediators? Underlying mechanisms?] Underlying mechanisms?

D. Valuation	Costs. What was the worth of the relative improvement of recipients? [Are the costs of the intervention or program, whether quantifiable or not, commensurate with the benefits of outcomes?]
E. Critical Review	<p>Locating unanticipated effects. What unexpected events and outcomes were observed? What are the limits to the findings? [e.g., limits of generalization]</p> <p>Interpreting. What are the implications of these findings? What do they mean in practical terms? [What decisions can they guide?]</p> <p>Fashioning recommendations. What recommendations do the findings imply for modifications in [or continuation of] program and policy?</p> <p>Policy analysis. What new policies and programmatic efforts to solve social problems do the findings support?</p>

*Note.* From “Actionable Evidence in Context,” by G. Julnes and D. Rog, 2015, In *Credible and Actionable Evidence* (pp. 225-226). Copyright 2017 SAGE Publications, Inc doi: 10.4135/9781483385839

If an evaluation is going to be designed in a manner that will produce credible evidence the methodological considerations need to be downstream of the development of the evaluation questions. Chelimsky (2017) stated, “No one method is uniquely suited to answering all questions, different approaches that are appropriate to both the specific questions and its purpose can be used together – one method’s strength bolstering another one’s weakness – to produce answers” (p.180). The perception towards lending credibility to evidence produced by an evaluation based on experimental design could have more to do with the individual receiving the evidence than the actual evaluation design.

Credibility is typically thought about the as the perception of the information as believable or plausible (Miller, 2017). This implies individuals receive information from the results of an evaluation and will form a judgment about the credibility of the

information. Individuals who are recipients of the information are unlikely to form judgments in the same way and based on the same criteria. Which criteria an individual uses to form an opinion depends on different factors, including their level of domain expertise, analytic skills possessed, and their world view (Miller, 2017).

An individual's level of expertise can be classified on a spectrum from novice, with little expertise about the evaluand, to expert, as someone with a great deal of expertise. If a recipient of evaluation information is a novice, they are able to assess surface characteristics of the evidence without domain knowledge and will, therefore, depend on those characteristics to determine the accuracy and credibility of the information (Lucassen & Schraagen, 2011). An expert, however, has a deeper level of content expertise relative to the evaluand and can distinguish between relevant and irrelevant information. However, Tyversky and Kahneman (2013) argued that experts rely on heuristics and intuitive devices, such as representativeness, availability of instances or scenarios, and adjustment from an anchor to arrive at their judgments about accuracy and credibility of the information. According to Tyversky and Kahneman (2013), these heuristic devices economical and effective, but they can lead to systematic and predictable errors. Miller (2017) opines this makes them more likely to ignore the surface characteristics, which are more noticeable to a novice.

Analytic skill pertains to the individual's ability to understand and evaluate the strength and weight of the evidence. "Weight refers to the reliability and consistency of evidence whereas the strength refers to the size of an effect or its position along a dimension, such as positive-negative continuum" (Miller, 2017, p. 46). An individual can evaluate the weight of the evidence by considering such things as the quality of the

sampling design and its execution. The strength of the evidence can be evaluated by considering such things as whether the clients liked the program or whether it was perceived to have made any positive or negative changes to the clients' well-being. Although the strength of evidence is often easily understood and captured by the information recipient, the weight of the evidence is not as intuitive and, as such, may often be ignored by someone without strong analytic skills (Miller, 2017).

The worldview of the recipient also plays a role in an individual's judgment of the credibility of evidence. If the evidence comes from a source similar to the worldview of the recipient it is more likely to be perceived as credible (Miller, 2017). As discussed by Mertens (2018), an individual's worldview is comprised of their subscription to certain ontology and epistemology, to which commonly associated methods are aligned. A worldview different than the worldview of the evaluator can result in a discounting of evidence due to the methods employed or the nature of the evidence itself. Birnbaum and Stegner (1979) and Birnbaum and Mellers (1983) demonstrated the effect of this sort of bias using a scale-adjustment model, which examined the effect of expertise on bias. In the experiment when the information came from a source that the individual trusted it was rated higher than when the information came from a source that was independent or a source that had a higher level of expertise. When the information was perceived as coming from an opposing perspective it was given the least amount of weight regardless of expertise. Birnbaum and Stegner (1979) provided evidence individuals are more likely to lend credibility to judgments when the individual perceives the source as in alignment with themselves and their interests.

Another consideration for the credibility of evidence is the inherent characteristics of the evidence itself. Miller (2017) posited that a core set of information characteristics exists that people use to determine whether or not information is credible. Examples of characteristics used by recipients to determine credibility are whether the information is accurate and free of errors of fact, the accuracy of the methods used to obtain the information, the saliency of the information, whether the information is current, fair, and impartial, which was echoed by Chelimsky (1998) and Rieh and Danielson (2007). Tseng and Fogg (1999) defined four types of credibility for consideration recipients may consider when judging credibility of evidence:

- Presumed Credibility – The extent to which a source conforms to the stereotype of a trustworthy provider of information;
- Reputed Credibility – The titles and affiliations associated with those who generate the information or where the information was generated;
- Surface Credibility – The outward appearance of the evidence, such as the quality of the prose, grammar, syntax, and spelling of the information; and
- Experienced Credibility – The firsthand or vicarious experience of the source's impartiality, fairness, competence, and honesty; such as observing the evaluation findings firsthand or through a peer-reviewed journal article

Greene (2017) offered a different perspective of the credibility of evidence as an earned phenomenon. From the perspective of democratic evaluation, Green (2017) stated “the credibility of evaluative evidence is not automatically granted via the use of particular empirical methodologies but rather is earned through inclusive, relational, and dialogic processes of interpretation and action that happen on the ground, in context,

and in interaction with stakeholders” (p. 206). From this perspective, credibility of evidence is earned through the use of the evaluation findings rather than solely through technical details. The factors which promote the use of the evaluation is the technical soundness of the evaluation design and methodologies used, the ability of the evaluation to generate information of import and consequence to the stakeholders in the contexts being evaluated, and the evaluation being directed towards the issues, concerns, and questions of meaning and relevance to a diversity of stakeholders, with special inclusion of the interests of stakeholders who are least well served in those contexts (Greene, 2017).

The perspective provided by Greene (2017) was inclusive of the other concerns regarding judgments of credibility but reserves the actual determination of credibility to whether or not the evidence is of use to the stakeholders. Greene (2017) concluded if the purpose of evaluation is to produce credible evidence upon which to base decisions for the purposes Donaldson, Christie, and Mark (2019) referred to as policies, practices, and concomitant resource allocations, then it is logical evidence is deemed credible if it is able to be used by stakeholders for those purposes. However, understanding the dynamics between the generation of evaluation results and use of those results is important as it will influence every step of the evaluation design and implementation.

### **Objectives-Oriented Evaluation**

Objectives-oriented evaluation was classified by Stufflebeam (2001) as a quasi-evaluation model which uses the objectives of an educational program as an advanced organizer to determine the extent to which the program’s objectives are being met. Early versions of objectives-oriented evaluation models were developed by Ralph Tyler

(Madaus & Stufflebeam, 1989) and Malcom Provus (Fitzpatrick, Sanders, & Worthen, 2011). Objectives oriented evaluation is included in the Methods branch of the Evaluation Theories Tree (Alkin & Christie, 2004) and the theorists represented in this group of evaluation models are rooted in the positivist/postpositivist paradigm and, as a result, importance is placed on objectivity and quantitative methods are commonly associated with these sorts of evaluations (Mertens, 2018).

The Tylerian (1931) evaluation approach was developed in the 1930s to work with undergraduate biology faculty at the Ohio State University. The goal was to develop achievement tests as a means of tracking students to help reduce failures and dropouts. The use of achievement tests rather than measurement as a means of educational assessment. The achievement tests would be grounded in the local course curriculum, whereas measurement was concerned with broad comparisons of programs at large (Wraga, 2017). Tyler (1931) defined a process of defining objectives to determine what students were to learn in a course and the student behavior that could be measured to determine whether or not the objectives were met. The discrepancy between what the student ought to have learned and what the student demonstrated as learned provided an evaluation of the course curriculum.

As a beginning step to the process Tyler (1931) had the biology faculty identify the objectives students were expected to achieve as a result of taking the course (Madaus & Stufflebeam, 1989); what kinds of behavior they should demonstrate, i.e. “what ways of thinking, feeling or acting; and with what content” (Madaus, Scriven, & Stufflebeam, 1983, p. 69). According to Fitzpatrick, Sanders, and Worthen (2011) “By stating objectives in terms of what students should be able to do, Tyler believed teachers

could more effectively plan their curricula and lessons to achieve those objectives” (p. 155). Once the objectives for the course were established, the next steps to complete the evaluation were to define the objectives in behavioral terms, find situations where achievement of the objectives could be shown, develop or select measurement techniques, collect the performance data, and compare the performance data with the behaviorally stated objectives (Fitzpatrick, Sanders, & Worthen, 2011).

An example of how the objectives would be used to evaluate the course and student performance provided by Tyler (1931), as expressed by Madaus and Stufflebeam (1989), were “An ability to formulate reasonable generalizations from the specific data of an experiment; the ability to apply general principles to new situations; use a microscope” (p. 10). From these course objectives the faculty developed tests and activities that would provide data on whether or not the students achieved the objectives. Depending on the behaviors defined in the objectives a multiple-choice test would be appropriate; however, some objectives required formats such as written responses or demonstrations that require observation. The data collected would then be analyzed to determine whether or not the course objectives were met. Discrepancies found from the data analysis would provide evidence of where gaps were between instruction and student learning.

Tyler (1931) advocated the use of multiple indicators to measure outcomes, including direct and indirect measures; Tyler (1931) opined evaluators should employ a broad array of data collection techniques to support the findings of the evaluation. Examples of additional sources of evidence beyond test scores included observation, self-reporting instruments, interviews, and examples of student products (Madaus &

Stufflebeam, 1989). This flexibility in the model laid foundations for additional objectives-oriented evaluation approaches to be designed.

The creation of the Discrepancy Evaluation Model (DEM) was credited to both Malcom Provus and Andres Steinmetz (Fitzpatrick, Sanders, & Worthen, 2011; Madaus, Scriven, & Stufflebeam, 1983). Provus' DEM is an objectives-oriented approach to evaluation, in the Tylerian tradition, which uses similar design but is used outside of the curriculum evaluation context by developing standards to use as comparison. This model was originally developed by Provus (1973) as means of providing information to the administration in the Pittsburgh Public School system as a way to support the administration's decision making. According to Madaus, Scriven, and Stufflebeam (1983), the process of the DEM was similar to Tyler's objectives development, that a standard would be determined for a given object (person or program). Once the standard is agreed on it is compared to the actual state of the evaluand, referred to as the performance, and the difference between the standard and the actual performance of the evaluand provides the discrepancy.

According to Madaus, Scriven, and Stufflebeam (1983), Andres Steinmetz contributed to DEM by providing a scheme with which to respond to the task of evaluating educational programs. Steinmetz (1983) described the goal of DEM as "to make judgements about the worth or adequacy of an object based on the discrepancy information between the standard and the performance" (p. 80) and as a result provided the formula  $S - P = D$ , where  $S$  is the standard,  $P$  is the performance, and  $D$  is the discrepancy. The role of the evaluator in a DEM evaluation is to work with the client to determine the standards the performance will be compared to and to seek out technical

expertise to measure the performance. Once the performance data is collected the evaluator works with the client to compare the performance to the standards and determine the discrepancy. Madaus, Stufflebeam, and Scriven (1983) provide a worked example of DEM (p. 80-84):

If wanting to evaluate a motorcycle for purchase the evaluator might meet with a mechanic to determine what the standards should be for determining whether or not the motorcycle is a good value. Characteristics the motorcycle should possess: Cost: should not cost over \$800; Power: should be able to cruise at 60 mph; stability: should be large and heavy enough to stay on the road; Noise: should be quiet; Appearance: should have classic "World War I" look; Mechanical Condition: should be in good condition and not presently need repairs

Provided in Table 5 are elements of the evaluation.

**Table 5**

*Example of DEM Evaluation Framework*

Evaluation Question	Standard	Source Information	of Instrument	Data Collection
1. How much does the motorcycle cost?	The motorcycle should not cost more than \$800.	Seller motorcycle	of Interview	Evaluator
2. Does the motorcycle cruise at 60 mph?	It should maintain 60 mph with 2 passengers on a straightaway all day.	Engineering specifications	Review engineering specs	of Evaluator
	It should maintain 60 mph with 2 adults up the mile-long hill on Rt. 629	Motorcycle	Road test	Evaluator

The strengths of objectives-oriented evaluation approaches frequently cited are related to its ease of use and narrow focus, and the methods are easily understood and

easy to follow (Fitzpatrick, Sanders, & Worthen, 2011). This makes it a desirable model for evaluations completed internally and appropriate for programs and projects which “have clear, supportable objectives” (Stufflebeam, 2001, p. 16). Completing the evaluation requires that stakeholders involved in the evaluation reflect on the program and its intentions. The sources of data, which inform the outcomes of the program, need to be dependent upon the nature of the objective (Madaus & Stufflebeam, 1989), which allows for broad considerations of both qualitative and quantitative data collection and analysis.

The role of the evaluator in objectives-oriented evaluation models focuses on working with subject matter experts, program directors, or other program stakeholders to define the standards or objectives the program is designed to achieve and developing the means to compare them with the actual program outcomes (Fitzpatrick, Sanders, & Worthen, 2011; Madaus, Scriven, & Stufflebeam, 1983; Wraga, 2017; Madaus & Stufflebeam, 1989). The evaluator is not establishing the objectives or the standards of the program or judging any discrepancy resulting from the evaluation. This is left to the stakeholders who helped establish the program objectives or standards as part of the evaluation.

Steinmetz (1983) identified inherent bias as a limitation of the discrepancy model and may influence the results, “Whatever conversations might be held about S are usually jargon loaded and assume knowledge over the very conditions or phenomenon which, as client, you don’t have, and which led to you turning to someone else in the first place. Thus, the specific S brought to bear on certain performance information may remain unknown to you and, to some extent because they are not articulated by the

expert, less than consciously known to them” (p. 82). However, this limitation can be mitigated depending on how broad the stakeholder group is that provides input on the objectives and any determined discrepancy. According to Wraga (2017) early curriculum reforms using this sort of model advocated for widespread faculty participation.

Additional limitations of objectives-oriented evaluation approaches often cited are the limitations imposed by the single-minded focus of the evaluation on the objectives or standards of a program or course (Fitzpatrick, Sanders, & Worthen, 2011). These self-imposed blinders, critics claim, can cause evaluators to ignore other program outcomes or make incomplete judgments about the program. As a result, the outcomes of the evaluation can be seen as leading to terminal information that is neither timely nor pertinent to help advance or improve the program (Stufflebeam, 2001).

### **Participatory Evaluation**

Participatory oriented evaluation models were first developed in the 1970s, out of growing concerns evaluators did not really understand the phenomena which they were evaluating (Fitzpatrick, Sanders & Worthen, 2011). Emerging models stressed “experience with program activities and settings as well as involvement of program participants, staff, and managers in the program evaluation” (Fitzpatrick, Sanders, & Worthen, 2011, p. 190). Cousins (2005) described participatory evaluation as “members of the evaluation, community and members of the stakeholder groups, relative to the evaluand each participate in some or in all of the shaping and/or technical activities required to produce evaluation knowledge leading to judgments of merit and worth and support for program decision making” Toal (2009) posited this description represented a broad range of evaluation approaches such as:

“fourth generation evaluation (Guba & Lincoln, 1989), stakeholder collaborative evaluation (O'Sullivan & O'Sullivan, 1998), empowerment evaluation (Fetterman, 2001), participatory evaluation (Cousins & Earl, 1992; Cousins & Whitmore, 1998), participatory monitoring and evaluation (Gaventa, Creed, & Morrissey, 1998), inclusive evaluation (Mertens, 1999), value-engaged evaluation (Greene, 2005), and democratic deliberative evaluation (House & Howe, 1999) (p. 350).

The theorists included in this group are representative of the use, values, and social justice branches of the Evaluation Theories Tree (Alkin & Christie, 2004).

Participatory evaluation as an evaluation model, as opposed to the category of models, grew out of the stakeholder-based evaluation model. It was an early attempt to move toward evaluation results which were more responsive to the needs of the users of the results (Cousins & Earl, 1992). In the stakeholder-based model the evaluator serves as the principal investigator and stakeholders are used in a consultative fashion; whereas, in a participatory evaluation model the evaluator is seen as a technical expert and facilitator, and the stakeholders largely take on the responsibility of the evaluation activities (Cousins & Earl, 1992). Cousins and Whitmore (1998) delineated participatory evaluation into two distinct streams: Practical-Participatory Evaluation (PPE) and Transformative-Participatory Evaluation (TPE). Moreau (2017) provided a description which aides in distinguishing between the two streams (p. 335):

Although these two streams are not exclusively distinct and a participatory evaluation may exude some element of each, the stream guides the overarching rationale of the evaluation, who is engaged in the planning and implementation, as well as these individuals' levels of participation. The primary purpose of P-PE is practical in nature and is to foster decision making, problem solving, and the use of evaluation processes and findings (Cousins 2005). In P-PE the evaluator works in partnership with stakeholders who have the power to use future evaluation findings and subsequent recommendations. Conversely, the main purpose of T-PE is political in form and is to empower individuals or groups, transform power relations, and promote social justice (Cousins & Whitmore, 1998).

Transformative-Participatory Evaluation endeavors to invoke participatory principles and actions in order to democratize social change (Cousins & Whitmore, 1998) and as such its theoretical underpinnings would be connected to the Social Justice branch of the Evaluation Theories tree, introduced by Mertens and Wilson (2012). Practical-Participatory Evaluation represents the Use branch of the Evaluation Theories tree and are therefore rooted in the Pragmatism paradigm, which considers knowledge in relation to the study and matches methods to the research questions (Mertens, 2018).

For the purposes of this study Transformative-Participatory Evaluation is not an appropriate evaluation model. Henceforth, participatory evaluation will be in reference to the Practical-Participatory Evaluation stream. Increasing evaluation utilization is one of the primary benefits of participatory evaluation. A core premise of this model is stakeholder participation in the evaluation will enhance the evaluation's relevance, ownership, and thus utilization (Cousins & Whitmore, 1998). Training stakeholders to complete evaluation activities promotes wider use of evaluative thinking within the organization. Patton (2018) referred to this as "process use" and described it as "individual changes in thinking and behavior, and program or organizational changes in procedures and culture, that occur among those involved in evaluation as a result of the learning which occurs during the evaluation process" (p. 20).

Criticisms of participatory evaluation models focused on the increased need for resources to complete the evaluation and the credibility of the evaluation to those who are external to the evaluation itself (Fitzpatrick, Sanders, & Worthen, 2011). Having those who are closest to the program completing the evaluation increases the risk of

inserting bias into the process. Daigneault and Jacob (2009) also identified ambiguous conceptualization of participatory evaluation as a concern.

The implementation of a participatory evaluation will vary based on the context and program being evaluated. However, the common element which need to be present for a participatory evaluation to be carried out is partnership between trained program evaluators and program stakeholders, as Moreau (2017) noted, “More specifically, the program evaluators (who have technical expertise in evaluation theory and methods) collaborate with program stakeholders (who have an understanding of a program and how it is supposed to operate) to evaluate a program” (p. 334). The stakeholders should represent not only those responsible for the development and delivery of a program but also those who are intended to participate and benefit from the program. Evaluation team size and makeup of stakeholders will vary depending on the scope of the evaluation and the size of the program.

Stakeholder involvement in completing the evaluation is one of the hallmarks of participatory evaluation, with the intent to promote evaluation use, evaluative thinking, and evaluative capacity. Stakeholders are “actively involved in the evaluation design, data collection, analyses, interpretation, and dissemination efforts” (Moreau, 2017, p. 334). Participatory evaluations should begin with the training of the stakeholders who will make up the evaluation team to provide them with the necessary knowledge and skills to successfully complete the various evaluation tasks. Training should be a first step in a participatory evaluation and should continue to occur throughout the evaluation as needs and skill gaps present themselves.

Following training, the stakeholders begin the evaluation by developing an evaluation framework (Moreau, 2017). The evaluation framework provides an overview of the major evaluation questions and establishes what will be evaluated, who will be involved in the evaluation activity when the activities will take place, what evaluation and analysis methods will be used, and how the findings will be consolidated and the results shared (Canadian International Development Agency, 2001). The framework developed is unique to each evaluation. The process will vary depending on the context of the evaluation and the makeup of the evaluation team.

Once the evaluation framework is determined the evaluation can be carried out by the team. All participants in the evaluation should be completing the evaluation activities with support from the trained evaluator. Participatory evaluation is method agnostic, so the evaluator should be prepared to support participants' analysis of data using qualitative, quantitative, and mix-methods (Moreau, 2017). This will also be the case when participants in the evaluation develop a report of the findings from the evaluation. Participants may need additional training or additional resources to develop a report, which properly summarizes and communicates the evaluation results to the standards of the field.

King and Fitzpatrick (2009) provided an example of a participatory evaluation of the Anoka-Hennepin School District's Special Education program. The evaluation team was made up of more than 100 people, which included a team of three consulting evaluators, three district evaluators and the district's special education administrators, and a large self-study team made up of as many stakeholder groups the team was able to identify. The evaluation framework described by King (2007) was developed from

working sessions with the self-study team meetings in which the over 100 members of the self-study team would engage in table discussions regarding concerns and experiences with the special education program in the district. The session would end with self-study team members completing post-session evaluations which included plusses, wishes, and questions. The information from the post-session evaluations was then analyzed by the district's evaluation team and led to a list of concerns and major evaluation questions.

### **Defining and Measuring Participation**

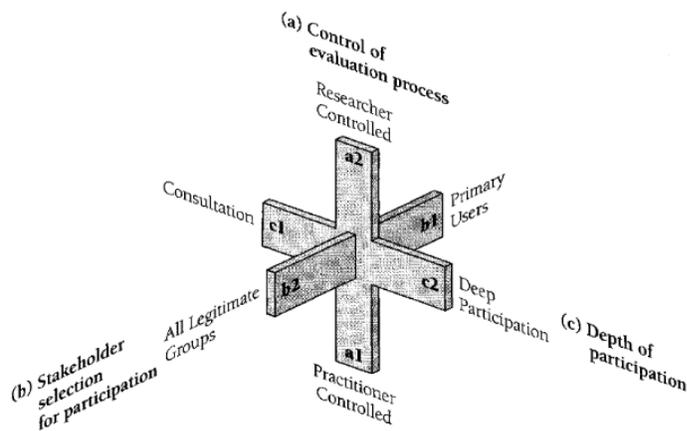
To some extent all evaluations include some level of participation. The evaluator must interact with various stakeholder groups to develop and complete the evaluation (King, 2007). Determining what type of participation in an evaluation qualifies as participatory evaluation is an ambiguity which theorists have struggled with. Daigneault and Jacob (2009) discussed the difficulties with this question. The issues associated with participatory evaluation were posited to be "insufficient and/or inadequate conceptualization and operationalization" (p. 331). This results in ambiguity around who gets to participate, or who is considered to be a stakeholder. There is also ambiguity around what sorts of activities are included as participation and at what point the participation is adequate to be considered participatory.

Cousins and Whitmore (1998) set forth a framework for participatory evaluation which identified three distinct process dimensions for participatory inquiry: a) Control of the evaluation process, b) Stakeholder selection for participation, and c) Depth of participation. Each dimension is independent of the other dimensions and exists on a spectrum of the extent to which the participation in the evaluation activities is limited or

extensive. Provided in Figure 2 is a visual representation of each dimension and the extremes of the dimensions

## Figure 2

### *Dimensions of Form in Collaborative Inquiry*



*Note.* From “Framing Participatory Evaluation,” by J. B. Cousins and E. Whitmore, 1998, *New Directions for Evaluation*, 80(Winter), 87–105.

Building from the conceptual framework provided in Cousins and Whitmore (1998), Daigneault and Jacob (2009) conceptualized and operationalized participatory evaluation by outlining the key decision points defined in participatory evaluation:

1. Evaluation questions and issues definition/methodological design – characterized as the moment when a decision is made about the framing of the evaluation including selection of evaluation questions, theoretical framework methods, techniques, and instruments.
2. Data collection and analysis - characterized by making decisions about how to concretely collect, assemble, code, and analyze data as well as carrying these

tasks out.

3. Judgements and recommendations formulation – characterized by making decisions and determinations of merit and worth of a program and formulating suggestions for future action.
4. Report and dissemination of evaluation findings – characterized by the decisions about reporting and diffusion of evaluation findings and implication (Daigneault & Jacob, 2009, p. 339).

Each decision point is considered a dichotomous indicator of participation and typed as involvement of nonevaluative stakeholders in the task (presence of the indicator) or no involvement of nonevaluative stakeholders in the task (absence of the indicator) (Daigneault & Jacob, 2009, p. 339).

Daigneault and Jacob (2009) developed coding schemes, using the m of n rule, where  $m = 1$  and  $n = 4$ , giving a weight of .25 to each indicator. These schemes are shown in tables 6, 7, and 8 for each of the dimensions of participatory inquiry posited by Cousins and Whitmore (1998) and are used to establish cut-off points for determining whether an evaluation can be considered participatory.

**Table 6**

*Coding Scheme for Extent of Involvement*

Number of Tasks Nonevaluative Stakeholders are Involved In	Level of Membership	
	Intuitive Label	Numerical
0	No involvement	.00
1	Limited/weak involvement	.25
2	Moderate involvement	.50
3	Substantial/strong involvement	.75
4	Full involvement	1.00

*Note.* From “Toward Accurate Measurement of Participation,” by P. M. Daigneault and S. Jacob, 2009, *American Journal of Evaluation*, 30(3), p. 340.

**Table 7***Coding Scheme for Diversity of Participants*

Number of Tasks Stakeholder Types Involved	Level of Membership	
	Intuitive Label	Numerical
0	No diversity	.00
1	Limited/weak diversity	.25
2	Moderate diversity	.50
3	Substantial/strong diversity	.75
4	Full diversity	1.00

*Note.* From “Toward Accurate Measurement of Participation,” by P. M. Daigneault and S. Jacob, 2009, *American Journal of Evaluation*, 30(3), p. 340.

**Table 8***Coding Scheme for Control of the Evaluation Process*

Level of Membership	Numerical
Intuitive Label	
Exclusive control by evaluator and/or nonparticipating evaluation sponsor	.00
Limited/weak control by nonevaluative participants	.25
Shared control between nonevaluative participants and evaluator and/or nonparticipating evaluation sponsor	.50
Substantial/strong control by nonevaluative participants	.75
Exclusive control by nonevaluative participants	1.00

*Note.* From “Toward Accurate Measurement of Participation,” by P. M. Daigneault and S. Jacob, 2009, *American Journal of Evaluation*, 30(3), p. 340.

With each dimension defined and a scale to measure participation developed, a threshold for whether or not an evaluation could be considered participatory was developed. According to Daigneault and Jacobs (2009), the presence of all three fundamental attributes of participation is required, and each category must have at least a score of .25 or greater. Daigneault and Jacob (2014) revised the process by adding a final step of averaging the scores across the three dimensions as an indicator a measurement of overall participation.

An additional attempt to operationalize and conceptualize participatory evaluation was the introduction of the Evaluation Involvement Scale (EIS) by Toal, (2009). The scale was developed as a response to a call for instruments designed to measure the level of involvement of stakeholders in a participatory evaluation in multisite settings. Due to the limitations of the program design the EIS only measures the third dimension of participatory inquiry posited in Cousins and Whitmore (1998), depth of participation. The scale was developed using key evaluation decision points posited in Burke (1998) as a general framework. Each key decision point was assigned to the three evaluation stages: 1) evaluation planning, 2) implementation, and 3) communication of results, as well as additional items the development team added to provide clarity or distinction to the decision points, resulting in thirteen item scale shown in Table 9 (Toal, 2009). Participants respond to each item rating the extent to which they were involved in the different evaluation activities using a 1:4 scale: 1 = No; 2 = Yes, a little; 3 = Yes, some; 4 = Yes, extensively (Toal, 2009, p. 354).

**Table 9**

*Evaluation Involvement Scale Items*

- 
1. Discussions that focused the evaluation
  2. Identifying evaluation planning team members
  3. Developing the evaluation plan
  4. Developing data collection instruments
  5. Developing data collection processes
  6. Collecting data
  7. Reviewing collected data for accuracy and/or completeness
  8. Analyzing data
  9. Interpreting collected data
  10. Writing evaluation reports
  11. Reviewing evaluation reports for accuracy and/or completeness
  12. Presenting evaluation findings (e.g., to staff, to stakeholders, to an external audience)
  13. Developing future project plans based on evaluation results
- 

*Note.* From “The Validation of the Evaluation Involvement Scale for Use in Multisite Settings,” by S. Toal, 2009, *American Journal of Evaluation*, 30(3), p. 356.

**Stakeholder**

Early in the evaluation planning stages the evaluation stakeholders or stakeholder groups are identified. However, stakeholder and stakeholder groups can be defined in different ways so, depending on the context of the program, oftentimes the evaluator needs to find some means of balancing between defining the stakeholders broadly or narrowly (Fitzpatrick, Sanders, & Worthen, 2011). The risk with defining the stakeholders and stakeholder groups too broadly is the additional cost and time it has the potential to add to completing the evaluation. The risk of defining it too narrowly is the chance important perspectives or input will be excluded from the evaluation, as well as the risk of disenfranchising groups or populations that were excluded.

The individuals or groups initially requesting the evaluation are often considered the primary audience for the evaluation and are frequently thought of as the stakeholders for an evaluation; however, there are additional stakeholders, which could be considered beyond this group (Fitzpatrick, Sanders, & Worthen, 2011). Scriven (1991) broadly defined a stakeholder as “one who has substantial ego, credibility, power, futures, or other capital invested in the program and thus can be held to some degree at risk with it” (p. 334). Scriven (1991) broadened the concept of stakeholders to include those who might be opponents of a program and have something to be gained in its failure, those who have stock in a program who might not be aware they are invested in it, and although quite removed, taxpayers or other types of investors.

Scriven (2015) provided a different framing for stakeholders and stakeholder groups. In this framing stakeholders were referred to as program consumers and impactees. Impactees were comprised of three different groups at different stages of a program and were potentially far reaching. The different groups discussed were

upstream impactees, midstream impactees, and downstream impactees. Table 10 provides examples of potential members each group. Upstream impactees included any individual or group who has an investment or interest in the program but who is not directly impacted by it. Commonly considered in this group would be funders of a program, supporters, or opponents of a program.

Also included in this group were anticipators. Anticipators include those who are not invested in a program but would perhaps have some sort of reaction to the announcement or planning of a program, such as real estate agents (Scriven, 2015). Midstream impactees included program staff for which the impact of the program is almost always different. The remaining impactees were the downstream (direct) impactees which were comprised of the primary recipients of the service or the program being evaluated and the downstream (indirect) impactees which included others who are not primary consumers of a program but were impacted by those around them who were, including family members, co-workers, or neighbors in the community where the program resided (Scriven, 2015).

The language used to refer to stakeholders posited by Scriven (2015) denoted attention evaluators must pay to the dynamics, power, and privilege in the evaluation context, as there is a connotation of power or authority if one has 'stake' in a program. Impactees are impacted by a program whether they choose to be or not and impact can be either positive and negative depending on the outcomes of the program and whether the impactee is upstream, midstream, and/or downstream. Scriven (2015) provided important considerations for language choice when referring to stakeholder groups:

Do not use or allow the use of the term 'beneficiaries' to refer to the impactees, since it carries with it the completely unsupported assumption

that all the effects of the program (or all the important effects) are beneficial, when of course the unintended effects may be deleterious and become deal-breakers. It is also misleading to use the term 'recipients' for this purpose, since many impactees are not receiving anything but are nevertheless being affected (p. 17).

**Table 10**  
*Groups of Impactees*

Upstream Impactees	Midstream Impactees	Downstream (Direct) Impactees	Downstream (Indirect) Impactees
Funding agencies		Recipients of a service	Recipients' family
Taxpayers	Program staff	Users of a service	Recipients' co-workers
Political supporters		Products of the program	
Anticipators			

Greene (2005) provided four stakeholder categories, which were similar to the groups identified by Scriven (2015) (Scriven's categories listed in parentheses for the purposes of comparison), but were more specific to roles commonly found in the context of educational and social welfare-oriented programs:

- a) People who have decision authority over the program, including other policy makers, funders, and advisory boards (upstream impactees)
- b) People who have direct responsibility for the program, including program developers, administrators in the organization implementing the program, program managers, and direct service staff (midstream impactees)
- c) People who are the intended beneficiaries of the program, their families, and their communities (downstream (direct and indirect) impactees); and
- d) People disadvantaged by the program, as in lost funding opportunities (p. 398)

The inclusion of individuals or groups who are disadvantaged by a program is a consideration unique to this framing of who is considered as stakeholders.

Fitzpatrick, Sanders, and Worthen (2011) used the categories identified by Greene (2005) and the suggestions provided in Scriven (2015) Key Evaluation Checklist to develop the Checklist of Potential Stakeholders and Audiences (shown in Table 11) which is meant to be a comprehensive representation of possible stakeholders for evaluators to use to ensure they are being inclusive in consideration of stakeholders and audiences. They acknowledged not every group represented in the checklist would be appropriate for every evaluation; however, the checklist provided a comprehensive list of groups to consider as well as those who might use or disseminate the evaluation results.

**Table 11**

*Evaluation Audience Checklist*, source: Fitzpatrick, Sanders, and Worthen, 2011, p. 289

<i>Evaluation Audience Checklist</i>					
<i>Entity to Be Evaluated</i>	<i>(Check all appropriate boxes)</i>				
<i>Individuals, Groups, or Agencies Needing the Evaluation's Findings</i>	<i>To Make Policy</i>	<i>To Make Operational Decisions</i>	<i>To Provide Input to Evaluation</i>	<i>To React</i>	<i>For Interest Only</i>
Developer of the program					
Funder of the program					
Person/agency who identified the local need					
Boards/agencies who approved delivery of the program at local level					
Local funder					
Other providers of resources (facilities, supplies, in-kind contributions)					
Top managers of agencies delivering the program					
Program managers					
Program deliverers					
Sponsor of the evaluation					
Direct clients of the program					
Indirect beneficiaries of the program (parents, children, spouses, employers)					
Potential adopters of the program					
Agencies who manage other programs for this client group					
Groups excluded from the program					
Groups perceiving negative side effects of the program or the evaluation					
Groups losing power as a result of use of the program					
Groups suffering from lost opportunities as a result of the program					
Public/community members					
Others					

*Note.* From *Program evaluation: Alternative Approaches and Practical Guidelines* (4th ed.) by J. Fitzpatrick, J. Sanders, and B. Worthen, 2011, p. 289. Copyright 2011 by Pearson.

In the context of a participatory evaluation, stakeholder inclusion is a key component and the diversity of the typology of stakeholders included in the evaluation is important (Cousins & Whitmore, 1998; Daigneault & Jacob, 2009). Daigneault and Jacob (2009) gave special consideration to the inclusion of nonevaluative stakeholders when conceptualizing and operationalizing participatory evaluation, Daigneault and Jacob (2009) posited four typologies for nonevaluative stakeholders to define and measure the stakeholder diversity: 1) Policy makers and decision makers; 2) Implementers and deliverers; 3) Target populations and intended beneficiaries; indirect beneficiaries and injured parties; and 4) Civil society and citizens. Table 12 provides a description and examples for each typology for further clarification.

Although resources dictate how broadly the evaluation is able to consider different stakeholder groups for inclusion in the evaluation design and implementation, the manner in which stakeholders are defined for an evaluation should be carefully considered. It is rare a program only impacts those immediately involved in the program as an implementer or as a recipient of the program. By considering the broader definitions and descriptions of stakeholders provided in Scriven (2015), Greene (2005), Fitzpatrick, Sanders, and Worthen (2011), and Daigneault and Jacob (2009) the design of the evaluation can take into consideration not only those directly involved and impacted by the program but also downstream impactees who experience secondary or tertiary effects of the program, whether intended or unintended.

**Table 12***Typologies of Nonevaluative Stakeholders*

Types	Description	Examples
Policy makers and decision makers	People politically, legally, and organizationally accountable for the program and its evaluation	Elected and appointed officials, high ranking civil servants, chief executive officers of nonprofit private foundations, think tanks, etc.
Implementers and deliverers	People responsible for the midlevel management and implementation of the program and the delivery of the intervention and/or services	Lower level program managers; street level civil servants, frontline staff, and professionals
Target populations and intended beneficiaries; indirect beneficiaries and injured parties	People toward which the program is directed to modify their behavior and or improve their well-being; local people indirectly and/or potentially affected by the program, either positively or negatively	Juvenile offenders, members of the LGBTQ community, university students with mental health issues, large families with violence problems, K-12 girls
Civil society and citizens	People and organizations having a political interest in the program and its evaluations	Interest groups, unions, think tanks, NGOs, professional associations, private firms, intellectuals, political parties, scientists

*Note.* From “Toward Accurate Measurement of Participation,” by P. M. Daigneault and S. Jacob, 2009, *American Journal of Evaluation*, 30(3), p. 340.

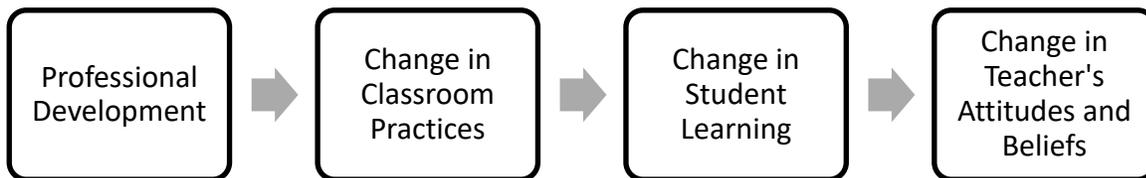
### **Professional Development**

Professional development (also referred to as faculty development and educational development) are the “processes and activities designed to enhance the personal knowledge, skills, and attitudes of educators so they might, in turn improve the learning of students” (Guskey, 2000, p. 16). The presence of professional development activities in the field of education have been growing since the 1970s (Centra, 1978) and

have become common place in the PK-16 and graduate education settings, often with the goal of promoting affective change on the part of the instructor, such as changes in the participants' attitudes, beliefs, or dispositions (Guskey, 2000). Guskey (2000) stated "the premise of trying to change attitudes, beliefs, or dispositions directly is that these affective changes will lead to change in school or classroom practices will ultimately result in improved learning for students" (p. 138-139). Shown in Figure 3 is a model of teacher change demonstrating the assumption of how professional development will impact student learning.

### Figure 3

#### *A Model of Teacher Change*



*Note.* From *Evaluating Professional Development*, by T. Guskey, 2000, p. 139.

Copyright 2000 from Corwin Press, Inc.

The activities included in professional development can be planned and implemented at an organizational level or at an individual level and include activities such as workshops, conferences, peer-observation, mentoring, study groups, inquiry action research, organizational improvement committees (Centra, 1978; Guskey, 2000; Borko, Jacobs, & Koellner, 2010). Participation in professional development activities is sometimes a requirement for maintaining licensing, certification, a stipulation for employment, or to maintain accreditation.

In higher education, professional development plays a key role, because teaching is an important activity for faculty and there is a need to provide meaningful and instructionally sound courses. According to Bergquist and Phillips (1975), teaching is frequently not a serious concern in the training or hiring of college faculty and is often neglected in issues of promotion and tenure. Professional development provides an opportunity to prepare faculty to fulfill their teaching and course design responsibilities. There is a concern that requiring the completion of professional development of this nature implies educators are doing an inadequate job and is seen as an indication they are seen as deficient which can result in resentment and devaluing of professional development activities (Guskey, 2000). However, to others it is recognition that education is a complex and dynamic professional field and educators must embrace an attitude of lifelong learning to keep abreast of an emerging knowledge base and continually refine their practice (Guskey, 2000)

The focus of traditional models of professional development is classified into three categories: 1) instructional development, 2) personal development, and 3) organizational development (Bergquist & Phillips, 1975; Centra, 1978). Professional development focused on instructional development includes activities to develop content knowledge, pedagogical skills, and teaching diagnosis. Personal development focused on professional development includes activities to promote interpersonal skills and career counseling. Organizational development focuses on improving the institutional environment for teaching and decision-making and often includes development for faculty, staff, and administration.

The types of delivery formats typically utilized with a traditional model are courses, workshops, or series designed to transmit a specific set of ideas, techniques, or content (Stein, Schwan, & Silver, 1999). Although these activities introduce new skills into teacher's existing repertoire, there are criticisms. They may be fragmented, disconnected, and decontextualized from the classroom. Teaching is considered routine and technical. Opportunities limit the ability of professional development to be translated into practice (Stein, Schwan, & Silver, 1999; Borko, Jacobs, & Koellner, 2010). According to Borko, Jacobs, and Koellner (2010) more contemporary models of professional development represent a movement toward the idea of building capacity and developing professional cultures of collaboration and use multiple professional development strategies to build capacity for understanding subject matter, pedagogy, and student thinking.

New models of professional development are emerging which are more closely aligned with constructivist and situative theories, grounded in classroom practice, are focused (at least in part) on students' learning, and engages teachers in inquiry about concrete tasks and provides them with opportunities to make connections between their learning and classroom instruction (Borko, Jacobs, & Koellner, 2010). In this sense, the context in which learning and teaching occurs is of greater importance than in traditional professional development models. The experience and perspective of the teacher plays a greater role and professional development is seen as ongoing and a collection of activities rather than single workshop or series.

Koellner and Jacobs (2015) discussed that traditional professional development events still serve an important purpose and can evoke learning, but more adaptive and

flexible models of professional development are available. Adaptive professional development involves ongoing and sustained time and in-house resources (Koellner & Jacobs, 2015). Specified approaches to professional development are going to have goals, content resources, and facilitation materials to ensure predetermined experiences (Koellner & Jacobs, 2015). Koellner and Jacobs (2015) described models of professional development as existing on a spectrum between adaptive and specified formats. Highly adaptive professional development models are meant to be more responsive to goals, resources, and circumstances and therefore are based on the context in which the professional development is occurring within; whereas highly specified models are going to be anchored to rigorous empirical studies and perhaps be the focus of larger scale outcome-focused investigations (Koellner & Jacobs, 2015). The adaptive vs. specified spectrum demonstrates the emergence of contemporary models of professional development, which are models of professional development that serve different purposes beyond content. In these contemporary models the strategies of focus, context, structure, and duration contribute to the outcomes.

Similar to the adaptive vs. specified spectrum presented by Koellner and Jacobs (2015), Hoessler, Godden, and Hoessler (2015) differentiated professional development opportunities as formal planned opportunities, facilitated opportunities, and spontaneous encounters. Formal planned opportunities consist of programming with specified goals or outcomes, such as training, workshops, or orientation. Facilitated opportunities would be those that are organized within the organization but not with any intended outcomes such as access to journals, organized mentorship, annual reviews, online resources available for voluntary consumption, or attending conferences. Spontaneous encounters

consist more of impromptu learning from unplanned interactions with peers, superiors, or just-in-time training related to specific issues. The inclusion of spontaneous encounters expands our understanding of what can be considered professional development and recognizes interactions within professional networks and self-directed learning as professional development, which are more aligned with the context in which teaching and learning is occurring and, as such, will directly inform instructional development and personal development.

Across contemporary models of professional development informal supports, such as socialization and communication networks, and the interplay between the individuals and their environments are emerging as key components (Knight, Tait, & Yorke, 2006; Borko, Jacobs, Koellner, 2010; Koellner & Jacobs, 2015; Hoessler, Godden, & Hoessler, 2015). Knight, Tait, and Yorke (2006) argued professional learning is systemic and is an interplay between individuals and their environment. Livingstone (2001), posited “We learn while we act continuously,’ referring to tacit informal learning and intentional informal learning as a means to gain new understanding and become able to do new things, or to do old things better, without being aware of it” (p. 321).

Borko, Jacobs, and Koellner (2010) argued high quality-professional development content needs to be situated in their classroom practice and in ways that help them to make connections between students’ learning and the classroom instruction. Additionally, high quality professional development has the characteristics of placing teachers in the role of learners and provides opportunities for them to serve as active participants collaborating in the development of the professional development. There was an evolution of the understanding of professional development from the “processes

and activities designed to enhance the personal knowledge, skills, and attitudes of educators so they might, in turn improve the learning of students” (Guskey, 2000, p. 16) to being understood as learning that occurs “in many different aspects of practice, including their classrooms, their school communities, and professional development courses or workshops (Borko, 2004, p. 4).

### **Evaluation of Professional Development**

As the prevalence of professional development has been growing since the mid-1970s, so to have the calls for rigorous evaluations of faculty development programs (Hoyt & Howard, 1978; Levinson-Rose & Menges, 1981; Kucsera & Svinicki, 2010; Brooks, Marsh, Wilcox, & Cohen, 2011; Bamber & Stefani, 2015). Those who are concerned with the lack of published rigorous evaluations of professional development programs point to the need for accountability of these programs to ensure they are meeting the needs of their organizations and being good stewards of the resources they are afforded. Gaff and Morstain (1978) opined, “it has become increasingly important to determine what faculty actually gain from [professional development] efforts and what benefits accrue to the institutions” (p. 73).

Kucsera and Svinicki (2010) referred to the limited scholarly base for faculty developers to reference as a lack of evidence that programs offered as professional development will indeed improve teaching. This leads to limited evidence for decision making purposes, which Hoyt and Howard (1978) identified as a necessity for the future of professional development programs; to be guided by “rationality and sound information rather than by emotion or political considerations” (p. 37). The limited availability of rigorous evaluations is a result of a variety of barriers. Historically, faculty

developers have not been provided with the resources to engage in the time-consuming process of evaluating programs, many developers do not have expertise in evaluation methodologies, and some see it as external to their responsibilities (Guskey, 2000). The changes to attitudes and behaviors of faculty, which often are the types of changes professional development programs are seeking to promote, are inherently difficult to evaluate and often occur over long periods of time due to the fact that there are many factors which impact faculty or staff adopting conceptual changes in their beliefs and practices (Sinatra & Pintrich, 2002; Kuscera & Svinicki, 2010).

According to Luque (2002) intentional conceptual change requires three prerequisites (p. 138):

- 1) Individuals need to be aware of the need to change and to be able to know what to change.
- 2) Individuals must want to change. They must consider change as a personal goal, and not as something imposed by others.
- 3) Individuals must be able to self-regulate their process of change; that is, they must be able to plan, monitor, and evaluate their process of change.

As a result, true change is not something that happens in one session, one semester, or in some cases one year. This is compounded by the fact that there is limited interest in institutions engaging in longitudinal studies or delaying participation in an effort to implement true randomization in the service of evaluation (Kucsera & Svinicki, 2010).

These barriers are evidenced by systematic reviews of the literature completed over the last several decades. Levinson-Rose and Menges (1981) completed a critical review of published research which included inter-institutional projects and campus-wide

programs, grants to support faculty projects, workshops and seminars, feedback from ratings by students, practice with feedback, microteaching and mini-courses, and concept-based training. The review identified five variables as evidence of change: 1) teacher attitude from self-report, 2) teacher knowledge from tests or observer, 3) teacher skill from observer, 4) student attitude from self-report, and 5) student learning from tests or observer reports. “The strongest evidence for most interventions is impact on students (the last two categories), and the weakest is self-reported opinion of participants (first category)” (p. 403). Even though self-reported attitude provides the weakest evidence, Levinson-Rose and Menges (1981) noted “much of this research fails to go beyond data collected on the spot from participants,” (p. 403).

These findings were supported by Chism and Szabo (1997) with the results of a survey completed by 97 faculty developers. The majority of developers used self-reported measures for evaluation and a very low percentage evaluated the effects of faculty development interventions on actual teaching or student learning outcomes. Finally, using the same variables as Levinson-Rose and Menges (1981), Kuscera and Svinicki (2010) completed a systematic review of the literature on faculty development focusing on instructional improvement in nine leading publication sources between the years 1992 and 2007. Their inclusion criteria for appropriate or rigorous evaluation was informed by best practices such as the U.S. Congress’s definition of scientifically valid educational evaluation. The resulting inclusion criteria addressed concerns raised by Levinson-Rose and Menges (1981) including requirements such as:

- The research must include an attempt to evaluate improvement in teaching or student learning,

- Studies which include self-reported measures must include at least one secondary source of measurement, and
- Studies must mention strategies to address reliability and validity of findings.

The result was only 10 out of 90 evaluation studies on faculty instructional improvement met the inclusion criteria. Of those that were included in the review, the majority of research designs included quasi-experimental design, mixed method approaches, and one randomized design. Kucsera and Svinicki (2010) also noted descriptions of how data were analyzed was “somewhat vague and, consequently, raised uncertainty in how much confidence should be given to the study’s results” (p. 7).

Recommendations from systematic reviews of the literature of professional development have indicated the need to include different forms of inquiry than quantitative methods, experimental research methodologies, and “embrace more broadly based definitions of rigor currently being developed in education and other social science fields” (Kucsera & Svinicki, 2010, p. 8). Kucsera and Svinicki (2010) indicated there are characteristics of faculty development, such as developers working with constructs which do not lend well enough to scalable instrumentation, highly individualistic participants, and the contextualized nature of education, which might be better suited for qualitative inquiry methods.

Bamber and Stefani (2015) argued there is a need to “reconceptualize impact of professional development programs as ‘evidencing value’” to reframe the discourse and “release us from inadequate or instrumental approaches to evaluation” (p. 242). Changing the discourse involves avoiding the desire to focus on direct impacts of faculty development programs, but rather attempt to impute indirect impact and finding ways to

evidence outcomes. Harper and Nicolson, (2013) referred to these as soft outcomes, “such as raised awareness of the scholarship of teaching, increased levels of confidence, perhaps even improved collegiality. Such soft outcomes are much less amenable to meaningful measurement of “impact” but may have profound consequences in practice gains” (p. 244).

The need for rigorous evaluations of professional development activities has been well documented; however, the characteristics of professional development make it such that evaluating changes to instructors’ skills, attitudes, and beliefs, which then lead to changes in student performance, make it difficult to evaluate using experimental or quasi-experimental designs. This provides rationale as to why reviews of the literature on evaluation of professional development programs have pointed to a paucity of evaluations, which utilize quantitative methods to demonstrate the effect of professional development on student learning outcomes. This led to calls by those such as Kuscera and Svinicki (2010) and Bamber and Stefani (2015) to reconsider rigor of professional development evaluation as adhering to the criteria to be scientifically valid rather than particular methods and to select methods that are more appropriate given the unique characteristics and contexts of faculty development.

### **Evaluative Thinking**

Evaluative thinking (ET) is identified as a desired outcome of participatory evaluation models (Cousins & Whitmore, 1998); however, like participatory evaluation, evaluative thinking as a construct is ill defined. King (2007) included evaluative thinking as an element of evaluation capacity building (ECB) through process use, “Identify and

support evaluation champions who will nurture evaluative thinking in themselves and others” (p. 49). This step is further explained:

First, you must identify individuals who understand or intuitively get evaluation, that is, people who are willing to spend time with you discussing options, thinking about how to involve others, and eventually making sense of data. These may be your clients or those holding positional authority, but not necessarily. In every organization I have worked with, I have met people who simply enjoy the evaluation process, either because they understand it intuitively and are interested in learning more or because they have studied it somewhere, often in a degree program in the guides of research methods course (p. 49).

The description provides a set of activities and characteristics that can be identified in individuals who engage in ET; however, the descriptions do not provide a definition or a useful construct for ET.

Buckley, Archibald, Hargraves, and Trochim (2015) defined and operationalized ET for the purpose of ECB through a review of the literature on critical thinking. As a result of the literature reviewed ET was defined as:

Critical thinking applied in the context of evaluation, motivated by an attitude of inquisitiveness and a belief in the value of evidence, that involves identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection and perspective taking, and informing decisions in preparation of action (Buckley, Archibald, Hargraves, & Trochim, 2015, p. 378).

Recognizing that ET, like critical thinking, is a skill that must be developed Buckley, Archibald, Hargraves, and Trochim (2015) presented five guiding principles for promoting ET:

1. Promoters of ET should be opportunist about engaging learners in ET processes in a way that builds on and maximizes intrinsic motivation,
2. Promoting ET should incorporate incremental experiences, following the developmental process of “scaffolding,”
3. ET is not a born-in skill nor does it depend on any particular education background; therefore, promoters should offer opportunities for it to be intentionally practiced by all who wish to develop as evaluative thinkers,
4. Evaluative thinkers must be aware of – and work to overcome – assumptions and belief preservation, and
5. In order to best learn to think evaluatively, the skill should be applied and practiced in multiple contexts and alongside peers and colleagues (p. 380-381).

There still is a need for a construct of ET to be able to measure it.

Fierro, Codd, Gill, Pham, Grandjean Targos, and Wilce (2018) utilized a multi-step process to develop indicators of ET, including a priori indicator identification, document review, interviews, and focus groups. The result of the data- analysis process were five indicators of ET: 1) Reflecting, 2) Perspectives, 3) Projecting, 4) Valuing Evaluation, and 5) Use (Fierro, Codd, Gill, Pham, Grandjean Targos, & Wilce, 2018). Provided in Table 13 are definitions and descriptions of the five indicators. It should be noted this research does not present a validated construct, rather an attempt to “move the concept of ET

from the intangible to the tangible” (Fierro, Codd, Gill, Pham, Grandjean Targos, & Wilce, 2018, p. 69).

**Table 13**

*Indicators of Evaluation Thinking*

Indicator	Definition
Reflecting –	Deliberatively giving critical attention to various aspects of a program, including its context and its evaluation; suggests a willingness to apply a critical lens reflectively
Thoughtful Questions	Organizational environment/individual attitudes reflect a stance of “fearless inquiry.” Asking questions that look at the big picture; suggest an openness to discovery and a willingness to encounter negative findings and unintended consequences; examines deeply held beliefs.
Deeper Understanding	Ongoing tendency to seek insights and probe for further information and explanation about the evaluation findings.
Describing Thinking	Communicating ideas and concepts as well as the logic or thought processes on which they are based, including describing or defining words or situations in one’s own words, or illustrating a concept visually (e.g. logic model, TOC). This activity is both a learning tool (working out the connections, ideally with others) as well as an indication of internal (personal) comprehension and communication device.
Identifying Assumptions	Identifying and articulating underlying concepts and beliefs.
Considering Context	Demonstrating awareness and responsiveness to the prevailing context and to changes occurring in their environment. Program and evaluation designs are flexible and adaptive.
Evaluation Review	Program makes efforts to examine or judge the quality of its evaluation work and/or assess the performance of evaluators.
Perspectives. –	Incorporating information and priorities from multiple viewpoints
Multiple Perspectives	Soliciting a diverse range of stakeholder views and perspectives on the evaluand.
Additional Points of View	Consideration of views and perspectives on the evaluand beyond stakeholders who can directly engage.
Participatory Evaluation	Declarations of the value of/commitment to broad participation in evaluation activities beyond the evaluator and program leadership.
Explicating Values	Considerations of the values pertinent to the evaluation and can/should influence the evaluation process, the program

---

theory (e.g., outcomes). What constitutes credible evidence, and the findings/recommendations.

---

Projecting – Envisioning success and the path to achieving it

---

Criteria of Success	Clearly articulating criteria for success
Linking Activities to Outcomes	Articulating the intended/expected connections among various elements of the program theory or program implementation.
Scaling	Planning for the ultimate outcomes of a program, as with reference to state-wide or population-level outcomes or program/outcome sustainability.
Suite of Evaluation Activities	Evaluation activities are conducted as a suite of studies as opposed to individual events. While each evaluation has its unique focus, the intent is to gradually bring clarity to the bigger picture (e.g., pieces of a puzzle)

---

Valuing Evaluation – Statements indicating belief in the importance and utility of evaluation

---

Value of Evaluation	Statements indicating belief in the importance and utility of evaluation.
Value of Evidence	Judgements values, and assertions about the program are informed by findings attained through systematic inquiry and credible methodology. When such findings do not exist, efforts are made to attain them.
Intent to Engage in ECB	Concrete documented efforts or plans to design and implement teaching and learning strategies to: (1) help individuals, groups, and the organization learn about what constitutes effective, useful, and professional evaluation practice and (s) support sustainable evaluation practice through such things as changes to organizational norms or infrastructure (e.g., evaluation policies, technology)
Distributed Responsibility	People in a variety of roles at various levels of implementation are responsible for conducting evaluations and using the findings.

---

Use – Impact or intended impact of the evaluation on the evaluand, stakeholders, and/or society

---

Planning for Use	Evaluation planning includes explicit consideration for the ways in which evaluation findings will be used to support subsequently thoughts and actions about the program and who the users will be. May also include specific actions taken to lay the groundwork for use.
------------------	---

---

---

Instrumental Use	Instances where the evaluation findings were used to modify the program in some way that improves alignments with program theory, will assist in the achievement of outcomes, and/or will mitigate negative consequences. Frequency may vary; may occur with interim or final findings.
Integration	Evaluation is embedded into the routine practices of the organization. People in the organization expect to engage with each other in clarifying key concepts, examining the quality of evidence available about effectiveness, and supporting their opinions and judgements with evidence.
Process Use	Instances in which program staff (may have) learned about or made changes to the program as a result of the evaluation process rather than the evaluation findings.

---

*Note.* From “Evaluative Thinking in Practice,” by L. Fierro, H. Codd, S. Gill, P. Pham, P. Targos, and M. Wilce, 2018, *New Directions for Evaluation*, (158), pp. 60-64.

## CHAPTER 3 METHODS

### Design

The purpose of this mixed-methods study is to explore whether or not an evaluation model that has the narrowed focus of a traditional objectives-oriented evaluation model but implemented in participatory evaluation method can be considered participatory. An additional purpose of the study is to explore whether or not the evaluation model is able to achieve a desired benefit of participatory evaluations, promoting evaluative thinking in non-evaluative stakeholders, and thereby builds the evaluative capacity within an organization. Therefore, the questions addressed in this study are:

Q1. How does the participation of a diverse set of stakeholders influence the development of the evaluation framework?

Q2. When limiting the scope of a participatory evaluation using an objectives-oriented evaluation model as an advanced organizer, does the evaluation meet the criteria provided by Daigneault and Jacob (2009) using the Participatory Evaluation Measurement Instrument (PEMI) and Evaluation Involvement Scale (EIS) provided by Toal (2009)?

Q3. Does the use of a participatory evaluation model promote evaluative thinking with the non-evaluative stakeholders involved in the evaluation?

According to Johnson, Onwuegbuzie, and Turner (2007) “mixed methods research combines elements of qualitative and quantitative research approaches for the purposes of breadth and depth of understanding and corroboration” (p. 123). Creswell

and Plano Clark (2018) defined the role of the researcher in mixed methods research as one who:

- Collects and analyzes both qualitative and quantitative data rigorously in response to research questions and hypotheses,
- Integrates (mixes or combines) the two forms of data and their results,
- Organizes these procedures into specific research designs that provide the logic and procedures for conducting the study, and
- Frames these procedures within theory and philosophy (p. 5).

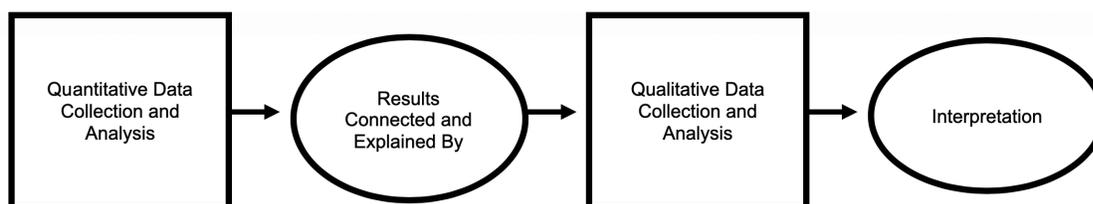
The benefit of mixed methods research is the combination of both quantitative and qualitative methods which provides opportunities for the limitations of one method to be compensated for by the strengths of the other method.

The mixed methods research design employed in this study is an explanatory sequential design, which consists of two distinct interactive phases illustrated in Figure 4 (Creswell & Plano Clark, 2018). The notation for this specific design is

QUAN → qual indicating there are two strands in the sequence, the quantitative method occurring first and receiving the greater emphasis in addressing this study's purpose, and the qualitative methods followed to help explain the quantitative results (Creswell & Plano Clark, 2018).

#### **Figure 4**

##### *Explanatory Sequential Design*



## Setting Description

Baker College is a private non-profit system of eight on ground campuses geographically dispersed across the state of Michigan and one virtual campus. The college is accredited by the Higher Learning Commission and offers programs that award certificates, associates, bachelors, masters, and doctoral degrees (HLC, 2019). The faculty population for Baker College consists of adjunct and full-time faculty with a ratio of approximately 1:11 full-time to part-time faculty (College Navigator, n.d.).

The First Year Faculty Experience (FYFE) program is used to onboard new faculty and provides them with the necessary preparation to teach their assigned course(s). The program is delivered via the Canvas Learning Management System and consists of required reading, pre-recorded videos and tutorials, and one required week of discussion board engagement. The program description is as follows:

During this five-day orientation you will learn about Baker College's professional expectations of faculty, experience our learning management system (Canvas), and familiarize yourself with academic operational practices typically performed as an instructor at Baker College. At the conclusion of this session, you will be able to:

- Navigate Canvas (the learning management system you are currently in)
- Integrate Baker College's policies and procedures
- Provide ideas about good instructional practices
- Locate resources that are available to you
- Use instructional resources to enhance learning
- Apply features within Canvas that offer effective communication and course engagement (Baker College FYFE, n.d.)

The FYFE has been in place since fall 2016 and has been identified as having to undergo a program evaluation as a component of a continuous quality improvement process within the Division of Academic Affairs.

## Mixed-Methods Design

### Phase One: Quantitative

#### Sampling Plan

A convenience sample was used, comprised of the stakeholders designing and implementing the program evaluation. Diversity in the evaluation team is not defined in demographic terms, but rather as inclusive of program stakeholders who are not traditionally included in program evaluation planning and implementation. The stakeholders were leadership, administrative staff, program officials who have faculty from their program complete the FYFE, adjunct faculty who completed the FYFE, and students whose instructors participated in the FYFE. Stakeholder types are categorized as traditional and non-tradition in Table 14. This sampling plan was applied to Phase I and Phase II of the mixed-methods design.

**Table 14**

*Traditional and Nontraditional Stakeholder Types*

Traditional Stakeholder	Non-Traditional Stakeholder
Program Leadership	Academic Program Directors
Program Staff	Adjunct faculty consuming the program
	Students of the adjunct faculty

#### Instrument

The quantitative data was collected using two instruments and one survey. The validated instruments developed to measure the level of participation in participatory evaluations are the Participatory Evaluation Measurement Instrument (PEMI) and the Evaluation Involvement Scale (EIS) (Daigneault & Jacob, 2009; Toal, 2009). The survey

was designed by the researcher to collect data on participants' confidence with evaluative thinking at three separate points throughout the process.

The PEMI is a nonnormative instrument, which measures the evaluation participants perception of participation on the three dimensions of Participatory Inquiry outlined by Cousins and Whitmore (1998) (Daigneault & Jacob, 2009; Daigneault & Jacob, 2014). The instrument was validated using the Instrument Development and Construct Validation process developed by Onwuegbuzie, Bustamante, and Nelson (2010), a 10-phase mixed-methods validation process (Daigneault & Jacob, 2014). Each dimension in the instrument has the same scale with different intuitive labels for the respective domain, 0 = .00, 1 = .25, 2 = .50, 3 = .75, 4 = 1.00. For an evaluation to be considered participatory it must receive at least an average of .25 on each domain. Finally, the scores from the three domains are averaged to measure overall participation (Daigneault & Jacob, 2014). Therefore, this instrument addresses the question of whether or not an evaluation can be considered participatory.

The second instrument used to collect quantitative data was the EIS. Using Messick's unitary concept of validity as a framework, the EIS was validated using data collected from a multi-site evaluation, with a resulting  $\alpha$  of .94, suggesting high internal consistency (Toal, 2009). The scale was developed using the evaluation key decision points posited in Burke (1998) as a general framework. Each key decision point was assigned to the three evaluation stages: 1) evaluation planning, 2) implementation, and 3) communication of results, as well as additional items the development team added to provide clarity or distinction to the decision points, resulting in thirteen item scale shown in Table 20 (Toal, 2009). Participants responded to each item rating the extent to which

they were involved in the different evaluation activities using a 1:4 scale: 1 = No; 2 = Yes, a little; 3 = Yes, some; 4 = Yes, extensively, or “I don't think this activity took place” (Toal, 2009, p. 354). The EIS addresses the question of how participatory the participatory evaluation was.

The third source of quantitative data will be a survey designed by the researcher to gather information from the participants about their confidence related to evaluative thinking. Using the five indicators of evaluative thinking established by Fierro, Codd, Gill, Pham, Grandjean Targos, and Wilce (2018) a 22-item survey was developed. Respondents rated their level of agreement with each statement using a five-point Likert-type scale (5 = Strongly agree, 4 = Agree, 3 = Unsure, 2 = Disagree, and 1 = Strongly disagree) (Fraenkel, Wallen, & Hyun, 2015).

### **Data Collection**

Two surveys were developed to collect the data. The survey developed to capture the respondents' confidence related to evaluative thinking represents the first survey. There were three submissions of this survey by the evaluation participants: 1) prior to the beginning of the evaluation, 2) following the completion of the evaluation training, and 3) following the conclusion of the evaluation. An additional survey was developed comprised of questions from the PEMI and the EIS. This survey was sent to the participants following the conclusion of the evaluation. The data was collected electronically using Qualtrics to email the surveys to the study participants and collect the results.

## Data Analysis

The results collected from the surveys were uploaded into an SPSS database. A separate form of analysis was completed on each data set. All statistical analyses was conducted at the nominal alpha level = 0.05.

### Participatory Evaluation Measurement Instrument (PEMI)

Descriptive statistics, including mean and standard deviations, was generated for each of the three domains of the PEMI. The Kruskal-Wallis test (Wilcox, 2011) was used to determine if there are differences in mean responses between the three domains represented in the PEMI.

$$H_0 : F_{(\text{Extent of Involvement})} = F_{(\text{Diversity of Participants})} = F_{(\text{Control of the Evaluation Process})}$$

$$H_1 : F_{(\text{Extent of Involvement})} \neq F_{(\text{Diversity of Participants})} \neq F_{(\text{Control of the Evaluation Process})}$$

According to Creswell and Plano Clark (2018), in an explanatory sequential design the results of the quantitative analysis are used to identify the results that need to be explored further through the use of qualitative methods. The results of the one-way analysis of variance provided evidence of differences between the three categories of the PEMI and was used for further exploration using qualitative methods.

### Evaluation Involvement Scale (EIS)

For each question on the EIS descriptive statistics, including mean and standard deviations, were generated for all question responses on the EIS. Additionally, the Sign Test was completed to calculate the median value of each question for the population. For a sample size  $n = 10$  the critical region is  $S \leq 1$  (Neave & Worthington, 1988).

$$H_0: = 3$$

$$H_1: \neq 3$$

## **Evaluative Thinking Questionnaire**

Descriptive statistics, including mean and standard deviations, were generated for all question responses on the questionnaire. To explore the change in confidence of the participants' evaluative thinking between the three repeated submissions of the questionnaire the Friedman's Test was used (Neave & Worthington, 1988).

$H_0$  = There are no differences in the means of the five different indicator categories of evaluative thinking (ET) between the three submissions of the survey

$H_1$  = There are some differences in the means of the five different indicator categories of evaluative thinking (ET) between the three submissions of the survey

## **Phase II: Qualitative**

### **Transcript Analysis**

The transcripts from the evaluation development sessions were analyzed to identify the unique contributions of the members of the evaluation team. The contributions from the members of the team who represent non-traditional stakeholders were used to respond to research question one: How does the participation of a diverse set of stakeholders influence the development of the evaluation framework? It is assumed that the unique contributions made by each member of the group would not have been made if they were not present.

### **Interview Protocol Development**

The interview protocol was developed to complete semi-structured interviews with the members of the evaluation planning committee to explain their responses on

the two instruments and their responses to the survey on evaluative thinking. The analysis of the quantitative data was used to identify which results are significant and needed to be further explained through qualitative data collection (Creswell & Plano Clark, 2018). Additionally, the findings from the transcript analysis were used in the development of the interview protocol to further explain the changes of the evaluative thinking survey results across the three submissions.

### **Data Collection**

The interviews with the research participants were conducted in person or virtually via WebEx. The interviews were recorded and transcribed by the researcher.

### **Data Analysis**

A general inductive analysis approach was used for analyzing the data collected from the semi-structured interviews and the evaluation planning sessions. Thomas (2006) posited “the primary purpose of the inductive approach is to allow research findings to emerge from the frequent, dominant, or significant themes inherent in raw data, without the restraints imposed by structured methodologies” (p. 238). Provided in Table 15 is an overview of the inductive approach identifying the purpose, analytic strategies, and analytic tasks.

**Table 15***General Inductive Analysis Approach*

Purpose	Analytic Strategies	Analytic Tasks
1. Condense raw text data into brief, summary formats	1. Multiple readings and interpretations of the raw data.	a) Rigorous reading and coding of documents/transcripts to allow major themes to emerge
2. Establish links between research objectives and summary findings derived from raw data	2. Categories are identified from the raw data into a framework or mode with key themes and processes identified	a) Identify text segments related to research questions b) Label text segments (categories) c) Create a model incorporating most categories
3. Develop model or theory about underlying structure evident in the data	3. Multiple interpretations are made from the raw data resulting in findings	a) Similarities across groups explored as applicable b) Summary of findings resulting from following analytic strategies described

*Note.* From *A Case Study of The Impact of a Systematic Evaluation Process in a Graduate Medical Education Residency Program* (Ph.D., Wayne State University). By H. Kromrei, 2014, p. 55

(<http://search.proquest.com/docview/1531330752/abstract/1B60B9C01F6748F4PQ/1>)

Reliability is defined as responding to the question “can the results of the research be checked independently?” (p. 42). In the field of naturalistic inquiry, Beuving and Vries (2014) defined validity as responding to the question “does the research measure what it claims to measure?” (p. 42). Addressing issues of reliability and validity in qualitative research is to address concerns of subjectivism in the collection and interpretations of qualitative data. Beuving and Vries (2014) identified four tools that can

be used in naturalistic inquiry to limit subjectivism, promote transparency, and ensure validity and reliability of the research findings: 1) use of systematic procedures employed in grounded theory, 2) triangulation of findings through multiple data collection methods, 3) documenting theoretical reflections about the data, and 4) member checks with the people whom are providing the data. Table 16 lists these four tools as well as strategies for implementing the tools to address concerns of validity and reliability, all of which were used in this study.

**Table 16**

*Tools to Address Validity and Reliability in Naturalistic Inquiry*

Tool	Description	Strategies
Systematic procedures employed in grounded theory	Employing the twin procedures of comparison and open coding.	Comparison takes place when new evidence is compared against existing evidence.  Open coding is sharing or making coding available to broader academic community and open to verification
Triangulation of findings through multiple data collection methods	Triangulation is confronting the same empirical situation with different research methods through iteration.	Ask questions to already collected material, formulating new propositions in new phase of data collection.  Contrast new propositions with propositions already formulated
Documenting theoretical reflections about data	Note taking and journaling to document and confront theoretical reflections as a means of confronting the researchers own predispositions and inclinations.	Consistently adopt a self-critical stance/reflexive attitude.  Make particular views about society explicit and treat them as testable

hypotheses.

Member checks

Share findings and interpretations with the people involved in them to check your progressive understanding of the situation and the people involved in them.

Following interviews share the summary of the interview with the interviewee to gain their perspective on its accuracy.

Share the analysis and conclusions of the findings with the informants.

**Table 17**

*Summary of Research Questions, Variables, Collection, and Analysis*

Research Questions	Variables/ Key Factors	Sample/ Participants	Data Collection	Data Analysis
Q1. How does the participation of a diverse set of stakeholders influence the development of the evaluation framework?	Unique contributions by members of the evaluation team to the development of the evaluation plan	All members of the evaluation team	Recording and transcription	Qualitative
Q2. When limiting the scope of a participatory evaluation using an objectives-oriented evaluation model as an advanced organizer, does the evaluation meet the criteria provided by Daigneault and Jacob (2009) using the Participatory Evaluation Measurement Instrument (PEMI) and Evaluation	PEMI, EIS	All team members	Survey, interview	Quantitative: Descriptive statistics, Kruskal-Wallis one-way analysis of variance by ranks  Qualitative: Analysis of semi-structured interview using general inductive

Research Questions	Variables/ Key Factors	Sample/ Participants	Data Collection	Data Analysis
Involvement Scale (EIS) provided by Toal (2009)?				analysis
Q3. Does the use of a participatory evaluation model promote evaluative thinking with the non-evaluative stakeholders involved in the evaluation?	Evaluative thinking survey results over 3 administrations	All team members	Survey, interview	Descriptive statistics, Friedman's Test  Qualitative: Analysis of semi-structured interview using general inductive analysis

## Chapter 4 Results

### Phase I: Quantitative

#### Participatory Evaluation Measurement Instrument

Following the completion of the program evaluation, participants completed the Participatory Evaluation Measurement Instrument (PEMI) (Daigneault & Jacob, 2009) via Qualtrics (ver. XM). All participants ( $N = 9$ ) completed the PEMI, rating the extent of their involvement in the First Year Faculty Experience (FYFE) program evaluation in the Extent of Involvement, Diversity of Participants, and Control of the Evaluation Process domains. The responses were converted to a numerical scale ranging from .00 to 1.00. SPSS (ver. 25) was used to generate descriptive statistics and complete analysis of the data. Cronbach's alpha was calculated to assess the reliability of the PEMI. The instrument was found to have acceptable reliability (3 items;  $\alpha = .76$ ).

The mean response for the Extent of Involvement domain ( $M = .6944$ ,  $SD = .1667$ ) indicates participants opined the extent of involvement was between moderate involvement and substantial/strong involvement. The mean response for the Diversity of Participants domain ( $M = .6667$ ,  $SD = .2165$ ) indicates the diversity of the evaluation participants was between moderate diversity and substantial/strong diversity. The mean response for the Control of the Evaluation Process domain ( $M = .5278$ ,  $SD = .2319$ ) indicates the control of the evaluation process was shared between the nonevaluative participants and the evaluator.

The Kruskal-Wallis test was used to determine if there were statistically significant differences in mean responses between the three domains represented in the PEMI. All analyses here and below were completed with nominal  $\alpha = 0.05$ . There

was not a statistically significant difference between the mean rating for the three domains of the PEMI ( $H(2) = 2.506, p = .286$ ) with a mean rank of 16.39 for the Extent of Involvement domain, 14.56 for the Diversity of Participants domain, and 11.06 for the Control of the Evaluation Process domain.

**Table 18**

*PEMI Descriptive Statistics and Mean Ranks*

Domain	N	M	SD	Min	Max	M Rank
Involvement	9	.6944	.1667	.50	1.00	16.39
Diversity	9	.6667	.2165	.50	1.00	14.56
Control	9	.5278	.2319	.00	.75	11.06

**Table 19**

*Kruskal-Wallis Test Statistics*

Kruskal-Wallis H	2.506
df	2
Asymp. Sig.	.286

Based on the results of the analysis of the PEMI the following questions have been generated and added to the semi-structured interview protocol for the qualitative phase of the study to further explain the responses that participants provided:

1. Reflecting on the level of membership in the evaluation, what were the considerations for you when selecting your response?
2. Reflecting on the diversity of participants in the evaluation, what were the considerations for you when selecting your response?
3. Reflecting on the control over the evaluation process during the evaluation, what were the considerations for you when selecting your

response?

### **Evaluation Involvement Scale**

Following the completion of the program evaluation, participants completed the Evaluation Involvement Scale (EIS) (Toal, 2009) via Qualtrics (ver. XM) All participants ( $N = 9$ ) completed the EIS indicating the extent they participated in the thirteen evaluation activities listed in the instrument, on a scale of 1 = No involvement to 4 = Yes, extensive involvement (Toal, 2009). The descriptive statistics for each item on the instrument are provided in Table 20. Cronbach's alpha was calculated to assess the reliability of the EIS. The instrument was found to have good reliability (13 items;  $\alpha = .83$ ).

Although the Sign test is not available in SPSS, the one sample Wilcoxon Signed-Rank Test, a functional equivalent, was used to analyze the results of the EIS to identify which activities on the instrument resulted with ratings that were statistically significant (Neave & Worthington, 1988). The hypothesized median rating for each activity was three. Six activities had ratings that were statistically significant: 1) EIS2: Identifying evaluation planning team members ( $M = 1.56$ ,  $SD = .726$ ,  $p = .009$ ), 2) EIS4: Developing data collection instruments ( $M = 2.22$ ,  $SD = .833$ ,  $p = .038$ ), 3) EIS10: Writing evaluation reports ( $M = 1.67$ ,  $SD = 1.00$ ,  $p = .015$ ), 4) EIS11: Reviewing evaluation reports for accuracy and/or completeness ( $M = 1.56$ ,  $SD = 1.014$ ,  $p = .012$ ), 5) EIS12: Presenting evaluation findings ( $M = 1.22$ ,  $SD = .667$ ,  $p = .005$ ), and 6) EIS13: Developing future project plans based on evaluation results ( $M = 1.89$ ,  $SD = 1.167$ ,  $p = .028$ ).

**Table 20***EIS Descriptive Statistics and Levels of Significance*

Item	N	Min	Max	M	SD	Sig.
EIS1	9	2	4	3.11	.601	.564
EIS2	9	1	3	1.56	.726	.009
EIS3	9	1	3	2.33	.866	.063
EIS4	9	1	3	2.22	.833	.038
EIS5	9	1	3	2.67	.707	.180
EIS6	9	1	4	2.67	1.000	.276
EIS7	9	2	3	2.78	.441	.157
EIS8	9	2	4	2.89	.782	.655
EIS9	9	1	4	2.78	1.093	.458
EIS10	9	1	4	1.67	1.000	.015
EIS11	9	1	4	1.56	1.014	.012
EIS12	9	1	3	1.22	.667	.005
EIS13	9	1	4	1.89	1.167	.028

A review of the individual submissions for the items with significant results demonstrates that these activities were rated at higher levels than the other activities as “No involvement” or “Yes, a little involvement,” by the participants. A sum of these two ratings for each item on the instrument is provided in Table 21.

**Table 21**

*EIS Items and Sum of No or Yes, a little Ratings*

Item	No or Yes, a little Ratings
EIS1	1
EIS2	8
EIS3	4
EIS4	5
EIS5	2
EIS6	2
EIS7	2
EIS8	3
EIS9	2
EIS10	8
EIS11	8
EIS12	8
EIS13	6

Based on the responses to the Evaluation Involvement Scale, the items with ratings that stood out as being significant were:

- Identifying evaluation planning team members
- Developing data collection instruments
- Writing evaluation reports
- Reviewing evaluation reports for accuracy and/or completeness
- Presenting evaluation findings (e.g., to staff, to stakeholder, to an external audience)
- Developing future project plans based on evaluation results
- Looking over your responses to the instrument, what were the factors you considered when providing your responses to these questions?

To determine the extent of involvement for each task included within the instrument the coding scheme posited by Toal (2009) is applied based on the mean rating:

- 1 = No Involvement
- 1.01 – 2.00 = Little Involvement
- 2.01 – 3.00 = Some Involvement
- 3.01 – 4.00 = Extensive Involvement (p. 355)

The coding scheme for each task is provided in Table 22 based on the responses from the study participants.

**Table 22***Coding Scheme Applied to EIS Results*

Item	Involvement Item	M	Involvement Level
EIS1	Discussions that focused the evaluation	3.11	Extensive
EIS2	Identifying evaluation planning team members	1.56	Little
EIS3	Developing the evaluation plan	2.33	Some
EIS4	Developing data collection instruments	2.22	Some
EIS5	Developing data collection processes	2.67	Some
EIS6	Collecting data	2.67	Some
EIS7	Reviewing collected data for accuracy and/or completeness	2.78	Some
EIS8	Analyzing data	2.89	Some
EIS9	Interpreting collected data	2.78	Some
EIS10	Writing evaluation reports	1.67	Little
EIS11	Reviewing evaluation reports for accuracy and/or completeness	1.56	Little
EIS12	Presenting evaluation findings (e.g., to staff, stakeholders, an external audience)	1.22	Little
EIS13	Developing future project plans based on evaluation results	1.89	Little
Overall		2.26	Some

**Evaluative Thinking Survey**

The Evaluative Thinking Survey was completed by each participant of the FYFE program evaluation three times during the study: 1) Before completing the program evaluation tutorial, 2) after completing the program evaluation tutorial, and 3) after completing the FYFE program evaluation. Participants responded to statements, rating their level of agreement on a five-point Likert scale, with the ends of the scale defined as 1 = Strongly disagree and 5 = Strongly agree. The statements were focused on five domains of evaluative thinking posited by Fierro, Codd, Gill, Pham, Grandjean Targos, and Wilce (2018): Reflecting, Perspectives, Projecting, Valuing Evaluation, and Use.

Cronbach's alpha was calculated to assess the reliability of the Evaluative Thinking Survey. The survey was found to have good reliability (22 items;  $\alpha = .86$ ).

Friedman's Test was used to explore the between the mean rating for each domain across the repeated submissions to explore the change in confidence of the participants' evaluative thinking. Provided in Table 23 are descriptive statistics for each of the three submissions of the Evaluative Thinking Survey as well as the mean rank. For the Reflecting domain, the Friedman's Test of differences among the repeated measures rendered a Chi-square value of 7.280 which was significant ( $p = 0.026$ ). For the Perspectives domain the Friedman's Test of differences among the repeated measures rendered a Chi-square value of 7.467 which was significant ( $p = 0.024$ ). For the Projecting domain the Friedman's Test of differences among the repeated measures rendered a Chi-square value of 8.706 which was significant ( $p = 0.013$ ). For the Valuing Evaluation domain the Friedman's Test of differences among the repeated measures rendered a Chi-square value of 7.697 which was significant ( $p = 0.021$ ). For the Use domain the Friedman's Test of differences among the repeated measures rendered a Chi-square value of 7.818 which was significant ( $p = 0.020$ ).

**Table 23***Evaluative Thinking Survey Descriptive Statistics and Mean Ranks*

Domain	N	M	SD	Min	Max	M Rank
RMean1	9	4.0556	.39965	3.50	4.83	1.50
RMean2	9	4.2778	.47871	3.50	5.00	1.94
RMean3	9	4.444	.5270	3.5	5.0	2.56
PMean1	9	3.9444	.55590	3.25	5.00	1.33
PMean2	9	4.3333	.48412	3.75	5.00	2.22
PMean3	9	4.5000	.41458	3.75	5.00	2.44
PRMean1	9	3.7222	.55120	2.75	4.25	1.22
PRMean2	9	4.1667	.58630	3.25	5.00	2.33
PRMean3	9	4.2778	.61802	3.25	5.00	2.44
VMean1	9	3.8889	.61379	3.25	4.75	1.28
VMean2	9	4.2778	.55120	3.50	5.00	2.33
VMean3	9	4.3611	.43501	4.00	5.00	2.39
UMean1	9	3.5556	.54167	2.50	4.25	1.28
UMean2	9	4.0556	.67056	3.00	5.00	2.28
UMean3	9	4.1389	.61379	3.00	5.00	2.44

**Table 24***Friedman's Test Statistics*

Domain	N	Chi-Square	Degrees of Freedom	Sig.
Reflecting	9	7.280	2	.026
Perspectives	9	1.467	2	.024
Projecting	9	8.706	2	.013
Valuing Evaluation	9	7.697	2	.021
Use	9	7.818	2	.020

Based on the results of the analysis of the Evaluative Thinking Survey, the following questions were generated and added to the semi-structured interview protocol

for the qualitative phase of the study to further explain the responses that participants provided:

1. On the Evaluative Thinking Survey, look over the statements listed in the “Reflecting” section. Based on your responses, how would you summarize how your thinking has changed in this area based on your experiences participating in the program evaluation?
2. On the Evaluative Thinking Survey, look over the statements listed in the “Perspectives” section. Based on your responses, how would you summarize how your thinking has changed in this area based on your experiences participating in the program evaluation?
3. On the Evaluative Thinking Survey, look over the statements listed in the “Projecting” section. Based on your responses, how would you summarize how your thinking has changed in this area based on your experiences participating in the program evaluation?
4. On the Evaluative Thinking Survey, look over the statements listed in the “Valuing Evaluation” section. Based on your responses, how would you summarize how your thinking has changed in this area based on your experiences participating in the program evaluation?
5. On the Evaluative Thinking Survey, look over the statements listed in the “Use” section. Based on your responses, how would you summarize how your thinking has changed in this area based on your experiences participating in the program evaluation?

## **Phase II: Qualitative**

For Phase II, follow-up interviews were held individually with the members of the evaluation team after completing the program evaluation ( $N = 9$ ). The interviews were conducted utilizing a semi-structured interview protocol comprised of the follow-up questions identified in Phase I as a result of the quantitative data analysis. The interviews were recorded and transcribed by the researcher.

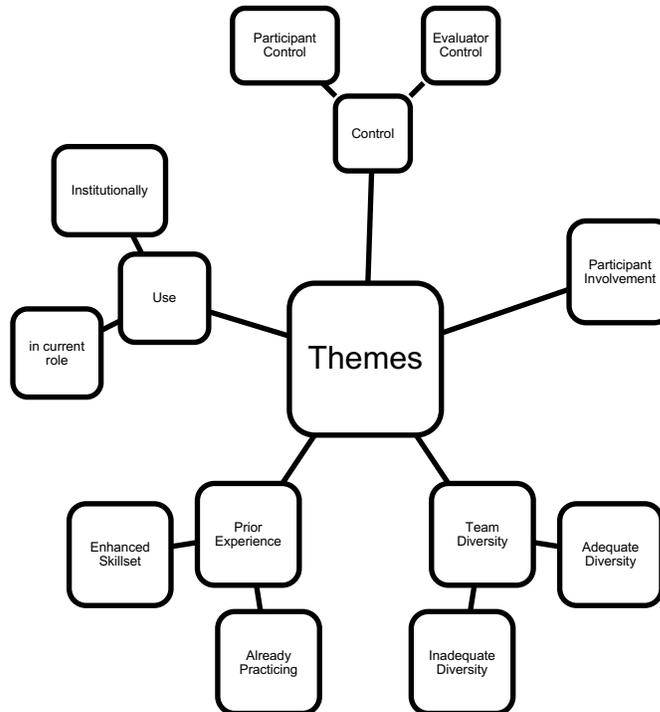
A general inductive approach (Thomas, 2006) was employed for analyzing the transcripts. Three transcripts were initially analyzed to develop a code book with descriptions and examples for each theme that would then be used for coding of the remaining transcripts. The transcripts and code book were shared with two additional coders to validate the codes and demonstrate reliability of coding. Coder #1 results were used as confirmation of the researchers coding and Coder #2 results were used to identify additional follow-up coding for the researcher and Coder #1.

From the analysis five major themes were identified:

1. Control of the evaluation process,
2. Participant involvement,
3. Evaluation team diversity,
4. Participants' prior experience, and
5. Use

**Figure 5**

*Major Themes and Subthemes Represented in Qualitative Data*



Provided in Figure 5 is a visual of the major themes and subthemes that emerged from the qualitative data.

### **Control**

The control theme emerged as a discussion of the role that the trained evaluator fulfilled throughout the evaluation process and shared control with the participants. Two subthemes were identified through the analysis, 1) evaluator control and 2) participant control. Control was discussed in context of the selection of the members of the evaluation team, leading of the discussion, the role of the evaluator as guide versus decision maker, and assignment of evaluation tasks.

Some responses which indicated control by the evaluator identified where the evaluator increased presence of the conversation to retain focus, such as participant 001 stating, “You were guiding the discussion.” Or Participant 007 stating, “If we had grandiose ideas or something, you kind of redirected, so we stayed on task rather than completely rewriting the FYFE. You kept us on track as far as evaluating rather than recreating.” Whereas examples of statements where participants indicating they felt they had control over the process were Participant 009 stating, “At some points we had notes and feedback where people kind of were focusing on an asset or aspect that wasn’t really part of what we’re doing, but we talked about it anyways, which was really great.” Or Participant 006 stated, “I think this felt like a true committee group that was brought together to do this work without any oversight necessarily from an upstream stakeholder. This really just felt like it was a very organic process that we were educated on and then allowed to actually practice without anyone else’s agenda.”

### **Participant Involvement**

The participant involvement theme emerged as a discussion of the various activities that the participants completed throughout the evaluation and the level of involvement in the process. The types of activities that were identified through the follow-up interviews were collecting data, reviewing data, collaborating with other members of the team to complete data collection, and providing feedback to other members of the team.

### **Team Diversity**

The team diversity theme emerged as a discussion about the adequacy of the diversity of the roles represented on the evaluation team as well as academic

backgrounds. This theme had two subthemes of adequate diversity and inadequate diversity. Participants who identified the diversity of the team as adequate felt that there was adequate representation of the different roles of individuals within the Academics division of the college. Participant 002 stated “I saw the group was made out of people, which represented various parts of the organization. So, there were people deeply involved in faculty development, but then there were Directors of Academic Affairs, Deans, faculty, adjunct faculty, so I thought that we had brought together good diversity of perspectives.”

Participants that identified the diversity as inadequate identified desires for increased diversity such as different divisions within the college, such as Student Affairs, OneStop, or students be involved. Inadequate diversity was also identified as part-time versus full-time employees. Participant 001 stated, “There were multiple faculty members, and so that is why I said moderate. You know, that it wasn’t incredibly diverse with one person from each role.” Or Participant 007 stated, “I was taking into account the different roles of people on the team. Okay, so that is moderately diverse. I felt as the one of the lone part-time people and non-program directors. That I was out of the norm rather than in the norm.”

### **Prior Experience**

The prior experience theme emerged as a discussion of how the participants’ prior experiences related to their perception of their evaluative thinking skills and development of evaluation skills as a result of participating in the program evaluation. This theme had two subthemes: 1) already practicing, and 2) enhanced skillset. Already practicing pertained to participants who identified, that due to their current role

or their background, they were already practicing certain aspects of evaluative thinking or applying a related skill. Examples of these types of statements were from participant 008 who stated, "Given the nature of the positions I've held in the past, I felt like a lot of this I have already done and kind of understand." Or, Participant 003 stated, "I was drawing my answers on other similar situations that I think I asked questions that are thoughtful. So, I was calling on a pattern of what I believe to be true in other work situations."

Enhanced skillset pertained to participants expressing that they developed new skills or that their evaluative thinking/skills had improved as a result of participating in the program evaluation. Responses representative of this theme were ones such as Participant 009 stated, "Thinking about my own field, I was like, Oh, I do quite a lot of this within my own scholarship and my own classrooms. I think that has changed a lot as I realized that these were relating back to program evaluation." Or Participant 006 stated, "I think it was the learning part, the true learning about the process of evaluation. I've been involved in evaluating things before and assessment."

## **Use**

The use theme emerged as a discussion of how participants identified ways or desires to increase the use or prevalence of program evaluation. This theme had two subthemes: 1) use in their role within the institution, and 2) desire for increased use as an institutional practice. Use in their current role within the institution pertained to the participant's desire or expressed ideas to practice program evaluation as a regular routine within their own program or department to collect data and use it for reflection on program or process improvements. Responses which represent this subtheme

were ones such as 003 stated, “I definitely, moving forward, can see myself saying, I’m trying to make a determination about this. So how do we want to measure that, and what data points can we pull to start making a decision on it?” Or Participant 008 stated, “My thought has changed from, I don’t have time to try to integrate this, to oh yes, I can incorporate this into smaller components that will eventually result in me being able to evaluate pieces into my current position.”

The desire for increased use as an institutional practice pertained to the participants expressing a desire to follow a similar process to explore program changes, justify policy decisions, and evaluate current practices for better alignment with institutional outcomes. Examples of statements which represent this subtheme are 002 stated, “[The organization] should not do evaluation just for the purpose of doing it, but you want to utilize that for future projects and plans, and you want to use it to improve them.” Or 006 stated, “I do feel like the organization needs to focus on this if we are truly going to make data driven decisions that we do need to build capacity and make this part of our culture process. We are in constant change, and I don’t know that we even collect data long enough to know what we’re doing with it before we change again.”

To respond to the first research question “Q1: How does the participation of a diverse set of stakeholders influence the development of the evaluation framework?” the initial evaluation planning meeting was recorded and transcribed. The members of the program evaluation team were categorized as either being a traditional non-traditional member of an evaluation team, shown in Table 25.

**Table 25***Traditional and Non-Traditional Evaluation Team Members*

Traditional	Non-Traditional
003 – Faculty Developer, Program Facilitator	001 – Program Director, HUS
006 – Campus Director of Academic Affairs	002 – Dean, C&IT
008 - Faculty Developer, Program Facilitator	004 – Faculty, HS
	005 – Director, Instructional Design
	007 – Practicum Coordinator, HUS
	009 – Program Director, Eng

The transcript was divided into the four primary activities of development of the program evaluation: 1) identifying stakeholders, 2) developing a logic model, 3) defining program objectives, and 4) developing the program evaluation framework (Madaus & Stufflebeam, 1989). A general inductive approach (Thomas, 2006) was employed for analyzing the transcripts for each of the four activities, paying attention to the unique contributions from members of the program evaluation team identified as non-traditional members.

From the identifying stakeholders activity one theme emerged from the unique contributions of the non-traditional participants, which was expanded comprehensiveness of the identified stakeholders. Participant 004 identified the following downstream stakeholders, which were not identified by other members of the team: existing faculty members, parents of students enrolled in the college, and various advisory boards or boards of education. Participant 002 identified the College Provost as an upstream stakeholder. Participant 001 identified student focused organizations within the college as downstream stakeholders, such as Student Affairs and OneStop.

From the logic model development activity one theme emerged from the unique contributions of the non-traditional participants, which was an increase in context and climate of the organization.

Participant 007 identified the delivery process – The process for how it is going to work out. Like, when is registration, who's responsible? When does the first module get delivered? Just the whole process.

Participant 001 identified the need to promote an awareness of resources – A lot of faculty don't have familiarity with the Center for Teaching Excellence. The faculty will be familiar with the faculty developers.

Participants 002 and 004 discussed the importance of student persistence as an outcome – 004 - Students feeling positive about the instructor. 002 – Yes, I put graduating and they move on.

From the analysis of the final two activities, defining program objectives and developing the evaluation framework, the theme organizational learning emerged. The non-traditional participants increasingly asked clarifying questions, questioned assumptions, and promoted discourse around clarifying the program objectives and the selection of the evaluation indicators.

Example 1:

Participant 004 – Navigating Canvas, this will be including the features of the Canvas for communication?

Participant 002 – Or familiarity with it?

Participant 004 – Well, so navigate Canvas, that will be including like introducing the options for communication. So, you will be able to navigate Canvas. This will include the communication strategies inside the Canvas.

Participant 005 – Yeah, watch your verb. So navigating is just really clicking around. And you know how to get to find things such as the grade book and discussion threads.

Example 2:

Participant 005: Yeah, but what does effectively mean?

Participant 007 – Right, because I'm thinking with effective communication that would naturally mean they're going to use it effectively to promote student engagement and good communication.

Participant 001 - Yeah, I guess to me, effective communication is so vague.

Participant 002 – There can be feedback. When you build the course you've been through, do certain things, which would want to clarify what you mean by effective.

Example 3:

Participant 007 – I know that quantitative is helpful, because you could use the SpeedGrader or the clickable rubrics and not put any comments. But is that actually teaching effectively or using it effectively? You're just going through and clicking five out of ten. You don't tell them why. Yeah, they used it, it would check as being done. Yeah you posted in discussion, but you did what I call virtual high fives, "Great comment student." Is that effective? So, if we're talking effective, I don't think you can quantitatively count it.

Example 4:

Participant 002 – And applying the [policies and procedures]

Participant 001 – What do we mean by applying? What are we integrating them into?

Participant 005 – No idea.

After a list of policies and procedures is comprised

Participant 005 – Aren't we really asking if faculty can follow these policies and procedures? That they really can, I don't think I mean, you know, I'm going to say understand. I don't know how you're going to measure that so.

## CHAPTER 5 DISCUSSION AND CONCLUSION

The primary purpose of this research study was to explore whether a program evaluation design which borrows the use of program objectives as an advanced organizer to guide the development of an evaluation framework, while using a participatory-oriented model, could draw on the benefits of both models. The secondary purpose was to determine if this evaluation method has the additional benefit of promoting evaluative thinking by non-evaluative stakeholders who take part in the program evaluation. The research questions for this study were:

1. How does the participation of a diverse set of stakeholders influence the development of the evaluation framework?
2. When limiting the scope of a participatory evaluation using an objectives-oriented evaluation model as an advanced organizer, does the evaluation meet the criteria provided by Daigneault and Jacob (2009) using the Participatory Evaluation Measurement Instrument (PEMI) and Evaluation Involvement Scale (EIS) provided by Toal (2009)?
3. Does the use of a participatory evaluation model promote evaluative thinking with the non-evaluative stakeholders involved in the evaluation?

A mixed-methods explanatory sequential design (Creswell & Plano Clark, 2018) was employed. The first phase was the collection and analysis of quantitative data in the form of survey responses by the participants who took part in the program evaluation. The surveys were comprised of an evaluative thinking survey developed by the researcher and administered three times throughout the duration of the study. A second survey was administered, and it was comprised of questions from the

Participatory Evaluation Measurement Instrument (PEMI) (Daigneault & Jacob, 2009) and the Evaluation Involvement Scale (EIS) (Toal, 2009). Following the analysis of the data, a semi-structured interview protocol was developed. This protocol was employed in Phase II to collect qualitative data to further explain the significant and non-significant findings from the quantitative data analysis.

The study was conducted during the Fall 2019 semester and involved a diverse set of stakeholders who came together to complete a program evaluation of the First Year Faculty Experience (FYFE) program at Baker College. The evaluation team was comprised of nine participants who were nominated by both campus and system leaders, the researcher, as well as from recommendations by other participants in an attempt to represent as many stakeholders as possible on the evaluation team. Prior to the start of the program evaluation, participants completed an asynchronous tutorial on program evaluation created by the researcher. In an attempt to minimize bias in the curriculum, the tutorial was based on the Framework for Program Evaluation available from the Centers for Disease Control (2018).

The program evaluation consisted of three meetings. The first meeting was held in-person and was comprised of four activities: 1) defining program stakeholders, 2) developing a logic model for the program, 3) defining program objectives, and 4) developing the program evaluation framework. Two subsequent meetings were held virtually via WebEx to review the data collected and discuss the reporting and communication of the evaluation results.

## **Discussion**

### **Development of Evaluation Framework**

The first question analyzed in this study was “How does the participation of a diverse set of stakeholders influence the development of the evaluation framework?” To answer this question, the transcript from the first meeting was analyzed using a general inductive analysis approach. Participants were identified as being traditional or non-traditional based on whether someone in their role within the hierarchy of the organization would typically be included in the program evaluation based on their relationship with the program. The transcript was divided into the four different activities completed during the first meeting and the contributions of the non-traditional participants were analyzed for major themes.

The themes which emerged were 1) an expanded comprehensiveness of the identified stakeholders, 2) an increase in context climate awareness of the organization, 3) the non-traditional participants promoted the asking of clarifying questions, questioning assumptions, and promoted discourse around clarifying the program objectives and selection of evaluation indicators. These findings reinforce the merits that evaluation methods represented in the Use branch of the evaluation theory tree are posited to provide, i.e. organizational learning, better alignment between the evaluation and the organizational context, and likely use of the evaluation results (Cousins & Whitmore, 1998; Fitzpatrick, Sanders, & Worthen, 2011; Alkin & Christie, 2004).

The second question was “When limiting the scope of a participatory evaluation using an objectives-oriented evaluation model as an advanced organizer, does the evaluation meet the criteria provided by Daigneault and Jacob (2009) using the

Participatory Evaluation Measurement Instrument (PEMI) and Evaluation Involvement Scale (EIS) provided by Toal (2009)?” The requirement posited by Daigneault and Jacob (2009) for an evaluation to be considered participatory is that the three domains of the PEMI: Extent of Involvement, Diversity of Participants, and Control of the evaluation process, each receive an average rating of at least .25, on a scale of 0.00 – 1.00, when completed by the evaluation team participants. The mean responses from the evaluation participants for each domain were: Extent of Involvement ( $M = .6944$ ), Diversity of Participants ( $M = .6667$ ), and Control of the Evaluation Process ( $M = .5278$ ), indicating that the evaluation was considered to be participatory.

Although there were no statistically significant differences between the mean ratings of the three domains, interview questions were added to the interview protocol to further understand how the participants were interpreting the concepts of involvement, diversity, and control as they relate to the program evaluation and why they selected the rating for each. Participant involvement, team diversity, and control emerged as major themes from the qualitative analysis. Across participants, there was a strong consensus that they felt they were involved in the evaluation process by their engagement in the various evaluation activities. Examples shared of activities included developing data collection instruments, data collection, interpreting results, reviewing work of other participants, collaborated with other participants, and engaged in conversations. No participant responded they were involved in all of the evaluation activities, or none of the evaluation activities. Instead, expressed that they contributed where they saw an opportunity and where they felt comfortable.

The theme of team diversity emerged from responses and descriptions of how the individual participants reported their thinking of diversity of the team as it pertained to the roles represented by the makeup of the team. The roles represented on the team were director of academic affairs, dean, program director, faculty developer, instructional designer, faculty, and practicum site coordinator. Those who reported feeling that there was adequate diversity on the team consistently reported they felt it was representative of the different roles within the Academics division of the college. Those that felt there could have been more diversity of roles on the team identified students and other divisions within the college, such as Student Affairs, as other roles that would have been appropriate to include. Additional ways in which diversity was discussed were the number of individuals representing each of the roles, full-time and part-time employees, and academic/professional background.

The theme of control emerged from responses discussing the level of control shared between the trained evaluator and the non-evaluative participants. Control was frequently expressed as being shared between the two roles. The various ways in which the non-evaluative participants' control was described were not limited but also not substantial, an ability to access materials, the ability to pick the task they would need to complete and who they collaborated with, and there being a lack of a sense of oversight. The ways in which the evaluator's level of control was described were guiding the process and discussion, keeping the team on task, and helping the team make decisions.

The results of the analysis of the EIS (Toal, 2009) resulted in each evaluation activity item being coded as a little involvement, some involvement, or extensive

involvement. The overall mean for the instrument ( $M = 2.26$ ) indicates that the overall evaluation experience for the participants was some involvement. Additionally, the one sample Wilcoxon Signed-Rank Test resulted in six activities having ratings being statistically significant: 1) Identifying evaluation planning team members ( $M = 1.56$ ,  $SD = .726$ ,  $p = .009$ ), 2) Developing data collection instruments ( $M = 2.22$ ,  $SD = .833$ ,  $p = .038$ ), 3) Writing evaluation reports ( $M = 1.67$ ,  $SD = 1.00$ ,  $p = .015$ ), 4) Reviewing evaluation reports for accuracy and/or completeness ( $M = 1.56$ ,  $SD = 1.014$ ,  $p = .012$ ), 5) Presenting evaluation findings ( $M = 1.22$ ,  $SD = .667$ ,  $p = .005$ ), and 6) Developing future project plans based on evaluation results ( $M = 1.89$ ,  $SD = 1.167$ ,  $p = .028$ ). These tasks were represented in the responses within the control and participant involvement themes. Three different participants reported being asked to provide recommendations of participants to add to the evaluation team, which would account for the mean rating between 1 – No involvement and 2 – Yes, a little. There were several data collection instruments, such as surveys, checklists, and tables, developed as part of the evaluation framework. This provided ample opportunity for participants to be involved in this specific activity. It is also a response which is prevalent in the Participant Involvement theme, which provides insight into why it received a mean rating between 2 – Yes, a little and 3 – Yes, some.

The final series of activities which were statistically significant were of particular interest due to the relatively high standard deviations compared to the other activities participants responded to on the instrument. It was apparent in the follow-up interview responses, represented in the participant involvement and participant control themes, that participants interpreted these activities differently. Some of the participants

interpreted these activities as the final reports, presentations, and planning which occurs at the end of the evaluation process, which had not yet occurred at the time they completed the survey. Whereas others interpreted these activities as the preliminary sharing of information which was occurring as the team progressed through the evaluation process, such as documenting findings of specific indicators, sharing data collection instruments with members of the group for feedback, and discussing possible program revisions as sidebars during the program evaluation meetings.

The third question was “Does the use of a participatory evaluation model promote evaluative thinking with the non-evaluative stakeholders involved in the evaluation?” To answer this question, the participants were administered the Evaluative Thinking Survey, designed to collect data on participants’ confidence with evaluative thinking, was administered at three separate points throughout the process. Participants completed it prior to their engagement with the program evaluation tutorial, after completing the tutorial, and after completing the final scheduled program evaluation team meeting. The survey was comprised of 22 questions which mapped onto five domains of evaluative thinking posited by Fierro, Codd, Gill, Pham, Targos, and Wilce, (2018): reflecting, perspectives, projecting valuing evaluation, and use. The mean for each domain were compared across the three separate administrations of the survey to identify domains where changes across administrations were significant. The changes in all of the domains were found to be statistically significant.

Responses to the follow-up interview questions pertaining to the findings from the Evaluative Thinking Survey predominantly were found in the prior experience and use themes. The two subthemes which emerged from the prior experience major theme

were already practicing and enhanced skillset. The responses by participants when asked to reflect about how their thinking changed for each of the domains was either their thinking had not changed because they already thought in the manner expressed by the survey item, already held the belief expressed by the survey item, or practiced in the manner expressed by the survey item, in which case they rated themselves consistently across the three administrations. Those who reported a change in their responses expressed feeling as though they came onto the team already having a good understanding of program evaluation but then their feelings changed once they completed the program evaluation tutorial, indicating that there was more to it than they had thought based on their prior experiences with other forms of evaluation, in which case their responses decreased between the first and the second administration. Additionally, several participants responded they felt more confident in their understanding as they engaged in the actual program evaluation, in which case their responses increased between the second and third administration.

The use theme had two subthemes of in current role and institutionally. After participating in the program evaluation, the participants expressed a mix of thoughts that they could see themselves employing evaluation in their own work or components of the practice that they experienced. Additionally, participants expressed that they see the value of incorporating it as a regular practice within the organization or expressing that they wished the organization was better taking it more seriously as a practice.

## **Conclusions**

Due to the prevalence of educational programming and the complexity of the different educational programs which exist in educational settings, it makes the

likelihood of program evaluations unlikely as institutions are faced with finite resources to invest in their programs. However, institutions are being expected to demonstrate data informed decision making and directors of programs are often expected to demonstrate continuous improvement of their programs (Fitzpatrick, Sanders, & Worthen, 2011). A solution to address this need and support the use of program evaluation within an organization while also being responsible stewards of resources provided to support programs is to engage stakeholders using a participatory evaluation methodology while limiting the scope of the evaluation using program objectives as an advanced organizer.

It is suggested, based on the findings of this study, limiting the scope of a program evaluation using the program objectives helps maintain a focused program evaluation which is approachable for non-evaluative stakeholders. Additionally, implementing the evaluation utilizing a participatory evaluation methodology, has demonstrated that it provides benefits to the participants and the organization by providing training and experience completing a program evaluation with the support of a trained evaluator. While limiting the scope of the evaluation to the program objectives, it may limit the ability of the evaluation team to change the focus or direction of the evaluation; however, it does not prevent the evaluation from being participatory. Additionally, it has the added benefit of promoting the capacity of evaluative thinking within the organization through the participants who had the opportunity to engage in the program evaluation. This has the effect of impacting the participants' thinking, skills, and beliefs about programs and how they are evaluated and revised.

## **Limitations**

The study involving participants at Baker College was limited in scope. The diversity of stakeholders represented in the evaluation team was limited to those that were willing to volunteer or those whose full-time responsibilities include this type of work. As such, the inclusion of students or additional adjunct-faculty failed to occur as they would not be compensated for their time. This also limited the sample size of the participants available for analysis.

The timeline available to complete the program evaluation was also abbreviated, only allowing for three meetings of the evaluation team over the span of one month. As a result, the activities of developing a final report and dissemination of results were delayed, which may have influenced the ratings that provided on some of the data collection instruments. It also could have influenced the level that participants perceived the evaluation to be participatory as the evaluator had the responsibility of keeping meetings within the time frame provided. This also may have had the impact of cutting off conversations prematurely, influencing perceived participation.

## **Recommendations for Future Research**

Based on the current study's findings and limitations, the researcher has developed the following recommendations for future research:

1. Given the small sample size, the study should be replicated with other programs and other institutions to attempt to replicate the findings.
2. A study should be conducted to develop a validated instrument to assess evaluative thinking.
3. A study should be conducted to explore the relationship between prior

experience and evaluative thinking.

4. A study should be conducted to explore the tension between roles of the trained evaluator and non-evaluative stakeholders when completing a program evaluation utilizing a participatory evaluation methodology.
5. A longitudinal study should be conducted to further understand the benefits and return on investment of promoting evaluation experience and evaluative thinking with non-evaluative stakeholders within the context of educational organizations.

## APPENDIX

### Appendix A: Evaluation Framework Development

#### Logic Model – Describe the Program

- Gather information available on the program, including but not limited to:
  - Mission and vision
  - Goals and objectives
  - Current program descriptions such as websites, program descriptions, fact sheets
  - Strategic plans
  - Business, communication, and marketing plans
  - Existing/previous logic models
  - Existing performance measures and/or program reviews
- Review the information and extract from it to create a two-column table including:
  - Column 1: Activities – What the program its staff do
  - Column 2: Outcomes = Who or what beyond the program and its staff needs to change and how
    - In generating outcomes, it helps to identify the target audiences for program activities and the action they must take in order for the activities to be successful
    - Within the list in column 2, identify the most distal outcome: What is the biggest impact on student success you aim to address with your program?
- Clarify the activities and outcomes with stakeholders to ensure:
  - Appropriate classification; no activities are actually outcomes and no outcomes listed are actually activities
  - No major redundancy in list of activities or list of outcomes
  - No major missing activities or outcomes
- Decide whether the activities should be ordered sequentially. If so:
  - Think about the “logical” relationship among the activities – which may or may not be the same as how they unfold over time – and determine if some activities need to occur before others can be implemented
  - Order the activities within the columns into earlier or later activities to reflect the sequential relationships
- Decide whether the outcomes should be ordered sequentially
  - Think about the “logical” relationship among the outcomes – will some outcomes logically need to occur before others can be achieved?
  - Move the outcomes into columns to reflect the sequence in which the outcomes should occur. Label the columns as needed (i.e., short-, mid-, long-term; or [proximal, intermediate, distal])
- Check in with your stakeholders
  - To ensure the activities and outcomes reflect their understanding of the program to ensure:

- There are no major missing activities or outcomes
    - The logical progression of activities
    - The logical progression of the outcomes
  - To (re)affirm the intended uses of the logical model (i.e., assess implementation, assess effectiveness, performance measurement, strategic planning)
 

The intended uses of the logic model will determine which, if any, of the elaborations below would make the logic model more useful.
- If depicting the program logic in a roadmap format is desirable, then:
  - Write each of the existing activities and outcomes on a sticky note, or equivalent
  - Move the notes around to allow for drawing lines to depict logical relationships
  - Draw in lines remembering that lines may go from:
    - One or more activities to subsequent activity
    - One or more activities to an outcome
    - One or more proximal outcomes to a more distal outcome
- If outputs are desired because stakeholders would like clarification of the direct result of the activities, then using the logic model table or (better) the roadmap:
  - Identify the activities for which outputs are desired
  - Identify the link between those activities and their successor activities or outcomes
  - Thinking about the logical link, what are the key attributes of the activity that must be present for it to produce its successor activity or outcome?
  - Place the outputs in the appropriate place in the logic model table or roadmap
- If inputs are desired because stakeholders would like the clarification of necessary resources to implement the program, then:
  - Identify the key inputs without which the program cannot be implemented. Think about broad categories such as staff, equipment, data, funds, and partnerships.
  - Place the inputs into a column to the left of the activities in the logic model
  - If it is important to see the link between each input and the activity it affects, then draw arrows from each input to the related activity
- Review and affirm the elaborations of the logic model with stakeholders to ensure it accurately represents the program and the relationships among the components
- Create a narrative to with the logic model. A one-page logic model will not be able to capture all the nuances of the program. The narrative will help explain the components of the logic model and how they work together to accomplish the outcomes. The narrative should include the following:
  - An expanded description of the activities, outcomes, and other components of the logic model
  - Any key linkages between activities, between activities and outcomes, and between different outcomes
  - Attribution v. contribution to outcomes, etc.

- Stakeholder expectations for what will be accomplished, etc.

### Focus the Evaluation

- The standards help you assess and choose among options at every step of the framework, but some standards are more influential for some steps than others. The two standards most important in setting the focus are “utility” and feasibility.” Ensure that all stakeholders have common understandings of the phrases (formative/summative) and types of evaluations (needs assessment/process/outcome/impact).
- Using the logic model, think through where you want to focus your evaluation, using the principles in the “utility” standard:
  - Purpose(s) of the evaluation: implementation assessment, accountability, continuous program improvement, generate new knowledge, or some other purpose
  - User(s): the individuals or organizations that will employ the evaluation findings
  - Use(s): how will users employ the results of the evaluation, e.g., make modifications as needed, monitor progress toward program goals, make decisions about continuing/refunding
  - Review and refine the purpose, user, and use with stakeholders, especially those who will use the evaluation findings
- Identify the program components that should be part of the focus of the evaluation, based on the utility discussion:
  - Specific activities that should be examined
  - Specific outcomes that should be examined
  - Specific pathways from activities to specific outcomes or outcomes to more distal outcomes
  - Specific inputs or moderating factors that may or may not have played a role in success or failure of the program
- Refine/expand the focus to include additional areas of interest, if any, identified in Steps 1 and Step 2
  - Does the focus address key issues of interest to important stakeholders?
  - Did the program description discussion identify issues in the program logic that may influence the program logic?
  - Are issues of cost, efficiency, and/or cost-effectiveness important to some or all stakeholders?
- Refine/expand the focus to include additional areas of interest based on the propriety and accuracy evaluation standards
  - Are there components of the program – activities, outcomes, pathways, or inputs/moderators that must be included for reasons of “ethics” or propriety?
  - Are there components of the program – activities, outcomes, pathways, or inputs/moderators that must be included to ensure that the resulting focus is “accurate”?
- “Reality check” the expanded focus using the principles embedded in the “feasibility” evaluation standard

- The program's stage of development: Is the focus appropriate given how long the program has been in existence?
  - Program intensity: Is the focus appropriate given the size and scope of the program, even at maturity?
  - Resources: Has a realistic assessment of necessary resources been done? If so, are there sufficient resources devoted to the evaluation to address the most desired items in the evaluation focus?
  - At this point the focus may still be expressed in very general terms – this activity, this outcome, this pathway. Now, convert those into more specific evaluation questions. Some examples of evaluation questions are:
    - Was [specific] activity implemented as planned?
    - Did [specific] outcomes occur and at an acceptable level?
    - Were the changes in [specific] outcomes due to activities as opposed to something else?
    - What factors prevented the activities in the focus from being implemented as planned? Were [specific inputs and moderating factors] responsible?
    - What factors prevented (more) progress on the outcomes in the focus? Were [specific moderating factors] responsible?
    - What was the cost for implementing the activities?
    - What was the cost=-benefit or cost-effectiveness of the outcomes that were achieved?
  - Consider the most appropriate evaluation design, using the four evaluation standards – especially utility and feasibility – to decide on the most appropriate design. The three most common designs are:
    - Experimental: Participants are randomly assigned to either the experimental or control group. Only the experimental group gets the intervention, Measures of the outcomes of interest are (usually) taken before and after the intervention in both groups.
    - Quasi-experimental: Same specifications as an experimental design, except the participants are not randomly assigned to a “comparison” group.
    - Non-experimental: Because the assignment of subjects cannot be manipulated by the experimenter, there is no comparison or control group. Hence, other routes must be used to draw conclusions, such as correlation, survey or case study.
- Some factors to consider in selecting the most appropriate design include:
- With what level of rigor must decisions about “causal attribution” be made?
  - How important is ability to translate the program to other settings?
  - How much money and skill are available to devote to implementing the evaluation?
  - Are there naturally occurring control or comparison groups? If not, will selection of these be very costly and/or disruptive to the programs being studied?
- Start the draft of the evaluation plan. You will complete the plan in step 4. But at this point begin to populate the measurement table with

- Program component from logic model (activity, outcome, pathway)
- Evaluation questions for each component
- Review and refine the evaluation focus and the starter elements of the evaluation plan with stakeholders, especially those who will use the evaluation results.

Source: Centers for Disease Control and Prevention Program Performance and Evaluation Office. (2018). CDC Program Evaluation Framework Checklist. Retrieved March 10, 2019, from <https://www.cdc.gov/eval/steps/index.htm>

## Appendix B: Evaluative Thinking Survey

Directions: Rate your level of agreement with each statement.

Scale: 1 = Strongly disagree  
 2 = Disagree  
 3 = Neutral  
 4 = Agree  
 5 = Strongly Agree

Indicator: Reflecting						
Thoughtful questions	I ask questions that are thoughtful of what we are doing and why.	1	2	3	4	5
Deeper understanding	When reviewing information and evidence I'm looking for opportunities for further questions and investigation.	1	2	3	4	5
Describing thinking	I am able to communicate my ideas as well as the logic or thought processes behind them.	1	2	3	4	5
Identifying assumptions	I am able to identify when I or other members of my team are working off of an assumption.	1	2	3	4	5
Considering context	I am aware and am responsive to the context in which the evaluation is taking place.	1	2	3	4	5
Evaluation review	I reflect on the quality of the work and assess my own performance and the performance of my team.	1	2	3	4	5
Indicator: Perspectives						
Multiple perspectives	I solicit a diverse range of views and perspectives on the evaluand.	1	2	3	4	5
Additional points of view	I take into consideration the views and perspectives of others, beyond stakeholders.	1	2	3	4	5
Participatory evaluation	I value and am committed to seeking broad participation in evaluation activities.	1	2	3	4	5
Explicating Values	I am able to consider the values that are pertinent to the evaluation and can promote them in the evaluation process.	1	2	3	4	5
Indicator: Projecting						
Criteria of success	I am able to articulate the criteria for a successful evaluation.	1	2	3	4	5

Linking activities to outcomes	I am able to articulate connections between various elements of the program theory and program implementation.	1	2	3	4	5
Scaling	I am able to think about program activities in ways that connects them to the highest level of outcomes for the program.	1	2	3	4	5
Suite of evaluation activities	I am able to see the evaluation activities as being interconnected rather than isolated activities.	1	2	3	4	5
Indicator: Valuing evaluation						
Value of evaluation	I am able to articulate the importance and utility of an evaluation.	1	2	3	4	5
Value of evidence	The judgements, values, and assertions I make about a program are informed by findings attained through systematic inquiry.	1	2	3	4	5
Intent to engage in ECB	I believe it is important for organizations to promote evaluation capacity building with non-evaluative professionals within the organization.	1	2	3	4	5
Distributed responsibility	People in a variety of roles at different levels of implementation of a program should be responsible for conducting evaluations and acting on the findings.	1	2	3	4	5
Indicator: use						
Planning for use	As part of evaluation planning I include considerations for ways in which the evaluation findings will be used to support subsequent thoughts and actions about a program.	1	2	3	4	5
Instrumental use	As part of use of evaluation results, I think about how the findings can be used to modify the program in some way that improves alignment with program theory, will assist in the achievement of the outcomes, and/or will mitigate negative consequences.	1	2	3	4	5
Integration	I attempt to embed evaluation into the routine practices of my department and/or organization.	1	2	3	4	5
Process use	I look for opportunities to make changes using the evaluation process rather than findings.	1	2	3	4	5

Source: Fierro, L. A., Codd, H., Gill, S., Pham, P. K., Targos, P. T. G., & Wilce, M. (2018).  
Evaluative Thinking in Practice: The National Asthma Control Program. *New Directions  
for Evaluation*, 2018(158), 49–72. <https://doi.org/10.1002/ev.20322>

**APPROVED**

JUL 16 2019

**WAYNE STATE UNIVERSITY  
INSTITUTIONAL REVIEW BOARD**

## Appendix C: Semi-Structured Interview Protocol

Advancing Objectives-Oriented Evaluation with Participatory Evaluation Methodology – A Mixed-Methods Study

Semi-Structured Interview Protocol V 2.0

### Participatory Evaluation Measurement Instrument (PEMI)

1. Reflecting on the level of membership in the evaluation, what were the considerations for you when selecting your response?
2. Reflecting on the diversity of participants in the evaluation, what were the considerations for you when selecting your response?
3. Reflecting on the control over the evaluation process during the evaluation, what were the considerations for you when selecting your response?

### Evaluative Thinking Survey

1. On the Evaluative Thinking Survey, look over the statements listed in the “Reflecting” section. Based on your responses, how would you summarize how your thinking has changed in this area based on your experiences participating in the program evaluation?
2. On the Evaluative Thinking Survey, look over the statements listed in the “Perspectives” section. Based on your responses, how would you summarize how your thinking has changed in this area based on your experiences participating in the program evaluation?
3. On the Evaluative Thinking Survey, look over the statements listed in the “Projecting” section. Based on your responses, how would you summarize how your thinking has changed in this area based on your experiences participating in the program evaluation?
4. On the Evaluative Thinking Survey, look over the statements listed in the “Valuing Evaluation” section. Based on your responses, how would you summarize how your thinking has changed in this area based on your experiences participating in the program evaluation?
5. On the Evaluative Thinking Survey, look over the statements listed in the “Use” section. Based on your responses, how would you summarize how your thinking has changed in this area based on your experiences participating in the program evaluation?

### Evaluation Involvement Scale

1. Based on the responses to the Evaluation Involvement Scale, the items with ratings that stood out as being significant were:
  - Identifying evaluation planning team members
  - Developing data collection instruments
  - Writing evaluation reports
  - Reviewing evaluation reports for accuracy and/or completeness
  - Presenting evaluation findings (e.g., to staff, to stakeholder, to an external audience)
  - Developing future project plans based on evaluation results

Looking over your responses to the instrument, what were the factors you considered when providing your responses to these questions?

**APPROVED**

SEP 30 2019

**WAYNE STATE UNIVERSITY  
INSTITUTIONAL REVIEW BOARD**

## Appendix D: Participatory Evaluation Measurement Instrument (PEMI) Items

### Participator Evaluation Measurement Instrument (PEMI) Items

Coding Scheme for Extent of Involvement, source: Daigneault & Jacob, 2009, p. 340

Number of Tasks Nonevaluative Stakeholders are Involved In	Level of Membership	
	Intuitive Label	Numerical
0	No involvement	.00
1	Limited/weak involvement	.25
2	Moderate involvement	.50
3	Substantial/strong involvement	.75
4	Full involvement	1.00

Coding Scheme for Diversity of Participants (referred to as stakeholder selection for participation by Whitmore & Jacob, 2009), source: Daigneault & Jacob, 2009, p. 341

Number of Tasks Nonevaluative Stakeholder Types Involved	Level of Membership	
	Intuitive Label	Numerical
0	No diversity	.00
1	Limited/weak diversity	.25
2	Moderate diversity	.50
3	Substantial/strong diversity	.75
4	Full diversity	1.00

Coding Scheme for Control of the Evaluation Process, source: Daigneault & Jacob, 2009, p. 343

Level of Membership	
Intuitive Label	Numerical
Exclusive control by evaluator and/or nonparticipating evaluation sponsor	.00
Limited/weak control by nonevaluative participants	.25
Shared control between nonevaluative participants and evaluator and/or nonparticipating evaluation sponsor	.50
Substantial/strong control by nonevaluative participants	.75
Exclusive control by nonevaluative participants	1.00

**APPROVED**

JUL 16 2019

WAYNE STATE UNIVERSITY  
INSTITUTIONAL REVIEW BOARD

## Appendix E: Evaluation Involvement Scale Items

### Evaluation Involvement Scale Items

1:4 scale: 1 = No; 2 = Yes, a little; 3 = Yes, some; 4 = Yes, extensively (Toal, 2009, p. 354).

- 
1. Discussions that focused the evaluation
  2. Identifying evaluation planning team members
  3. Developing the evaluation plan
  4. Developing data collection instruments
  5. Developing data collection processes
  6. Collecting data
  7. Reviewing collected data for accuracy and/or completeness
  8. Analyzing data
  9. Interpreting collected data
  10. Writing evaluation reports
  11. Reviewing evaluation reports for accuracy and/or completeness
  12. Presenting evaluation findings (e.g., to staff, to stakeholders, to an external audience)
  13. Developing future project plans based on evaluation results
- 

**APPROVED**

JUL 16 2019

WAYNE STATE UNIVERSITY  
INSTITUTIONAL REVIEW BOARD

## Appendix F: Qualitative Code Book

Code ID	Code Name	Description/Definition	Example
A	Evaluator role and control	Items that discuss/describe the role that the trained evaluator played during the program evaluation and what control the evaluator held during the program evaluation.	<p>“You dictated some of the things.”</p> <p>“You acted as a guide.”</p> <p>“You were leading the discussion.”</p> <p>“You decided the team members.”</p>
B	Participant control	Items that discuss the level of control that the participants held during the program evaluation.	<p>“We all had a say.”</p> <p>“It was collaborative, I did provide input.”</p> <p>“I definitely had a presence in the process. But you all had the ability to, to kind of make the decisions as we were going along.”</p>
C	Participant Involvement	Items that discuss the type of or level of activity that participants had during the program evaluation.	<p>“We did everything.”</p> <p>Discusses data collection activities.</p> <p>Discusses survey/questionnaire design.</p>
D	Team Diversity +/-	<p>Items that discuss the make-up of the evaluation team</p> <p>+ indicates that the diversity of the team was positive or adequate</p> <p>- indicates that the diversity of the team was inadequate</p>	<p>- “We all play very similar roles.”</p> <p>- “It wasn’t incredibly diverse, there was one person from each role.”</p> <p>+ “The group was made out of people which represented various parts of the organization.”</p>
E	Prior Experience +/-	<p>Participants relate prior experiences to the program evaluation activity and</p> <p>+ indicates that they identify that participating in this program evaluation expanded or changed their thinking or skill set</p> <p>- Indicates that participating didn’t expand/change thinking or skill set but did affirm</p>	<p>- “I’m able to communicate my ideas, I don’t think that improved or got worse.”</p> <p>- “I was doing that in general.”</p> <p>+ “I have a better overview of the process.”</p> <p>+ “It allowed me to look at how A was connected to B”</p>

		existing skills/ thinking	
F	Use	Participant identified ways or desires to increase use/prevalence of program evaluation in current responsibilities or institutionally.	“Moving forward I can think of me trying to determine what data points I can pull to measure something.” “I think there would be value bringing in additional people.”

## REFERENCES

- Alkin, M., & Christie, C. (2004). *Evaluation Roots* (By pages 13-65).  
<https://doi.org/10.4135/9781412984157>
- Bamber, V., & Stefani, L. (2015). Taking up the challenge of evidencing value in educational development: from theory to practice. *International Journal for Academic Development*, 21(3), 242–254. <https://doi.org/10.1080/1360144X.2015.1100112>
- Bergquist, W. H., & Phillips, S. R. (1975). Components of an effective faculty development program. *The Journal of Higher Education*, 46(2), 177–211.  
<https://doi.org/10.2307/1980880>
- Beuving, J., & de Vries, G. (2014). *Doing qualitative research : The craft of naturalistic inquiry*. Retrieved from  
<http://ebookcentral.proquest.com/lib/wayne/detail.action?docID=1934288>
- Birnbaum, M. H., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology*, 45(4), 792–804.
- Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, 37(1), 48–74.
- Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. In P. Peterson, Baker, E., & McGraw, B. (Eds.), *International Encyclopedia of Education* (pp. 548–556). Oxford: Elsevier.

- Borko, Hilda. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3–15.
- Brooks, C. D., Marsh, L., Wilcox, K., & Cohen, B. (2011). Beyond satisfaction: Toward an outcomes-based, procedural model of faculty development program evaluation. *Journal of Faculty Development*, 25(3), 5–12.
- Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation*, 36(3), 375–388. <https://doi.org/10.1177/1098214015581706>
- Burke, B. (1998). Evaluating for a change: Reflections on participatory methodology. *New Directions for Evaluation*, 1998(80), 43–56. <https://doi.org/10.1002/ev.1116>
- Canadian International Development Agency. (2001, March). *How to perform evaluations: Participatory evaluations*. Retrieved from <https://www.oecd.org/derec/canada/35135226.pdf>
- Centers for Disease Control and Prevention Program Performance and Evaluation Office. (2018). CDC Program evaluation framework checklist. Retrieved March 9, 2019, from <https://www.cdc.gov/eval/steps/index.htm>
- Centra, J. A. (1978). Types of faculty development programs. *The Journal of Higher Education*, 49(2), 151–162. <https://doi.org/10.2307/1979280>
- Chelimsky, E. (1998). The role of experience in formulating theories of evaluation practice. *American Journal of Evaluation*, 19(1), 35–55. <https://doi.org/10.1177/109821409801900104>

- Chelimsky, E. (2007). Factors influencing the choice of methods in federal evaluation practice. *New Directions for Evaluation*, 2007(113), 13–33.  
<https://doi.org/10.1002/ev.213>
- Chelimsky, E. (2017). *Credibility, policy use, and the evaluation synthesis* (S. Donaldson, C. Christie, & M. M. Mark, Eds.). <https://doi.org/10.4135/9781483385839>
- Chism, N., & Szabo, B. L. (1997). Teaching Awards: The problem of assessing their impact. *To Improve the Academy*, 16, 181–200.
- College Navigator—Baker College. (n.d.). Retrieved October 29, 2019, from <https://nces.ed.gov/collegenavigator/?id=168847>
- Cousins, J. B. (2005). Will the real empowerment evaluation please stand up? A critical friend perspective. In Fetterman, D. M. & Wandersman, A. (Eds.), *Empowerment evaluation principles in practice* (pp. 183–208).
- Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, 14(4), 397–418.
- Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, 80(Winter), 87–105.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (Third). Thousand Oaks, California: SAGE Publications, Inc.
- Daigneault, P. M., & Jacob, S. (2009). Toward accurate measurement of participation: Rethinking the conceptualization and operationalization of participatory evaluation. *American Journal of Evaluation*, 30(3), 330–348.  
<https://doi.org/10.1177/1098214009340580>

- Daigneault, P. M., & Jacob, S. (2014). Unexpected but most welcome: Mixed methods for the validation and revision of the participatory evaluation measurement instrument. *Journal of Mixed Methods Research, 8*(1), 6–24.
- Donaldson, S. (2017). Examining the backbone of contemporary evaluation practice credible and actionable evidence. In *Credible and Actionable Evidence: The Foundation for Rigorous and Influential Evaluations* (Second, pp. 3–26). Los Angeles: SAGE Publications, Inc.
- Donaldson, S., Christie, C., & Mark, M. (2019). *What counts as credible evidence in applied research and evaluation practice?* <https://doi.org/10.4135/9781412995634>
- Donaldson, S., & Lavelle, J. M. (2018, June). *Basics of using theory to improve evaluation practice*. Presented at the American Evaluation Association Summer Workshop, Atlanta, GA.
- Fang, Y. (2016). *The effect of dialectical thinking on attribute framing: The roles of processing modes and associations* (PhD Thesis). Retrieved from <https://proxy.lib.wayne.edu/login?url=https://search-proquest-com.proxy.lib.wayne.edu/docview/1850135978?accountid=14925>
- Fetterman, D. (2001). *Foundations of empowerment evaluation*. Thousand Oaks: Sage.
- Fierro, L. A., Codd, H., Gill, S., Pham, P. K., Grandjean Targos, P. T., & Wilce, M. (2018). Evaluative thinking in practice: The national asthma control program. *New Directions for Evaluation, 2018*(158), 49–72. <https://doi.org/10.1002/ev.20322>
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B., R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Boston: Pearson.

- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2015). *How to design and evaluate research in education*.
- Gaff, J. G., & Morstain, B. R. (1978). Evaluating the outcomes. *New Directions for Higher Education*, 1978(24), 73–83. <https://doi.org/10.1002/he.36919782408>
- Gaventa, J., Creed, V., & Morrissey, J. (1998). Scaling up: Participatory monitoring and evaluation of a federal empowerment program. *New Directions for Evaluation*, 1998(80), 81–94. <https://doi.org/10.1002/ev.1119>
- Greene, J. (2017). *What counts as credible evidence in applied research and evaluation practice?* (By pages 153-167). <https://doi.org/10.4135/9781412995634>
- Greene, J. C. (2005). *Encyclopedia of evaluation*.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, Calif.: SAGE Publications, Inc.
- Guba, E., Lincoln, Y., & Lynham, A. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research*. Thousand Oaks: Sage. /z-wcorg/.
- Guskey, T. R. (2000). *Evaluating professional development*. Thousand Oaks: Corwin Press, Inc.
- Harper, F., & Nicolson, M. (2013). Online peer observation: Its value in teacher professional development, support and well-being. *International Journal for Academic Development*, 18(3), 264–275. <https://doi.org/10.1080/1360144X.2012.682159>
- Henry, G. T. (2017). *When getting it right matters: The struggle for rigorous evidence of impact and to increase its influence continues* (S. Donaldson, C. Christie, & M. M. Mark, Eds.). <https://doi.org/10.4135/9781483385839>

Higher Learning Commission. (n.d.). Retrieved February 24, 2019, from

<https://www.hlcommission.org/component/directory/?Action=ShowBasic&Itemid=192&instid=1977&lang=en>

Hoessler, C., Godden, L., & Hoessler, B. (2015). Widening our evaluative lenses of formal, facilitated, and spontaneous academic development. *International Journal for Academic Development*, 20(3), 224–237. <https://doi.org/10.1080/1360144X.2015.1048515>

House, E. R., & Howe, K. R. (1999). *Values in evaluation and social research*. Retrieved from

<http://proxy.lib.wayne.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=e00xna&AN=474093&site=ehost-live&scope=site>

Hoyt, D. P., & Howard, G. S. (1978). The evaluation of faculty development programs. *Research in Higher Education*, 8(1), 25–38. <https://doi.org/10.1007/BF00985854>

Johnson, R. B. (2017). Dialectical pluralism: A metaparadigm whose time has come. *Journal of Mixed Methods Research*, 11(2), 156–173.

<https://doi.org/10.1177/1558689815607692>

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112–133.

<https://doi.org/10.1177/1558689806298224>

Johnson, R. B., & Schoonenboom, J. (2016). Adding qualitative and mixed methods research to health intervention studies: Interacting with differences. *Qualitative Health Research*, 26(5), 587–602. <https://doi.org/10.1177/1049732315617479>

- Johnson, R. B., & Stefurak, T. (2013). Considering the evidence-and-credibility discussion in evaluation through the lens of dialectical pluralism. *New Directions for Evaluation*, 2013(138), 37–48. <https://doi.org/10.1002/ev.20056>
- Julnes, G., & Rog, D. (2017). *Actionable evidence in context: Contextual influences on adequacy and appropriateness of method choice in evaluation* (S. Donaldson, C. Christie, & M. M. Mark, Eds.). <https://doi.org/10.4135/9781483385839>
- Karsh, E., & Fox, A. S. (2014). *The only grant-writing book you'll ever need : Top grant writers and grant givers share their secrets*. Retrieved from <http://ebookcentral.proquest.com/lib/wayne/detail.action?docID=894918>
- King, J. A. (2007). Developing evaluation capacity through process use. *New Directions for Evaluation*, 2007(116), 45–59. <https://doi.org/10.1002/ev.242>
- King, J., & Fitzpatrick, J. (2009). Eight: evaluation of the special education program at the Anoka-Hennepin School District: An interview with Jean A. King. In J. Fitzpatrick, C. Christie, & M. Mark (Eds.), *Evaluation in Action*. <https://doi.org/10.4135/9781412990288>
- Knight, P., Tait, J., & Yorke, M. (2006). The professional learning of teachers in higher education. *Studies in Higher Education*, 31(3), 319–339. <https://doi.org/10.1080/03075070600680786>
- Koellner, K., & Jacobs, J. (2015). Distinguishing models of professional development: The case of an adaptive model's impact on teachers' knowledge, instruction, and student achievement. *Journal of Teacher Education*, 66(1), 51–67. <https://doi.org/10.1177/0022487114549599>

Kromrei, H. T. (2014). *A case study of the impact of a systematic evaluation process in a graduate medical education residency program* (Ph.D., Wayne State University).

Retrieved from

<http://search.proquest.com/docview/1531330752/abstract/1B60B9C01F6748F4PQ/1>

Kucsera, J. V., & Svinicki, M. (2010). Rigorous evaluations of faculty development programs. *The Journal of Faculty Development; Stillwater*, 24(2), 5–18.

Levinson-Rose, J., & Menges, R. J. (1981). Improving college teaching: A critical review of research. *Review of Educational Research*, 51(3), 403–434.

<https://doi.org/10.2307/1170213>

Livingstone, D. W. (2001). *Adults' informal learning: definitions, findings, gaps and future research* [Working Paper]. Retrieved from Centre for the Study of Education and Work, OISE/UT website: <https://tspace.library.utoronto.ca/handle/1807/2735>

Lucassen, T., & Schraagen, J. M. (2011). Factual accuracy and trust in information: The role of expertise. *Journal of the American Society for Information Science and Technology*, 62(7), 1232–1242. <https://doi.org/10.1002/asi.21545>

Luque, M. L. (2002). The role of domain-specific knowledge in intentional change. In G. M. Sinatra & Pintrich, P. R. (Eds.), *Intentional Conceptual Change* (pp. 135–172).

Retrieved from <http://ebookcentral.proquest.com/lib/wayne/detail.action?docID=356320>

Madaus, G. F., Scriven, M., & Stufflebeam, D. L. (1983). *Evaluation models: Viewpoints on educational and human services evaluation*. Boston; Hingham, MA, U.S.A.: Kluwer-Nijhoff; Distributors for North America, Kluwer Academic Publishers. /z-wcorg/.

Madaus, G. F., & Stufflebeam, D. L. (Eds.). (1989). *Educational evaluation: Classic works of Ralph W. Tyler*. Boston: Kluwer Academic Publishers.

- Mertens, D. M. (1999). Inclusive evaluation: Implications of transformative theory for evaluation. *American Journal of Evaluation*, 20(1), 1–14.  
<https://doi.org/10.1177/109821409902000102>
- Mertens, D. M. (2018). *Mixed methods design in evaluation*. Thousand Oaks, California: SAGE Publications, Inc.
- Mertens, D. M., & Wilson, A. T. (2012). *Program evaluation theory and practice : A comprehensive guide*. Retrieved from  
<http://ebookcentral.proquest.com/lib/wayne/detail.action?docID=869331>
- Miller, R. L. (2017). *How people judge the credibility of information lessons for evaluation from cognitive and informational sciences* (S. Donaldson, C. Christie, & M. M. Mark, Eds.). <https://doi.org/10.4135/9781483385839>
- Moreau, K. A. (2017). Twelve tips for planning and conducting a participatory evaluation. *Medical Teacher*, 39(4), 334–340.
- Neave, H. R., & Worthington, P. L. (1988). *Distribution free tests*. London: Unwin Hyman Inc.
- O’Sullivan, R. G., & O’Sullivan, J. M. (1998). Evaluation voices: Promoting evaluation from within programs through collaboration. *Evaluation and Program Planning*, 21(1), 21–29.  
[https://doi.org/10.1016/S0149-7189\(97\)00042-6](https://doi.org/10.1016/S0149-7189(97)00042-6)
- Patton, M. Q. (2018). A historical perspective on the evolution of evaluative thinking. *New Directions for Evaluation*, 2018(158), 11–28. <https://doi.org/10.1002/ev.20325>
- Provus, Malcom. (1973). *Educational evaluation: theory and practice* (B. R. Worthen & J. R. Sanders, Eds.). Worthington, Ohio: C.A. Jones Pub. Co. /z-wcorg/.
- Richlin, L. (2006). *Blueprint for learning : Constructing college courses to facilitate, assess, and document learning*. Sterling, Va.: Stylus Pub. /z-wcorg/.

- Rieh, S. Y., & Danielson, David R. (2007). Credibility: A multidisciplinary framework. *Annual Review of Information Science and Technology*, 41(1), 307–364.
- Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park, Calif.: Sage Publications. /z-wcorg/.
- Scriven, M. S. (2015, August 15). *Key evaluation checklist*. Retrieved from [http://www.michaelscriven.info/images/MS\\_KEC\\_8-15-15.doc](http://www.michaelscriven.info/images/MS_KEC_8-15-15.doc)
- Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19(1), 1–19. <https://doi.org/10.1177/109821409801900102>
- Sinatra, Gale M., & Pintrich, P. R. (2002). The role of intentions in conceptual change learning. In G. M. Sinatra & Pintrich, P. R. (Eds.), *Intentional Conceptual Change* (pp. 1–18). Retrieved from <http://ebookcentral.proquest.com/lib/wayne/detail.action?docID=356320>
- Sorinola, O. O., & Thistlethwaite, J. (2013). A systematic review of faculty development activities in family medicine. *Medical Teacher*, 35(7), e1309–e1318. <https://doi.org/10.3109/0142159X.2013.770132>
- Stein, M. K., Schwan, M., & Silver, E. A. (1999). The development of professional developers: Learning to assist teachers in new settings in new ways. *Harvard Educational Review; Cambridge*, 69(3), 237.
- Steinmetz, A. (1983). The discrepancy evaluation model. In G. F. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation Models*. Boston: Kluwer-Nijhoff ; Distributors for North America, Kluwer Academic Publishers.
- Stufflebeam, D. L. (2001). Evaluation models. *New Directions for Evaluation*, 89(Spring).

Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, and applications*.

Retrieved from

<http://ebookcentral.proquest.com/lib/wayne/detail.action?docID=1801247>

Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*.

/z-wcorg/.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), 237–246.

<https://doi.org/10.1177/1098214005283748>

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

<https://doi.org/10.1177/1098214009337031>

Tseng, S., & Fogg, B. J. (1999). Credibility and computing technology. *Commun. ACM*,

42(5), 39–44. <https://doi.org/10.1145/301353.301402>

Tversky, A., & Kahneman, D. (2013). Judgement under uncertainty: Heuristics and biases.

In W. T. Ziemba & L. C. MacLean (Eds.), *Handbook Of The Fundamentals Of Financial Decision Making (In 2 Parts)* (pp. 261–268). Retrieved from

<http://proxy.lib.wayne.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=e020mna&AN=605574&site=ehost-live&scope=site>

Tyler, R. W. (1931). A Generalized technique for conducting achievement tests. *Educational Research Bulletin*, 10(8), 199–208.

Wathen, C. N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology*, 53(2), 134–144.

Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies*. Upper Saddle River, N.J.: Prentice Hall. /z-wcorg/.

Wilcox, R. (2011). *Modern statistics for the social and behavioral sciences: A practical introduction*. Retrieved from <https://ebookcentral.proquest.com>

Wraga, W. G. (2017). Understanding the Tyler rationale: Basic principles of curriculum and instruction in historical context. *Espacio, Tiempo y Educación*, 4(2), 227–252.

<https://doi.org/10.14516/ete.156>

**ABSTRACT****ADVANCING OBJECTIVES-ORIENTED EVALUATION WITH  
PARTICIPATORY EVALUATION METHODOLOGY – A MIXED METHODS STUDY**

by

**DUSTIN SAALMAN****May 2020****Advisor:** Dr. Shlomo Sawilowsky**Major:** Educational Evaluation and Research (Quantitative)**Degree:** Doctor of Philosophy

The ability to complete program evaluations of educational programming is typically restricted by the availability of resources, such as time, money and a trained evaluator. A mixed methods study was completed to explore the use of a participatory evaluation program evaluation with the use of the program objectives as an advanced organizer. Participatory evaluation is purported to increase organizational learning and promote evaluative thinking within an organization (Cousins & Whitmore, 1998). Objectives oriented evaluation is an easily understood evaluation method which provides a refined focus program outcome (Madaus & Stufflebeam, 1989). An explanatory sequential design was employed utilizing quantitative findings to collect qualitative data to further explore the participants' experiences completing the program evaluation.

The findings indicated that this combined evaluation methodology met the criteria posited in Daigneault and Jacob (2009) and Toal (2009) to be considered

participatory in its implementation. It also involved participants in ways which provided them experiences which helped develop evaluative thinking, skills, and beliefs.

## **AUTOBIOGRAPHICAL STATEMENT**

Dustin Saalman has served as the Director for the Center for Teaching Excellence at Baker College since 2019. He has served in similar positions related to faculty development at Wayne State University, School of Medicine and Baker College, Online and Center for Graduate Studies. Prior to transitioning to higher education, Dustin served in several roles in the Michigan K-12 Public Education system from 2006 to 2014, such as special education teacher, technology teacher, technology director, and curriculum director.

Dustin earned his Bachelors of Science, Elementary Education and Special Education from Eastern Michigan University in 2006. He then earned his Masters of Arts, Education and Instructional Technology from Saginaw Valley State University in 2012.