



8-2019

An Exploratory Investigation into the Role of Evaluation Theory in Evaluative Thinking

Lyssa Wilson Becho

Western Michigan University, lyssa.becho@gmail.com

Follow this and additional works at: <https://scholarworks.wmich.edu/dissertations>



Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation

Becho, Lyssa Wilson, "An Exploratory Investigation into the Role of Evaluation Theory in Evaluative Thinking" (2019). *Dissertations*. 3484.

<https://scholarworks.wmich.edu/dissertations/3484>

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.



AN EXPLORATORY INVESTIGATION INTO THE ROLE OF
EVALUATION THEORY IN EVALUATIVE THINKING

by

Lyssa Wilson Becho

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Interdisciplinary Ph.D. in Evaluation
August 2019

Doctoral Committee:

Chris L. S. Coryn, Ph.D., Chair
Jennifer Greene, Ph.D.
Bianca Montrosse-Moorhead, Ph.D.

Copyright by
Lyssa Wilson Becho
2019

ACKNOWLEDGEMENTS

Throughout the writing of this dissertation I have received a great deal of support and assistance. I would first like to thank my advisor, Dr. Chris Coryn, for his instrumental guidance and expertise throughout the process of formulating my research topic and methodology. Second, I would like to thank my committee members, Drs. Jennifer Greene and Bianca Montrosse-Moorhead, for their time, guidance, and thoughtful feedback. Their invaluable insight helped to increase the quality and thoughtfulness of this dissertation.

I would like to acknowledge the support of friends and colleagues at The Evaluation Center and Interdisciplinary Ph.D. in Evaluation. I have been continually inspired and encouraged by my fellow students throughout my time at Western Michigan University. I would like to particularly thank my colleagues Erica Dodds, for countless weekly calls and productive dissertation retreats, and Kelly Robertson, for her encouragement and advice. Additionally, I would like to thank my colleagues Lori Wingate and Emma Leeburg for cheering me on along the way.

Finally, I'm extremely grateful to my family, my parents, and my sister, for always believing in me. I cannot begin to express my thanks to my husband, Carlos Becho, who has been there for me in countless big and small ways. He has never wavered in his support and I could not have completed this dissertation with my sanity intact without him.

Lyssa Wilson Becho

AN EXPLORATORY INVESTIGATION INTO THE ROLE OF EVALUATION THEORY IN EVALUATIVE THINKING

Lyssa Wilson Becho, Ph.D.

Western Michigan University, 2019

The field of evaluation generally agrees that evaluation theory is important. Evaluation theorists and scholars have written prolifically about the role of evaluation theory in practice. However, the empirical literature associated with how and why evaluation theory is important is still emerging. In particular, how evaluation theory fosters high quality evaluation practice through transparent logical reasoning, attention to values and valuing, and contextual responsiveness, and how these inform evaluative thinking are understudied phenomena in the field.

This study intends to provide insights into the relationship between evaluation theory and evaluation practice to answer two major research questions. The first research question asked to what extent American Evaluation Association (AEA) and Canadian Evaluation Society (CES) evaluators are trained in evaluation theory. The second asked what role evaluation theory plays in evaluative thinking. A sequential, two-phase, mixed-methods design was used to investigate these questions, including a cross-sectional survey of AEA and CES members and in-depth one-on-one interviews. The online survey asked a random sample of AEA and CES evaluators about their training in evaluation theory, including the modalities of training and

their familiarity with specific approaches. The interviews were designed to unearth the role of evaluation theory in evaluators' thinking and decision-making.

Findings from the survey reveal that 80% of AEA and CES evaluators had some type of training in evaluation theory. This training generally took the form of short presentations or webinars, graduate courses, and 1–4 day professional workshops. Overall, evaluators were most familiar with participatory evaluation, developmental evaluation, and utilization-focused evaluation, and least familiar with deliberative democratic evaluation, consumer-oriented evaluation, and constructivist or fourth-generation evaluation. Emergent themes from interviews showed that evaluators were responsive to a variety of contextual factors in their practice regardless of their level of training or familiarity with evaluation. Evaluators with all levels of training in and familiarity with evaluation theory endorsed the importance of use as an indicator of successful evaluation, and privileged stakeholder engagement as a factor in facilitating and increasing use. However, evaluators with more training in and familiarity with evaluation theory were more intentional in considering their options and more explicit in justifying the reasoning underlying their actions. Their decisions were generally backed by personal or professional values. Evaluators with less training in and familiarity with evaluation theory were more influenced by preferences or requirements of funders, their disciplinary area, or their own organization. Implications and avenues for future research of this study include identifying the best ways to teach evaluation theory, particularly in light of growing conversations around the professionalization of evaluation and the inclusion of evaluation theory in both the AEA and CES evaluator competencies.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER 1.....	1
Introduction	1
Statement of the Problem	1
Good Evaluation	2
Evaluative Thinking.....	4
Conceptual Framework	6
Purpose Statement and Research Questions.....	10
Study Significance and Implications	11
Definition of Key Terms	12
Dissertation Organization	14
CHAPTER 2.....	15
Introduction	15
Evaluation Theory.....	15
Defining and Situating Evaluation Theory	15
Evaluation Theory in Theory.....	18

Table of Contents - continued

Evaluation Theory in Practice	18
Evaluation Practice	19
The Gap Between Theory and Practice	20
Decision Making in Practice	22
Factors Influencing Decision Making in Evaluation	22
Evaluative Thinking.....	25
Differing Definitions	26
Characteristics of Evaluative Thinking	30
Summary	33
The Current Study.....	35
CHAPTER 3.....	38
Methodology	38
Overall Research Design	38
Quantitative Phase	40
Design	40
Population and Sampling.....	40
Instrumentation.....	41
Procedure	42
Sample Characteristics	43
Data Processing and Analysis	45
Qualitative Phase.....	45
Design	45

Table of Contents - continued

Population and Sampling.....	46
Instrumentation.....	46
Procedure	47
Sample Characteristics	47
Data Processing and Analysis	51
Positionality	51
CHAPTER 4.....	53
Findings	53
Training in and Familiarity with Evaluation Theory	54
Training in Evaluation Theory Among AEA and CES Evaluators.....	54
Short Webinars and Presentations on Evaluation Theory.....	56
Professional Development Workshops and Courses on Evaluation Theory.....	57
University Courses on Evaluation Theory	59
Informal Training in Evaluation Theory	59
Summary	59
Familiarity with Evaluation Theory Among AEA and CES Evaluators.....	60
Summary	61
Role of Evaluation Theory.....	62
Perceived Role of Evaluation Theory.....	63
Evaluation Theory in Practice	64
Evaluation Theory in Evaluative Thinking.....	68
Summary	70
Implicit Role of Evaluation Theory.....	72

Table of Contents - continued

Contextual Responsiveness	72
Logical Reasoning	83
Valuing.....	91
Values	97
Chapter Summary.....	101
CHAPTER 5.....	103
Summary	103
Conclusions	104
Training in and Familiarity with Evaluation Theory	104
Role of Evaluation Theory	105
Discussion.....	107
Lack of Standard Vocabulary	108
Revisiting the Conceptual Framework.....	109
Limitations	110
Future Research.....	111
References.....	115
Appendix A: Survey Questions for AEA and CES Members	126
Appendix B: CES Member Sample Demographics Compared to Population Demographics..	134
Appendix C: Interview Protocol	137

Table of Contents - continued

Appendix D: HSIRB Approval.....	141
--	------------

LIST OF TABLES

2.1.	A comparison of evaluative thinking definitions.....	36
3.1.	Overall survey sample characteristics.....	44
3.2.	Overall interview sample characteristics.....	50
4.1.	Comparison of respondents who attended formal training in evaluation theories, models, or approaches and those who did not.....	55
4.2.	Comparison of evaluators with low, average, and high evaluation theory.....	62
4.3.	Comparison of exemplary quotes around the perceived role of evaluation theory.....	71
4.4.	Comparison of exemplary quotes around contextual responsiveness in practice.....	81
4.5.	Comparison of exemplary quotes around logical reasoning in practice.....	89
4.6.	Comparison of exemplary quotes around valuing in practice.....	95

LIST OF FIGURES

1.1.	Conceptual framework of evaluation theory's role in evaluative thinking and practice.....	6
3.1.	Sampling procedures and final response rates for the survey and interview phases.....	49
4.1.	Type of formal training in evaluation theories, models, and approaches and the percent of respondents who reported attending each type ($n = 209$).....	55
4.2.	Type of short webinar or presentation in evaluation theories, models, and approaches attended by respondents ($n = 117$).....	56
4.3.	Percent of webinars or presentations on evaluation theory that focused on a single theory, multiple theories, or whether it depended on the training ($n = 117$).....	56
4.4.	Training types and whether evaluation theory was considered the sole, major, or minor focus of the training.....	57
4.5.	Training types and whether they focused on a single evaluation theory, between 2 – 5 theories, or more than 5 theories.....	58
4.6.	Instruction delivery platform for training types, whether online, mix of in-person and online, or in-person only.....	58
4.7.	AEA and CES evaluators' overall familiarity with evaluation theories ($n = 209$).....	60
4.8.	AEA and CES evaluator's familiarity with specific evaluation theories ($n = 186$).....	61

CHAPTER 1

Introduction

Evaluation theory aims to foster good evaluation practice, which is much more than simply technical rigor and validity. Good evaluation practice also includes transparent logical reasoning, the inclusion of values and valuing, and the artful decisions that come with being responsive to contextual factors (House, 1980; Schwandt, 2014). How evaluation theory fosters transparent logical reasoning, attention to values and valuing, and contextual responsiveness in evaluation practice, and how these inform active and intentional evaluative thinking, are understudied issues in the field. It is the intention of this study to lend empirical insight into these relationships. This first chapter introduces the context associated with good evaluation and the position and role of evaluation theory in practice; it provides the rationale underscoring the necessity and significance of this dissertation study.

Statement of the Problem

Overall, the field of evaluation generally agrees that evaluation theory is important (Alkin, 2012; Mark, 2018; Schwandt, 2014; Shadish, Cook, & Leviton, 1991). Shadish (1998) expands on the numerous ways in which evaluation theory creates a core identity for evaluators: providing the profession with a common language, values, knowledge base, and a public face to distinguish professional evaluators from non-evaluators. Shadish asserts that evaluation theory provides evaluators an identity separate from that of other social scientists, something more than just applied research, the heart of evaluation as a unique field. The importance given to evaluation theory can be seen in textbooks devoted to the different paradigms, histories, and characteristics of theories and approaches (Alkin, 2012; Fitzpatrick, Sanders, and Worthen 2011; Mertens & Wilson, 2012; Shadish, Cook, & Leviton, 1991; Stufflebeam & Coryn, 2014), as well as in the dedicated Topical Interest Group (TIG) in the American Evaluation Association (AEA)—the theories of evaluation TIG. AEA and the Canadian Evaluation Society (CES) have both identified knowledge of evaluation theory as a necessary component for being considered a competent evaluator in the draft AEA evaluator competencies (competency 1.4; AEA, 2017, p. 1) and the CES evaluator competencies

(competency 1.1, 1.8, and 2.5; CES, 2018). At the individual level, evaluation theory is intended to guide good evaluation practice; theorists hope their writings serve as a model for decision making while conducting evaluations (Shadish, Cook, & Leviton, 1991). While the importance of evaluation theory is generally agreed upon, details about how, and whether, evaluation theory leads to good practice, in what ways, and under what conditions, is a fruitful area for further research.

Good Evaluation

High-quality evaluation can be defined in numerous ways (House, 1980; Scriven, 2011; Stufflebeam, 1999; Yarbrough, Shulha, Hopson, & Caruthers, 2010). All agree that, while technical rigor and validity are essential, good evaluation practice is also committed to the utility of evaluation findings through transparent logical reasoning, contextual responsiveness, and adherence to ethical practices. Alongside the standard of accuracy, the Joint Committee on Standards for Educational Evaluation recognizes utility, feasibility, and propriety (Yarbrough, et al., 2010). Similarly, the draft AEA Evaluator Competencies (2017) identify technical skills needed for quality evaluation, while also recognizing the necessity of using evidence and logic to make evaluative judgments (competency 1.5), responding to unique contexts (competency 3.2), and planning for evaluation use and influence (competency 4.5). And the CES Competencies for Canadian evaluation practice similarly include establishing a credible link between recommendations and findings (competency 2.9), responding to natural contexts (competency 3.1), and encouraging use of findings (competency 2.10). All of these guidance documents for quality evaluation, along with numerous expert evaluators and theorists (Scriven, 1980, 1990; Shadish, Cook, & Leviton, 1991; Stake, 2004), concur that attention to logical reasoning and contextual responsiveness is an important part of distinguishing good evaluation.

Logical reasoning is at the core of evaluation through the process of supporting and defending a conclusion or claim with warrants and backings (Fournier, 1995; Scriven, 1980; Hurteau, Houle, & Mongiat, 2009). Fournier (1995) writes, “the means to developing, strengthening, and clarifying reasoning that leads to legitimate evaluative conclusions is the

crux of successful evaluation theory and practice” (p. 15). If evaluation is to be used as a platform for decision making, stakeholders need to find the evidence presented in evaluative findings credible and actionable (Donaldson, Christie, & Mark, 2015). A strong, logical, and clearly explained line of reasoning lends credibility and trust to evaluative findings. Fournier and Smith (1993) note that “when evaluation findings are challenged or utilization has failed, it was because stakeholders and clients found the inferences weak or the warrants unconvincing” (p. 322). A strong logical argument contains a claim (or conclusion) that stems from evidence. Warrants help explain why the connection between the claim and the provided evidence is trustworthy; warrants are then supported by backings (Toulmin, 1964). Scriven (1980) applied this logical reasoning to evaluation, enumerating four main steps in the logic of evaluation: (1) establish criteria, (2) set standards, (3) collect and analyze data, and (4) synthesize into a final judgment. Fournier (1995) expanded on Scriven’s logic of evaluation to identify more specific working logics from the general logic he put forth.

More recently, Hurteau, Houle, and Mongiat (2009) investigated “legitimate and justified judgments” in program evaluation reports. They determined evaluation findings were “legitimate judgments” if they referred to the evaluation questions or goals, are supported by criteria, and supported by a standard (p. 312). Further, they looked at whether these judgments were justified, defined as whether there was justification (warrants and backings) provided for the chosen criteria and standards, and a documented process used to synthesize information (p. 312). Without transparent and legitimate reasoning to justify evaluative conclusions, evaluation will not be taken seriously by stakeholders.

In addition to logical reasoning, the credibility and utility of evaluative results rely on the contextual responsiveness of the evaluation. Not all evaluands are the same, therefore neither should all evaluations be the same. In their discussion of evaluative reasoning, Nunns, Peace, and Witten (2015) discuss the importance of context in creating justified evaluation findings. They point to the contextual dependency of valuing (Patton, 2012) and the inability to detach contextual factors from evaluative reasoning (p. 145). They write,

at a practical level, context determines the choice of criteria, how they are developed and by whom (Henry, 2002), the nature and extent of argument, the warrants selected

(Smith, 1995), and the complexity of the evaluative conclusion/judgement resulted (Julnes, 2012). (p. 145)

The importance of responding to context in evaluation practice goes beyond influencing the elements of a logical argument and influences almost all decisions made throughout the evaluation planning, implementation, and reporting process. For example, an evaluator must consider the right methodology for the right disciplinary area, historical location, or even underlying epistemological assumptions. The ability to be intentional and nuanced about responding to contextual factors requires an understanding of the evaluation methodologies available and an awareness of the situational contexts that might influence choosing the ‘right’ one.

Evaluative Thinking

The concept of evaluative thinking captures the concepts of logical reasoning, values and valuing, and contextual responsiveness (Vo, 2013; Vo, Schreiber, & Martin, 2018). Using the Delphi technique with evaluation experts to determine the key components of evaluative thinking, Vo (2013) defined evaluative thinking as “a process by which one marshals evaluative data and evidence to construct arguments that allow one to arrive at contextualized value judgments in a transparent fashion” (p. 107). In her study, Vo put forth a conceptual model of evaluative thinking as an individual outcome as well as an input for social change (p. 15). In this model, she connects individual inputs of time, resources, and interest, leading to the activities of education, training, or mentoring, which result in evaluative thinking. Evaluative thinking then leads to certain values, beliefs, and practices, which ultimately changes programs, organizations, and society (Vo, 2012, p. 15).

One of the activities identified in Vo’s (2013) model leading to evaluative thinking is education and training. It is hypothesized in this dissertation that training in evaluation theory is a particularly important activity in cultivating productive evaluative thinking. Schwandt’s (2014) view of evaluation theory lends itself to this hypothesis. He writes, “These bodies of knowledge [evaluation theory] are repertoires of concepts, insights, explanations, and tools that professional practitioners can use as heuristics, tools ‘to think with’” (p. 234). Schwandt

sees evaluation theory as a direct tool for evaluative thinking, different ways to explore evaluation practice and tailor to the evaluand's context. Mark (2018) connects the ideas of evaluation theory, thinking through varied choices, and being intentional in reasoning and responsiveness:

Evaluation theory can serve as a guide to practice. It can help navigate the choices associated with different 'schools of thought' and varied method options. ... What the field needs is people who know when, where, why, and how different methods could and should be used in evaluation practice. Theory tells us that. (p. 134)

Mark implies that evaluation theory helps us consider different schools of thought in evaluation practice and how they might best fit into everyday decision making in evaluation. For example, evaluation theory allows evaluators to answer the questions of what approach would work best for what situation, when should which stakeholders be involved, and why and when is it worthwhile to create a logic model. It is hypothesized in this dissertation that evaluation theory is a driving force in evaluators' ability to think divergently, complexly, and intentionally about the practical problems in evaluation practice. Vo et al. (2018) write, evaluative thinking "reflects the ability to creatively arrive at evaluation-specific solutions by determining what combination of methods, tools, techniques, and theories are appropriate in light of contextual practices" (p. 40). A broad understanding of evaluation theory allows evaluators to draw from a divergent pool of options.

Previous studies have looked at evaluation practitioners' proclaimed theory in comparison to theorists' (Christie, 2003), practitioners' application of theory in practice (Coryn, et al, 2011; Miller & Campbell, 2006; Fitzpatrick, 2004), and decision making in evaluation practice (Kundin, 2010; Tourmen, 2009). While these studies have built a foundation of empirical understanding of evaluation theory in practice, there is still need for more. There have been numerous calls to increase the empirical literature around the theory to practice connection in evaluation (Cousins & Earl, 1999; Scriven, 1991; Shadish, Cook, & Leviton, 1991; Smith, 1994). A deep dive into the cognitive processes of evaluative thinking and the role evaluation theory plays has not been fully examined. It is the intention of this dissertation to explore patterns in evaluative thinking for those who have training in evaluation theory relative

to those who do not. A more complete understanding of this process can lead to a deeper understanding of how to encourage good evaluation practice. Currently, there is no research that can show the extent and type of training in evaluation theory among evaluators.

Conceptual Framework

To better understand the role of the evaluation theory and its place in influencing evaluative thinking and evaluation practice, a conceptual framework was developed based on a review of the literature (illustrated in Figure 1.1). As this framework intends to explore the role of evaluation theory in evaluative thinking, ultimately leading to good evaluation practice, the final product of good evaluation practice lies at the far right of the model, with evaluation theory on the far left.

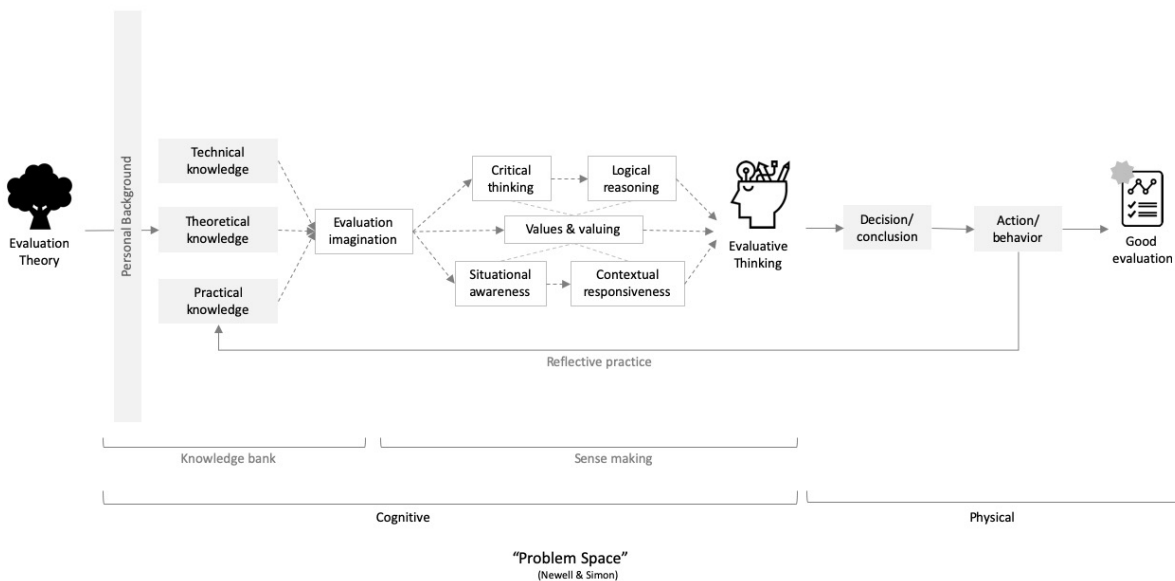


Figure 1.1. Conceptual framework of evaluation theory's role in evaluative thinking and practice.

This framework recognizes that every problem space includes cognitive aspects as well as physical aspects of problem solving (Newell & Simon, 1972). The acknowledgement of both the cognitive and physical phases – the thinking that takes place, which then results in action – mirrors Vo's (2013) conceptualization of evaluative thinking. She writes "evaluative thinking

does not simply happen in the mind; rather, this cognitive process is manifested as a problem-solving practice – the doing of evaluation” (p. 106). Evaluative thinking cannot inherently be separated from the conclusion and action that emerge from the cognitive process. This idea is also reflected in Vo, Schreiber, and Martin’s (2018) conceptual model of evaluative thinking, which acknowledges both the cognitive aspects and application of evaluative thinking. Both the cognitive and physical phases are reflected in this conceptual framework. The cognitive portion of the problem space is divided into knowledge bank and sense making in an attempt to separate information in a pure form from when that information is entangled with the previous beliefs, assumptions, and biases of the individual. This is an admittedly overly simplistic diagram of the thought process. Feedback loops likely exist between almost every concept in the model and occur within milliseconds as neurotransmitters jump between nerve synapses in the brain. However, the model is therefore an intentional oversimplification as this is not a dissertation on neurocognition.

Evaluation theory, as it exists in a textbook or other external form, enters an individual’s cognition and is immediately interpreted through the lenses and perspectives of their personal background. This includes their historical experiences, geographic location, and demographic characteristics, previous education in other disciplines, or epistemological and ontological assumptions, whether they are cognizant of this or not. After passing through this initial filter, the evaluation theory sits in a theoretical knowledge bank, alongside other technical knowledge and practical knowledge. Technical knowledge refers to an individual’s ability to carry out the steps necessary to conduct the chosen methods of evaluation, whether this is data collection, analysis, or reporting. For example, an evaluator may know that an ANOVA is the correct inferential test to run in a certain situation, yet they might not possess the specific knowledge to conduct the test. Practical knowledge is gained through the doing and experiencing of evaluation. While these can be separated theoretically, they are quickly mixed together and combined in the process of thinking.

Each area of knowledge influences an individual’s evaluative imagination. This refers to the ability to consider the various pathways, opportunities, and choices an evaluator can use in approaching a specific situation. Schwandt (2014) writes about the evaluative imagination as a

kind of “inventory”, noting that “to the extent that the inventory is sparse, my response [to identifying evaluative solutions to practical problems] will be limited” (p. 234). Without a broad evaluative imagination “decisions in practice [are made] only out of habit” (p. 234). This conceptual framework posits that an expansive evaluative imagination allows evaluators to consider a broad array of evaluation approaches, instead of always relying on default methods. The ability to think divergently about the possibilities of tools available to the evaluator allows them to be more intentional and transparent in their logical reasoning, valuing, and contextual responsiveness.

In this dissertation, it is hypothesized that learning evaluation theory would broaden one’s evaluative imagination by providing more tools to think with and a greater understanding of possible approaches and pathways in evaluation. Evaluation theories present different ways of thinking, different ways of collecting and interpreting data, and different assumptions regarding stakeholder information needs and involvement. For example, an evaluator trained in Campbell’s (Campbell, 1969; Campbell, 1991; Shadish, Cook, & Campbell, 2002) theoretical perspective would lean towards the use of experimental and quasi-experimental designs, and in effect, be a more quantitatively oriented evaluator. Similarly, an evaluator trained in Stake’s (1995; 2003) theory would lean towards the use of case studies and qualitative methods. Each of these is a valid and appropriate approach to conducting evaluation, however, they may be appropriate only in differing contexts. An evaluator who is trained in both theories and can think critically about the evaluative context they are addressing, can step back and determine which approach is most appropriate given the environment, stakeholder perspectives, and information needs. Learning a broad spectrum of evaluation theories introduces evaluators to different possible shapes evaluation practice can take and what contexts those shapes are appropriate for.

The framework integrates the aspects of logical reasoning, values and valuing, and contextual responsiveness as they are important aspects of evaluative thinking and quality evaluation. The concept of contextual responsiveness acknowledges the necessity for situational awareness, the ability of an evaluator to be aware of the contextual factors that may influence important decisions and actions. Additionally, logical reasoning is preceded by critical

thinking. While earlier definitions of evaluative thinking categorized it as an applied area of critical thinking, Patton (2018), Schwandt (2018), and Vo, Schreiber, and Martin (2018) acknowledge this as shallow and limiting. Vo, Schreiber, and Martin (2018) write,

Making reasoned, evidence-based choices set critical thinking and evaluative thinking apart from judgments made based only on deeply held and unchallenged beliefs (e.g., stereotypes). Making a reasoned choice about value and being able to defend it is what distinguishes [evaluative thinking] from critical thinking. (p. 40)

Given this distinction, the conceptual framework considers critical thinking as part of evaluative thinking, but not all encompassing. Additionally, it lends itself to the inclusion of values and valuing within the framework. Scriven's (1980) logic of evaluation rests on the demonstration of *value* with evidence. Within an evaluation, it is necessary for an evaluator to identify whose and which values will be included, addressed, or argued. These three main concepts – logical reasoning, values and valuing, and contextual responsiveness – are in alignment with the recent literature on evaluative thinking (Vo, 2013; Vo, Schreiber, & Martin, 2018; Patton, 2018).

This conceptual framework intends to model evaluators' decision making on a micro scale, outlining the cognitive elements that influence evaluative thinking, which ultimately leads to a decision or conclusion, leading to an action or behavior. Figure 1.1 only visualizes a single conclusion as a representation of all the micro and macro decisions that make up an evaluation. This could be decisions around resources, stakeholder engagement, methods or methodology, evaluation approaches, inclusion of values, reporting decisions, and every decision in between. A multitude of decisions and actions make up the final evaluation.

The feedback loop of reflective practice is also integrated, represented by an arrow from action/behavior, back to practical knowledge. Practical knowledge is the most obvious point where reflecting and learning from ones' own experience plays a role. Scholars in decision making sciences have shown the importance of practical knowledge in making decisions in practice (Guzman, 2009). Evaluation scholars have also pointed to the role of practical knowledge in responsive evaluation (Benkofske, 1995; Fitzpatrick, 2004; Greene, 2006). Tourmen (2009) aimed to illuminate the relationship between theory and practice in the decisions evaluators make, however, found that practical knowledge is more pronounced as an

influential factor. Tourmen compared the activities of novice and expert evaluators to find that experienced evaluators have different conceptualizations of situations based on action rules or implicit theories stemming from practical knowledge (p. 20). The concept of implicit theories – or internal theories, folk theories, or working models – appears often in the conversation about evaluation theory and practice (Barela, 2005; Christie, 2003; Greene, 2006; Shadish, Cook, and Leviton, 1991). Shön (1983) coined the term reflection-in-action and wrote, “When someone reflects-in-action, he becomes a researcher in the practice context. He is not dependent on the categories of established theory and technique but constructs a new theory of the unique case” (p. 68). Some argue that these implicit theories of evaluation practice may be stronger than formal theories (Schwandt, 2002, 2003; Tourmen, 2009; Fitzpatrick, 2004), however, there is a need for more empirical research around the creation and use of informal theories (Christie, 2003). As mentioned earlier, the visual representation for reflective practice as a single arrow is a simplified understanding, as reflecting on previous experience influences all aspects of thought and reasoning. This entire process is recursive and cyclical in nature, allowing for multiple feedback loops to influence every aspect of this conceptual model.

This conceptual framework was created through study of the current literature on evaluation theory and practice, evaluative thinking, evaluative reasoning, responsive practice, decision making and reflective practice. It is presented as an initial attempt to make sense of the complex cognitive processes that connect evaluation theory, evaluative thinking, and good evaluation practice. Through the process of this dissertation, this framework may be revised and reordered to better align with the results of the study. It is not the intent of this dissertation to validate this framework as a true model, but it is open to change by the end of the study.

Purpose Statement and Research Questions

The objective of this dissertation is to investigate the role of evaluation theory in practice by first examining the extent of training in evaluation theory among practitioners and then how evaluation theory influences practitioners’ evaluative thinking. As an exploratory study, the following research questions will be investigated:

1. To what extent and in what ways are evaluation practitioners in the AEA and CES communities trained in evaluation theory?
 - a. What types of training in evaluation theory are practitioners receiving, if any?
 - b. To what extent are evaluation practitioners educating themselves in evaluation theory outside of formal training?
2. What role does evaluation theory play in evaluative thinking?
 - a. How do patterns of evaluative thinking differ in evaluators with training in evaluation theory compared to those without?
 - b. How do evaluators with training in evaluation theory differ in their patterns of thinking in relation to:
 - i. Logical reasoning
 - ii. Values and valuing
 - iii. Contextual responsiveness

Study Significance and Implications

This study has practical significance and implications for the field of evaluation. First, a greater understanding of the role of evaluation theory in practice would affect how evaluators are trained. As the field of evaluation has matured, not all evaluators have found their way to the profession through formal training in evaluation methods and theories. For many practitioners, evaluation theory seems a thought experiment for academics in ivory towers. Adding to the evidence base for the practical implications of learning evaluation theory may encourage more evaluators to seek out theoretical knowledge or training. Additionally, further evidence around the cognitive processes of evaluation theory in evaluation thinking and decision making may encourage training institutions to place more emphasis on the teaching of evaluation theory.

Second, given the conversations occurring in the profession around questions of competencies and certification of qualified evaluators, it is important to understand and add clarity to the role of evaluation theory in good evaluation practice. If evaluation as a profession

is going to draw boundaries around the training required to become a competent evaluator, the results of this study could provide the beginning of a basis for the extent evaluation theory should be included in those requirements.

Definition of Key Terms

While Shadish (1998) commends evaluation theory as a means for providing a common language around evaluation, there is still significant variation in the language used to discuss many concepts in the field, and often, these concepts have different meanings in different contexts. These definitions serve to clarify the language used throughout this particular study.

As used throughout this dissertation, **evaluation theory** refers to a body of knowledge that guides evaluators with regard to what evaluation is and how it ought to be conducted; evaluation theory is an organized collection of ideas, models, and approaches that provide practitioners with heuristic “tools ‘to think with’” (Schwandt, 2015, p. 33). In its current state, most evaluation theory is prescriptive, instead of descriptive or predictive; it suggests how evaluation should be done, rather than how it is done (Alkin, 2012).

Evaluation practice refers to “how evaluators conduct their work” (Rog, 2015). In one sense, practice can be considered “the technical expression or application of theoretical principles” (Chouinard et al., 2016). In another, evaluation practice is influenced by a combination of practical knowledge, theoretical knowledge, and contextual factors (Chouinard et al., 2016; Schwandt, 2015). Evaluation practice inherently includes practical problems. Schwandt (2014) points to two fundamental characteristics of practical problems: (1) they are ill structured, meaning there is uncertainty around the goals, means, or rules for addressing the problem, and there is a possibility of multiple solutions, possibly no solutions, and (2) they demand an action as a response (p. 232). Evaluation practice certainly encompasses this uncertainty, complexity, and demand for action.

Evaluative thinking can be defined as “... a particular kind of critical thinking and problem-solving approach that is germane to the evaluation field. It is the process by which one marshals evaluative data and evidence to construct arguments that allow one to arrive at contextualized value judgments in a transparent fashion” (Vo, 2013, p. 105). Evaluative thinking

can be conceptualized at the individual or organizational level. Evaluative thinking is considered a separate concept from evaluative practice, or the action of doing evaluation. For one can think evaluatively without actually carrying out an evaluation. Additionally, one can conduct aspects of evaluation practice without a deep consideration of evaluative thinking. Vo, Schreiber, and Martin (2018) acknowledge that while the term evaluative thinking has a rich history in the literature, it has been used as more of a “catch phrase rather than a phenomenon worthy of meaningful consideration” (p. 31).

Scriven and Paul (1987) defined **critical thinking** as “... the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action” (p. 1). Critical thinking, considered both a process of inquiry and an outcome, is essential to the ability to construct a logical argument. **Logical reasoning** is the ability to construct a warranted argument. In the context of evaluation, this specifically refers to the justification of evaluative conclusions based on identified values and reasoned evidence (Fournier, 1995; Scriven, 1980).

The term **context** is used throughout this dissertation as more than simply the sociopolitical, historical, and geographic situation in which an evaluation takes place. It also accounts for the information needs of stakeholders; constraints of resources such as time, money, and data; and the epistemological assumptions and preferences of stakeholders and evaluators. Greene (2005) identifies five dimensions of context, including a setting’s (a) descriptive and demographic character, (b) material and economic features of a setting, (c) institutional and organizational climate, (d) interpersonal dimensions, and (e) political dynamics. An evaluator is **contextually responsive** when their choices are sensitive and reactive to these aspects of an evaluand’s context. In order to be contextually responsive in practice, evaluators must be aware of important situational factors and how they might influence the evaluation (**situational awareness**).

Dissertation Organization

The remainder of this dissertation consists of four chapters that provide more context around the study, as well as a summary of findings and conclusions. Chapter 2 provides a review of the literature on evaluation theory, evaluative thinking, and additional, related concepts to provide an ample understanding of how this study fits into the current landscape of literature. Chapter 3 describes the methods and methodology of the study, including a developmental mixed methods approach utilizing a quantitative survey to create a sampling frame for exploratory interviews. Chapter 4 presents findings, analyzing main themes from the data. Finally, Chapter 5 provides a discussion of the findings in light of previous literature. Additionally, the chapter will include a reflection on the limitations of the current study and suggestions for future research.

CHAPTER 2

Introduction

This study explores the role of evaluation theory in evaluative thinking in order to determine its influence on evaluators' decision making in practice. To position this study in the current literature, this review will focus on several bodies of relevant literature. First, the concept of evaluation theories will be discussed. This includes definitions of evaluation theory and suppositions of importance by early theorists. Second, the call for more empirical research on evaluation and evaluation theories has shown that practice is not always aligned with theory. This gap between evaluation theory and practice and its relevance to decision making in evaluation practice. will be examined. Third, relevant literature about decision making in practice will be reviewed. This includes the influence of implicit theories, practical experience, epistemological preferences, and contextual constraints on the decisions made by evaluators. With an in depth understanding of theory and of practice, this review will discuss the concept of evaluative thinking and how it bridges the gap between theory and practice. Finally, this review will discuss what distinguishes good evaluation and how evaluative thinking and evaluation theory fit into encouraging good evaluation practice.

Evaluation Theory

Defining and Situating Evaluation Theory

There are numerous definitions for evaluation theory, however, they all suggest that the collection of writing and theoretical knowledge considered to be evaluation theory serves as a "framework to guide the study and practice of program evaluation" (Stufflebeam & Coryn, 2014, p. 50; see also Alkin, 2012; Fitzpatrick, Christie, & Mark, 2008; Shadish, 1998; Shadish, Cook, & Leviton, 1991). While Scriven (1991) acknowledges that evaluation theories "include a wide range of efforts", Mathison (2005) captures a relatively straight forward definition: "evaluation theory serves to provide a plausible body of principles that explain and provide directions to the practice of evaluation" (p. 142). Acknowledging the messy world of theory in the field of evaluation, Shadish (1998) wrote "with few exceptions, evaluation theory is neither concise nor axiomatic; and it is not a single theory but rather a set of diverse theoretical

writings held together by the common glue of having evaluation practice as their target” (p. 2). While the boundaries of what is evaluation theory are not neatly defined, it is worthwhile to wade through the mess because of the importance of evaluation theory to the field and to the individual practitioner.

The evaluation literature discusses *theories*, *approaches*, and *models* to capture the variety of writings that have been considered under the umbrella of evaluation theory. Alkin (2004) asserted that while the term *theories* is “conventionally used in evaluation literature, in some ways, it would be more appropriate to use the term *approaches* or *models*” (p. 5). Most of what is considered evaluation theory can more accurately be considered prescriptive models, “guiding frameworks that specify what a good or proper evaluation is and how evaluation should be done” (Alkin, 2004, p. 5). These prescriptive models have little empirical evidence to support the underlying claims, particularly causal claims, about what the evaluation approach purports to achieve. A descriptive, empirical, or predictive model would be a portrayal of how evaluation is conducted in practice, or tested predictions for specific outcomes (Alkin, 2003, 2004). Currently, there are very few theories in the field of evaluation that could be considered descriptive, empirical, or predictive. Potentially the most researched evaluation theory is that of evaluation use, the centerpiece of utilization-focused evaluation; some empirical support does exist for claims of utilization (King & Alkin, 2018).

This categorization sets evaluation theory apart from other types of theories. For example, Donaldson and Lipsey (2006) defined social science theories as an “attempt to provide generalizable and verifiable knowledge about the principles that shape social behavior” (p. 62). Social science theories are generally empirically derived, descriptive or predictive theories of social phenomenon. Evaluation theory, on the other hand, is a prescriptive model of practice. Evaluation theory is also often conflated with program theory. A predictive type of theory, program theory outlines the underlying assumptions of how a program or intervention works and how it expects to bring about change (Donaldson & Lipsey, 2006; Lipsey, 1993). While these types of theories can play an important role in evaluation practice, evaluation theory serves a more tangible purpose in providing approaches and models of how evaluation should be

practiced, although little empirical backing for these approaches or models exist (Fitzpatrick, Christie, & Mark, 2008).

The breadth of evaluation theories, approaches, and models have been captured in a number of frameworks to explain their similarities and differences. Shadish, Cook, and Leviton (1991) described the historical development of theories as a series of stages. In the first stage, theorists focus on methodologies and the production of credible knowledge, relying heavily on experimental designs and avoiding bias. The second stage of theorists acknowledge the importance of attending to the use and utility of findings, considering the life of evaluative data once in the hands of stakeholders. The third stage combines learnings of stage one and two theorists for more contextually responsive approaches. Shadish, Cook, and Leviton (1991) took great care to compare and contrast the epistemological, ontological, methodological, and practical differences amongst theorists.

Another categorization framework of evaluation theories that also follows a historical development pattern is Alkin and Christie's (2004; 2012) evaluation theory tree. Anchoring the tree roots in the three primary evaluation aims of social accountability, fiscal control, and social inquiry, the three main branches on the tree include theories that emphasize valuing, use, and methods. The secondary branches, and leaves in later versions, represent different theorists, allowing readers to compare similarities and differences across theoretical perspectives.

Building on Alkin and Christie's tree, Mertens and Wilson (2012) based their tree branches on epistemological assumptions. In addition to the methods branch (based on postpositivism), use branch (based on pragmatism), and values branch (based on primarily on constructivism), they added a social justice branch that includes theorists operating from a transformative paradigm. Other categorization frameworks for organizing evaluation theories center around the extent to which evaluation theories focus on judging a program's value (Stufflebeam & Coryn, 2014), and the primary use and user of the evaluation (Fitzpatrick, Sanders, and Worthen, 2004). These different categorization frameworks help to illustrate the diversity and breadth of evaluation theories and approaches that exist, as well as the different areas of practice they hope to inform.

Evaluation Theory in Theory

One of the most cited proponents of evaluation theory and its importance to the field of evaluation is William Shadish, memorably proclaiming that evaluation theory “is who we are” as evaluators (1998). Shadish argued that evaluation theory is particularly important to the field as a whole because it provides a common language for evaluators, encompasses the values and issues evaluators care about, provides an identity for evaluators different than that of other professions, presents a face to the outside world, and forms the knowledge base that defines the profession. According to Shadish, evaluation theory is the glue that holds disparate practitioners, fields, and methods together as a single discipline.

Evaluation theory is equally important at the level of the individual evaluator. Evaluation theory is said to give directives to practitioners on how evaluation could or should be carried out (Alkin, 2012; Chelimsky, 2012; Fitzpatrick, Christie, and Mark, 2008; Shadish, Cook, & Leviton, 1991). Authors of evaluation theories have “sought to develop approaches that assist evaluators in designing and carrying out useful, defensible program evaluations” (Stufflebeam & Coryn, 2014, p. 46). Many theoretical writings on the role of evaluation theory espouse evaluation theory as a blue print for evaluators in decision making, providing guidance on what to consider and which direction to pursue. Evaluation theory tells us the “when, where, why, and how different methods could and should be used in evaluation practice” (Mark, 2018, p. 134). Ideally, theory and practice are symbiotic, each informing and improving the other. Chelimsky (2012) wrote “... evaluation theory and practice are interdependent: each one learns from the other and, in that learning process, both are inspired to stretch, to bend a little, and to grow” (p. 91). Chelimsky also pointed out that early theorists were rather naïve in their assumption that their theoretical writings would have immediate application and usefulness to practitioners. The relationship of evaluation theory and practice has shown to be more complex than even Chelimsky believed.

Evaluation Theory in Practice

There have been numerous calls to increase research on evaluation, particularly to increase empirical studies on the link between evaluation theory and practice (Chelimsky, 2012;

Christie, 2003; Cousins & Earl, 1999; Henry & Mark, 2003; Mark, 2008; Miller, 2010; Scriven, 1991; Shadish, Cook, and Leivton, 1991; Smith, 1993; Smith, 2010; Szany, Azzam, & Galen, 2012). A greater empirical understanding of how theory informs practice, and practice informs theory, will lead to further improvements of theories and their accessibility to practitioners (Datta, 2003). Research that has been conducted shows a more complex picture about the role of evaluation theory in practice than early theorists presumed.

Evaluation Practice

To further understand the disconnect between theory and practice, it is important to characterize the nature of practice. Guzman (2009) defined practice as “a set of planned or unplanned actions that can be common, habitual and frequent, or emergent, casual and unexpected” (p. 87). The practice of evaluation happens in complex contexts with moving pieces and sometimes opaque reasoning, aptly described by Schwandt (2005) as “rough grounds” (p. 99). Rog expanded,

practice encompasses all the activities we consider under the evaluation umbrella, from framing, scoping, and designing the study; determining who participates and how much; scoping the roles people have; choosing the methods for collecting data; analyzing, sharing, and disseminating data; and fostering use. (2015, p. 226)

The practice of evaluation encompasses the individual actions and choices that are made in planning, conducting, and carrying out an evaluation. Practical decisions also rely on larger assumptions, including the role of evaluation and evaluators in social programming, the purpose or objectives of the evaluation, and epistemological and value assumptions.

In many ways, the practice of evaluation can be equated to problem solving in dynamic systems. Schwandt (2014) described practical problems as having two fundamental characteristics, “they are ill structured and they demand an action as a response” (p. 232). Schwandt expanded,

They are ill structured (Mitroff & Silvers, 2010) in that goals, means, or constraints are not particularly clear; there is the possibility of multiple solutions (or no solution at all); and, there is uncertainty about which concepts, principles, or rules are necessary to

address the problem. They are problems that demand an action because they take the general form “what should I do now, in this situation, facing these circumstances?”

Taking action in a thoughtful and wise way, in turn, requires reasoned reflection to decide what one is to do. Practical problem solving demands critical judgment for action. (2014, p. 232)

Just as the boundaries of evaluation theory are messy, evaluation practice can also be messy. With choices being made every step of the way, from the purpose of the evaluation, to stakeholder involvement, methods and methodology choices, and use and utility of findings, evaluation practice is highly variable and requires numerous on-the-spot decisions (Fitzpatrick, 2004; Tourmen, 2009). Between all of the possible choices and possible combinations, Cronbach noted “evaluation practice is more art than science” (as cited in Shadish, Cook, & Leviton, 1991), and more recently, Chouinard et al. (2016) pointed out “evaluation is more than a technical craft requiring more than application of methodological rigor and skill” (p. 11). With the “rough grounds” of evaluation practice, it is no surprise that evaluation theory doesn’t provide a straight forward map.

The Gap Between Theory and Practice

Christie's (2003) mapping of theory to practice and comparing the practice of theorists and everyday evaluators gave the field an insightful look into the connection between theory and practice in real life situations. Two of Christie’s main findings are particularly important in understanding the actual state of evaluation theory in practice. First, Christie (2003) found that only a small proportion (10%) of her sample of practicing evaluators reported using a particular theory to guide their work. Additionally, an even smaller proportion (6%) identified a specific text used to direct their evaluation practice. These findings suggest practicing evaluators may not have extensive training in or familiarity with evaluation theory. A study conducted by Shadish and Epstein in 1987 found a similar lack of familiarity with evaluation theory among evaluation practitioners. In their discussion, they worry “such findings suggest there is a danger of scholarly illiteracy in evaluation about its own writings and concepts” (p. 586). Using LaVelle’s (2014) research, it can be calculated that less than one-fourth of American Master’s

degrees, PhD's, and certificates in evaluation require students to take a course specifically about evaluation theory. This low level of evaluation theory in evaluation training, in combination with low levels of formal evaluation training for the average practitioner (Datta, 2003; Grob, 2018), support the findings by Shadish and Epstein (1987) and Christie (2003). A representative survey of evaluators has not been conducted to truly know the extent of knowledge or training around evaluation theory. However, if evaluation theory is to be a guide to practice, these low rates of knowledge and awareness call into question whether the connection is as strong as theorists' aspirations.

Christie's (2003) second important finding addresses this concern directly, suggesting a lack of congruence between the evaluation practice of theorists and everyday practitioners. Christie found a wide diversity of practice that did not align with theoretical writings, concluding that a gap remains between evaluation theory and practice. Other studies have shown similar misalignment of average evaluation practice and theoretical writings. Looking at evaluation reports as a proxy for practice, Coryn, Noakes, Westine, and Schroter's (2011) study on theory-driven evaluation and Miller and Campbell's (2010) study on empowerment evaluation found inconsistencies between what was put forth in theoretical literature and what was conducted in practice. Additionally, Fitzpatrick's (2004) in-depth discussions with exemplary evaluators on their practice found that practice rarely follows or relies on theory. Further complicating matters, and showing just how complex practice can be, among those that have been trained in and do explicitly draw from evaluation theory, it is unclear the extent to which they draw from a single theory or from various aspects of theory (Bledsoe & Graham, 2005; Leeuw & Donaldson, 2015). Interestingly, Williams (1989) found that evaluation theory is more complex and discordant than evaluators' actual practice. While Williams did not dig into the influencing factors on evaluators' practice, other studies have shown decision making in evaluation practice to rely on other aspects, as well as and perhaps in place of evaluation theory (Barela, 2005; Christie, 2003; Fitzpatrick, 2004; Tourmen, 2009).

Decision Making in Practice

Research on the process of decision making in evaluative practice identifies a number of influential factors outside of theoretical knowledge. These influential factors can be categorized into external contextual factors of the evaluand, internal characteristics of the evaluator, and practical knowledge resulting from reflection in action.

Kundin's (2010) conceptual framework for decision making in evaluation provides a scaffold for understanding influential factors in evaluation practice. Pulling together various theories into an umbrella theory of practical decision making, Kundin's framework is founded on three elements practitioners consider in decision making. In the first element, decisions are guided by situational awareness and the context of evaluands. Here Kundin draws on work from Endsley (1997) and Greene (2005). The second element references Scriven's (1991) general logic of evaluation as well as Fournier's (1995) working logic to address the practical reasoning required in an evaluation in regard to the phenomenon, questions, problem, and values of an evaluation. The final element stresses the importance of reflection in every day practice, incorporating Schön's (1983) concept of reflection on action and Schwandt's (2002, 2003, 2005) writings on evaluation practice. Acknowledging the multiple influence on decisions in practice, Kundin (2010) wrote,

Evaluators develop working 'theories' about which methods to use based on their assumptions, expertise, values, and judgment. These 'theories' are likely informed by a combination of academic study, disciplinary socialization, professional expertise, and evaluators' insight and understanding about a program's context. (p. 354)

Kundin's framework acknowledges the complexity of situations and the merging of theoretical understanding and tacit knowledge, as well as paints a comprehensive understanding of the current state of evaluator's decision making.

Factors Influencing Decision Making in Evaluation

As previously discussed, evaluators are continually making decisions and actively solving problems in a changing and complex environment (Chelimsky, 2012; Kundin, 2010; Schwandt, 2014). It is evident that these contextual factors would play a crucial role in the decisions

evaluators make. These influential contextual factors may include the intended purpose of the evaluation, needs and wants of stakeholders, political and cultural setting, practical factors of access to people and data, as well as resource constraints including limitations on time and money (Benkofske, 1995; Fitzpatrick, 2004; Hood, 2005; Greene, 2006; Kundin, 2010; McGuire, Zorzi, and Perrin, 2002; Schwandt, 2003; Patton, 2008). Braverman and Arnold (2008) also pointed out the following influencing factors in their study about methodological decisions: funder requirements, professional standards and trends in the field, requirements for formal program recognition and legitimation (e.g., designs required by What Works Clearing House), organizational culture, and the burden of increasing methodological rigor on program participants as well as staff. Examples of contextual factors influencing decisions might be an evaluator deciding on a certain methodology because of its perceived validity in that field (Grob, 2018), or following a certain strategy for stakeholder involvement to support the client organization's values. Aligning decisions in evaluation practice with contextual factors can increase the overall quality and utility of evaluations (Patton, 2004; Shadish, Cook, & Leviton, 1991; Stake, 1980; Vo & Christie, 2015).

In addition to external influencers, the personal characteristics and background of evaluators play a role in their decision making. Evaluators often endorse particular epistemologies, potentially unknowingly. These philosophical underpinnings can be influenced by educational background, field of study, or inherent beliefs in the nature of truth and knowledge. They, in turn, influence preferred and sometimes default evaluation approaches and practices (Benkofske, 1995; Fitzpatrick, 2004; Kundin, 2010). Evaluators have also been shown to be influenced by their mentors, advisors, professors, or supervisors (Chandler, 2001). Mentors could have an effect on personal preferences or expertise, another influential factor in evaluators decision making (Braverman & Arnold, 2008).

The third important factor influencing evaluators' decision making is their practical knowledge. Practical knowledge can be conceptualized as "the ability to put into effect previously acquired knowledge in specific circumstances ... the blend of explicit and tacit procedural knowledge with explicit and tacit practice" (Guzman, 2009, p. 93). Schön (1983) refers to the use of practical knowledge in decision making as reflection-in-action. He writes

“when someone reflects-in-action, he becomes a researcher in the practice context. He is not dependent on the categories of established theory and technique, but constructs a new theory of the unique case” (p. 68). Drawing on their previous experience with what has worked for them and in which circumstances, it is as if evaluators conduct an on the spot experiment to determine which is the best choice for each decision. Kandin (2010) explained, “through reflection, practitioners build up a collection of images, ideas, examples, and actions that they can draw upon when making practice decisions” (p. 354). Some have referred to this new theory created through reflective practice as mental models, implicit theories, or folk theories (Barela, 2005; Greene, 2006; Vo & Christie, 2003).

The power of practical knowledge and implicit models can be seen when comparing the practice of expert and novice evaluators. Tourmen’s (2009) study comparing evaluators decision making in a simulated case study found differences in the process of decision making of evaluators who were just beginning their practice compared to evaluators with years of experience. She observed that experienced evaluators have built pragmatic conceptualizations and action rules through their previous evaluations. These experiences change their perceptions of the situational diagnosis and evaluation analysis, allowing the evaluators to be more comfortable with the complexity that surrounds the evaluand and their evaluation. While her observations do note this pragmatic knowledge is “not completely disconnected from formal theoretical knowledge” (p. 8), it is “built” through experience.

Tourmen (2009), in fact, started her investigation with the intention of exploring how evaluation theory influenced decision making in practice. However, her findings barely discuss evaluation theory, instead focusing on practical knowledge and experience as influencing decisions. While not explicitly stated, readers can infer that evaluation theory did not feature prominently in evaluators’ reflections on their practice. Other studies have also directly refuted the importance of evaluation theory in practice. Through interviews with practitioners, Chandler (2001) found that “evaluation theory or philosophizing about evaluation is not immediately recognized as important to practice decisions” (p. 3), noting that respondents instead discussed their values about evaluation, client needs, and contextual factors. Scriven, a founding father of evaluation theory himself, has questioned practitioners’ need for evaluation

theory, calling it “something of a luxury” (1991, p. 360). House (2003), another prominent evaluation theorist, is not concerned by the gap between theory and practice, considering “one would want the theorists, who spend time searching for anomalies, problems, and new concepts, to have different ideas than practitioners. In a sense, theorists earn their keep by creating ideas that are not common” (p. 56). Studies of evaluation practice and experts’ careful consideration of practice have certainly agreed on one fact, that the influence of theory on practice is complex.

If evaluation theory does not provide a straight forward map for practice, perhaps it plays a more nuanced role in the process of evaluation. Like problem solving, evaluation as a practice unfolds as a process of thinking, weighing decisions, and then acting. Perhaps evaluation theory is infused into these steps in a way that is not immediately obvious. In Schwandt’s (2014) careful consideration of evaluation theory, he wrote

When we use the term theory in evaluation we could be talking about any of various bodies of conceptual or theoretical knowledge that can serve as an aid in thinking through options in a situation that a practitioner faces. These bodies of knowledge are repertoires of concepts, insights, explanations, and tools that professional practitioners can use as heuristics, tools “to think with.” They are aids to the evaluation imagination, as practitioners come to understand the problems before them and how those problems might be solved. (p. 234)

Schwandt takes a nuanced view of connection between evaluation theory and practice, recognizing that it may be more diffuse. In this view, the utility of evaluation theory is closer to Weiss’s (1988) enlightenment use rather than Patton’s (1988) instrumental use. Perhaps, as Schwandt puts forth, evaluation theory is more of a tool for evaluative thinking than a decision tree to follow.

Evaluative Thinking

Evaluative thinking as a higher order of thinking is inherent in human cognition outside of formal evaluation (Buckley, Archibald, Hargraves, & Trochim, 2015; House, 2016; McKegg & King, 2014; Patton, 2018; Wehipeihana & McKegg, 2018). As House (2016) points out, “human

thought is fundamentally evaluative” (p. 104). Both Bloom (1956) and Kuhn (2005) included evaluative thinking in their hierarchy of thought and learning in the field of education. In the context of formal evaluation, evaluative thinking becomes an essential aspect of a high-quality evaluation (Buckley et al., 2015; Vo & Archibald, 2018). However, the term *evaluative thinking* was rarely used in the literature until 2001, when it began to gain traction (Buckley et al., 2015). Prior to the growing literature base around evaluation thinking in evaluation capacity building, the idea of thinking evaluatively was only alluded to by writers. For example, Weiss (1998) referred to the *evaluative cast of mind*, writing that collaborative evaluation helps “program people reflect on their practice, think critically, and ask questions about why the program operates as it does. They learn something of the evaluative cast of mind – the skeptical questioning point of view, the perspective of the reflective practitioner” (p. 25). Even earlier, House (1977) refers to “a dialogue in which [evaluators and stakeholders] are free to employ their reasoning” (p. 48). Both of these refer to key characteristics of the modern concept of evaluative thinking – reflective practice and critical thinking.

As with many terms in the social sciences, the concept of evaluative thinking does not have a single agreed-upon definition. Authors have explored the different possible boundaries and implications of what it means to think evaluatively (Vo & Archibald, 2018). In order to acknowledge the varied conceptualizations of evaluative thinking, this review will contrast the differing definitions, as well as review the major characteristics generally agreed to be associated with evaluative thinking.

Differing Definitions

The various definitions in the literature for evaluative thinking can be compared across three main dimensions: the scope of implications (at the societal, organizational, or individual level); whether the thinking is done by an evaluator or non-evaluator; and whether the intended outcome is to produce an evaluative judgment or for the purpose of learning. A summary of definitions can be found in Table 2.1.

At the macro level, evaluative thinking has implications for society and evaluation as a field. Patton (2018) likens evaluative thinking to Frierean pedagogy, as a liberator of minds to

be critical questioners of the world. Patton (2018), Schwandt (2018), and House (2016) recognize that evaluative thinking promotes democratic values and spreads critical dialogue. Dahler-Larsen (2012) also contributed to this macro level thinking about evaluation and evaluative thinking. Declaring, “we live in the age of evaluation” (p.1), Dahler-Larsen (2012) argued that evaluative thinking has become intertwined with modern society. As evaluative thinking has an impact on society at large, there are also implications for evaluation as a discipline. Dedicated to reasoned evidence and critical thought, the field of evaluation is steeped in the values of evaluative thinking. Evaluative thinking can be considered a process unique to evaluation, providing a distinctive identity for the field. Schwandt (2018) wrote, what sets evaluative thinking apart from the kind of thinking demanded in other practices, is that it is about the reasoning involved in making evaluative judgments. In other words, evaluative thinking is what evaluators do and it involves constructing defensible judgments about the value of things, ideas, arguments, actions, and so on. (p. 127)

At the meso level, Schwandt (2018) asked us to think about evaluative thinking as a social practice. Acknowledging that thinking and reasoning are a social phenomenon, he challenged evaluators to focus on the group level and to recognize the influence of shared reasoning, cultural histories, systems, and norms. This social view of evaluative thinking is justified by three elements: first, a recognition that evaluation is increasingly a group effort, not just the actions of a single individual; second, research that shows sense making to be a “collective undertaking rooted in social interaction and negotiation” situated in specific contexts (Schwandt, 2018, p. 130-131); and third, the perception of evaluators’ responsibility to serve the greater public, not just their clients (Schwandt, 2018). Along this line, evaluative thinking has been considered important in building evaluation capacity at the organizational level (Bennett & Jessani, 2011; Buckley et al., 2015; Preskill & Boyle, 2008). Baker and Bruner, at the Bruner Foundation, take the position that evaluative thinking should be considered at the organizational level. In Baker and Bruner’s (2012) definition of evaluative thinking they wrote, “it is an approach that fully integrates systematic questioning, data, and action into the organization’s work practices” (p. 1). Their guide for nonprofit and philanthropic organizations,

Integrating Evaluative Capacity into Organizational Practice, discussed the importance of thinking evaluatively in all aspects of the organization from service delivery to human resources. The Bruner Foundation's Evaluative Thinking Assessment Tool assesses the evaluative thinking capacity of an organization in sixteen domains including leadership, governance, staff development, and finances. This detailed and wholistic instrument embodies the belief that evaluative thinking is an organizational endeavor, a habit of mind that can be infused into an organization's culture.

At the micro level, most definitions of evaluative thinking are at the level of an individual, what Schwandt (2018) refers to as "the evaluative thinker." Here, evaluative thinking is the cognitive process of an individual weighing their options and making decisions. Different definitions call evaluative thinking a process, a habit, a capacity, and a skill, all referring to an individualized undertaking (Buckley et al., 2015; Schwandt, 2018; Vo, 2013; Vo et al., 2018; Volkov, 2011). Even proponents of evaluative thinking at the level of organizations acknowledge its individualized nature. Buckley et al. (2015) explain "an organization that 'thinks evaluatively' might be better described as an organization in which people at all levels of the organization are evaluative thinkers and are engaged in the evaluation of what their organization does in some capacity" (p. 383). In some ways, this acknowledgement serves to bridge the meso and micro views.

At this micro, individual level, evaluative thinking can be conceptualized in two ways: as primarily a process undertaken by evaluators and as a habit of mind for evaluators and non-evaluators alike. The latter stance is generally held by those in the field of evaluation capacity building, acknowledging the benefits for clients, funders, program staff, and even those not directly tied to a formal evaluation (Archibald, Sharrock, Buckley & Young, 2018; Bennett & Jessani, 2011; Buckley et al., 2015; Preskill & Boyle, 2008). The former, evaluative thinking mainly as a process for evaluators, can be seen in the definitions but forth by Vo (2013). Through the use of a delphi study, Vo (2013) developed the following definition, characterizing evaluative thinking as "...the process by which one marshals evaluative data and evidence to construct arguments that allow one to arrive at contextualized value judgments in a transparent fashion" (p. 105). While Vo's definition is open to transference to evaluators and

non-evaluators alike, there is a preference for individuals conducting evaluations which require reasoned value judgments.

In Schwandt's (2018) explanation of this individualized approach to evaluation, he cites a definition which encapsulates the position of evaluative thinking as a process by evaluators presented by Eileen Stryker on EvalTalk. Stryker wrote,

thinking evaluatively is about *how we arrive at* or account for judgments about value or quality. Evaluative thinking means being tuned into value judgments people make, and questioning how those judgments were arrived at and what evidence may exist to substantiate the value claim. (as cited in Schwandt, 2018, p. 127)

These conceptualizations of the evaluator as the evaluative thinker reinforce ideas that evaluative thinking provides a unique identity for the field of evaluation. Vo (2013) specifically states that evaluative thinking is "... a particular kind of critical thinking and problem-solving approach that is germane to the evaluation field" (p. 105).

A third area of tension for defining evaluative thinking is the intended outcome or purpose of thinking evaluatively. The advocates of evaluative thinking in capacity building work tend to see the purpose as learning: "it can contribute to improved learning and decision making" (Baker & Bruner, 2012, p. 1); "evaluative thinking is learning for change" (Bennet & Jessani, 2011, p. 24); evaluative thinking can help "with processes of continuous reflection and learning" (Archibald et al., 2018, p. 80); evaluative thinking leads to "increased understanding and clarity" (Earl, 2013, p. 188). As many of these writers described evaluative thinking as a *habit of mind*, there are not always direct outputs or deliverables from the process other than learning and reflection. Alternatively, Vo (2013) saw the purpose of evaluative thinking to arrive at reasoned and evidence-based value conclusions. Patton (2018) noted that in a way, Scriven anticipated the modern concept of evaluative thinking "in explicating the logic of evaluation, and the reasoning processes that undergird reaching evaluative conclusions and rendering evaluative judgments" (p. 18). While this more traditional definition of evaluation requiring a value judgment is still acknowledged, alternative purposes of evaluation have also been accepted, including the purpose of program or organization improvement and the development of knowledge (Mark, Henry, & Julnes, 2000).

While these definitions differ in their implications, audiences, and intentions of evaluative thinking, they are reflecting different sides of a dice rather than fundamentally disagreeing. In their concluding chapter of the *New Directions for Evaluation* focused on evaluative thinking, Vo and Archibald (2018) made a point to explain that the intention of the issue was not to get the *right* definition of evaluative thinking, rather they were trying to “take a step back and start with simply trying to ‘get it,’ and to understand and appreciate evaluative thinking for what it is in practice, theory, and research” (p. 139). As they pointed out, a unified definition of evaluative thinking is not entirely necessary. As with many concepts in social sciences, a singular operationalized definition is not required for the concept of evaluative thinking to be useful in practice and research. Looking at these different sides of the dice, it is almost surprising that at their core, these varied views share some similar characteristics of what evaluative thinking entails.

Characteristics of Evaluative Thinking

Most definitions of evaluative thinking inevitably touch on three main characteristics: the necessity of reflective practice, critical thinking, and creative thinking. Patton (2018) expanded, writing that “rigorous evaluative thinking combines critical thinking, creative thinking, inferential thinking, and practical thinking” (p. 21). These analogies and comparisons have important implications in how evaluative thinking shows up in practice and how one might go about measuring the concept.

It is evident from the literature that whether a process or a habit of mind, evaluative thinking involves reflective practice. Baker and Bruner (2012) call evaluative thinking “a type of reflective practice” (p. 1), while other definitions refer to “pursuing deeper understanding through reflection” (Buckley et al., 2015) and an “ongoing process of questioning, reflecting, learning and modifying” (Bennett & Jessani, 2011, p. 24). Fierro and colleagues (2018) set out to identify indicators of evaluative thinking within an organizational setting. Through in-depth case studies they found twenty-two indicators of evaluative thinking, which they grouped into five main categories. The first of these categories is reflection. Fierro and colleagues (2018) defined reflecting as “deliberately giving critical attention to various aspects of a program, including its

context and its evaluation; suggests a willingness to apply a critical lens reflexively” (p. 59). This definition encompasses the ability to reflect on one’s own actions – building practical knowledge – as well as the ability to critically consider contextual factors.

Critical thinking frequently appears in definitions and discussions about evaluative thinking. Again, scholars define evaluative thinking in terms of critical thinking: evaluative thinking is “a particular kind of critical thinking” (Vo, 2013, p. 105); “evaluative thinking is critical thinking applied in the context of evaluation” (Buckley et al., 2015). Critical thinking’s importance in evaluative thinking is characterized by a general reverence for logical and evidenced reasoning, as well as attention to values and valuing. Vo (2018) acknowledged these influences when she compared the two; she wrote,

...making reasoned, *evidence-based* choices set critical thinking and evaluative thinking apart from judgments made based only on deeply held and unchallenged beliefs (e.g., stereotypes). Making a reasoned choice about *value* and being able to defend it is what distinguishes [evaluative thinking] from critical thinking. (p. 40)

She contends the inclusion of value is what sets evaluative thinking apart from critical thinking, but both make use of reasoned evidence in the decision making process. Others do not necessarily agree, noting that “evaluative thinking doesn’t exclusively, or even primarily, result in value judgments,” instead expressing the general purpose of learning (Buckley, personal correspondence). Critical thinking involves the cognitive skills of interpretation, analysis, inference making, and explanation (Facione, 1990; Schwandt, 2018), all at the core of evaluative thinking.

Evaluative thinking’s attention to values goes beyond crafting value judgements to also include attention and respect for differing cultural and social values. Wehipeihana and McKegg (2018) discussed the importance of acknowledging indigenous values and ways of knowing and incorporating these into the definition of evaluative thinking, rather than always relying on western perspectives. They also called on evaluators to acknowledge personal biases and assumptions that may influence decision making in ways that are not always aligned with rational thinking. Archibald, Neubauer, and Brookfield (2018) saw the integration of values into evaluative thinking as a way to integrate social justice. Recognizing the educational and

sociopolitical roots of critical theory, they argued “equating evaluative thinking with critical thinking can be more meaningful and useful – in a pragmatic, analytic, and political sense – if the emancipatory potential of criticality as a tool for social justice is readmitted to the equation” (p. 110). Similarly, Patton (2017, 2018) recognized the influence of Frierean principles as a pedagogy for evaluation. While certain approaches to evaluation are considered “value-driven”, such as those identified on the social justice branch of Mertens and Wilson’s (2012) evaluation theory tree, this conceptualization of evaluative thinking pushes all kinds of evaluations to consider the implications of social justice and other values in practice.

The third characteristic to emerge from the literature on evaluative thinking is the integration of creative thinking. While not as prevalent as reflective practice or critical thinking, Bennet and Jennasi (2011), Vo et al. (2018), and Patton (2018) recognized the role creativity plays in thinking evaluatively. Vo et al. (2018) recognized that evaluative thinking “reflects the ability to creatively arrive at evaluation-specific solutions by determining what combination of methods, tools, techniques, and theories are appropriate in light of contextual particulars” (p. 40). Later, speaking to the rough ground of practice, Vo et al. (2018) also wrote that evaluative thinking “reflects the ability to creatively navigate uncertainty, ambiguity, and complexity” (p. 44). Thinking evaluatively not only requires consideration of multiple perspectives (Fierro et al., 2018), but also the consideration of multiple options. Early psychometric theories of creativity equate creative thinking with divergent thinking, the ability to identify a plethora of options and pathways (Guilford, 1950; Mednick, 1962). Newell, Shaw, and Simon (1962) merged creative thinking and problem solving, speculating that creative individuals could identify a multitude of solutions to a problem and then compare them to isolate the best solution in a specific context. While confluence models of creativity (Sternberg & Lubart, 1996) recognize the multifaceted nature of creativity, the ability to creatively solve problems in the uncertain and complex contexts of evaluation certainly plays a role in evaluative thinking. Admittedly, the concepts of reflective practice, critical thinking, and creative thinking may not be mutually exclusive when it comes to evaluation practice. Vo and Archibald (2018) wrote “creativity clearly plays an important role in evaluative thinking, yet it remains unclear if these types of thinking (i.e., critical thinking, evaluative thinking, reflective practice, etc.) are partially overlapping circles in

a Venn diagram, or if one of them is a subset of the other, or if there is some other conceptual relationship between these clearly similar concepts” (p. 141). But recognizing creative thinking as distinct from reflective and critical thinking acknowledges scholars who portray evaluation as more art than science (Chouinard et al., 2016; Shadish, Cook, & Leviton, 1991).

Together, these characteristics paint the image of an evaluative thinker who is reflective of their own practices, responsive to contextual factors, critically inquisitive, committed to using evidence and logical reasoning to come to conclusions, attentive to values and valuing, and mindful of different options; *and* who further uses these skills and dispositions to creatively solve problems in ambiguous, uncertain, and complex environments (Archibald et al., 2018; Baker & Bruner, 2012; Bennett & Jessani, 2011; Buckley et al., 2015; Fierro et al., 2018; Schwandt, 2018; Vo, 2013; Vo et al., 2018; Wehipeihana & McKegg, 2018). All of these characteristics have also been considered qualities of good evaluators. Buckley et al. (2015) noted that evaluative thinking and good evaluation are one in the same. However, they also categorize evaluative thinking, “in combination with evaluation knowledge and skills”, as “essential for high-quality evaluation practice” (p. 378).

Summary

Both evaluation theory and evaluative thinking are, ideally, committed to the pursuit of good evaluation. While technical rigor and validity or credibility are agreed upon components of a good evaluation, scholars also emphasize the importance of utility, transparent reasoning, contextual responsiveness, attentiveness to values, and adherence to ethical practices (House, 1980; Scriven, 2012; Stufflebeam, 1999; Yarbrough, Shulha, Hopson, & Caruthers, 2010). The Committee on Standards for Educational Evaluation put forth the standards for evaluation as utility, accuracy, feasibility, propriety, and evaluation accountability (Yarbrough et al., 2010). Similarly, competencies for evaluators by both the Canadian Evaluation Society and the American Evaluation Association recognize the importance of using evidence and reasoning to come to a conclusion, responding to unique contexts, and attending to the utility of an evaluation (American Evaluation Association, 2018; Canadian Evaluation Society, 2010). Scholars of evaluation reiterate the intricacy of what makes a good evaluation (Chouinard et al.,

2006; House, 1980; Scriven, 1980, 1990; Shadish, Cook, & Leviton, 1991; Stake, 2004). The elements of a good evaluation are the same ones that evaluation theory attends to and attempts to instil.

Scriven's (1980) logic of evaluation builds a foundation for the argument that logical reasoning is at the core of evaluation (Fournier, 1995; Hurteau, Houle, & Mongiat, 2009). Using evidence and reason to construct arguments is part of evaluative thinking and plays a role in evaluators' decision making in practice (Hurteau, Houle, & Mongiat, 2009; Tourmen, 2009; Vo, 2013; Vo et al., 2018). In the absence of transparent and strong reasoning to justify conclusions, evaluations will not be taken seriously by stakeholders (Fournier, 1995; Fournier and Smith, 1993).

Additionally, the ability to respond to contextual factors – including values, stakeholder needs, resource restrictions, and environmental requirements – has been lauded as essential in good evaluation practice (Fitzpatrick, 2012; Kirkhart, 2010; Patton, 2012; Rog, 2012; Schwandt, 2014). Evaluative thinking attends to contextual responsiveness in its emphasis on reflective practice as well as critical awareness of situations and biases. Evaluation theory attempts to provide guidance for these differing conditions. Although, as Smith (1993) and Vo and Christie (2015) point out, evaluation theory could do a better job with further empirical study of evaluation practice in order to more fully understand the influence of context on practice. More research on evaluation contingency theories would better prepare theory for providing practitioners with decision maps (Miller, 2010; Shadish, 1998).

Creativity is not always discussed as a requirement for good evaluation, but it is clear that creativity helps differentiate excellent evaluation. Smith's (2016) idea of evaluative fit represents this concept well. Smith argued that eloquent evaluations fit the right approaches and methodologies for a particular context, in a way that becomes an expressive form of art. House's (1999) concepts of evaluation validity requiring truth, beauty, and justice also conjure this idea, as well as Cronbach's reference to evaluation as "art" and Chouinard's (2016) reference to it as a "craft".

Together, these elements of good evaluation seem to bind evaluation theory and evaluative thinking. While previous studies of evaluation theory in practice had alluded to its

role in evaluation thinking (Barela, 2005; Christie, 2003; Fitzpatrick, 2004; Tourmen, 2009), an empirical foundation for understanding this link is in the early stages. In this dissertation study, the concepts of logical reasoning, contextual responsiveness, and values and valuing are investigated as effects of evaluation theory.

The Current Study

The literature reviewed here on evaluation theory, decision making, evaluative thinking, and good evaluation have all substantially informed the creation of this study, as illustrated in the conceptual framework discussed in Chapter 1. This investigation sought to deepen understanding of training in evaluation theory by evaluators, as well as the role evaluation theory plays in the evaluative thinking of practicing evaluators. These were investigated in a two-phase mixed methods study. The first phase consisted of a cross-sectional survey of Canadian Evaluation Society and American Evaluation Association members to determine the prevalence of formal and informal training in evaluation theory. The second phase included in-depth interviews of evaluators to explore the factors that influenced their decision making in practice. The reviewed literature provided valued context and additional understanding during data collection and analysis.

Table 2.1. A comparison of evaluative thinking definitions.

Authors	Definition of Evaluative Thinking	Scope of Definition	Primary Thinker	Purpose of Evaluative Thinking
Preskill & Boyle (2008)	"about getting people in organizations to look at themselves more critically through disciplined processes of systematic inquiry ... about helping people ask these questions and then go out and seek answers" (p. 148)	Organizational	Non-evaluators	Learning
Bennett & Jessani (2011)	"evaluative thinking is ... an ongoing process of questioning, reflecting, learning and modifying. Evaluative thinking is an inherently reflective process, a means of resolving the 'creative tension' between our current and desired levels of performance. It allows us to define the lessons we want to learn, to determine the means for capturing those lessons, and to design systems to apply them in improving our performance. By going beyond the more time- and activity-bound processes of monitoring and evaluation, evaluative thinking is learning for change." (p. 24)	Organizational	Non-evaluators	Learning for change
Volkov (2011)	"a person's or organization's ability, willingness, and readiness to look at things evaluatively and to strive to utilize the results of such observations" (p. 38)	Individual and Organizational	Non-evaluators	Rationed decision making
Vo (2013)	"... a particular kind of critical thinking and problem-solving approach that is germane to the evaluation field. It is the process by which one marshals evaluative data and evidence to construct arguments that allow one to arrive at contextualized value judgments in a transparent fashion" (p. 105)	Individual	Evaluators	Reasoned, evidence-based value judgment

Authors	Definition of Evaluative Thinking	Scope of Definition	Primary Thinker	Purpose of Evaluative Thinking
Earl (2013)	"an inquiry habit of mind ... [in which individuals] consider a range of evidence and keep searching for increased understanding and clarity by engaging in a spiral of systematic analysis of the situation, professional reflection, action and reanalysis" (p. 188)	Individual		Learning
Baker & Bruner (2012)	"Evaluative thinking is a type of reflective practice that uses the same five evaluation skills listed above in other than programs, strategies, and initiatives. It is an approach that fully integrates systematic questioning, data, an action into an organization's work practices. It can contribute to improved learning and decision making" (p. 1)	Organizational	Non-evaluators	Learning and decision making
Patton (2018)	"People must also know how to use the information, that is, to weigh evidence, consider the inevitable contradictions and inconsistencies, articulate values, interpret findings, deal with complexity, and examine assumptions, to note but a few of the things meant by 'thinking evaluatively'" (p. 16) "rigorous evaluative thinking combines critical thinking, creative thinking, inferential thinking, and practical thinking" (p. 21)	Individual	Evaluators	Evaluative conclusions
Eileen Stryker (n.d.)	"thinking evaluatively is about how we arrive at or account for judgments about value or quality. Evaluative thinking means being tuned in to value judgments people make ... and questioning how those judgments were arrived at and what evidence may exist to substantiate the value claim"	Individual	Evaluators	Judgements about value or quality

CHAPTER 3

Methodology

It is the intention of this dissertation to explore the connection between evaluation theory and evaluative thinking. While previous studies have looked at evaluation theory in practice, few have taken a qualitative, exploratory approach to understand the nuanced connections throughout the cognitive process. In this dissertation, a mixed methods approach was taken in order to answer the two main research questions: (1) to what extent and in what ways are evaluation practitioners in the AEA and CES communities trained in evaluation theory? and (2) what role does evaluation theory play in evaluative thinking? This chapter will review the research design used to answer these questions, along with details about sampling, instrumentation, and analysis.

As evaluation is an applied practice, many theorists have suggested the use of qualitative studies to further explore and understand the connection between theory and practice (Alkin, 2003; Christie, 2003; Henry & Mark, 2003). Existing studies of evaluation theory have used different qualitative approaches, including in-depth interviews with practitioners to describe their practice and illuminate different theoretical influences on their practice (Fitzpatrick, 2004; Fitzpatrick, Christie, & Mark, 2009; Barela, 2005); direct observations of evaluators in practice (Barela, 2005); and simulated case studies in hopes of reproducing practice in action (Alkin & Christie, 2005; Tourmen, 2009). Other approaches to studying evaluation theory have included reflective case studies (e.g., Bledsoe & Graham, 2005); surveys of theorists and practitioners (Christie, 2003; Harnar, 2012; Shadish & Epstein, 1987; Williams, 1989); and systematic reviews of evaluation reports or case studies (Coryn, et al., 2011; Hurteau, Houle, & Mongiat, 2009; Miller & Campbell, 2006). This dissertation adds to the literature on connecting evaluation theory to practice, while simultaneously examining the extent of training in evaluation theory among current evaluation practitioners.

Overall Research Design

The mixed methods approach employed in this study consists of a quantitative phase followed by a qualitative phase (Tashakkori & Teddlie, 2003; Creswell & Plano Clark, 2018). This

study is considered a developmental mixed method design due to the use of a quantitatively oriented survey to construct the sampling frame for qualitative interviews (Greene, Caracelli, and Graham, 1987). Greene, Caracelli, and Graham (1987) defines a developmental mixed method design as one that “seeks to use the results from one method to help develop or inform the other method,” including for sampling purposes. For the purpose of this study, quantitative data is defined as information that is categorized into numerical values for the purpose of answering generalizable questions like “what” and “how many.” Qualitative data, on the other hand, is considered descriptive qualities or experiences about the lived experience of participants. These data answer questions about understanding and describing. This study mixes quantitative and qualitative data based on methods and methodologies, not on the level of paradigms. Overall, this study takes a predominately post-positivist stance throughout. Inquiry logics stemming from the research questions called for mixed methods methodology (Greene, 2008). The first research question asked for generalized proportions of a population, preferring a more quantitative approach, while the second research question asked for description of experiences and motivations in practice, identifying the lived experience of evaluators, favoring a qualitative approach. Quantitative and qualitative data are mixed during the interpretation stage, allowing comparisons between the lived experiences of groups (collected through interviews) based on survey responses. A mixed method approach was used in order to provide more insightful understanding and comprehensive findings, not to triangulate data across paradigms on the same constructs (Greene, et al. 2001).

The sequential quantitative then qualitative method placed a greater emphasis on the qualitative study. Mixed methods designs are typically explained using notation to convey the timing and weighting of each method (Creswell & Plano Clark, 2018; Tashakkori & Teddlie, 1998). While the design of this study does not fit neatly into the major mixed methods designs outlined by Creswell and Plano Clark (2018), using the same notation format the design of this study could be written as

Quan → QUAL

The appearance of quantitative prior to the arrow, followed by qualitative, represents the sequential order of the methods. The capitalization of the qualitative method indicates the

major emphasis of the findings was derived from the qualitative data. While this sequence is the same as an explanatory sequential design, this study's design was not for the explicit purpose of explaining the initial quantitative results in more depth (Creswell & Plano Clark, 2018). The design of this study was closer to a development mixed method design as explained by Greene, Caracelli, and Graham (1989) because the quantitative phase was used in order to develop a sampling frame for the qualitative phase. The quantitative data informs the discussion and implications of the qualitative data, but findings from the two phases were not formally integrated as in other mixed methods designs (Creswell & Plano Clark, 2018; Tashakkori & Teddlie, 1998). The rationale for this approach is that the quantitative data provides an understanding of the extent evaluation theory is studied by evaluators (mainly answering research question #1), while the qualitative data allows for a deeper understanding of the motivations behind decision making in practice (mainly answering research question #2).

In order to fully discuss the details of each phase of this study, the quantitative phase and the qualitative phase will be discussed separately and sequentially, mirroring the order of the study.

Quantitative Phase

Design

The quantitative phase consisted of a cross-sectional survey of American Evaluation Association (AEA) and Canadian Evaluation Society (CES) members.

Population and Sampling

The intended population for this research was evaluators in North America. Given the difficulty in identifying everyone who conducts evaluation work, membership in the American Evaluation Association (AEA) or Canadian Evaluation Society (CES) served as a proxy for this population. The AEA membership is diverse, consisting of people who not only conduct evaluation, but also those who manage, fund, or oversee evaluation efforts. In order to limit the sampling frame to those who practice evaluation, only those who identify themselves as an "evaluator" in their AEA profile were included in the sampling frame. This was intended to

increase the response rate as well as reduce oversampling of AEA members who do not conduct or carry out evaluations. This suggestion comes from a study by Coryn et al. (2019) in which respondents indicated that they do not answer most surveys sent to the AEA community because the surveys do not pertain to their work. Some respondents noted they cannot adequately provide responses to the survey questions because they are not directly involved in evaluation practice. With these criteria, the AEA sampling frame included approximately $N = 3,078$ (as reported in Coryn et al., 2019; exact counts for the final sampling frame were not provided by AEA representatives). The CES membership is diverse in the languages spoken by its membership. Only English speaking CES members who have agreed to be contacted regarding survey research were included in the sampling frame, for a final sampling frame of $N = 634$ of CES members.

Two simple random samples, one of AEA members and another of CES members, were drawn to construct the final survey sample. Necessary sample sizes were calculated to estimate proportions using equations from Schaeffer, Mendenhall, Ott, and Gerow (2012. p. 93), conservatively assuming a 50% chance of having training in evaluation theory and with population sizes of $N_{AEA} = 3,078$ and $N_{CES} = 634$. An additional twenty-percent oversampling was added in order to account for an anticipated low response rate, making the final drawn samples of $n_{AEA} = 425$ and $n_{CES} = 313$.

Instrumentation

Items for the quantitative survey questionnaire were developed by the researcher for the purpose of answering research question #1 and creating a sampling frame for the qualitative interviews to address research question #2. The survey was constructed with the intention of eliciting information regarding respondents' (1) extent of training in evaluation theory, (2) type of training in evaluation theory, (3) familiarity with evaluation theory, and (4) extent of evaluation practice.

Both formal and informal training and education in evaluation theory was considered in the survey instrument. Formal training or education referred to learning opportunities that were conducted by a trained educator or content expert in the form of short webinars or

lectures, workshops or professional development training, and courses at a college or university. Informal training refers to self-study, casual conversations, or reading theoretical texts. For the purpose of this instrument, evaluation theory was operationalized and defined as “a coherent set of conceptual, hypothetical, pragmatic, and ethical principles forming a general framework to guide the study and practice of program evaluation” (Stufflebeam & Coryn, 2014, p. 50). With the added explanation that evaluation theory, models, and approaches provide explicit guidelines on how evaluation should or could be practiced. Finally, extent of practice was defined by years of practicing evaluation and number of evaluations worked on in the past two years.

Through a review of previous literature utilizing surveys to assess similar topics (Christie, 2003; Harnar, 2012; Shadish & Epstein, 1987), a list of relevant questions was compiled. These questions were then modified in order to fit the specific purpose and context of this study. The final instrument consists of 43 questions, of which 36 are close-ended and seven are open-ended. Twenty-six questions addressed respondents’ extent and type of training in evaluation theory, two questions addressed familiarity with evaluation theory, nine addressed the extent and type of evaluation practice, and six questions were about evaluation theory in practice. The survey was reviewed by twelve experts in evaluation theory and/or survey development to assess the clarity of questions and assess the face and content validity of the instrument. Finally, cognitive interviews were conducted with six evaluators with varying degrees of training in evaluation theory. Survey questions were altered according to the results of the expert reviews and cognitive interviews. The final survey questionnaire resulting from this process is provided in Appendix A.

Procedure

The survey was administered online using the Qualtrics Survey Software platform. Members of the AEA sample received an invitation to participate in the survey from the author’s University email address on December 3, 2018. They received three follow up reminders on December 12, December 17, and January 7. Mail merge from Outlook was used for all survey communication, as opposed to using the Qualtrics email dissemination platform in

an attempt to prevent survey invitations being blocked by firewalls or spam filters. Responses were tracked each week, with reminders only being sent to non-respondents. Members of the CES sample received an invitation to participate in the survey directly from a CES representative on December 3, 2018. They received one follow up reminder on December 17, again from a CES representative. Due to privacy restrictions, CES maintained control over administering survey links instead of invitations being sent and monitored by the author. The survey for both AEA and CES members closed on January 12, 2019. Once a respondent completed the survey, an automatic notification was sent to thank the participant for their time.

As a small token of appreciation and incentive for participation, survey respondents were provided with a summary of survey results. In a survey of AEA members conducted by Coryn et al. (2019) providing a summary of results was noted as one of the top motivations for responding to surveys.

Sample Characteristics

The final, usable sample for AEA members (after 8 undeliverable email addresses were removed) was $n_{AEA} = 417$ and $n_{CES} = 313$ for CES members. As CES survey invitations were sent directly from a CES representative no information is known about undeliverable emails. From the random sample of AEA members who identified as evaluators in their membership profile, a response rate of 27.5% ($n = 115$) was obtained. From the random sample of CES members who were English speaking and opted to participate in external research requests, a response rate of 30.0% ($n = 94$) was obtained. Combined, the overall response rate was 28.6%. Shown in Table 3.1 are some characteristics of the overall sample as gathered through the survey responses. Demographics of the sample in comparison to the overall population are available for CES members as shown in Appendix B. Overall, the characteristics of the CES sample matched closely with the CES population, making a strong case for the generalizability of survey results to the whole association. A comparison of sample demographics to the overall population is unavailable for AEA members, as this additional information was not provided by AEA representatives after being requested multiple times.

Table 3.1. Overall survey sample characteristics.

Characteristic	<i>n</i>	Percent	Valid Percent
Highest Degree of Evaluation Education			
Professional development workshops	39	19%	21%
Badges or certifications in evaluation	6	3%	3%
Graduate certificate in evaluation	9	4%	5%
Master's Degree in evaluation	12	6%	7%
Master's Degree in another field	66	32%	36%
Doctorate Degree in evaluation	12	6%	7%
Doctorate Degree in another field	39	19%	21%
Post doctorate degree	1	0%	1%
<i>Missing</i>	25	12%	
Hold a Credentialed Evaluator Designation			
Yes	28	13%	15%
No	162	78%	85%
<i>Missing</i>	19	9%	
Self-Identified Evaluation Proficiency			
Novice	25	12%	15%
Intermediate	63	30%	38%
Advanced	58	28%	35%
Expert	22	11%	13%
<i>Missing</i>	41	20%	
Evaluation Practitioner			
Yes	169	81%	89%
No	21	10%	11%
<i>Missing</i>	19	9%	
Years Practicing Evaluation			
5 years or less	63	30%	38%
6 - 10 years	39	19%	23%
11 - 20 years	40	19%	24%
21 years or more	25	12%	15%
<i>Missing</i>	42	20%	
Evaluations Conducted in Past Two Years			
5 or less	96	46%	60%
6 through 10	41	20%	26%

Characteristic	<i>n</i>	Percent	Valid Percent
10 or more	22	11%	14%
<i>Missing</i>	50	24%	
Area of Practice			
Nonprofit, non-governmental organizations, or foundations	94	45%	
Government evaluation	70	33%	
Public health	54	26%	
Higher education	47	22%	
Pre-K to 12 education	36	17%	
Another field	34	16%	
International development	26	12%	
Business and industry	12	6%	
Criminal justice	9	4%	

Data Processing and Analysis

Once the survey was closed, the data were downloaded as a CSV file and stored on the author's laptop in a password protected Microsoft Excel file. Data were de-identified and imported into the SPSS Statistical Software by IBM for analysis. Analysis of these data mostly consisted of descriptive statistics, in the form of percentages and averages, in order to address research question #1 (i.e., to what extent and in what ways are evaluation practitioners in the AEA community trained in evaluation theory?). Open-ended responses were analyzed through an emergent coding process using MAXQDA qualitative data analysis software.

Qualitative Phase

Design

The qualitative phase of the study consisted of one-on-one in-depth, semi-structured interviews with a purposeful sample of evaluators who responded to the survey in the first phase of the study.

Population and Sampling

The sampling frame for interviews was constructed from participants' responses to questions in the survey. It was the original intention to draw a purposeful sample from survey respondents (Patton, 2002) in order to identify participants based on criteria of training in evaluation theory (those who have had training and those who have not) and extent of practical experience (length of time practicing evaluation). Practical experience was a consideration in sampling for interviews because previous studies have pointed to the important influence of practical knowledge in decision making in evaluation (Fitzpatrick, 2004; Tourmen, 2009). At the end of the survey, respondents were asked whether they would be willing to participate in follow up interview; $n = 57$ respondents agreed (24 AEA members and 33 CES members). Using their survey responses, these participants were divided into four groups using the criteria of training in evaluation theory and extent of practical experience ($n_1 = 8$; $n_2 = 26$; $n_3 = 14$; $n_4 = 9$). Proportional samples were drawn for initial invitations to participate in interviews. Initially, twenty respondents were invited to participate in interviews. However, low response rates eventually led to all 57 of those who volunteered being invited to participate in interviews.

Instrumentation

The protocol for the interviews was constructed with the purpose of addressing research question #2 (i.e., what role does evaluation theory play in evaluative thinking?). The interview protocol was intended to be a semi-structured interview so as to allow for flexibility in the interviewing process (Charmaz, 2014). Nine questions asked interviewees to recall an in-depth example of an evaluation from their past that they were particularly proud of. This set of questions asked interviewees to reflect on the setting and context, approaches or methods, stakeholder involvement, justification of evaluative argument, interpretation and support of evaluative evidence, and reporting of evaluative conclusions. Follow up questions explored the factors influencing and motivating the choices behind these decisions. The next set of seven questions asked interviewees directly about the role evaluation theory played in their evaluation practice. While the first set of questions explored motivations behind practice

without an explicit mention of evaluation theory to determine whether interviewees volunteer this information, the second set of questions asked about these connections directly.

For the purpose of this study, evaluative thinking was operationalized around three main concepts in the cognitive process by which evaluators make decisions in their practice: (1) contextual responsiveness, (2) the inclusion and consideration of values and valuing, and (3) the use of logical reasoning to justify evaluative evidence and conclusions (Vo, 2013). Given that the second research question asks about the role of evaluation theory in evaluative thinking, interview questions were constructed in order to elicit motivating factors underlying these three areas of thinking.

In order to assess the clarity and face validity of questions, the interview protocol was reviewed by seven evaluation experts and piloted with three evaluators from the sample with varying levels of training in evaluation theory and extent of evaluation practice. Questions were revised according. The final interview protocol is provided in Appendix C.

Procedure

Participants invited to contribute to the qualitative phase were contacted via the email address they provided in their survey responses. This email invitation included an explanation of the study purpose, limitations, and benefits. Participants were asked to review informed consent documents prior to scheduling an interview. Interviews were conducted either via video conference using the Zoom Video Communications or via telephone. All interviews were recorded digitally through the Zoom platform or by a digital recording device. Audio files were stored in a password protected file on a password protected computer. Names and contact information for interviewees were not connected to the audio files.

Sample Characteristics

Of the 57 evaluators invited to participate in an interview, 23 did not respond to the email invitation, 10 were unavailable during the time frame, four scheduled an interview but canceled without ever rescheduling, and 20 participated in interviews. One interviewee's audio recording was corrupted, making it unusable. Therefore, 19 usable interviews were included in

the final analysis. The sampling process and response rates for both the survey and interview phases are depicted in Figure 3.1. General characteristics of interviewees are shown in Table 3.2.

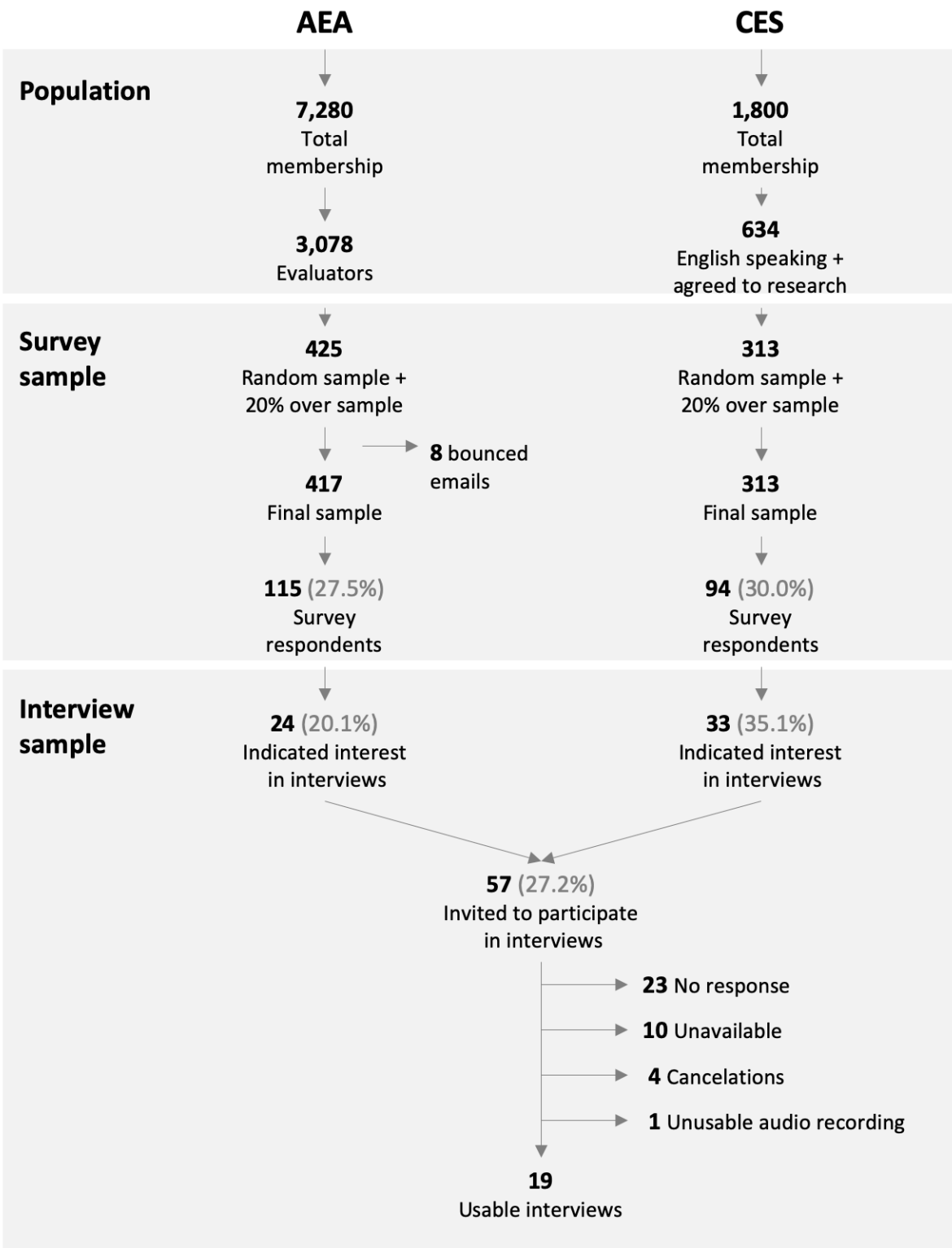


Figure 3.1. Sampling procedures and final response rates for the survey and interview phases.

Table 3.2. Overall interview sample characteristics.

Characteristics	<i>n</i>	Percent
Association		
AEA	6	32%
CES	13	68%
Highest Degree in Evaluation Education		0%
Professional development workshops, badges, certifications, or graduate certificate in evaluation	7	37%
Master's in evaluation or related field	6	32%
PhD in evaluation or related field	6	32%
Holds a Credentialed Evaluator Designation		0%
Yes	2	11%
No	17	89%
Self-Identified Evaluation Proficiency		
Novice evaluator	1	5%
Intermediate evaluator	7	37%
Advanced evaluator	8	42%
Expert evaluator	3	16%
Evaluation Practitioner		
Yes	19	100%
No	0	0%
Years Practicing Evaluation		
5 years or less	6	32%
6 – 10 years	6	32%
10 years or more	9	47%
Evaluations Conducted in Past Two Years		
5 or less	11	58%
6 through 10	6	32%
10 or more	1	5%
Area of Practice		
Nonprofit, non-governmental organizations, or foundations	11	58%
Public health	9	47%
Government evaluation	8	42%
Higher education	6	32%
Pre-K to 12 education	5	26%
International development	5	26%
Other field	4	21%

Characteristics	<i>n</i>	Percent
Business and industry	3	16%
Criminal justice	0	0%

Data Processing and Analysis

One audio file became corrupted, forcing this interview to be excluded from final analysis. All other interview files were transcribed into text files using the online transcription service Rev.com. Transcribed files were analyzed through multiple recurrent rounds of coding using the computer-assisted qualitative analysis software program MAXQDA. The initial coding cycle used process coding with the intention of following the thought process used to make decisions in evaluation practice (Charmaz, 2014; Saldaña, 2016). Focused coding was then used to aggregate initial codes into themes, or larger constructs of meaning (Charmaz, 2014). Special attention was paid to the a priori concepts of contextual responsiveness, values and valuing, and logical reasoning in order to identify the extent to which these were present in each interviewee's evaluation practice, in alignment with the sub-questions of research question #2. Data collected about the interviewees from the quantitative phase, including their training in evaluation theory and extent of evaluation practice, were integrated into the qualitative data in order to compare results across theoretical knowledge and practical knowledge. Patterns were compared across those with high, low, and average familiarity with evaluation theory, as well as extent of evaluation practice to determine if differences across these groups are present. Emergent themes and findings are reported in Chapter 4 of this dissertation.

Positionality

The process of interpreting and coding qualitative data rests on the assumptions and implicit biases of the researcher. Therefore, it is important to identify my background as it relates to demographics, education, evaluation practice, and interest in evaluation theory. I identify as an English-speaking, American-born, white woman with an educational background in sociology and graduate training in an interdisciplinary evaluation program. My work at

Western Michigan University has provided me with six years of contributing to, designing, and conducting evaluations at the local, national, and international levels. My epistemological assumptions tend to be post-positivist with a leaning towards pragmatic in my evaluation work. I value a mixed-methods approach in my evaluation practice and research. I have completed in-person, semester long, doctoral level courses focused on evaluation theory, along with self-study of many theoretical texts. As I have had a particular interest in evaluation theory, I am very familiar with major evaluation theories, along with some lesser known theories. My long-term commitment to evaluation theories is evident in my two years as Program Chair, in addition to two years as TIG Chair of the AEA Theories of Evaluation Topical Interest Group. Additionally, I have experience teaching evaluation theories to others through a graduate-level course on evaluation theories at Western Michigan University as well as multiple pre-conference workshops at the American Educational Research Association. These experiences knowingly and unknowingly inform this work.

CHAPTER 4

Findings

In this chapter, I present the quantitative and qualitative results obtained from the surveys and interviews described in Chapter 3. The implications and limitations of these results are discussed in Chapter 5. The research questions for this dissertation study were:

3. To what extent and in what ways are evaluation practitioners in the AEA and CES communities trained in evaluation theory?
 - a. What types of training in evaluation theory are practitioners receiving, if any?
 - b. To what extent are evaluation practitioners educating themselves in evaluation theory outside of formal training?
4. What role does evaluation theory play in evaluative thinking?
 - a. How do patterns of evaluative thinking differ in evaluators with training in evaluation theory compared to those without?
 - b. How do evaluators with training in evaluation theory differ in their patterns of thinking in relation to: contextual responsiveness, logical reasoning, valuing, and values.

Findings in this chapter are presented in two major sections in alignment with the research questions – training in and familiarity with evaluation theory and the role of evaluation theory. The first section presents findings from the survey regarding AEA and CES members' training in evaluation theory and familiarity with various approaches. The second section presents findings mainly from interviews with evaluators. This section first describes the explicit role evaluators described evaluation theory playing in their practice and evaluative thinking, then discusses the implicit role of evaluation theory as described by their reflections on practice. The implicit role of evaluation theory is organized around the four components of sub-question b: contextual responsiveness, logical reasoning, valuing, and values. These dimensions were identified in Chapters 1 and 2 as primary components of high-quality evaluation and evaluative thinking.

Training in and Familiarity with Evaluation Theory

Evaluation theory is frequently acknowledged in textbooks, professional discussions, and is even part of the AEA and CES evaluator competencies. However, the extent to which evaluators are receiving training in evaluation theory is largely unknown. According to Christie (2003), only 10% of evaluators in her research sample named an evaluation theorist they used in their practice. This has led many to presume that practicing evaluators' familiarity with evaluation theory is low. This research is the first generalizable study to look at training in and familiarity with evaluation theory since Shadish and Epstein (1987). In this section, findings from the cross-sectional survey of AEA and CES members around their training in and familiarity with evaluation theory are presented.

Training in Evaluation Theory Among AEA and CES Evaluators

Eighty percent (80%) of AEA and CES evaluators have received formal training in evaluation-specific theories, models, or approaches¹. For the purpose of this survey, evaluation-specific theories, models, or approaches were defined as “a coherent set of conceptual, hypothetical, pragmatic, and ethical principles forming a general framework to guide the study and practice of program evaluation” (Stufflebeam & Coryn, 2014, p. 50). With the added explanation that evaluation theory, models, and approaches provide explicit guidelines on how evaluation should or could be practiced. Those who reported attending formal training tended to self-identify as advanced or expert, identify a preferred evaluation approach, hold a graduate degree or higher, obtain the CES credentialed evaluator designation, or serve as an internal evaluator at a higher rate than those who did not receive formal training (see Table 4.1).

¹ Results in this section were obtained from a random sample of AEA and CES members with a overall response rate of 29%. More details, including a comparison of sample characteristics to population characteristics can be found in Chapter 3.

Table 4.1. Comparison of respondents who attended formal training in evaluation theories, models, or approaches and those who did not.

	Attended formal training in evaluation theories (<i>n</i> = 167)	Did not attend formal training in evaluation theories (<i>n</i> = 41)
Self-identified as advanced or expert	51%	35%
Identified a preferred evaluation approach	59%	44%
Graduate degree or higher	74%	61%
CES Credentialed evaluator	17%	5%
Internal evaluator	39%	29%
Evaluation practitioner	90%	85%
External evaluator	31%	47%

Approximately half (56%) of all respondents reported they attended a short webinar or presentation on evaluation theories, while half (50%) reported attending at least one graduate-level class on evaluation theories and slightly less than half (43%) reported attending a professional development workshop of 1 – 4 days in length. As shown in Figure 4.1, far fewer attended professional development courses between 1 – 2 weeks long (13%) or undergraduate courses focused on evaluation theories (7%). The next sections describe these modalities of training in more detail.

Most evaluators received training in evaluation theory from a short webinar or presentation (56%), graduate course (50%), or 1 – 4 day PD workshop (43%).

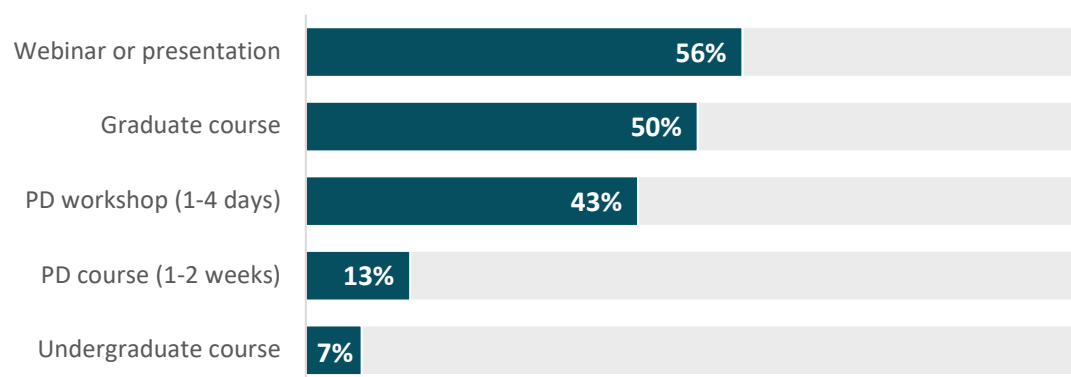


Figure 4.1. Type of formal training in evaluation theories, models, and approaches and the percent of respondents who reported attending each type (*n* = 209).

Short Webinars and Presentations on Evaluation Theory

Short webinars and presentations on evaluation theory were the most frequently reported modes of training by evaluators (56% of all survey respondents). These short presentations were typically conference presentations (75%) or AEA sponsored webinars (75%) (see Figure 4.2). Other short webinars and presentations attended by evaluators were sponsored by employees, private groups, University lecture-series, or local affiliate groups. As shown in Figure 4.3, more than half (56%) of evaluators reported the webinars or presentations they attended focused on a single evaluation theory as opposed to multiple theories.

Short webinars and presentations on evaluation theory tended to be conference presentations or AEA sponsored webinars.

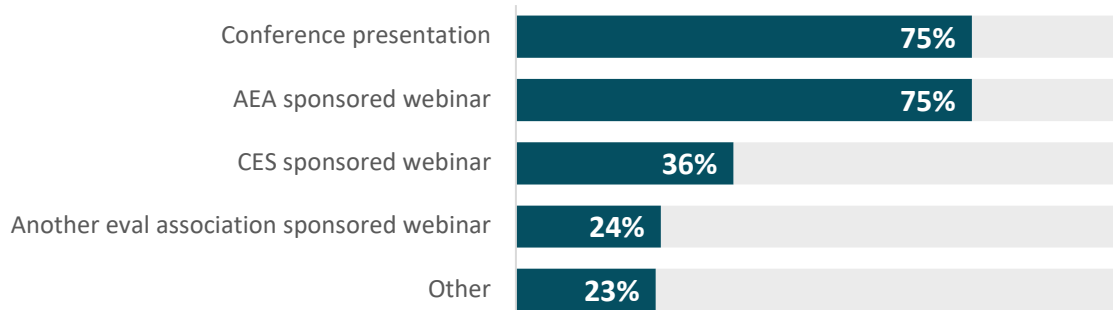


Figure 4.2. Type of short webinar or presentation in evaluation theories, models, and approaches attended by respondents ($n = 117$).

Fifty-six percent of webinars and presentations attended focused on a single evaluation theory.

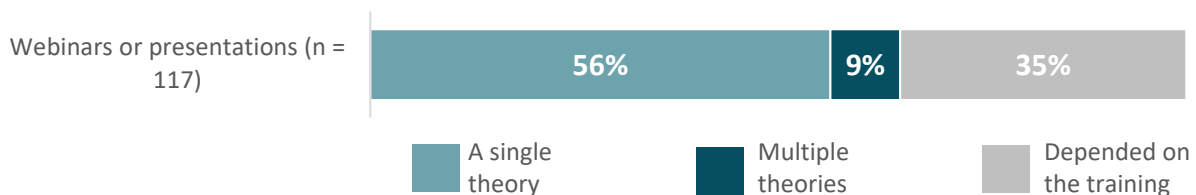


Figure 4.3. Percent of webinars or presentations on evaluation theory that focused on a single theory, multiple theories, or whether it depended on the training ($n = 117$).

Professional Development Workshops and Courses on Evaluation Theory

Forty-three percent (43%) of evaluators indicated they have attended a professional development workshop between 1–4 days in length. Of these respondents, 38% indicated these professional development workshops focused solely on evaluation theory, while 43% indicated evaluation theory was a major focus (see Figure 4.4). These professional development workshops rarely focused on more than five theories (2%) and were predominately in-person (71%) (see Figures 4.5 and 4.6).

Thirteen percent (13%) of evaluators indicated they have attended a professional development course between 1–2 weeks in length. Of these respondents, only fourteen percent (14%) indicated these professional development courses focused solely on evaluation theory, while 64% indicated evaluation theory was a major focus (see Figure 4.4). Professional development courses attended by respondents generally focused on more theories than shorter professional development workshops (61% covered between 2 and 5 theories; see Figure 4.5). The majority of these professional development courses were conducted in-person (61%), however, 18% were conducted online and 21% conducted both online and in-person (see Figure 4.6).

Evaluation theory was only the sole focus of trainings 38% or less of the time.

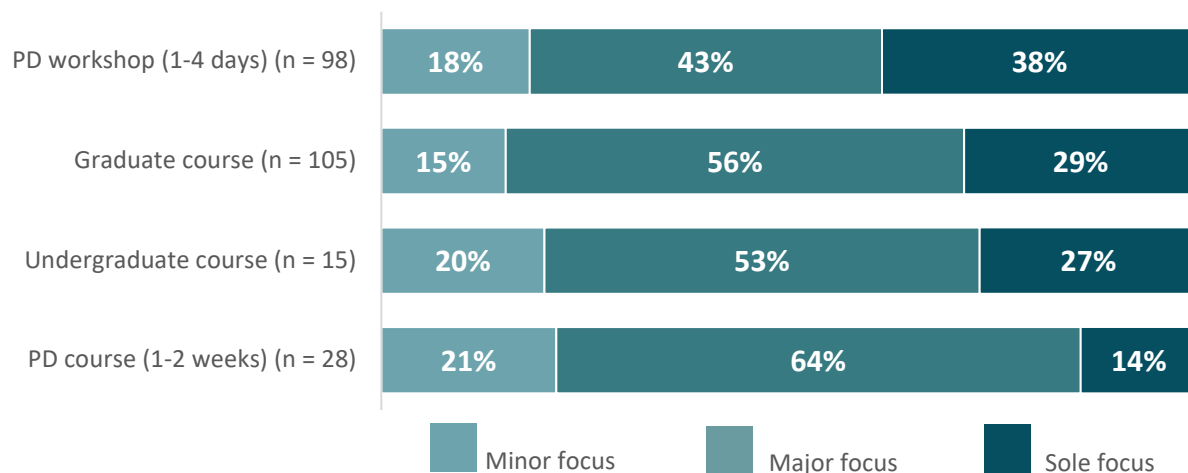


Figure 4.4. Training types and whether evaluation theory was considered the sole, major, or minor focus of the training.

University courses tended to focus on multiple theories, while professional development workshops were more likely to focus on a single evaluation theory.

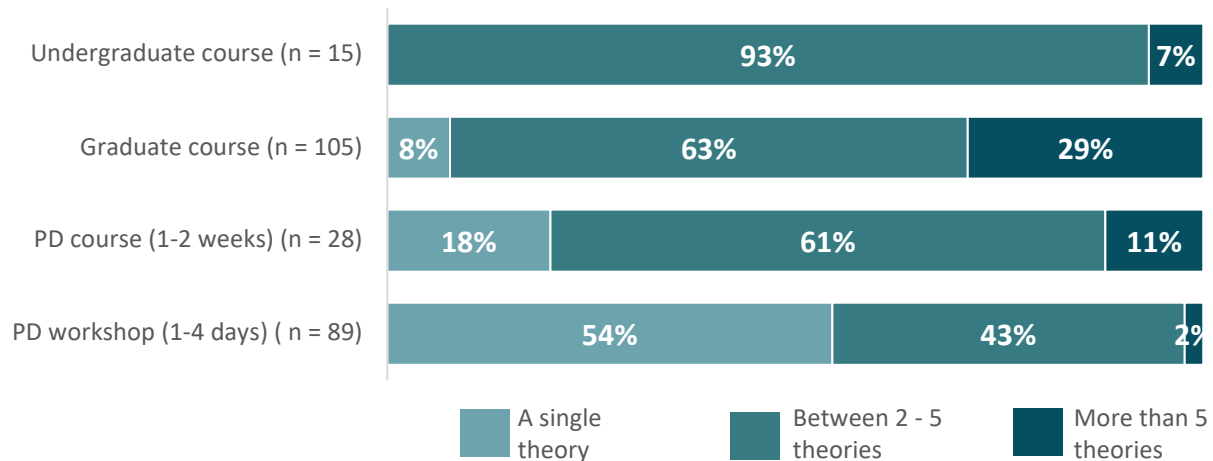


Figure 4.5. Training types and whether they focused on a single evaluation theory, between 2 – 5 theories, or more than 5 theories.

Most trainings on evaluation theory were conducted in-person.

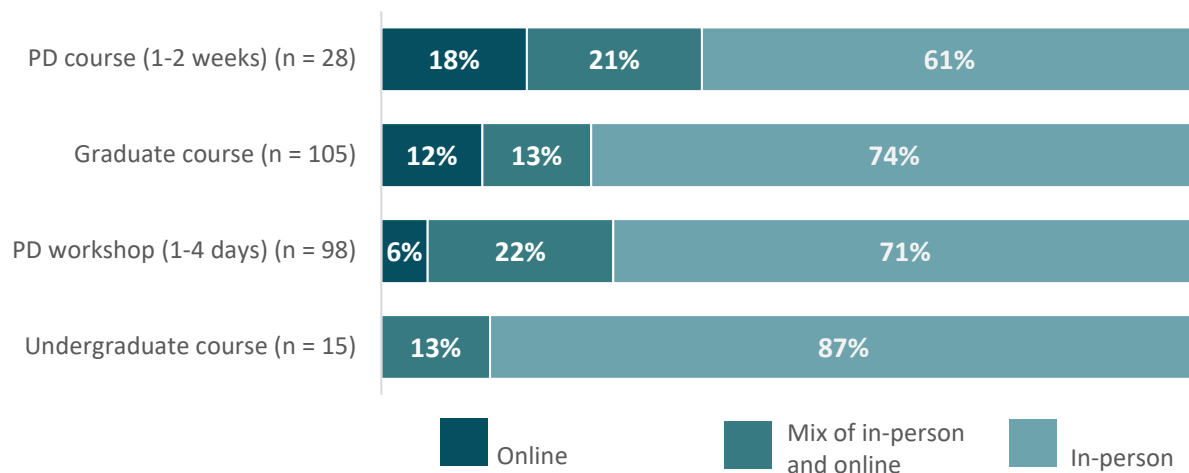


Figure 4.6. Instruction delivery platform for training types, whether online, mix of in-person and online, or in-person only.

University Courses on Evaluation Theory

Half of survey respondents (50%) indicated they attended at least one graduate-level course on evaluation theory, while only 7% indicated they attended an undergraduate-level course on evaluation theory. Of the undergraduate courses on evaluation theory, approximately half (53%) reported evaluation theory was a major focus and almost all (93%) said their undergraduate course focused on between 2 and 5 theories. The majority of undergraduate courses were in-person (87%).

Of the graduate-level courses on evaluation theory, approximately half (56%) reported evaluation theory was a major focus and almost all (91%) said their graduate course focused on more than one evaluation theory. Seventy-four percent (74%) of respondents indicated their graduate course was in-person, with 12% online and 13% a mix of in-person and online.

Informal Training in Evaluation Theory

Seventy percent (70%) of all AEA and CES evaluators indicated they participated in self-study outside of formal training on evaluation theories, while 30% indicated they had received mentoring around evaluation theories. Most evaluators who reported doing self-study on evaluation theory kept up with published literature from journals like the *American Journal of Evaluation* and the *Canadian Journal of Program Evaluation*, reading textbooks, as well as using online resources such as Better Evaluation and USAID. Evaluators who reported receiving mentoring in evaluation theory obtained mentoring from people such as their graduate advisors, peer learning circles, or senior colleagues.

Summary

Overall, 80% of the AEA and CES evaluators who responded to this study's survey reported receiving formal training in evaluation-specific theories, models, or approaches. The most dominant modality for this training was webinar or short presentations (56%) and graduate level courses (50%). Most trainings had a major or sole focus on evaluation theory and were conducted in-person. The longer the training modality, the more evaluation theories were

generally covered during the training, with 54% 1—4 day professional development workshops covering only a single theory.

Familiarity with Evaluation Theory Among AEA and CES Evaluators

One-third (33%) of AEA and CES evaluators reported being somewhat familiar with only a few evaluation theories. Twenty-six percent (26%) reported being somewhat familiar with many and twenty percent (20%) reported being very familiar with only a few. Eighteen percent (18%) of AEA and CES evaluators reported being very familiar with many evaluation theories. Only three percent (3%) of evaluators are not familiar with evaluation theory at all (see Figure 4.7). The most well-known evaluation theories are participatory evaluation (94% are at least somewhat familiar), developmental evaluation (89% are at least somewhat familiar), and utilization-focused evaluation (85% are at least somewhat familiar). The least well-known evaluation theories are deliberative democratic evaluation (43% are at least somewhat familiar), consumer oriented (42% are at least somewhat familiar), and constructivist or fourth-generation evaluation (31% are at least somewhat familiar) (see figure 4.8).

Most evaluators (97%) are at least somewhat familiar with a few evaluation theories.

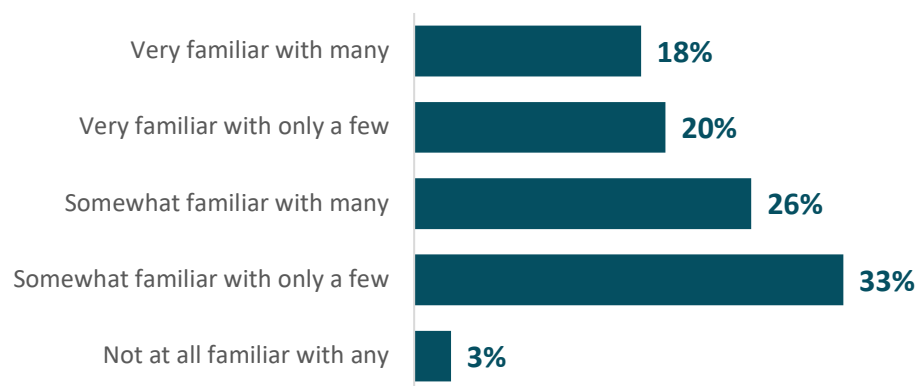


Figure 4.7. AEA and CES evaluators' overall familiarity with evaluation theories ($n = 209$).

Summary

Most evaluators are at least somewhat familiar with a few evaluation theories, with only 18% reporting to be very familiar with many. AEA and CES evaluators were most familiar with participatory evaluation, developmental evaluation, and utilization-focused evaluation. They were least familiar with deliberative democratic evaluation, consumer-oriented evaluation, and constructivist or fourth-generation evaluation.

AEA and CES evaluators are most familiar with participatory, developmental, and utilization-focused evaluation. They are least familiar with consumer oriented and deliberative democratic evaluation.

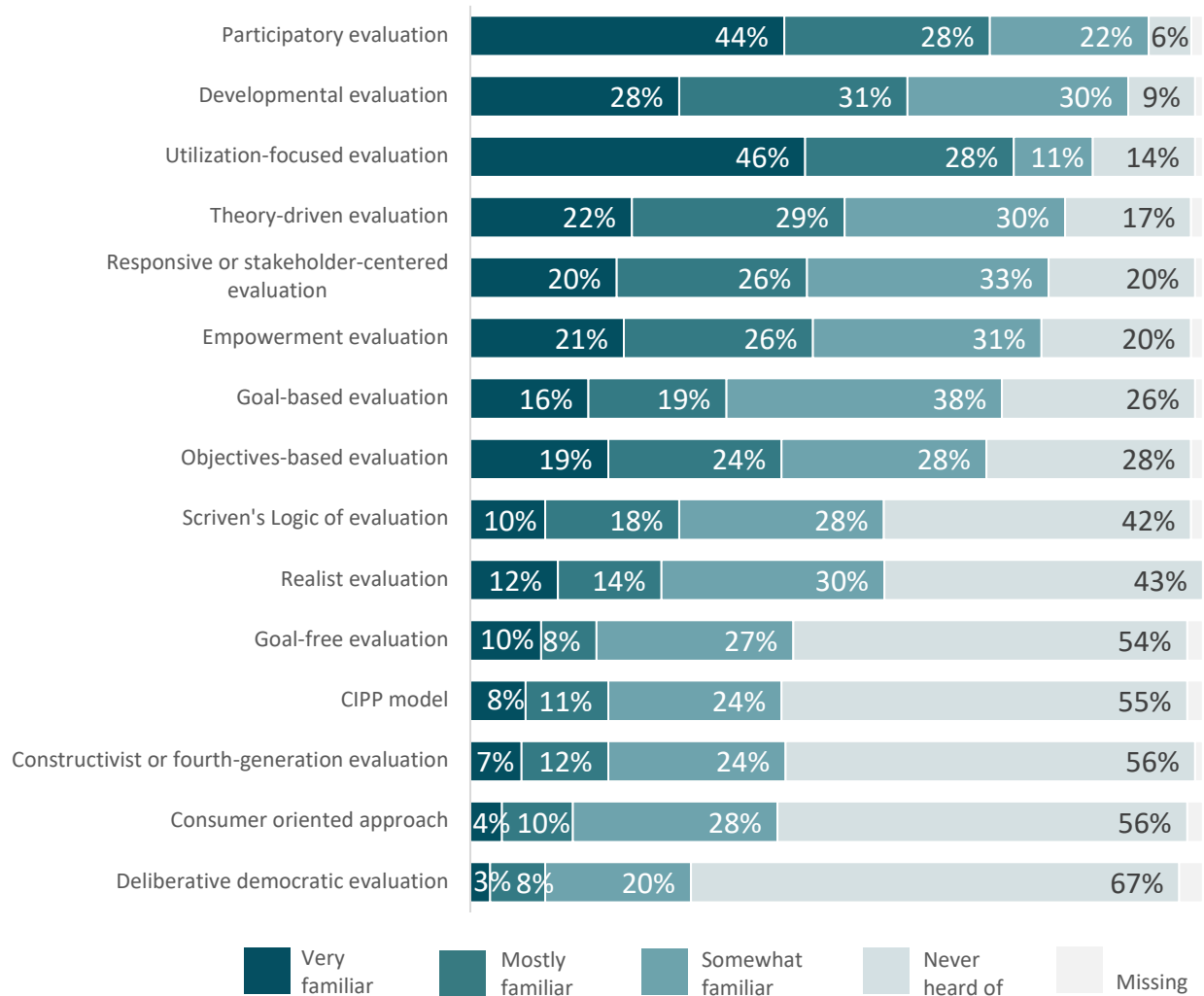


Figure 4.8. AEA and CES evaluator's familiarity with specific evaluation theories ($n = 186$).

Role of Evaluation Theory

Evaluators who participated in interviews were asked to describe an evaluation they were particularly proud of, followed by questions about their experience with evaluation theory and its perceived importance in their work. Evaluation theory was intentionally not asked about until after they described their practice in order not to influence evaluator's recall of practice. As discussed in Chapter 3, 20 evaluators agreed to participate in interviews, with 19 final usable transcripts. These evaluators varied in their training in and familiarity with evaluation theory. In order to explore the influence of evaluation theory on evaluative thinking, survey responses were used to partition participants into three groups: high, average, and low training in and familiarity with evaluation theory. Items included in this grouping included evaluators' attendance in training on evaluation theory as well as their rated familiarity with evaluation approaches.

Table 4.2. Comparison of evaluators with low, average, and high evaluation theory.

	Low Evaluation Theory (n = 5)	Average Evaluation Theory (n = 9)	High Evaluation Theory (n = 5)
Training in Evaluation Theory			
Attended webinars or presentations	60%	78%	100%
Attended professional development workshops	40%	56%	60%
Attended professional development courses	0%	11%	40%
Attended undergraduate courses	0%	0%	20%
Attended graduate courses	20%	78%	80%
Familiarity with Evaluation Theories			
Somewhat familiar with only a few	100%	11%	0%
Somewhat familiar with many	0%	44%	20%
Very familiar with only a few	0%	44%	0%
Very familiar with many	0%	0%	80%
Average number of theories at least somewhat familiar	7.6	12.3	12.8
Average number of theories very familiar with	0.6	2.5	7.0
Years of Experience			

	Low Evaluation Theory (<i>n</i> = 5)	Average Evaluation Theory (<i>n</i> = 9)	High Evaluation Theory (<i>n</i> = 5)
6 or more years	80%	44%	100%
5 or less years	20%	56%	0%

The main findings from these interviews are presented below. First, evaluator's perceived role of evaluation theory is discussed, followed by an analysis of patterns in their evaluative thinking as expressed through their descriptions of practice. Throughout, the reflections of evaluators with high and low levels of training in and familiarity with evaluation theory are compared to address the second research question:

2. What role does evaluation theory play in evaluative thinking?
 - a. How do patterns of evaluative thinking differ in evaluators with training in construe evaluation theory compared to those without?
 - b. How do evaluators with training in evaluation theory differ in their patterns of thinking in relation to:
 - i. Logical reasoning
 - ii. Values and valuing
 - iii. Contextual responsiveness

Perceived Role of Evaluation Theory

When asked explicitly about the role evaluation theory played in their practice and thinking about evaluation, many evaluators interviewed provided general descriptions of evaluation theory as their "foundation," "guiding compass," or "groundwork of everything that [they] do." A few expressed their difficulty in seeing the practical application of evaluation theory, even if they did see aspects of their work that have been influenced by theoretical concepts. Some evaluators spoke about their use of evaluation theory in justifying evaluation to non-evaluators or to justify their decisions in practice. In this section, major themes around evaluator's perceived use of evaluation theory will be discussed. In other words, this section describes how evaluators assume they use evaluation theory in practice. The next section

examines whether these same uses of evaluation theory were expressed through evaluators' reflections on their practice.

Evaluation Theory in Practice

Evaluators who acknowledged a positive influence of evaluation theory on their practice discussed how evaluation theory provided “ground rules” for their practical decisions, giving them a “direction” and “purpose” for their actions. Evaluators of all levels of training in and familiarity with evaluation theory frequently discussed evaluation theory as an essential part of their evaluation toolbox and how evaluation theory helped them to be more intentional about their evaluation activities. Only a few evaluators—those reporting low training in and familiarity with evaluation theory—acknowledge formal evaluation theory had no role in their work.

At times, evaluators portrayed evaluation theory as ubiquitous in their evaluation practice. One evaluator stated that “[evaluation theory has] affected the work that I do in evaluation, virtually every single case that I’ve ever done.” While another described evaluation theory as “the groundwork of everything that I do.” A third evaluator expanded on the integration of utilization focused evaluation (UFE) into all aspects of her work; “I think UFE is so essential to what we do, and it's my guiding compass in terms of all the evaluations that we do, and really all the work that I do.... It just completely guides how I approach my interactions with people and then how I design evaluations and how I collect information and how I report on my findings.” These evaluators who described evaluation theory as having a foundational role in their evaluation work tended to have average to high training in and familiarity with evaluation theories. Many evaluators were more explicit in how evaluation theory aided their work.

Evaluation theory as a tool to better address stakeholder needs. Evaluators discussed how knowledge of multiple evaluation theories helped them to match evaluation approaches and activities to the information needs of the stakeholder. They explained how evaluation theory helped them when planning an evaluation, acknowledging that each evaluation theory or approach gives evaluators a “different set of ground rules” that guide their practice. Evaluators highlighted how evaluation theory can provide multiple perspectives to a situation, allowing evaluators to choose the one that best suits a situation. As one evaluator explained,

I don't just come in with a single perspective or single set of tools to do an evaluation.

I'm much better prepared to think more broadly about the context, and I can think more critically to what the stakeholders, the sponsors, want, and then I can start that idea for data, or ask the right questions to shape the data collection, or the methods behind it.

Another pointed to the importance of a varied toolbox of evaluation approaches,

I think knowing multiple evaluation approaches is the way to get to successful evaluations. The evaluators have come in with, would pretend to have a toolbox of 12 or 15 approaches, but actually have one or two. It only goes so far and just think of it in a kind of statistical way. The likelihood of missing the boat increases when you start narrowing the toolbox or narrowing the number of tools you can use.

Evaluators with a varied evaluation theory toolkit acknowledged their ability to consider multiple perspectives and offer stakeholders "the best fit" to the context.

The "toolkit" metaphor was frequently used to describe the utility of evaluation theory. As one evaluator stated, "theory is all part of my toolkit so I think when a project is presented to me, I often find myself going back to the theory to make sense of the question at hand before I present any ideas or ways forward to each of the projects so I think it set a foundation for me. It will always be part of my toolkit." Given the emphasis evaluators placed on being contextually responsive and meeting stakeholders' needs (discussed in depth later in this chapter), the toolbox metaphor is particularly useful to describe the role of evaluation theory in addressing the many situations that arise for evaluators. One evaluator described the utility of evaluation theories as identifying potential pathways to solving problems,

It can be easy to kind of fall in to the same sort of solutions and patterns. Knowing the different approaches and theories, I think it's incredibly useful, having models for how it's actually been executed and sort of studies that document what projects like that look like really expands your toolkit for when new problems and opportunities arise, or for even being able to see opportunities where people hadn't seen them before.

Evaluation theory was often described as an essential tool for planning and problem solving.

Evaluators with average to high training in and familiarity with evaluation theories placed more

emphasis on evaluation theory as tools in their toolbox allowing them to address issues that arise in practice.

Evaluation theory as an aid in decision making. Finally, evaluators who identified a strong presence of evaluation theory in their practice discussed how evaluation theory aided their decision making. One evaluator talked about the difficulties in scoping an evaluation, but how “once you land on a theory, it gives you that anchor” from which to base decisions. Another evaluator described their use of a heuristic device to remind them of the variety of evaluation approaches to choose from;

I actually have a periodic table of evaluation right beside me.... And it just reminds me of the array of approaches. And when you have that in your back pocket, ... It almost is faster decisions because you know the context of each of those approaches, so you can sort of eliminate, or investigate faster.

One evaluator was insistent on the need for evaluators to learn evaluation theory, stating

It's so informative, I think in this day and age we get a lot of researchers who land in evaluation who don't have any training in evaluative thinking, and I think that's dangerous. Because we need to be able to design and make decisions about evaluation, use the most efficient tools to be able to inform decision making and come to a very clear, succinct conclusion that's consumable by whoever needs to consume that information.

Evaluation theory was described as an aid to decision making, but also in justifying those decisions to others.

Evaluators discussed how evaluation theory helped them justify the choice of an approach or method to stakeholders. One evaluator described “preparing” or “arming” themselves with evaluation theory before meetings with clients in order to better support their choice of methods. Both internal and external evaluators discussed the importance of referencing evaluation theory in order to build credibility of choices in evaluations, or even the intention behind evaluating at all. One evaluator stated, “[it] provided a lot of framework to go in, and confidence to be like, ‘yes, there is a process to evaluation. We can call it something.’” Evaluation theory as a way to justify evaluative decisions to those who may not be familiar with

or trusting of evaluation was primarily emphasized by those with high training in and familiarity with evaluation theory.

Non-use and confusion over evaluation theory. These notions of evaluation as a guiding compass to practice, a tool in the evaluation toolkit, and an aid to decision making align with evaluation theorists' ideals of the symbiotic relationship between theory and practice. However, not all evaluators found evaluation theory to be practical. A few evaluators questioned the practicality of evaluation theory or rebuffed the idea that it had a direct impact on their evaluation practice. Most who struggled with the practicality of evaluation theory had an average amount of training in and familiarity with evaluation theory. One of these evaluators expressed an exasperation that evaluation theory did not fit the messiness of the real world, at one point while describing a real-life situation saying "oh gosh, this is a disaster. You know, Michael Quinn Patton's not going to help me now!" Teaching a course on evaluation, the same evaluator told her students "Evaluation is messy, very messy, and it's not exact. So, don't take what we're doing in class as gospel truth.... because it doesn't really work that way in practice." Even as a teacher of evaluation, this evaluator was still hesitant about the direct practicality of evaluation theory. This lack of translation to practice was generally accompanied by a confusion among the multitude of theories.

Evaluators with all levels of training in and familiarity with evaluation theory expressed confusion regarding the large number of evaluation theories and approaches, as well as confusion between what would be considered a theory and what would be considered a methodology. As one evaluator mentioned, "I think I use more evaluation methodology. Not so much the theory." In many interviews, there was an extended discussion on what should be considered evaluation theory for the purpose of the interview, with some wanting to discuss social science theories, and others preferring to call evaluation theories "approaches." Evaluators with more familiarity with evaluation theories tended to be more positive about the role evaluation theory played in their practice, while evaluators with less familiarity in evaluation theories expressed confusion between too many theories or general non-use of evaluation theory. One evaluator with a low exposure to evaluation theory set himself apart

from those who studied evaluation theory. They described themselves as more of a practitioner, saying

there's a whole theoretical side, like people who study evaluation, probably like yourself, and write journal articles and it's like sometimes I get lost in all that stuff and, for me, one evaluation is much more hands on. I go out, I'll do interviews, I will do focus groups, I will do surveys. Which I guess is probably the theory, but I don't ... I guess, maybe this is also because I didn't study evaluation; I just kinda fell into it as more a practitioner. I don't read Quinn Patton or The Theory and I can't tell you the differences between utility focus versus ... I understand a little bit about it but it doesn't really ... I almost see it as the ivory tower of academia about what their talking about all the time, I don't even know, and you go and they're talking in abstract and I can just talk about the practical things that I've done.

This evaluator found a real disconnect between their work and evaluation theory, to the point where they didn't even find it valuable to pursue learning about it more.

Evaluation Theory in Evaluative Thinking

Many evaluators acknowledged that evaluation theory had an influence on how they thought about evaluation and their evaluative thinking. Most who were hesitant to agree that evaluation theory had a direct impact on their practice, discussed its more indirect impact on their thought process. One evaluator noted, "I would say it's always in the back of your mind." Another evaluator indicated that evaluation theory

influences my thinking more than my practice, yes. I would say that evaluation theory and just thinking about those concepts and having them floating around in the back of my head, like quotes from Bob Stake, you know looking at different aspects of evaluation, or Michael Scriven, I mean, just some of those concepts.

Evaluators generally found it difficult to express in words the exact role evaluation theory plays. One evaluator described evaluation theory as "tangled together" with other aspects of methodology and practical experience, saying "I feel like I absorb a lot of information maybe about different theories and then it just comes out as a big, like it gets all tangled together and

that's how I perform my work." Some just came to a loss of words when pressed for more detailed explanations.

One evaluator described how they saw evaluators use evaluation theory in their work without realizing they were using theory, "I know there's a lot [of evaluators] who don't look at theory in their work.... I think a lot of their work is incorporating theory without calling it theory." Reflecting on other's work, she noted that many evaluators have logical, well thought out evaluation plans but they wouldn't call the influence of evaluation theory by name. Instead, she noted, "I think there's anecdotal theories or unintentional theories that are driving those decisions." The idea of an underlying set of assumptions about how or why certain decisions are made invokes Vo and Christie's (2003) writing on folk theories.

Some evaluators described evaluation theory's influence on their ability to be self-reflective of their own work. One evaluator more recently exposed to evaluation theory explained,

now, I can be so much more reflective and think so much more deeply about what I do because I'm aware that there has actually been people thinking about these things and these concepts and writing about them and there's actual words to use to describe the phenomena that I see and the context that I experience. And so, knowing that there's theory there that you can dive into and fall back on, it makes me be able to reflect about situations in a totally different way. It kind of expands my realm of experience in some ways because I just have a new lens with which to think about the different scenarios that I've been in in the past and different scenarios that I encounter now in present day and will continue to.

This evaluator points to evaluation theory as something to "fall back on" but also to expose new ways of thinking, in order to "expand the realm of experience." Another evaluator acknowledged evaluation theory's role in their critical reflection, specifically making them more aware of their biases, "it's enhanced my ability to reflect on not just my practice, but why I do what I do, and be more aware. Pay more attention, be more aware of my own biases." Evaluation theory seemingly plays a number of different roles in evaluative thinking, from

influencing folk theories, and spurring critical reflection, to exposing evaluators to varied perspectives.

Summary

Most evaluators interviewed acknowledged that evaluation theory played a role in either their evaluation practice or thinking. Only one evaluator completely eschewed the idea of needing evaluation theory in order to practice evaluation. Evaluators with a higher level of training in and familiarity with evaluation theory were more likely to discuss evaluation theory's influence on their decision making, and how they use evaluation theory as a way to justify the conduct of evaluation or their choices throughout an evaluation. Evaluators with a lower level of training in and familiarity with evaluation theory acknowledged that evaluation theory played a role in their practice but tended to talk solely of utilization focused evaluation or participatory evaluation. Exemplary quotes organized according to level of training and familiarity with evaluation theory are displayed in Table 4.3.

Table 4.3. Comparison of exemplary quotes around the perceived role of evaluation theory.

	Low Evaluation Theory	Average Evaluation Theory	High Evaluation Theory
Role of evaluation theory in practice	<p>"I think UFE is so essential to what we do, and it's my guiding compass in terms of all the evaluations that we do, and really all the work that I do."</p> <p>"I don't know if I look to evaluation theories specifically when I'm developing evaluation and maybe that's just because I've been doing this work so long, I don't need to."</p>	<p>"it was evaluation theory that drew me into evaluation in the first place, so it's probably the groundwork of everything that I do."</p> <p>"Yeah, it sets a foundation, it helps that initial engagement and that direct to the approach moving forward"</p>	<p>"There's different types of theories. But there's still that upfront scoping, and context, and situational analysis, where everything is ambiguous. Everything is kind of up in the air, and then you land on the theory, and then you move forward in my world."</p>
Tool to address stakeholder needs		<p>"It just armed me with more, yeah, tools in the toolbox, I guess because I think if I didn't have that then what would be telling me that I wasn't doing... they gave me that broader, that broader foundation, that stronger foundation"</p>	<p>"But because of the theories and the tools that I have in my toolkit I can explain the program to you in ways that these people have never heard before and really reveal the program to them and sometimes for the first time ever."</p>
Aid in decision making	<p>"I mean definitely if we're talking about just the philosophy around utilization focused evaluation that definitely drive a lot of what we do and a lot of what I do,"</p>	<p>"I prepare myself or better arm myself before I enter these client meetings so I can actually pull out, when they bring something up, "Oh yes, so this is how it's been used before or this is some options." What ... Do I think they always stay in my head and are readily available? No, unfortunately not, I find myself always having to go back and remind myself."</p>	<p>"It's so informative, I think in this day and age we get a lot of researchers who land in evaluation who don't have any training in evaluative thinking, and I think that's dangerous. Because we need to be able to design and make decisions about evaluation, use the most efficient tools to be able to inform decision making and come to a very clear, succinct conclusion that's consumable by whoever needs to consume that information."</p>
Role of evaluation theory in thinking	<p>"I feel like I absorb a lot of information maybe about different theories and then it just comes out as a big, like it gets all tangled together and that's how I perform my work."</p>	<p>"Evaluation theory influences my thinking more than my practice, yes. I would say that evaluation theory and just thinking about those concepts and having them floating around in the back of my head..."</p>	<p>"I think so because you know the pieces of the theory, the steps involved in ... So, if you understand multiple theories, you know each one of them offers different pieces. And if you don't know many theories, then you may be limited in your approach, or lens."</p>

Implicit Role of Evaluation Theory

This section looks at evaluators' descriptions of their practice to determine whether evaluation theory plays a role in their evaluative thinking. While many evaluators mentioned the importance of evaluation theory in the abstract, it was rarely mentioned outright when describing their practice. This section describes emergent themes around the underlying reasoning behind decision making and patterns in evaluative thinking across four main dimensions: contextual responsiveness, logical reasoning, and the inclusion of values and valuing in order to address the second research question.

In one-on-one interviews, evaluators were asked to describe an evaluation they were particularly proud of, including a description of the evaluand, their evaluation design, methods, and methodologies, engagement with stakeholders, inclusion and creation of criteria and standards, as well as reporting and communicating findings to stakeholders. It is important to recognize that responses by evaluators are considered the most salient or important to that evaluator in a particular context. Their lack of discussion of a particular topic is not necessarily considered a disagreement, but instead the topics they focused on are considered to have more weight than others.

Overall, evaluators with low, average, and high training in and familiarity with evaluation theory were responsive to contextual factors in their work. However, evaluators with higher training in evaluation theory are more eloquent in describing the values that underly and justify their practical decision making. Evaluators with lower training in and familiarity with evaluation theory were more influenced by requirements and assumptions from funders, disciplinary areas, and their own organizations, whereas evaluators with higher evaluation theory were less constrained by these elements and relied more on their values as a driving force.

Contextual Responsiveness

Many evaluation experts discuss the importance of contextual responsiveness as a characteristic of high-quality evaluation. And, contextual responsiveness was the most commonly expressed influencing factor on evaluators' decision making in practice. Evaluators

with high, average, and low training in and familiarity with evaluation theory placed great importance on responding to stakeholder needs as well as selecting appropriate methods for a given situation. Considering these values, it is not surprising that contextual factors came up most often when describing their evaluation practice. To discuss these many factors, this section will report the emergent themes around contextual responsiveness organized into Greene's (2005) five dimensions of context: (a) descriptive and demographic character, (b) material and economic features of a setting, (c) institutional and organizational climate, (d) interpersonal dimensions, and (e) political dynamics.

Descriptive and demographic character. While describing their evaluation practice, a primary concern in evaluators' decision making was to align their evaluation choices with stakeholder needs. Evaluators also talked about the influence of the geographic location of a project and the characteristics of participants, as well as the influence of practical constraints.

Multiple evaluators interviewed discussed how geographic distance between sites included in their evaluation affected their chosen approach and data collection methods. Evaluators gave examples of sites being so far apart that implementing a participatory approach became necessary for feasibility; "I've learned that in rural Alaska you cannot be every place all the time. You need to have participatory evaluation." Geography even affected this evaluator's choices in reporting. They discussed their creation of an e-book to disseminate evaluative findings that included a lot of photos. They explained, "people don't understand what's happening in rural Alaska. It's like, invisible. Most people don't go there. You can't drive there. So, they don't know... it doesn't really exist for most people, especially in the legislature."

Demographic characteristics of the participant population was also a factor in considering methods and data collection techniques. One evaluator explained how they modified an online survey into an arts-based feedback session in order to better respond to the participant population,

we really wanted to ensure that everyone had an opportunity to be more involved. This segment of the population was also maybe, a lot of people that were maybe not as literate or not having access to computers and that kind of thing, so we wanted to make

sure that people could just participate any way they felt they were able to and comfortable.

Instead of using the same data collection methods as used previously, this evaluator was creative in responding to the needs of participants.

Evaluators frequently talked about the importance of matching their chosen designs, methods, and reporting to stakeholder's information needs and adequately addressing evaluation questions. Evaluators described being attentive to what stakeholders would find "most meaningful" and using this to guide their choices. Evaluators recognized that choices throughout an evaluation "really depends on the situation." Referring to good evaluators knowing multiple evaluation theories and methodologies, this evaluator expressed, "a good evaluator can do that cause they know." Tailoring an evaluation to the needs of stakeholders and settings were generally considered to be a desirable characteristic of a good evaluation. As one evaluator put it, "the key thing is that the evaluator team knows what they're doing and knows what their responsibilities are and knows what they need to be responsive to." One evaluator describes how their team used a developmental approach in order to meet stakeholder needs, "even in our proposal stage we could see what they needed and based on the RFP the need was there, even if it wasn't identified as purely developmental.... It's certainly what they needed." Evaluators talked about engaging stakeholders to determine what they needed from an evaluation and then matching their methodological choices to the agreed upon evaluation questions. As one evaluator explains, "we consult with our stakeholder and what are their information needs, then in each area we apply the most useful methods. We try to answer the questions as comprehensively as we can with different multiple methods." Evaluators motivations behind methodological choices frequently came down to the types of questions they were trying to answer and the types of information necessary to address those questions. Many evaluators explained they chose a qualitative approach to increase the richness of their evaluation data or to complement quantitative data. As one evaluator illustrated,

I felt like qualitative tells the story more, and quantitative people just want a survey and you miss a lot from it. So, choosing these methods were like, here's where we're able to

actually hear what people are saying, and here's the observations of what happened, not just, here's how people reflected on it in five minutes before they left for the day. Like this evaluator, many described wanting to tell the whole story and capture a “diversity of perspectives,” which wasn’t always possible with a single method. Instead they found methods that “complimented each other well” and answered evaluation questions in full.

Material and economic features. Evaluators pointed to the influence of time and monetary constraints on their choices in planning and conducting evaluations. Time constraints affected the approaches considered at the beginning of an evaluation due to the shortened time period given to respond to RFPs, as well as data collection techniques based on the time needed for participant recruitment. One evaluator described how they steered stakeholders away from certain methods due to resource constraints,

I dissuaded people from thinking about it, because that starts to get incredibly expensive in terms of recruitment and finding people and everything, and if the questions that we have aren’t about the general, higher ed teaching community, then I don’t know why we would want to factor them in.

Some evaluators described their organization’s “frugal” nature and choice of methods based on the need to get a “bigger bang for their buck.”

Everyone who mentioned resource constraints influencing their practice acknowledged a balance between working within the realistic constraints while also obtaining the data needed to answer the evaluation questions. As one evaluator explained,

we didn’t have an unlimited amount of time, so we had to look at the resources that we have as well as the time we had to complete the evaluation, and we chose lines of evidence that would be feasible with our one-year evaluation. So that’s how we chose to proceed.

Time and monetary restrictions were by no means a primary source of decision making as described by the evaluators; however, they did play a role in bounding the approaches and methods that were considered by evaluators with all levels of training and familiarity with evaluation theory.

Institutional and organizational climate. Evaluators' evaluative thinking was influenced by the norms and assumptions from client organizations, disciplinary areas, and their own evaluation organizations. Assumptions about epistemology, ontology, and methodology, whether consciously or subconsciously preferred, swayed evaluators in their considerations and choices throughout evaluations.

Stakeholders' negative connotations of evaluation influenced how evaluators talk about their work and restricted their choices. Canadian evaluators in particular discussed client's hesitation toward evaluation. One evaluator explained, "it [evaluation] kind of triggers a lot of negative connotations about judging and judging them to be good or bad, or pass or fail, some of this kind of stuff, whereas I try to frame, when I'm talking to new people about it, it's just an assessment. It's acknowledging this organization." While another described, "I didn't want to call it evaluation, because I would be kind of threatening, and I'm coming in as a stranger. I just told them I'm fresh eyes and I'll have some observations." Evaluators focused more on building relationships with stakeholders in order to build their trust in evaluation. One evaluator explained, "So, there was resistance there, but it's really a lot of, I spent a lot of time talking and listening. Listening first, and then talking, really. That's how I've built those relationships. And some of its networking, I guess. Relationships and networking." Evaluators changed how they present themselves and their work to appear less threatening in environments of evaluation skepticism. Much of their practice was oriented towards strengthening evaluation capacity and understanding.

Some evaluators reported having stakeholders who had positive experiences and inherent trust in evaluation. These evaluators found they could spend more time on other aspects of the evaluation instead of convincing stakeholders about the value of evaluation. One evaluator described, "I think the respect right from the beginning from our client was different. Although there was some skepticism at first, and we did feel we had to justify ourselves and prove ourselves to the client, but there seemed to be a really some early wins in terms of ensuring their buy-in and their appreciation of us." Others found they had to use a participatory or collaborative approach because that's how their stakeholders already operated; "I guess, it's just out of the function of how this organization operates. They're very engaged with their

stakeholders, very engaged with funders.” The climate of stakeholder organizations – whether positively or negatively exposed to evaluation – had an influence on the practice of evaluators.

Disciplinary area norms and assumptions can also affect the choices made throughout an evaluation. Evaluators discussed decisions as being “how things are done” in their discipline. For example, one evaluator noted, “there’s such a big push within the health sector right now for patient-oriented or patient engagement in research.” Some of these choices were referred to as “traditional data” or “what is typically done.” These phrases reflect the idea of doing things because they are the way things are done, usually code for unexamined assumptions or ideological orientations. One evaluator with low training in evaluation theory explained how data collection and analyses were typically prescribed by their organization,

We traditionally use these methods in our department. It was not unique to this evaluation that we do interviews, that we do focus groups, that go out to the site, do site visits, that we crunch our administrative data, do statistical analyses. We pretty much do that whenever we can. Yeah.

Others were working in disciplines different from their own but recognized the importance of playing by the rules of that discipline to increase the use of findings. One evaluator explained her choice of Excel as a reporting platform to align with disciplinary norms of her clients, “I was very proud that I’d come up with this succinct table... because of course engineers, they want things very succinct. They love Excel, I think I put it in Excel. They wanted it, so.” Evaluators of all levels of training and familiarity with evaluation theory were influenced by the expectations set, knowingly or unknowingly, by the climates of stakeholders, discipline areas, or evaluation organizations.

Interpersonal dimensions. Evaluators intermittently pointed to interpersonal dimensions of a context as influential in evaluative decision making. Multiple evaluators discussed an individual who had an overwhelming influence on the direction of the evaluation. Evaluators talked about particularly influential higher-ups who pushed for evaluation elements, for example, the incorporation of qualitative data into a primarily quantitative evaluation or breaking up longer evaluation reports to be digestible by different audiences. One evaluator talked about a vice president who pushed their evaluation to incorporate more stories, saying

“he’s a member of the National Academies of Sciences, a very accomplished scientist, but he’s also really interested in stories and storytelling ... it really took [the program] to the next level ... people started really using our stuff.” Some of these influential individuals have procedural oversight on the evaluation or serve on an advisory committee, while others are simply proponents of one dimension or method. Evaluators of average training in and familiarity with evaluation theory talked about a particularly influential person located outside of the evaluation team.

Political dimensions. Evaluators frequently mentioned the influence of funder requirements or preferences on their evaluative thinking and decision making. Particularly in evaluations funded by federal governments or international organizations, where funders required certain choices about everything from approaches, and data collection methods, to measurement indicators, and reporting. Evaluators generally found these requirements flexible and felt they could be “creative” in order to meet both funder and program needs.

Evaluators talked about results frameworks and measurement indicators required by federal departments in the United States, as well as international organizations, such as the use of the Organisation for Economic Cooperation and Development’s DAC criteria. Evaluators voiced that sometimes criteria or indicators required by a funder are not valued by local stakeholders. One evaluator explained how required measurement indicators became “a major point of contention between the [funder] and the states.” They continued, “there’s a lot of information that they needed for their national evaluation that was not useful in terms of program improvement or in terms of, just being specific enough for their state needs.” This evaluator also found the required reporting to be “not entirely user-friendly and... too dense.” Acknowledging this tension between what the funder needed and what would be useful to the local stakeholders, the evaluator created separate reporting deliverables to communicate evaluative findings which each focused on different measurement indicators. Some funders required additional aspects of evaluations, including mandated patient engagement by a health care system, mandatory reviews of evaluation plans by ministerial offices, and approvals of baseline measurements by international development organizations.

While many evaluators, of both high and low training in and familiarity with evaluation theory, found these requirements “flexible” and could alter their evaluations to meet the requirements, funder expectations still play a role in how evaluators think about and approach evaluations. Funders’ explicit and implicit preferences for designs, such as randomized control trials or quasi-experimental designs, influenced evaluators to favor these designs and approaches. One evaluator explained her choice to integrate a quasi-experimental design (QED) into her evaluation plan, even though she had no prior experience conducting this type of evaluation. They explained,

we put [the QED] in the proposal and then [the funder] liked it and they were like ‘great, this sounds really good.’ So, we went with it, but the organization I work for has never done a QED before. And so honestly, it’s pretty challenging and one of my lessons learned is I’m going to be a lot more cautious before putting that into future proposals, to be honest.

This evaluator displayed a high regard for using a quasi-experimental design without really being critical about the tradeoffs the design presented. Given the power held by funders in evaluation, it is easy for evaluators to be overly influenced by their preferences without considering whether the design is the best fit for the larger context.

Summary. Evaluators with all levels of training in and familiarity with evaluation theory identified contextual factors as influencing their choices in evaluation practice. All evaluators were particularly responsive to stakeholders needs and attentive to aligning methodological choices with evaluation questions. Evaluators with low training in and familiarity of evaluation theory did not discuss the influence of monetary and economic factors on their evaluations, while evaluators with average and high evaluation theory did. Evaluators with low training in and familiarity with evaluation theory tended to work in organizations that had a highly structured and regular process to conducting evaluations. This led them to be heavily influenced by their organizational climate, while evaluators with average and high evaluation theory were more skillful at navigating contextual assumptions and what they thought would best fit the evaluation. Finally, evaluators with low training in and familiarity with evaluation theory were at times more swayed by the demands of funder’s preferences. All evaluators dealt

with the restrictions placed by funder requirements, however, evaluators with high evaluation theory were more adept at negotiating multiple influencing factors. Exemplary quotes organized according to level of training and familiarity with evaluation theory are displayed in Table 4.4.

Table 4.4. Comparison of exemplary quotes around contextual responsiveness in practice.

	Low Evaluation Theory	Average Evaluation Theory	High Evaluation Theory
Descriptive and demographic character	"We consult with our stakeholder and what are their information needs, then in each area we apply the most useful methods. We try to answer the questions as comprehensively as we can with different multiple methods."	"It was also, we really wanted to ensure that everyone had an opportunity to more involved.... This segment of the population was also maybe a lot of people that were maybe not as literate or not having access to computers and that kind of thing, so we wanted to make sure that people could just participate any way they felt they were able to and comfortable."	"I tend to choose approaches that I think work best for the client." "It was what the client wanted and certainly what they needed. They didn't call it a developmental evaluation, but certainly that's what very quickly emerged, that we identified even in the proposal stage.... So through those types of discussions, early on we identified this evaluation developmental approach."
Material and economic features		"I mean, you have to go in there with some cost considerations, so to kind of go in and say look, we've got \$30,000 we want to spend on this, we can blow it all on getting a survey panel to give you information about 5% of the people you actually want to talk to, or I can save you \$20,000 and just go in and do a targeted inquiry with these people."	"So, that we needed to be able to have the full spectrum, so we didn't have an unlimited amount of time, so we had to look at the resources that we have as well as the time we had to complete the evaluation, and we chose lines of evidence that would be feasible within our one year evaluation. So, that's how we chose to proceed." "There's a number of different factors. The first would be the context and use. I'm not gonna apply a very rigorous, or time intensive approach if it's a three-month evaluation, where they just want an infographic, or something quick. And it's more ... Okay, it comes down to context, budget, time, resources, and the use of it. Would be sort of my major ones right now."
Institutional and organizational climate	"We traditionally use these methods in our department. It was not unique to this evaluation that we do interviews, that we do focus groups, that go out to the site, do site visits, that we crunch our administrative data, do statistical analyses. We	"I didn't want to call it evaluations, because I would be kind of threatening, and I'm coming in as a stranger. I just told them I'm fresh eyes and I'll have some observations." "I think one of the nice things about our team is that, we really discussed data collection and	"I think it's very easy to come to false conclusions, so I think being able to look and understand that different stakeholders have different perspectives and different motives for participating in evaluative research. We

Interpersonal dimensions	<p>pretty much do that whenever we can. Yeah.”</p> <p>“Did we consider any alternative approaches? Like any things different that we didn’t do? No, not really. We have a very set way that we do it.”</p> <p>“I’d say in terms of the methods applied, the planning involved, it’s the same always.”</p>	<p>before we actually do it, especially either with the protocol or with our physicians, is we really gotta come to a consensus. It’s very much bouncing ideas off each other. I find that just really striking, just because it truly is a group activity.”</p>	<p>need to be able to look across many stakeholders to get various opinions.”</p> <p>“I guess, it’s just out of the function of how this organization operates. They’re very engaged with their stakeholders, very engaged with funders.”</p>
		<p>“Well, the executive director of the organization was very clear that there were a few things that he wanted done and so they definitely got added into the method and it was appropriate. So, he was very involved in that, which ... and fortunate”</p> <p>“I think it’s a perfect storm of a lot of really good people working together.”</p>	
Political dimensions	<p>“I think that was a major point of contention between the [funder] and the states, but I’ll say there’s a lot of information that they needed for their national evaluation that was not useful in terms of program improvement ... We did some other data collection for them that was related, but not reported to the [funder], but that was based on what the state needed, so that’s part of it. Another thing was that the types of reporting that they had to do for the [funder] was not entirely user-friendly and it was a little bit too dense. “</p>	<p>“I think that, I definitely think that the federal restrictions or guidelines or whatever you want to call them that have been used to define what evaluation could be has been a little bit detriment, has been a little bit of a cost to the field because I think a lot of times, people have had to spend a lot of money answering questions that they weren’t really interested in. I think that that’s really problematic.”</p> <p>“That’s kind of funny, but part of it is that the donor is a lot more comfortable with, in some situations, with qualitative data.”</p>	<p>“Well, in Canada we have a new policy on results, and that gives evaluators a lot more flexibility in terms of the scope of our evaluations as well as the methodologies that we use. So, treasury board is that top level who gives us the funding for the initiatives as well as approves ultimately our approach as well as our evaluation reports. They give us a lot of leeway, I think it’s more so now than ever before. So, I do feel that we have a lot of flexibility.”</p>

Logical Reasoning

If an evaluation is to be useful in decision making, a certain level of justification or logical reasoning behind evaluative findings should be considered credible by decision makers. Whether consciously or unconsciously acknowledged, logical reasoning provides the foundation of evaluative conclusions. Scriven (1980) and Fournier (1995), through their explanations of the general and working logic of evaluation, recognize the warrantability of evaluative conclusions as a core component of evaluation. To investigate evaluator's thinking around logical reasoning and the working logics they constructed in their evaluation practice, evaluators were asked to discuss whether stakeholders found evaluative conclusions trustworthy and why. This set of questions was intended to unearth the warrants evaluators used to bolster their findings.

Overall, evaluators found it very difficult to verbalize the logical reasoning underlying their conclusions. Evaluators did discuss the following building blocks of trust and credibility of evaluative conclusions: stakeholder engagement or participant validation; assumptions around data quality including objectivity, rigor, and triangulation; and finally, trust in the evaluator's reputation. The reasoning behind valuing is also an important aspect of evaluative logic and is discussed in the next section.

Participant validation. Evaluators with all levels of training in and familiarity with evaluation theory pointed to stakeholder engagement as an important factor in building trust in evaluative findings. Evaluators discussed the perceived credibility of their evaluative findings as a function of their engagement with stakeholders throughout the process of an evaluation. One evaluator explained, "A lot of them [stakeholders] were involved in it, so I think when it reflected what they did and they saw that their experience was reflected, it felt trustworthy." Evaluators described engaged stakeholders as eventual "champions" of the evaluations and a "supporter of [the] findings." Evaluators involved stakeholders in planning evaluation methods, recruiting participants, and interpreting data. As one evaluator explains, they involved stakeholders from the very beginning as a way to build a foundation of trust in the conclusions, "I think it's this upfront work of making sure that ... the data we're asking for, the approach that we're taking, and then presenting the initial findings, lays the groundwork for people to trust the end result." Another evaluator explained stakeholder trust as a function of engagement, "I

think people trust the results because they also did the knocking on doors or saw the process out there. Like, 'Yeah, this is what I saw too, and this is what I see in my community.' I still have questions on what we can do with it, but it seems to reflect their experience." Evaluators also described engaging stakeholders in the creation of recommendations. As one evaluator explains, their recommendations were accepted as credible because they were created in partnership with stakeholders,

recommendations totally reflected a shared interpretation with the client and represent our evolution of thinking around that recommendation because we talked about it and we understood different angles for why something might not be possible or might not be a preferred approach to take with a certain partner or whatever. Yeah, we felt definitely the recommendations were really well received.

Stakeholders' trust in evaluative findings increased as they understood the data collection process, when they saw their experience reflected in the data, and when they were involved with critical discussions around the implications for the project.

Hand-in-hand with stakeholder engagement was the importance placed on frequent communication with stakeholders. Multiple evaluators spoke to the power of open discussions between the evaluator and stakeholders. One evaluator credited an open and honest dialogue for the success of their evaluation, saying, "I don't think things would've happened at all, if it didn't happen that way." A key aspect of frequent communication with stakeholders was ensuring transparency throughout the evaluation. One evaluator noted transparency was the "number one factor" in building trust in the evaluative findings, noting "it was designed as such, we embraced that transparency, that participatory approach." The same evaluator described "everything was on the table" in their "highly transparent" evaluation. Another evaluator explained the power of transparency, saying "none of them [the stakeholders] were surprised because we'd always been in conversation with the client the whole way through so basically everything that was in the report had already been tested or floated to the client beforehand." Transparency was held in high regard by evaluators with all levels of training in and familiarity with evaluation theory.

Some evaluators discussed the concept of stakeholder ownership of the evaluation as more important than simply trust in the findings. One evaluator described, “Trust, yes, but more so, I think like ownership more. So, feeling like, yeah, we did this, and also kind of that, what steps can we take, and identifying those.” This evaluator described how ownership not only lead to trust in the results, but also, in her experience, more use of evaluation results. Another evaluator described their involvement of multiple stakeholder groups as a way to build ownership and eventual use of evaluative findings,

It was also important to give them the feeling that they are part of it, so they also feel the ownership. Later, when the findings are out, they take that into consideration. They don't tell us that, "Okay, we were not consulted. We did not know. We don't agree. Evaluators' comments and underlying motivations put voice to the assumption that frequent communication and engagement with stakeholders throughout the evaluation process builds stakeholder buy-in to evaluative findings, which leads to increased trust and use.

Quality of data. Evaluators frequently pointed to the objectivity, rigor, or triangulation of data as the reason behind their evaluation's credibility. One evaluator described stakeholder's trust in the evaluative conclusions as a reflection of the evaluator's “due diligence” in choosing sound methodologies. A lot of evaluators used the term “rigor” in explaining the trustworthiness of their findings. One evaluator explained, “when you asked about how the stakeholders might trust my findings, it's sort of like ... I don't know, I feel like I'm so ingrained and focused on the [participants]. It's sort of a secondary thought to... because I feel I know I operate with rigor.” When asked to expand on what was meant by rigor, evaluators mentioned strong methods, transparency, and triangulating between data sources.

One evaluator explained stakeholder's trust in findings as a function of their evaluation's “rigor,” which they then described as triangulating data from multiple sources,

I think we try to be rigorous, we try to take an approach to triangulate data, like we want, just because we can get some information from one source does not mean it's the only thing that can help us, so we try to use multiple methods. We almost always use multiple methods, always actually, I should say, to get information, different quality and

different quantity of information, pair it up and triangulate it to come up with a clear story.

Another evaluator explained their tendency to “over rigor when possible.” When asked to expand, they described “We do more things than we actually use. So, more data collection, or more. Just because we have multiple different accountabilities, so making sure that the best step forward out of the... I guess the array of information that we’re collecting at this phase.”

Other evaluators listed triangulation directly as a reason for credibility in their evaluation findings. Evaluators with all levels of training in and familiarity with evaluation theory discussed triangulation as a factor in the trustworthiness of their evaluation.

While evaluators with all levels of training in and familiarity with evaluation theory listed objectivity or rigor as reasons why their findings were trustworthy, those with low training in and familiarity with evaluation theory tended to list the objectivity or “rigor” of the data as the primary reason stakeholders trusted their findings in ways that equated evaluation with research. As one evaluator with low evaluation theory put it, “Because I think, number one, they’re rigorous. Our methodologies make sense and they have high validity, I think, in terms of measuring what they’re supposed to be measuring.” The evaluator continued, “We don’t just rely on opinion. We go out and do primary research. ... we have the power to go up and collect data and not just rely on secondary sources.” Evaluators with higher levels of evaluation theory tended to list rigor as among multiple factors that built credibility in their findings. However, some still relied on objectivist assumptions to serve as the backbone of their evaluative argument. As one evaluator with high evaluation theory described, stakeholders trusted their findings because “we had the data. The data was there. We tried to keep going back to the data, the data, ‘This is what the data tells us.’ And whatever your perceptions were of this program or the clientele or what was happening out there in the real world, we rely on the data.” This evaluator almost seems to elevate and reify the concept of “data” as being trustworthy by nature. Some evaluators displayed an inherent trust in rigor without critically examining underlying assumptions. The term rigor almost served as a blanket statement for sound methods.

Evaluators with average or high training in and familiarity with evaluation theory placed more importance on the representation and contextualization of their data as important factors in the credibility of their conclusions. One evaluator with high training in and familiarity with evaluation theory explained,

Context is everything in evaluation.... understanding the context is critical in order to provide something that's been useful and helpful to them making decisions. Because those are the situations that they face every day. Unless you put yourself in the manager's shoes, in their position and understand the situation, you can't be particularly helpful. So, it's all about the utility, the helpfulness of the evaluation and the work that you're doing. That's always paramount in my mind. So, understanding context is critical to doing that.

This evaluator stresses the contextualization of data in order to ensure the utility of evaluative findings. This change of focus to not only define validity as the trustworthiness of data, but also the utility of the data sets evaluators with high training in and familiarity with evaluation theory apart from those having less training in and familiarity of theory.

Reputation of the evaluator. Finally, a major emergent theme in why stakeholders placed trust in evaluators' findings as described by the evaluators was because of stakeholders' inherent trust in the evaluator. One evaluator, with low training in and familiarity with evaluation theory, described how the reputation of their organization was important in the implicit trust from stakeholders, "A lot of it had to do with the reputation of the firm, to be honest, more than the plan. I think that in an external evaluation, the reputation of the firm is big in the results being believed." This evaluator even discussed that evaluators' track record was "how people get hired to do the work in the first place." Another evaluator discussed similar reasoning for why their evaluation team was chosen, explaining, "I think there's a reputation that we deliver hard-hitting findings." Other evaluators described stakeholders trust in them stemming from their credentials, as opposed to their reputation. One evaluator described, "and so she trusted me. I think partly because of my background and credibility in learning and development. But she appreciated, I think, being an engineer, of course they're very scientific and evidence based, and she appreciated that." While some of this trust in the

evaluator was connected to their use of systematic and rigorous methods or their relationship with stakeholders, some was simply voiced as an inherent trust based on credentials.

Summary. Unearthing the logical reasoning behind evaluative findings proved to be very difficult through an interview format. Evaluators with all levels of training in and familiarity with evaluation theory were not adept at reflecting on the logic underlying their conclusions. Of the themes that emerged, all evaluators discussed the role of stakeholder engagement in building the trustworthiness of evaluative conclusions. Most evaluators who participated in this study were at least somewhat familiar with evaluation theories such as participatory (94%) and utilization focused evaluation (85%). While all evaluators talked about the rigor of their evaluation methods as a reason for the trust in their findings, evaluators with high training in and familiarity with evaluation theory also placed importance on contextualizing the data in order for it to be useful to decision makers. Exemplary quotes organized according to level of training and familiarity with evaluation theory are displayed in Table 4.5.

Table 4.5. Comparison of exemplary quotes around logical reasoning in practice.

	Low Evaluation Theory	Average Evaluation Theory	High Evaluation Theory
Trust is built by participant validation	<p>"None of them were surprised because we'd always been in conversation with the client the whole way through, so basically everything that was in the report had already been tested or floated to the client beforehand."</p>	<p>"Trust, yes, but more so, I think like ownership more.... A lot of them were involved in it, so I think when it reflected what they did and they saw that their experience was reflected, it felt trustworthy."</p> <p>"Because they've been involved from the beginning and because I'm going to walk them along the way.... So I think it's this upfront work of making sure that the data that we're getting, the approach that we're taking,... and then presenting the initial findings, lays the groundwork for people to trust the end result."</p>	<p>"They were involved, like I say, as the development evaluation in this case. They were involved in every aspect of the evaluation; everything was on the table. I mean, from the start as a developmental evaluation we kind of threw the door open to anyone on the staff who wanted to participate at any of the key stages in the evaluation, they were welcome to do so.... We embraced that transparency, that participatory approach."</p>
Trust is built by high quality data	<p>"I think that we don't just rely on opinion. We go out and do primary research."</p> <p>"Because I think, number one, they're rigorous. Our methodologies make sense and they have high validity, I think, in terms of measuring what they're supposed to be measuring"</p>	<p>"Because we want to triangulate, have more than one measure, and more than one way to measure, so we're more likely to have credible, accurate results."</p> <p>"I guess some of the theory touches on validating, like triangulating. That I fully support. I don't think there's a lot of people who'd argue with that. You don't want to be a one trick pony. You want to use more than one method, not necessarily a ton, but more than one method, and more than one stakeholder group, so that you can get that ... some validation. You have more likely ... your results are accurate."</p>	<p>"I think the fact that first of all it was transparent, second we did our homework, third we had the data, the data was there. We tried to keep going back to the data, the data, 'This is what the data tells us.' And whatever your perceptions were of this program or the clientele or what was happening out there in the real world, we rely on the data."</p> <p>"I think [the trustworthiness of conclusions] is reflective of my credibility as well, and my due diligence, and the methodological approach that I'm doing."</p>

	Low Evaluation Theory	Average Evaluation Theory	High Evaluation Theory
Trust is built by contextualizing the data			<p>“Well context is everything in evaluation.... understanding the context is critical in order to provide something that's been useful and helpful to them making decisions. Because those are the situations that they face every day. Unless you put yourself in the manager's shoes, in their position and understand the situation, you can't be particularly helpful. So, it's all about the utility, the helpfulness of the evaluation and the work that you're doing. That's always paramount in my mind. So, understanding context is critical to doing that.”</p>
Trust is built by the reputation of the evaluator	<p>“A lot of it had to do with the reputation of the firm, to be honest, more than the plan. I think that in an external evaluation, the reputation of the firm is big in the results being believed. Or the reputation of the person who's doing the work or whose name is on it.”</p> <p>“I think there's a reputation that we deliver hard-hitting findings.”</p>	<p>“And so she trusted me. I think partly because of my background and credibility in learning and development. But she appreciated, I think, being an engineer, of course they're very scientific and evidence based, and she appreciated that.... So, they trusted me, really, based on my experience, my credentials.”</p>	

Valuing

According to Scriven's (1980) logic of evaluation, the inclusion and justification of valuing the merit, worth, or significance of an evaluand is what distinguishes evaluation from other forms of (social science) research. His logic of evaluation requires the comparison of criteria measurements to warranted standards to determine how an evaluand is performing. Hurteau, Houle, and Mongiat (2009) consider an evaluative judgement justified when evaluators provide justification for the criteria chosen, justification of the standards compared to, and documentation of the process used to synthesize the judgment. Most evaluators, regardless of their training in and familiarity with evaluation theory, set criteria for their evaluations. However, fewer evaluators set standards for their criteria. Those that set and did not set standards did not differ depending on training in and familiarity with evaluation theory.

Criteria. Almost all evaluators described having identifying criteria (also described as measures or indicators) for their evaluation. Only one evaluator denied having any type of criteria. They explained,

Not specific criteria, no, and I'm being completely honest with you, no.... like I said, I'm a novice. It hasn't been a major part of my practice moving forward and just the way the conversation and the planning process evolved for this particular project... they didn't present the right timing to put that in place.

Their evaluation was more of a process evaluation, therefore, the evaluator stated most of the data was derived from observations. Of the evaluators who did set criteria for their evaluation, these criteria were generally created in collaboration with stakeholders, informed by relevant literature, or required by funders.

In alignment with other areas of evaluators' practice, evaluators of all levels of training in and familiarity with evaluation theory engaged stakeholders in identifying criteria. Some pulled information from the proposal – as one evaluator explained, “the first thing you always ask for is, okay, give me a copy of your proposal so I can see what you said and you were gonna do, and how you were gonna measure it.” – while others preferred discussions with stakeholders. As one evaluator explained,

And then for specific questions we have lots of discussions with our programs staff who are basically the clients or who they're actually users of this evaluation. Like what exactly we want to know. That basically guided our criteria and the evaluation questions, and then evaluation markers for this evaluation instead of indicators and what we have to do, and what we have to measure and look at during this process.

As another evaluator with high training in and familiarity with evaluation theory explained, they intentionally aligned criteria with stakeholders in order to enhance the utility of evaluative findings, "We always develop an evaluation design report, with indicators so that we can best inform decision making. So, we have indicators under program relevance, design and delivery, and impact." The importance of stakeholder engagement valued by evaluators with all levels of training in and familiarity with evaluation theory also played a role in selecting criteria for an evaluation.

Other evaluators discussed their criteria being informed by relevant literature. This review of the literature helped evaluators to identify important criteria for programs similar to their evaluand. As one evaluator described,

And so what we ended up doing to try to build out a roadmap and an evaluation plan for us was really to identify what it is that goes into successful partnerships and collaborations, and there's lots and lots of literature and theory out there on that. We used that to give us some goalposts and some things to be looking for as this initiative evolves.

Another evaluator discussed the use of meta-analysis studies to identify important criteria across multiple programs with similar activities and outcomes. Evaluators with low and average training in and familiarity with evaluation theory talked about using literature to select criteria. This was not a prominent reasoning process by evaluators with high evaluation theory.

Finally, some evaluators noted that certain criteria were required by the funder. These tended to be evaluations in the international development sphere, where sustainable development goals and DAC criteria were required, or federally funded evaluations where evaluations were meant to be aggregated upwards to a larger goal.

Standards. About a third of evaluators noted they did not include any type of standard in their evaluation, most of whom had average or low training in and familiarity with evaluation theory. Some did not include a standard, benchmark, or target because they felt it was not appropriate to their situation or appropriate for the evaluator to determine. One evaluator with low training in and familiarity with evaluation theory explained, “I think because of the bulk of our work on our end was qualitative there wasn’t a whole lot of benchmarking there.” Another evaluator with average training in and familiarity with evaluation theory described why they felt determining value was not their role as an evaluator,

I am the person who can help them gather the information and make sense of the information, but it’s still up to them to decide [what success is]... It was not for me to say, well this is really good and this was somewhat okay and this is like, what do you think makes sense based on what you know in your organization based on what you know about your sector based on what kind of resources you have to actually make change.

This evaluator felt stakeholders had more expertise to determine standards, and therefore the task should be left to them. A few evaluators noted that it was not within the scope of their evaluation, however, another evaluation team was to take lead on setting the standards. As one evaluator described, “There was a second team that did exactly what you said, was applying standards and benchmarks for performance once we’d worked out some of the basics on this thing.”

Evaluators who did set standards in their evaluation discussed basing them on research and negotiating benchmarks with stakeholders. Evaluators with low and average training in and familiarity with evaluation theory predominately turned to the literature to set standards for their evaluation. Some described using “industry standards benchmarks,” others explained their use of research-based reliability standards. Most evaluators who drew from research to determine standards intended to compare their evaluand’s results to others who used the same instrument. As one evaluator explains,

It was just a general well-being assessment scale because we wanted to look at some of those mental health outcomes. We used that, and it’s kind of a standardized thing, so

we used it to compare against other programs that were using similar scales. Yeah, we used that because it's more standardized and we could look at it more broadly. Some evaluators working in international settings described using other studies to determine standards, such as national benchmarks or international goals. A few evaluators used pre-program baseline measurements as their comparison point.

Evaluators also discussed negotiating standards with their stakeholders. One evaluator talked about using a rubric that was created in collaboration with program staff in order to rate levels of engagement and impact. Another evaluator described working with stakeholders to determine targets for criteria, "I still feel like we're almost scoping with our evaluability assessment to solidify [targets]. And then it will be finalized in our evaluation, our overarching plan for the expansion [of the program]." While standards created by stakeholders does not assure external validity, evaluators noted that discussing standards, benchmarks, and targets with stakeholders was a helpful activity in order to have both evaluation team and program staff think more deeply about the intended impact of the program. One evaluator discussed how she combined both stakeholder engagement and referring to the literature, "But they were kind of eyeballing it. They were round about. So at least I came at it with some kind of best practice reference."

Summary. While almost all evaluators acknowledged setting criteria for their evaluation, a third did not set standards, benchmarks, or targets. Evaluation criteria were generally created in collaboration with stakeholders, informed by the literature, or required by funders. Similarly, evaluation standards were informed by research or negotiated with stakeholders. Evaluators who did not set standards either did not feel it was within their responsibility, was not appropriate for the purpose of their evaluation, or cooperated with a second evaluation team that was responsible for setting standards. Exemplary quotes organized according to level of training and familiarity with evaluation theory are displayed in Table 4.6.

Table 4.6. Comparison of exemplary quotes around valuing in practice.

	Low Evaluation Theory	Average Evaluation Theory	High Evaluation Theory
Criteria created by or with client	“At the beginning of every evaluation we come up with a framework for the questions we’re going to ask as well as errors and success... We establish [this framework] with the program that we’re going to be evaluating or the department has already established whether they are already. When they design a program they’re supposed to come up with measurement indicators.”	“It’s not up to me to decide what is success of the program.... I set the criteria in that point of view of like, well what do you think? For example, I’d ask the managers, what evidence would you [want to see]? What would tell you that staff were moving towards a more employment focused or were more employment oriented?”	“the first thing you always ask for is, okay, give me a copy of your proposal so I can see what you said and you were gonna do, and how you were gonna measure it.” “We always develop an evaluation design report, with indicators so that we can best inform decision making. So, we have indicators under program relevance, design and delivery, and impact.”
Criteria informed by literature	“And so what we ended up doing to try to build out a roadmap and an evaluation plan for us was really to identify what it is that goes into successful partnerships and collaborations, and there's lots and lots of literature and theory out there on that. We used that to give us some goalposts and some things to be looking for as this initiative evolves, and I guess that's also the other key point of it was that this was an initiative that was just still getting off the ground.”	“Yeah, the literature. We did the big literature review, but I actually went and looked and found a few meta-analysis papers.”	
Criteria required by funder	“it was a combination of starting with what would be optimal to measure and then looking at how we could measure the feasibility rate of measuring what we want to measure and then also additional layer to that is what the CDC found appropriate to measures.”		“The grant obviously has certain objectives and indicators that they want you to put in. But I also am trying to align with the sustainable development goals in maternal health. And so larger global frames.”
Reasoning for not including criteria		“Not specific criteria, no, and I’m being completely honest with you, no.... like I said, I’m a novice. It hasn’t been a major part of my practice.”	

	Low Evaluation Theory	Average Evaluation Theory	High Evaluation Theory
Standards informed by research	<p>"The reliability is from some research by an expert, a [content area] expert.... It was just kind of like web searching and I found a bunch of his research that he had done.... He had developed a methodology for appraisal of rural transportation system, as well as some recommended indicators. So we based a lot of that on that."</p>	<p>"It was just a general well-being assessment scale because we wanted to look at some of those mental health outcomes. We used that, and it's kind of a standardize thing, so we used it to compare against other programs that were using similar scales. Yeah, we used that because it's more standardized and we could look at it more broadly."</p>	<p>"There is probably a baseline. It's usually what's in these programs. This is what's happening now. You know? And what's gonna happen when you put the program, those funds, towards doing something different?"</p>
Standards negotiated with stakeholders		<p>"But they were kind of eyeballing it. They were round about. So at least I came at it with some kind of best practice reference."</p>	
Reasoning for not including standards	<p>"We didn't go so far as to determine that. Instead, what we did was we were always giving feedback on those elements and observing what was happening that could be evidence of that element being there or not being there.... I think there wasn't ever a benchmark... I don't think we ever said, 'you have to have a certain number of these or a certain score on these elements to determine that this is successful.' We looked it and said, 'This is what you do. This is what you could do better or this is what you could do differently.'"</p>	<p>"I am the person who can help them gather the information and make sense of the information, but it's still up to them to decide [what success is]... It was not for me to say, well this is really good and this was somewhat okay and this is like, what do you think makes sense based on what you know in your organization based on what you know about your sector based on what kind of resources you have to actually make change."</p> <p>"we've reached out to various content or subject matter experts that are performing similar to see if they... really, we're all at the same stage in this patient's engagement and research so unfortunately there is maybe a lack of benchmark to compare results to"</p>	<p>"There was a second team that did exactly what you said, was applying standards and benchmarks for performance once we'd worked out some of the basics on this thing. But it was mainly standards that they were looking at. What is the standard we're trying to achieve with treatments of various types? So there was a team involved with that."</p>

Values

There are many ways to consider the use of values within evaluation practice: which and whose values are included, or what programmatic measures are valued. What stood out in this set of interviews was the use of values to justify decision making in practice. Evaluators with higher training in and familiarity with evaluation theory were driven by personal or professional values in their thinking and practical decision making. These values were very prominent in reflections on their evaluation practice and informed and influenced their decisions in relation to evaluation design, methodologies, stakeholder engagement, and reporting. These values were not always explicitly stated, and their connection to evaluation theory was not perfectly clear. The only time evaluators explicitly mentioned evaluation theory in their descriptions of practice was when an evaluator used a specific evaluation approach, such as developmental evaluation or participatory evaluation. Instead, references to evaluation theory were veiled references to values that are prominently discussed in evaluation theory, such as stakeholder engagement, the independence and objectivity of evaluators, or the utility of evaluative findings. Compared to those with lower evaluation theory, evaluators with high training in and familiarity with evaluation theory were more purposeful and intentional about their underlying reasoning for decisions and actions, relying on values as a driving force. On the other hand, evaluators with lower training in and familiarity with evaluation theory were more influenced by funder requirements and assumptions, evaluation processes set in place by the evaluation organization, or internal logic built through previous experience.

In all of the evaluations described by evaluators with high training in and familiarity with evaluation theory, there was a strong sense of intentionality in their thinking around evaluative decisions. In many ways, the different values that drove these different evaluators reflected the different branches of Alkin and Christie's (2012) evaluation theory tree, with emphases on methods, values, and use. For example, one evaluator continued to emphasize the importance of weaving multiple lines of evidence together to support the validity of conclusions in a complex situation. Illustrating their commitment to ensuring valid results they explained,

It's one of those basics when you learn evaluation, to be able to look at the different types of data that you need to collect and analyze to be able to come to meaningful

conclusions. Right, so every type of data has its own limitation, so ideally when you combine different types of data and conduct different forms of analyses, you will be able to minimize the risks of making a false claim, of coming to false conclusions about a program.

At multiples points in describing their evaluation, the evaluator emphasized the importance of using “a diversity of methodologies that could be used to be able to inform decision making, to come to support conclusions.” This commitment to “complex reasoning” and “minimizing risk of a false claim” informed their evaluation planning, data collection, and even reporting. When asked how they presented findings to stakeholders, the evaluator said,

So, that's a tricky one in terms of, do you wait to provide preliminary key findings across all lines of evidence, and so each evaluation is different, and I guess that decision is based on the complexity of the research. If you know that one line of evidence has many gaps that can only be filled by another line of evidence, then, in that case, you wait to present key finding across all lines of evidence.... I think in these cases, we need to look across lines of evidence so that we can say something that is more nuanced and more complete and not mislead decision makers to thinking that the evaluation will lead to specific conclusions that maybe one line of evidence suggests.

They described their use of multiple data collection methods as “safety nets” in order to triangulate and verify the validity of data. Even in talking about including a variety of stakeholders, they explained, “it’s very easy to come to false conclusions, so I think being able to look and understand that different stakeholders have different perspectives and different motivations for participating in evaluative research. We need to be able to look across many stakeholders to get various opinions.” The commitment to sound reasoning and valid methodology was present in every aspect of their reflection.

Another evaluator with high training in and familiarity with evaluation theory highlighted their commitment to including the voices of participants in Uganda. This valuing of participants’ perspective was a driving force throughout their evaluation. When describing why they took a participatory approach to the evaluation, the evaluator explained,

Partly because we are going into Uganda. And we are considered Westerners, and so we want to be embedded in the culture and embedded in the needs of the participants. And that's our primary purpose is to serve the mothers, and to serve the organizations. So, ensuring that their voice is heard throughout the entire process, we have [funders] in Uganda as well that sort of walk hand-in-hand with us, and in the hospitals, who are familiar with the context. So, including and engaging each primary user of the evaluation I think is the only way to approach this.

The evaluator described starting with an evaluability assessment in order to “provide more context” and “insight” to “either confirm my approach or sort of refute my original thoughts.”

The evaluator highly valued participants’ standpoint, saying

this evaluation is for them. I strongly believe whatever it is that I do, it’s for the participant, or for the people who would be using these results. If I go in and evaluate in my own way, in my own lens, it won’t resonate. It won’t be used, or it won’t be valuable.... I think I’m always driven by the people that I’m serving.

This reflection on behalf of the evaluator, that her lens or perspective might not resonate with the program participants, led her to involve a local evaluator throughout her evaluation process. She described this local evaluator as essential to increasing the credibility of their evaluation results; “I see her as a resource because she is embedded. She's in Uganda. She's lived and breathed the culture, and she might have insights to the approach, and how we might speak to people.” These descriptions of valuing participant perspectives were present in all of this evaluator’s reflections without being prompted. Promoting and valuing participant perspective as a means to ensure valid and useful findings was a common thread throughout their evaluative thinking.

As a final example of values informing evaluator’s thinking and eventual decision making, another evaluator with high training in and familiarity with evaluation theory was influenced by the end goal of program improvement through the use of evaluative findings. While many evaluators with all levels of evaluation theory mentioned the importance of evaluation use for program improvement, it was rarely as embedded in every aspect of the evaluation as it was here. Improving program activities through evaluation use was the driving

force behind all of this evaluators actions, from the choice of an evaluation approach, to the use of qualitative data over quantitative data, to engaging decision makers in order to get buy-in and ensuring that the evaluation team was immersed in the historical understanding of the area in order to contextualize the evaluative findings. The evaluator took care in working with stakeholders to align the required indicators from the funder to measure what is meaningful to the program. This was illustrated in how they discussed the disjuncture between required indicators and meaningful indicators. As they explained, “So, with this kind of weird indicators, like they don’t look whether what other systems we have, they have. No, we have to conduct it. So, we take that requirement, but we turn it into what is really useful for us, not just for the sake of conducting.” The evaluator’s commitment to program improvement and use of evaluative findings also led them to engage in capacity development of the stakeholders. The evaluator explained,

any evaluation should have a big component of the capacity development, so you just don’t conduct the evaluation. You really teach people, not teach but to help them to learn about different components. Why we apply this approach, why we apply this methodology, why we need to make it independent, why we should also show good sides and not very good sides of the program.

Commitment to use for program improvement continued even after the final report was submitted. When describing how evaluative findings were communicated to stakeholders, the evaluator described it as a “continuous process.” Stakeholders “listened to [the findings]. They provided some comments, but then it’s a continuous process. Evaluation doesn’t just end by finalizing the report.” The evaluation team continued to work with stakeholders to draft an action plan, “like what we are going to do, do they accept the recommendations, do they not accept, do they partially accept or agree with the recommendation, and what do they do to address that recommendation?” This evaluator’s value of program improvement through the use of evaluative findings was embedded in their evaluative thinking and overall conception of their role as an evaluator.

Most of the evaluators with lower training in and familiarity with evaluation theory were familiar with at least one or two evaluation theories. While they discussed some similar

influences on their practice including contextual restrictions, feasibility, and commitment to use, their practice did not come across as intentional or driven by values. Instead, they discussed their actions without much reflection on why. Some acknowledged their lack of reflection, “Oh boy, it’s hard to know for sure. That’s a good question but obviously I haven’t reflected on much.” Others worked within a very rigid process for conducting evaluation, either required by the funder or accepted practice by the evaluation organization. One evaluator described their evaluation plan, saying, “our results framework is pretty much based on the donor’s results framework. It’s not very different, but it aligns with the changes that we expected to do.” This evaluator acknowledged her evaluation organization wrote in indicators required by the funder without a clear plan of how to measure them.

Summary. Overall, the descriptions of evaluative thinking by evaluators with high training in and familiarity in evaluation theory were often intentional and highly influenced by a set of internal values. Their reasoning behind choices in practice were rarely because it was required or expected, but because it aligned with their values, both personal and professional.

Chapter Summary

The majority of AEA and CES evaluators have had some training on evaluation theory and are somewhat familiar with at least a few theories. Evaluators tend to get their training in evaluation theory from short webinars or presentations, graduate courses, or professional development workshops. Evaluators were most familiar with participatory evaluation, developmental evaluation, and utilization-focused evaluation. They were least familiar with deliberative democratic evaluation, consumer-oriented approach, and constructivist or fourth-generation evaluation.

When asked directly about the role of evaluation theory, most evaluators, even those with little familiarity with theory, speak positively of the influence on practice. Only a few expressed concerns about the practicality of evaluation theory. While many evaluators sung the praises of evaluation theory in the abstract, they acknowledged that in hindsight, evaluation theory did not play an active role in the evaluation they just discussed in detail for the

interview. When asked to reflect on a specific evaluation, evaluators rarely mentioned specific evaluation theories unless they were explicitly using them in their evaluation designs.

Evaluators with all levels of training in and familiarity with evaluation theory described being responsive to contextual factors. However, evaluators with higher training in and familiarity with evaluation theory were more skilled at negotiating multiple contextual factors at once. When it came to logical reasoning behind building trustworthiness of evaluative findings, evaluators with high evaluation theory placed more emphasis on contextualizing data in order to build stakeholder's buy-in and use of findings. Evaluators' justification for criteria or standards did not differ depending on training in or familiarity with evaluation theory. However, evaluators with high training in and familiarity with evaluation theory displayed more of a commitment to their personal or professional values in their evaluative thinking and decision making. These evaluators were more intentional in their choices and less reliant on contextual constraints or funder requirements.

CHAPTER 5

In this final chapter, a summary of the purpose, research questions, methods, and findings of this dissertation study is presented. Conclusions and discussion, as well as limitations and potential avenues of future research are also considered.

Summary

This study intended to provide insights into the relationship between evaluation theory and evaluation practice to answer two major research questions. The first research question asked to what extent American Evaluation Association (AEA) and Canadian Evaluation Society (CES) evaluators were trained in evaluation theory. The second asked what role evaluation theory played in evaluative thinking. While evaluation scholars have generally agreed that evaluation theory is important, this study adds to the growing empirical literature about evaluation theory in practice (Mark, 2018; Schwandt, 2014; Shadish, Cook, & Leviton, 1991; Smith, 1993).

A sequential, two-phase, mixed-methods design was used in this study, including a cross-sectional survey of AEA and CES members and in-depth one-on-one interviews. The online survey asked a random sample of AEA and CES evaluators about their training in evaluation theory, including the modalities of training and their familiarity with specific approaches. The interviews intended to unearth the role of evaluation theory in evaluators' thinking and decision making. As Vo (2013) acknowledges, evaluative thinking is not just a cognitive process, but "manifested as a problem-solving practice – the doing of evaluation" (p. 106). Therefore, to explore the entirety of evaluative thinking, it is necessary to consider both evaluators' thought processes and the decisions that arise, which results in action.

Findings from the survey found that 80% of AEA and CES evaluators had some type of training in evaluation theory. This training generally took the form of short presentations or webinars, graduate courses, and 1 — 4 day professional workshops. Overall, evaluators were most familiar with participatory evaluation, developmental evaluation, and utilization-focused evaluation, and least familiar with deliberative democratic evaluation, consumer-oriented evaluation, and constructivist or fourth-generation evaluation. Emergent themes from

interviews showed that evaluators with all levels of training in and familiarity with evaluation theory were responsive to a variety of contextual factors in their practice. Evaluators with all levels of training in and familiarity with evaluation theory endorsed the importance of use as a measure of a successful evaluation, and privileged stakeholder engagement as a factor in increasing use. However, evaluators with higher training in and familiarity with evaluation theory were more intentional in considering their options and more explicit in justifying the reasoning underlying their actions. Their decisions were generally backed by personal or professional values. For example, evaluators with greater knowledge of evaluation theory tended to rely on values (such as commitment to valid methodology, including participant voices, and increasing program improvement through evaluation use) that underlined their actions throughout their evaluation. Almost all reasoning or explanation for why decisions were made tied back to evaluators' values. While these evaluators never explicitly tied the prominence of these values back to evaluation theory, evaluators with lower training in and familiarity with evaluation theory did not show similar patterns in thinking. They were more influenced by preferences or requirements of funders, their disciplinary area, or their own organization.

Conclusions

Training in and Familiarity with Evaluation Theory

In Christie's (2003) study comparing practitioners' practice to evaluation theory, she notes that only 10 percent of her sample "indicated using a particular theory to guide their work" (p. 13). Her study also demonstrated that there is little overlap between prescribed evaluation theory and practitioner's work, therefore, she asserts that formal "theory is not requisite to evaluation practice" (p. 23). A survey of AEA members in 1987 by Shadish and Epstein found a similar lack of familiarity with evaluation theory among evaluation practitioners. However, results from this dissertation suggest that a much greater proportion of practicing evaluators are at least somewhat familiar with evaluation theory. Eighty percent of AEA and CES evaluators reported having some type of formal training in evaluation theory and almost all (97%) reported being at least somewhat familiar with a few evaluation theories. This

is in rather stark contrast to Shadish and Epstein's (1987) and Christie's (2003) results. It must be acknowledged that the sampling frame from these studies differ (Christie's sample was of evaluators of Head Start programs in California) and the studies were conducted 16 and 31 years apart. Whether this means that training around evaluation theory has advanced in the past 16 — 31 years, or that evaluators associated with professional memberships, such as AEA or CES, are more likely to be trained in evaluation theory, it acknowledges that most AEA and CES evaluators are no stranger to evaluation theory.

In particular, AEA and CES evaluators were most familiar with theories that emphasized stakeholder engagement and evaluation use (such as participatory evaluation and utilization-focused evaluation). This was reiterated in the commitment to use and stakeholder engagement by evaluators who were interviewed. AEA and CES evaluators frequently mentioned the influence of Michael Quinn Patton's writings on their evaluative thinking and practice. AEA and CES evaluators were least familiar with deliberative democratic evaluation, consumer-oriented evaluation, and constructivist or fourth-generation evaluation. Particularly interesting was AEA and CES evaluators' relative lack of familiarity with Scriven's logic of evaluation. Only 28% of evaluators were familiar (very or mostly) with the logic of evaluation, with 28% only somewhat familiar. This leaves 42% of AEA and CES evaluators who have never heard of Scriven's logic of evaluation. This is a rather startling number given the prominence most evaluation theorists and experts place on this meta-theory of evaluation. This percentage is lower than the percent of evaluators in 1987 who indicated they were familiar with the concept of evaluation as making value judgements by Scriven (42%; Shadish and Epstein, p. 577).

Role of Evaluation Theory

This study originally hypothesized that learning evaluation theory would broaden one's evaluative imagination by providing more tools to think with and a greater understanding of possible approaches and pathways in evaluation. AEA and CES evaluators with all levels of training in and familiarity with evaluation theory displayed a broad tool box of evaluation methods and were attentive to contextual factors. Influencing factors on methodological

decisions by evaluators were aligned with the findings from Braverman and Arnold (2008), emphasizing the importance of contextual factors. Evaluation theory did not play a role in evaluators' ability to think divergently about possible methods to use in their practice. However, evaluators with high training in and familiarity with evaluation theory were more intentional about their solutions to practical problems in evaluation. Instead of evaluation theory providing a guide to follow, evaluators with high training in and familiarity with evaluation theory showed enhanced reasoning and critical reflection in their practice. Evaluators with high training in and familiarity of evaluation theory also showed a strong commitment to values and used these values as justification for their decisions in practice. In their reflections on practice, these evaluators used their commitment to values as warrants for their decision making. They were able to be more critical when making decisions and considering why they chose one path over another. In this sense, evaluation theory served as more of a framework for considering and justifying decisions, rather than a model for making decisions. Evaluation theory does not provide a clear map showing which path to take, but instead provides the roots for critical thinking and logical reasoning. Evaluators with high training in and familiarity with evaluation theory didn't necessarily have more tools in their toolbox, but were simply more adept at considering complex reasoning about why and when it was appropriate to use particular tools.

Previous research on decision making in evaluation by Tourmen (2009) emphasized the importance of years of practice over exposure to evaluation theory. Tourmen concluded that being an expert in evaluation, compared to a novice, is what allows evaluators to be more situationally aware and more reflective in practice. This dissertation study took into account the practical experience of evaluators with both high and low training in and familiarity with evaluation theory. The practical experience of evaluators in both groups were approximately the same. Eighty-percent of evaluators with low training in and familiarity with evaluation theory identified as expert evaluators with 6 or more years of experience, while 100% of those with high training in and familiarity of evaluation theory identified as experts (see Table 4.1). This relatively equal distribution might suggest that differences in evaluative thinking between

these groups found by this study is, in part, due to evaluation theory and not length of evaluation practice.

Evaluation theory played a role in evaluators incorporation of values and logical reasoning into practice, but not their contextual responsiveness or valuing. Vo (2013) defines evaluative thinking as “a process by which one marshals evaluative data and evidence to construct arguments that allow one to arrive at contextualized value judgments in a transparent fashion” (p. 107). All of these aspects – contextual responsiveness, logical reasoning, valuing, and values – come together in evaluative thinking to produce high quality evaluation. Where evaluators with high training in and familiarity with evaluation theory tended to differ in their evaluative thinking was their nuanced attention to personal and professional values over contextual influences. They were also more attentive to the importance of contextualization of their data and findings to build stakeholder trust in findings.

Discussion

Evaluation theorists define evaluation theory as “guiding frameworks that specify what a good or proper evaluation is and how evaluation should be done” (Alkin, 2004, p. 5) and a “framework to guide the study and practice of program evaluation” (Stufflebeam & Coryn, 2014, p. 50). However, evaluators in this study did not seem to follow evaluation theory as a guide and rarely used evaluation theories explicitly in their practical decision making. Instead, their use of evaluation theory was implicit and perhaps at times subconscious. Evaluation theory seemed to provide a set of values evaluators could use to critically reflect on their thought process. The role of evaluation theory, as described by evaluators in this study, was closer to Mathison’s (2005) definition: “evaluation theory serves to provide a plausible body of principles that explain and provide directions to the practice of evaluation” (p. 142). These principles – or as emerged in this study, values – were not necessarily used by evaluators to decide what methods should be used or considered but were used to justify actions on issues like stakeholder engagement, building the trustworthiness of findings, and encouraging use of evaluative findings. Mark (2018) wrote evaluation theory tells us “when, where, why, and how different methods could and should be used in evaluation practice” (p. 134). This study

highlights evaluation theory's role in explicating not what evaluators should do, but helping to frame *why*. Evaluation theories seem to be, as Schwandt (2014) wrote, more “tools to think with” (p. 33) than directly applicable to practice. Instead of a strict guide to practice, evaluators with more training in and familiarity with evaluation theory are able to draw upon aspects of different theories to bolster their critical thinking and intentionality in solving practical problems.

Evaluation theorists portray evaluation practice as inextricable from evaluation theory. But as Shadish and Epstein's (1987) and Christie's (2003) studies demonstrated, there are evaluators conducting evaluations without knowledge of evaluation theory. Perhaps, evaluation theory is still playing a supporting role in these evaluations even though the evaluator has no personal knowledge of evaluation theory. In this study, evaluators, particularly those with low training in and familiarity with evaluation theory, discussed their reliance on funder requirements, organizational processes, and disciplinary assumptions. While the evaluators themselves may not have been trained in evaluation theory, these contextual structures may have been influenced by evaluation theory. The evaluation policy or process followed by evaluators with low training in or familiarity with evaluation theory may have originated from someone with knowledge of evaluation theory. Therefore, evaluation theory may still inform their practice indirectly. It is also possible that some aspects of evaluation theory have seeped into the general understanding of evaluation, even by those who have not formally studied evaluation theory. In particular, the ideas of use and stakeholder engagement were discussed by evaluators with all levels of evaluation theory. These elements seem to have become part of the basic understanding of what evaluation is, instead of a discrete and unique evaluation theory.

Lack of Standard Vocabulary

In his AEA Presidential Address, Shadish (1998) describes five main roles that evaluation theory plays in the profession. He claims evaluation theory provides a common language for evaluators, encompasses the values and issues evaluators care about, provides an identity for evaluators different from that of other professions, presents a face to the outside world, and

forms the foundational knowledge that defines the profession. While evaluation theory did seem to outline common values and provide justification of decisions to non-evaluators, evaluation theory did not seem to provide a common language even among the evaluators interviewed. Many evaluators discussed the difficulty in defining “evaluation theory” and agreeing on what approaches should be considered under its scope. Additionally, their difficulty in responding to questions about concepts like logical reasoning and valuing made it clear that not all evaluators spoke the same language. For example, there was a considerable difference in the use of terms such as criteria, indicator, and measures. The range of responses to questions about the use of evaluation standards also called into question whether there was a commonly agreed upon definition of standard. This may not be surprising, given that 43% of AEA and CES evaluators responded they had never heard of Scriven’s logic of evaluation. Ozeki (2016) found a similar lack of familiarity and utility of general logic among AEA evaluators. There was even variety in evaluators use of the term stakeholder engagement. Some discussed stakeholders’ role in the process of the evaluation while others referred to collecting data from stakeholder groups. While these differences in terminology were generally rectified with clarification, it does point to the lack of a standard vocabulary among evaluators and the failure of evaluation theory to standardize fundamental concepts and vocabulary across evaluators.

Revisiting the Conceptual Framework

In Chapter 1, a conceptual framework was outlined to illustrate the role of evaluation theory and its place in evaluative thinking and practice. Towards the end of this study, the general structure of this conceptual framework holds with a few additions. First, the framework was intended to be at the level of an individual. However, this greatly underestimates the collaborative nature of evaluations. Very rarely did an evaluator describe themselves as the sole decision maker. Instead, they described the influence of their fellow evaluation team members, their larger organization, or the involvement of stakeholders and funders. This tendency for collaborative decision making is reminiscent of Schwandt’s (2018) argument for evaluative thinking as a social practice, recognizing that decision making is often a group effort, not the actions of a single individual. This collaborative decision making and interaction between

multiple networks of evaluative thought both constrained and expanded evaluators' thinking about evaluation at multiple points.

Second, the framework mentions the influence of evaluators' personal background, without giving emphasis to any specific background characteristics. From interviews with evaluators it was clear that their personal beliefs about evaluation and values in their work played an influential role in how they interpreted situations and came to decisions. Evaluators' personal background was not just a filter at the beginning but served as a lens they saw everything through at every point in the evaluation. Finally, this conceptual framework positions the notion of evaluative imagination as evaluators' ability to consider various pathways, opportunities, and choices that an evaluator can use in a specific situation. It was posited that evaluation theory could expand these pathways and help evaluators to think more divergently. Instead, it was less about whether evaluators considered alternative approaches, but the intentionality behind why they considered them. Where the evaluative imagination could be ascertained, it was not about the boundaries of possibilities but why one choice rather than another. Evaluators with more training in and familiarity with evaluation theory were more intentional about choosing the path they walked and were more nuanced in their reflection about why. These evaluators aligned their actions with underlying values about evaluation, instead of simply following funder requirements or organizational procedures unquestioningly.

Limitations

Limitations to the survey phase of this study consist of an imperfect sampling frame, the possibility of biased responses, and potential misunderstanding of questions. This study intended to generalize to all practicing evaluators in North America. However, there are no comprehensive databases that include all evaluators within the United States and Canada. Therefore, the AEA and CES membership databases were used as a proxy. There are most likely many evaluators in both countries who conduct evaluations as part of their work that are not members of either of these professional organizations, therefore, they would be outside of the scope of inference for this study. These evaluators may be less likely to have training in

evaluation theory. Second, it is possible that social desirability played a role in how respondents answered survey questions. Respondents may have overestimated their familiarity with evaluation theory in their responses to be perceived as more adept or knowledgeable. Additionally, the survey questions relied on self-report without any triangulation of the depth of respondents' knowledge. As previous studies have shown, evaluators who claim they use a particular approach do not always follow the approach as prescribed by theorists (Coryn, et al., 2011; Miller & Campbell, 2006). Finally, respondents may have read quickly through the directions and might have not fully absorbed the definition of evaluation theory given in the beginning of the survey. Some evaluators may have responded to questions with a broader definition of what is considered evaluation theory than the study's definition. This may result in an over estimation of training in and familiarity with evaluation theory.

Initially, this study hoped to compare evaluators with high training in and familiarity with evaluation theory to those that had none. However, given the high proportion of training in evaluation theory reported by survey respondents and of those who agreed to participate in interviews, there was not an opportunity to compare evaluators with no evaluation theory. Therefore, evaluators with lower and higher training in and familiarity with evaluation theory were compared.

Finally, the qualitative phase of this study is limited by the subject matter itself. Reflecting on one's own evaluative thinking and explicating the cognitive process behind decision making in practice is not easily verbalized. Evaluators were not always adept at explaining themselves or clearly identifying the reasonings behind their actions. Sometimes this was because the evaluation they were recalling happened in the past, other times it was because they had not fully considered their own reasoning before. If time allowed, it might be beneficial to follow up with interviewees and see if they had additional insights after our initial conversation.

Future Research

In many ways, this exploratory study unearthed more research questions than it provided answers. With a general understanding of how many AEA and CES evaluators receive training in evaluation theory and with promising effects of learning evaluation theory on

practice, more research is needed into the most effective ways to teach evaluation theory to novice evaluators. While many evaluators in this study discussed the benefits of learning evaluation theory on their evaluative thinking and practice, many also discussed their confusion over the multitude of evaluation theories, models, and approaches. They criticized many evaluation theories for not being practical. Some evaluators described having practical experience before learning about evaluation theory as crucial to their internalization of theory's lessons. As one evaluator explained,

I think there's a lot to be said for designing educational programs that really give people the practice, the practical experience kind of early on first so that you can understand the practical world and how things work and then start introducing theory over time.

There is certainly discussion on integrating practice-based learning for new evaluators, however, more research is needed on the best ways to integrate practical experience and learning evaluation theory. Similar research might explore whether there is a best time to learn evaluation theory (e.g., after some experience with evaluation practice) or whether there are other ways besides practical experience to simulate this experience for more novice evaluators.

Additionally, a potential area for future research is about the effects of being able to compare perspectives across multiple approaches. Potential questions in this line of research might include: Does exposure to multiple evaluation theories with opposing perspectives increase your ability to think critically about your own evaluation practice? If so, what are the implications for how evaluation theory is taught and communicated throughout the field? If, as this study points to, evaluation theory aids evaluators by identifying underlining values to consider practical decision making, should evaluation theories be communicated in ways that clearly identify these values? Should evaluation theories be compared by the values exposed in each theory as is done by Shadish, Cook, and Leviton (1991) and to a certain extent by the evaluation tree (Alkin, 2012; Merten & Wilson, 2012)? How can heuristic devices help evaluators see more practical utility in evaluation theory and help evaluators become more critical, reflective practitioners?

With AEA's recent release of evaluator competencies and discussions about professionalization of the field, more research is needed into how evaluation theory fits into

these competencies and what is the minimum level to be considered competent. As seen in this study, many AEA and CES evaluators already have some familiarity with evaluation theory. But is that enough to fulfill the competency? An important line of research might look at how much evaluation theory is enough for good evaluation. This would, of course, require a minimum benchmark for what would be considered good evaluation. This dissertation discusses good evaluation along the lines of contextual responsiveness, logical reasoning, valuing, and values. Even evaluators with low training in and familiarity with evaluation theory were contextually responsive, could consider the trustworthiness of their findings, and at times used criteria and standards to explicate value. Is this necessary and sufficient to be considered a competent evaluator?

Most evaluators with formal training in evaluation theory identified their training as from a graduate course or short webinar or presentation. The ability to learn complex concepts like evaluation theory differs widely in these modalities ranging from one hour to four months. With expansion of evaluation competencies and professionalization, it may not be realistic to expect all evaluators to learn evaluation theory from a full semester graduate course. In that case, more research is needed on how to evaluation theory can be taught in one-hour segments.

Findings from this dissertation also began to wade into the territory of critical thinking in evaluation. Many education scholars are focused on defining and creating measures for critical thinking in public schools. But the area of measuring critical thinking within the evaluation field is a ripe area for exploration. In particular, more research is needed on how critical thinking and reflective practice in evaluation is, if at all, taught to new evaluators. What are the best ways to encourage active critical reflection in practice by evaluators?

Finally, this study alluded to the role of evaluation commissioners and funders in structuring decisions evaluators make in practice. Given the importance of evaluation commissioners' preferences, assumptions, and requirements, more research is needed into the benefits of increasing their knowledge of evaluation theory. Increasing the evaluative imagination of evaluation commissioners may strengthen request for proposals and better align evaluation findings with information needs. Conversely, having a limited knowledge of

evaluation theory may limit the kind of evaluations that are conducted and the type of evaluative findings that are produced. Further research into how evaluation theory could benefit evaluation commissioners and in effect the quality of evaluations would be beneficial to the overall quality of evaluations.

References

- Alkin, M. C. (2003). Evaluation theory and practice: Insights and new directions. In C. Christie (Ed.), *The practice-theory relationship in evaluation* (pp.81-89). San Francisco, CA: Jossey-Bass.
- Alkin, M. C. (Ed.). (2004). *Evaluation roots: A wider perspective of theorists' views and influences* (1st ed.). Thousand Oaks, CA: Sage.
- Alkin, M. C., & Christie, C. A. (Eds.) (2005). *Theorists' Model in Action. New Directions for Evaluation*, 106. San Francisco, CA: Jossey-Bass.
- American Evaluation Association. (2018). *The 2018 AEA evaluation competencies*. Retrieved from <https://www.eval.org/page/competencies>.
- Archibald, T., Neubauer, L. C., & Brookfield, S. D. (2018). The critically reflective evaluator: Adult education's contributions to evaluation for social justice. In A. T. Vo & T. Archibald (Eds.), *Evaluative Thinking. New Directions for Evaluation*, 158, 109–123.
- Archibald, T., Sharrock, G., Buckley, J., & Young, S. (2018). Every practitioner a “knowledge worker”: Promoting evaluative thinking to enhance learning and adaptive management in international development. In A. T. Vo & T. Archibald (Eds.), *Evaluative Thinking. New Directions for Evaluation*. 158, 73–91.
- Baker, A., & Bruner, B. (2012). *Integrating evaluative capacity into organizational practice: A guide for nonprofit and philanthropic organizations and their stakeholders*.
- Barela, E. M. (2005). How school district evaluators make sense of their practice. *Dissertation Abstracts International*, 67/02. (UMI No. 3208339).
- Benkofske, M. T. (1995). An examination of how evaluators and evaluation clients choose data collection methods: The factors and decision-making process. *Dissertation Abstracts International*, 57/03 (166). (UMI No. 9623619).
- Bennett, G., & Jessani, N. (Eds). (2011). *The knowledge translation toolkit: Bridging the know-do gap: A resource for researchers*. New Delhi, India: Sage.
- Bledsoe, K. L. & Graham, J. A. (2005). The use of multiple evaluation approaches in program evaluation, *American Journal of Evaluation*, 26(3), 302-319.

- Braverman, M. T., & Arnold, M. E. (2008). An evaluator's balancing act: Making decisions about methodological rigor. In M. T. Braverman, M. Engle, M. E. Arnold, & R. A. Rennekamp (Eds.), *Program evaluation in a complex organizational system: Lessons from Cooperative Extension. New Directions for Evaluation*, 120, 71–86.
- Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluation thinking: Insights from research on critical thinking. *American Journal of Evaluation*, 36(3), 375-388.
- Campbell, D. T. (1969). Reforms as experiments. *American psychologist*, 24, 409-429.
- Campbell, D. T. (1991). Methods for the experimenting society. *Evaluation Practice*, 12(3), 223-260.
- Canadian Evaluation Society. (2018). *Competencies for Canadian evaluation practice*. Retrieved from https://evaluationcanada.ca/sites/default/files/competencies_for_canadian_evaluation_practice_revised_edition_2018.pdf?v2
- Chandler, M. "How Evaluators Engage Theory and Philosophy in Their Practice." Paper presented at the American Evaluation Association Annual Conference, St. Louis, Mo., Nov. 2001.
- Charmaz, K. (2014). *Constructing Grounded Theory*, 2nd Ed. Thousand Oaks, CA: Sage.
- Chelimsky, E. (2012). Balancing evaluation theory and practice in the real world. *American Journal of Evaluation*, 34(1), 91–98.
- Chouinard, J., Boyce, A., Hicks, J., Jones, J., Long, J., Pitts, R., & Stockdale, M. (2016). Navigating theory and practice through evaluation fieldwork. *American Journal of Evaluation*, 1-13.
- Christie, C. A. (2003). Understanding evaluation theory and its role in guiding practice: Formal, folk, and otherwise. *American Journal of Evaluation*, 97, 91 – 93.
- Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. In C. A. Christie (Ed.), *The practice-theory relationship in evaluation* (pp. 7-35). *New Directions for Evaluation*, 97. San Francisco, CA: Jossey-Bass.

- Christie, C. A. & Rose, M. (2003). The language of evaluation theory: Insights gained from an empirical study of evaluation theory and practice. *Canadian Journal of Program Evaluation*, 18(2), 33-45.
- Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Schroter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990-2009. *American Journal of Evaluation*, 32(2), 199-226.
- Coryn, C. L. S., Wilson, L. N., Westine, C. D., Mateu, P. F., Hobson, K. N., Vo, A. T., Schroter, D. C., Fiekowsky, E. L., Abu-Obaid, R. N., & Ramlow, M. (2019). Do incentives increase response rates to internet surveys of American Evaluation Association members? Findings from a randomized experiment. *American Journal of Evaluation*.
- Cousins, B. J., & Earl, L. M. (1992). The case for participatory evaluation, *Educational Evaluation and Policy Analysis*, 14(4), 397-418.
- Cousins, B. J., & Earl, L. M. (1999) When the boat gets missed: Response to M. F. Smith, *American Journal of Evaluation*, 20(2), 309-317.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research*. 3rd Ed. Thousand Oaks, CA: Sage.
- Dahler-Larsen, P. (2012). *The Evaluation Society*. Stanford, CA: Stanford University Press.
- Datta, L. (2003). Important questions, intriguing method, incomplete answers. In C. A. Christie (Ed.), *The practice-theory relationship in evaluation* (pp. 37-46). *New Directions for Evaluation*, 97. San Francisco, CA: Jossey-Bass.
- Donaldson, S. I., & Lipsey, M. W. (2006). Roles for theory in contemporary evaluation practice: Developing practical knowledge. In I. Shaw, J. Greene, & M. Mark (Eds.), *Handbook of evaluation*. London: Sage.
- Donaldson, S. I., Christie, C. A., & Mark, M. M. (Eds.) (2015). *Credible and actionable evidence: the foundation for rigorous and influential evaluations*. Second Edition. Las Angeles: Sage.
- Earl, L. (2013). Looking forward: Evaluation in New Zealand education. In M. Lai & S. Kushner (Eds.), *A national developmental and negotiated approach to school self- evaluation*. Bingley, UK: Emerald Publishing.

- Endsley, M. R. (1997). The role of situation awareness in natural decision making. In C. E. Zsombok & G. Klein (Eds.), *Naturalistic decision making* (pp. 269-283). Mahwah, NJ: Lawrence Erlbaum.
- Facione, P. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. Millbrae, CA: The California Academic Press.
- 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. In A. T. Vo & T. Archibald (Eds.), *Evaluative Thinking. New Directions for Evaluation*. 158, 49–72.
- Fitzpatrick, J. L. (2004). Exemplars as case studies: Reflections on the links between theory, practice, and context. *American Journal of Evaluation*, 25(4), 541-559.
- Fitzpatrick, J. L., Christie, C., & Mark, M. M. (2009). *Evaluation in action: Interviews with expert evaluators*. Thousand Oaks, CA: Sage.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines* (3rd ed.). New York, NY: Longman.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). New York, NY: Longman.
- Fournier, D. M. (1995). Establishing evaluative conclusions: A distinction between general and working logic. In D. M. Fournier (Ed.), *Reasoning in evaluation: Inferential links and leaps* (pp. 15-32). *New Directions in Evaluation*, 68. San Francisco, CA: Jossey-Bass.
- Fournier, D. M., and Smith, N. L. (1993). Clarifying the merits or argument in evaluation practice. *Evaluation and Program Planning* 16(4), 315-323.
- Greene, J. C. (2005). Context. In S. Matheson (Ed.), *Encyclopedia of evaluation* (pp. 82-84). Thousand Oaks, CA: Sage.
- Greene, J. C. (2006). Method choice: Five discussant commentaries. *New Directions for Evaluation*, 113, 111-127.
- Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology? *Journal of Mixed Methods Research*, 2(1), 7-22.
- Greene, J. C., Benjamin, L., & Goodyear, L. (2001). The merits of mixing methods in evaluation. *Evaluation*, 7(1), 25-44.

- Greene, J. C., Caracelli, V. J. (1997) Defining and describing the paradigm issue in mixed-method evaluation in J. C. Greene and V. J. Caracelli (eds) *Advances in Mixed-Method Evaluation: The challenges and benefits of integrating diverse paradigms, New Directions for Evaluation No. 74*, 5-17. San Francisco, CA: Jossey-Bass.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), p. 255-274.
- Grob, G. F. (2018). Evaluation practice: Proof, truth, client relationships, and professional growth. *American Journal of Evaluation*, 39(1), 123-132.
- Guilford, J. P. (1950). Creativity. *American Psychologist*, 5, 444-454.
- Guzman, G. (2009). What is practical knowledge? *Journal of Knowledge Management*, 13(4), 88-98.
- Harnar, M. A. (2012). *Theory Building Through Praxis Discourse: A Theory- And Practice-Informed Model of Transformative Participatory Evaluation*. CGU Theses & Dissertations, 57.
- Henry, G. T., & Mark, M. M. (2003). Toward an agenda for research on evaluation. In C. Christie (Ed.), *The practice-theory relationship in evaluation (pp.69-80)*. San Francisco, CA: Jossey-Bass.
- Hood, S. (2005). Culturally responsive evaluation In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 96-102). Thousand Oaks, CA: Sage.
- House, E. R. (1977). *The logic of evaluative argument* (Monograph No. 7). Los Angeles, CA: Center for the Study of Evaluation, UCLA Graduate School of Education, University of California.
- House, E. R. (1977). The logic of evaluative argument. In *CSE monograph series in evaluation* (Vol. 7). Los Angeles, CA: UCLA, Center for the Study of Education.
- House, E. R. (1980). *Evaluating with validity*. Beverly Hills, CA: Sage.
- House, E. R. (2003). Stakeholder bias. In C. Christie (Ed.), *The practice-theory relationship in evaluation (pp.69-80)*. San Francisco, CA: Jossey-Bass.

- Hurteau, M., Houle, S., & Mongiat, S. (2009). How legitimate and justified are judgments in program evaluation? *Evaluation*, 15(3), 307-319.
- King, J. A., & Alkin, M. C. (2018). The centrality of use: Theories of evaluation use and influence and thoughts on the first 50 years of use research. *American Journal of Evaluation*, (currently available through OnlineFirst).
- Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation*, 31(3), 400-413.
- Kundin, D. M. (2010). A conceptual framework for how evaluators make everyday practice decisions. *American Journal of Evaluation*, 31(3), 347-362.
- LaVelle, J. M. (2014). An examination of evaluation education programs and evaluator skills across the world. (Unpublished doctoral dissertation). Claremont Graduate University, Claremont, CA.
- Leeuw, F. L., & Donaldson, S. I. (2015). Theory in evaluation: Reducing confusion and encouraging debate. *Evaluation*, 21(4) 467–480.
- Lipsey, M. W. (1993). Theory as method: Small theories of treatments. *New Directions for Program Evaluation*, 57, 5-38.
- Mark, M. M. (2007). Building a better evidence base for evaluation theory. In N.L. Smith & P.R. Brandon (Eds.), *Fundamental issues in evaluation (pp.111-134)*. New York, NY: Guilford Press.
- Mark, M. M. (2018). Strengthening links between evaluation theory and practice, and more: Comments inspired by George Grob's 2017 Eleanor Chelimsky Forum presentation, *American Journal of Evaluation*, 39(1), 133-139.
- Mathison, S. (Ed.). (2005). *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage.
- McGuire, M., Zorzi, R., & Perrin, B. (2002). Projet de la société canadienne d'évaluation à l'appui de la défense des intérêts et du perfectionnement professionnel, *Société Canadienne d'Evaluation*. <http://www.evaluation.francophonie.org>.
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, 69, 220-232.

- Mertens, D. M. & Wilson, A. T. (2012). *Program Evaluation Theory and Practice: A Comprehensive Guide*. New York, N.Y.: Guilford.
- Miller, R. L. (2010). Developing standards for empirical examinations of evaluation theory. *American Journal of Evaluation*, 31(3), 390-399.
- Miller, R. L. & Campbell, R. (2006). Taking stock of empowerment evaluation: An empirical review. *American Journal of Evaluation*, 27(3), 296-319.
- Newall, A., Shaw, C., & Simon, H. A. (1962). The process of creative thinking. In H. E. Gruber, G. Terrell, & M. Wertheimer (Eds.), *Contemporary approaches to creative thinking* (pp. 153-189). New York: Pergamon.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewoods, NJ: Prentice-Hall.
- Ozeki, S. (2016). Three empirical investigations into the logic of evaluation and valuing practices (Unpublished doctoral dissertation).
- Patton, M. Q. (1988). The evaluator's responsibility for utilization. *Evaluation Practice*, 9(2), 5-24.
- Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods*. 3rd Ed. Thousand Oaks, CA: Sage.
- Patton, M. Q. (2012). Contextual pragmatics of valuing. In G. Julnes (Ed.), *Promoting evaluation in the public interest: Informing policies for judging value in evaluation*. *New Directions for Evaluation*, 133, 97-108.
- Patton, M. Q. (2015). *Qualitative research and evaluation methods*, 4th ed. Los Angeles: Sage.
- Patton, M. Q. (2017). *Facilitating evaluation*. Thousand Oaks, CA: Sage.
- Patton, M. Q. (2018). A historical perspective on the evolution of evaluative thinking. In A. T. Vo & T. Archibald (Eds.), *Evaluative Thinking*. *New Directions for Evaluation*. 158, 11-28.
- Patton, M.Q. (2008). *Utilization-focused evaluation*, 4th edition. Thousand Oaks, CA: Sage.
- Preskill, H., & Boyle, S. (2008). Insights into evaluation capacity building: Motivations, strategies, outcomes, and lessons learned. *The Canadian Journal of Program Evaluation*, 23, 147-174.
- Rog, D. (2015). Infusing theory into practice, practice into theory. *American Journal of Evaluation*, 36(2), 223-238.

- Rog, D. J. (2012). When background becomes foreground: Toward context-sensitive evaluation practice. In D. J. Rog, J. L. Fitzpatrick, & R. F. Conner (Eds.), *Context: A framework for its influence on evaluation practice. New Directions for Evaluation*, 135, 25–40.
- Saldaña, J. (2016). *The coding manual for qualitative researchers. 3rd Ed.* Thousand Oaks, CA: Sage.
- Scheaffer, R. L., Mendenhall III, W., Ott, R. L., & Gerow, K. G. (2012). *Elementary Survey Sampling* (7th ed.). Belmont, CA: Thompson.
- Schwandt, T. (2005). The centrality of practice to evaluation. *American Journal of Evaluation*, 26(1), 95-105.
- Schwandt, T. (2014). On the mutually informing relationship between practice and theory in evaluation. *American Journal of Evaluation*, 35(2), 231–236.
- Schwandt, T. A. (2002). *Evaluation practice reconsidered*. New York, NY: Peter Lang Publishing.
- Schwandt, T. A. (2003). 'Back to rough ground!' Beyond theory to practice in evaluation. *Evaluation*, 9, 353-364.
- Schwandt, T. A. (2018). Evaluative thinking as a collaborative social practice: The case of boundary judgment making. In A. T. Vo & T. Archibald (Eds.), *Evaluative Thinking. New Directions for Evaluation*. 158, 125–137.
- Scriven, M. (1980). *The Logic of Evaluation*. Inverness, CA: Edgepress.
- Scriven, M. (1991). *Evaluation Thesaurus (4th ed.)*. Newbury Park, CA: Sage.
- Scriven, M. (2011). Evaluating evaluations: A meta-evaluation checklist. Retrieved from The Evaluation Center, evaluation checklists website: www.wmich.edu/evaluation/checklists.
- Scriven, M. (2012). The logic of valuing. In G. Julnes (Ed.), *Promoting valuation in the public interest: Informing policies for judging value in evaluation. New Directions for Evaluation*, 133, 17-28.
- Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19(1), 1-19.

- Shadish, W. R., & Epstein, R. (1987). Patterns of program evaluation practice among members of the evaluation research society and evaluation network. *Evaluation Review* 11(5), 555-590.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage Publications.
- Shön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books, Inc.
- Smith, N. L. (1993). Improving evaluation theory through the empirical study of evaluation practice. *American Journal of Evaluation*, 14(3), 237-242.
- Smith, N. (2010). Characterizing the evaluand in evaluating theory. *American Journal of Evaluation*, 31(3), 383-389.
- Smith, N. (2016). Evaluative fit as a principal meta-evaluation criterion. Paper presented at the *American Evaluation Association Conference*, Atlanta, GA.
- Stake, R. E. (1995). *The Art of Case Study Research*. Thousand Oaks, CA: Sage.
- Stake, R. E. (2003). *Standards-Based and Responsive Evaluation*. Thousand Oaks, CA: Sage.
- Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. *American Psychologist*, 51, 677-688.
- Stufflebeam, D. L. (1999). Program evaluations metaevaluation checklist. Retrieved from The Evaluation Center, evaluation checklists website: www.wmich.edu/evaluation/checklists.
- Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, & applications* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Szany, M., Azzam, T., Galen, M. (2012). Research on evaluation: A needs assessment. *The Canadian Journal of Program Evaluation*, 27, 39-64.
- Tashakkori, A. & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

- Tashakkori, A. & Teddlie, C. (2003). *Handbook of Mixed Methods in Social & Behavioral Research*. Thousand Oaks, CA: Sage.
- The Canadian Evaluation Society. (2010). *Competencies for Canadian Evaluation Practice*. Retrieved from https://evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice.pdf.
- Toulmin, S. E. The Uses of Argument. New York: Cambridge University, 1964.
- Tourmen, C. (2009). Evaluators' decision making: The relationship between theory, practice, and experience. *American Journal of Evaluation* 30(1), 7-30.
- Vo, A. (2013). Toward a definition of evaluative thinking (Unpublished doctoral dissertation). University of California, Los Angeles (UCLA).
- Vo, A. T., & Archibald, T. (Eds.). (2018). *Evaluative Thinking. New Directions for Evaluation*, 158.
- Vo, A. T., & Christie, C. A. (2015) Advancing research on evaluation through the study of context. In Paul R. Brandon (Ed.), *Research on evaluation. New Directions for Evaluation*, 148, 43-55.
- Vo, A. T., Schreiber, J. S., & Martin, A. (2018). Toward a conceptual understanding of evaluative thinking. In A. T. Vo & T. Archibald (Eds.), *Evaluative Thinking. New Directions for Evaluation*. 158, 29–47.
- Volkov, B. B. (2011). Beyond being an evaluator: The multiplicity of roles of the internal evaluator. In B. B. Volkov & M. E. Baron (Eds.), *Internal evaluation in the 21st century. New Directions for Evaluation* (pp. 25–42). San Francisco, CA: Jossey-Bass.
- Wehipeihana, N., & McKegg, K. (2018). Values and culture in evaluative thinking: Insights from Aotearoa New Zealand. In A. T. Vo & T. Archibald (Eds.), *Evaluative Thinking. New Directions for Evaluation*. 158, 93–107.
- Weiss, C. H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice*, 9(1), 5-19.
- Williams, J. E. (1989). A numerically developed taxonomy of evaluation theory and practice. *Evaluation Review*, 13(1), 18-31.

Yarbrough, D. B., Schula, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users, 3rd edition*. Thousand Oaks, CA: Sage.

Appendix A

Survey Questions for AEA and CES Members

Survey Questions for AEA and CES Members

Survey Preface

Thank you for participating in this survey. The following questions will ask about your familiarity with and training in evaluation-specific theory, models, and approaches.

You do not need to be familiar with these theories, models, and approaches to respond to this survey. Your responses are still very valuable.

Evaluation theories, models, and approaches provide explicit guidelines on how evaluation should or could be practiced. Evaluation-specific theory is considered separate from program theory – a conceptual understanding of how a program works – and social science theories – theories drawn from other fields such as economy, political science, and sociology.

By **evaluation specific theories, models, and approaches**, I am referring to “a coherent set of conceptual, hypothetical, pragmatic, and ethical principles forming a general framework to guide the study and practice of program evaluation” (Stufflebeam & Coryn, 2014, p. 50).

Some examples of evaluation specific theories, models, and approaches include: theory-driven evaluation, utilization focused evaluation, participatory evaluation, CIPP model, realist evaluation, culturally responsive evaluation, and empowerment evaluation.

Again, you do not need to be familiar with these theories, models, and approaches to respond to this survey. Your responses are still very valuable.

Survey Questions

1. How familiar are you with evaluation-specific theories, models, and approaches?
 - a. Not at all familiar with any
 - b. Somewhat familiar with only a few
 - c. Somewhat familiar with many
 - d. Very familiar with only a few
 - e. Very familiar with many
2. Have you had any formal training in evaluation-specific theories, models, and approaches? *For the purpose of this survey, formal training is defined as an intentionally structured learning experience led by an instructor or facilitator on the subject area, either in person or online. This may include webinars, lectures, online courses, professional development workshops, college courses, etc.*
 - a. Yes
 - b. No

The next few questions ask about your formal training in evaluation-specific theories, models, and approaches. Different kinds of training will be asked about separately. For example, you will be asked if you have attended any short webinars or presentations, professional development

workshops, and semester long courses. Please choose the response that best characterizes your experience.

3. [If YES to Q2] Have you attended any short webinars or presentations (**1 – 6 hours**) about evaluation-specific theories, models, or approaches?
 - a. Yes
 - b. No
 - c. [If YES to Q3] Which of the following best describes the webinars or presentations (1 – 6 hours) you have attended? *Select all that apply.*
 - i. Conference presentation
 - ii. American Evaluation Association (AEA) sponsored webinar (e.g., coffee break, eStudy, etc.)
 - iii. Canadian Evaluation Society (CES) sponsored webinar
 - iv. Another evaluation association sponsored webinar
 - v. Other, please specify [text box]
 - d. Within each short webinar or presentation (1 – 6 hours), how many evaluation-specific theories, models, or approaches did the short webinars or lectures typically cover?
 - i. A single theory, model, or approach
 - ii. Multiple theories, models, or approaches
 - iii. Depends on the webinar or lecture
4. [If YES to Q2] Have you attended a workshop or professional development training (**1 – 4 days**) that taught evaluation-specific theories, models, or approaches?
 - a. Yes
 - b. No
 - c. [If YES to Q4] Were these workshops or professional development trainings (1 – 4 days) you attended generally online or in person?
 - i. Online
 - ii. In Person
 - iii. Mix of online and in person
 - d. To what extent were evaluation-specific theories, models, and approaches the focus of the workshop(s) or professional development training(s) (1 – 4 days) you have attended?
 - i. The sole focus
 - ii. A major focus
 - iii. A minor focus
 - e. Within each workshop or professional development training (1 – 4 days) you have attended, how many evaluation-specific theories, models, and approaches did they typically cover?
 - i. A single theory, model, or approach
 - ii. Between 2 – 5
 - iii. More than 5

5. [If YES to Q2] Have you attended any short professional development course (**1 – 2 weeks**) that taught evaluation-specific theories, models, or approaches?
- a. Yes
 - b. No
 - c. [If YES to Q5] Were these short professional development courses (1 – 2 weeks) generally online or in person?
 - i. Online
 - ii. In Person
 - iii. Mix of online and in person
 - d. To what extent were evaluation-specific theories, models, or approaches the focus of the short professional development course(s) (1 – 2 weeks) you have attended?
 - i. The sole focus
 - ii. A major focus
 - iii. A minor focus
 - e. Within each professional development course (1 – 2 weeks) you have attended, how many evaluation-specific theories, models, or approaches did they typically cover?
 - i. A single theory, model, or approach
 - ii. Between 2 – 5
 - iii. More than 5
6. [If YES to Q2] Have you attended an undergraduate level course at a college or university (**semester long, 2 – 3 months**) that taught evaluation-specific theories, models, or approaches?
- a. Yes
 - b. No
 - c. [If YES to Q5] How many undergraduate level courses (semester long, 2 – 3 months) about evaluation-specific theories, models, or approaches have you taken?
 - i. [text box]
 - d. Were these undergraduate courses (semester long, 2 – 3 months) you have attended generally online or in person?
 - i. Online
 - ii. In Person
 - iii. Mix of online and in person
 - e. To what extent were evaluation-specific theories, models, or approaches the focus of the undergraduate course(s) (semester long, 2 – 3 months) you have attended?
 - i. The sole focus
 - ii. A major focus
 - iii. A minor focus
 - f. Within each undergraduate course (semester long, 2 – 3 months) you have attended, how many evaluation-specific theories, models, or approaches did they typically cover?
 - i. A single theory, model, or approach
 - ii. Between 2 – 5

- iii. More than 5
- 7. [If YES to Q2] Have you attended a graduate level course at a college or university (**semester long, 2 – 3 months**) that taught evaluation-specific theories, models, or approaches?
 - a. Yes
 - b. No
 - c. [If YES to Q5] How many graduate level courses (semester long, 2 – 3 months) about evaluation-specific theories, models, or approaches have you taken?
 - i. [text box]
 - d. Were these graduate courses (semester long, 2 – 3 months) you have attended generally online or in person?
 - i. Online
 - ii. In Person
 - iii. Mix of online and in person
 - e. To what extent were evaluation-specific theories, models, or approaches the focus of the graduate course(s) (semester long, 2 – 3 months) you have attended?
 - i. The sole focus
 - ii. A major focus
 - iii. A minor focus
 - f. Within each graduate course (semester long, 2 – 3 months) you have attended, how many evaluation-specific theories, models, or approaches did they typically cover?
 - i. A single theory, model, or approach
 - ii. Between 2 – 5
 - iii. More than 5
- 8. Have you done any self-study (outside of formal training) on evaluation-specific theories, models or approaches?
 - a. Yes
 - b. No
 - c. [If YES to Q4] What did this self-study look like? What activities did it include? Please be as specific as you can. *If your self-study included reading books, which books? If your self-study included watching videos or reading websites, which videos and websites?*
 - i. [text box]
- 9. Have you received any mentoring around evaluation-specific theories, models, or approaches?
 - a. Yes
 - b. No
 - c. [If YES to Q5] What did this mentoring look like? What activities did it include? *Please be as specific as you can.*
- 10. How familiar are you with the following evaluation-specific theories, models, or approaches? [Resposne options include Never heard of it, Somewhat familiar, Mostly familiar, Very familiar]
 - a. CIPP model

- b. Constructivist or fourth-generation evaluation
 - c. Consumer oriented approach
 - d. Deliberative democratic evaluation
 - e. Developmental evaluation
 - f. Empowerment evaluation
 - g. Goal-based evaluation
 - h. Goal-free evaluation
 - i. Scriven's Logic of evaluation
 - j. Objectives-based evaluation
 - k. Participatory evaluation
 - l. Realist evaluation
 - m. Responsive or stakeholder-centered evaluation
 - n. Theory-driven evaluation
 - o. Utilization-focused evaluation
 - p. Other, please specify [text box]
11. How does your knowledge of evaluation theory influence your evaluation practice and thinking about evaluation?
- a. [text box]
12. What is the highest degree of evaluation education or training in evaluation that have you completed? *Select only one.*
- a. Professional development workshops
 - b. Badges or certificates in evaluation
 - c. Graduate certificate in evaluation
 - d. Master's Degree in evaluation
 - e. Master's Degree in another field
 - f. Doctorate Degree in evaluation
 - g. Doctorate Degree in another field
 - h. Postdoctorate degree
13. Do you hold a credentialed evaluator designation from the Canadian Evaluation Society?
- a. Yes
 - b. No
14. Do you consider yourself an evaluation practitioner? *In other words, do you conduct evaluations or participate as part of an evaluation team?*
- a. Yes
 - b. No
 - c. [If NO to Q14] What do you consider your role in the field of evaluation? [text box] [If NO to Q14, end survey]
15. [If YES to Q14] Approximately how many years have you been practicing evaluation?
- a. [text box]
16. In the past 2 years, approximately how many evaluations have you worked on?
- a. [text box]

17. How would you describe yourself as an evaluator?

- a. Novice evaluator
- b. Intermediate evaluator
- c. Advanced evaluator
- d. Expert evaluator

18. Which content areas do you typically work in? *Select all that apply.*

- a. Criminal justice
- b. Government evaluation
- c. Higher education
- d. International development
- e. Nonprofit, non-governmental organizations, or foundations
- f. Pre-K to 12 education
- g. Public health
- h. Business and industry
- i. Other, *please specify* [text box]

19. Which statement best describes your role as an evaluator or as part of an evaluation team?

- a. I am external to the organizations I evaluate
- b. I am internal to the organizations I evaluate
- c. I conduct a mix of both external and internal evaluations

The last couple of questions ask you about your preferred approach. This survey acknowledges that the choice of evaluation approach or theoretical orientation is (1) not always singular, and (2) dependent on context.

20. Do you have a preferred evaluation approach(es) or theoretical orientation(s)? *While the choice of evaluation approaches may depend on the context, this question is asking if there is an approach that you tend to lean towards or tend to rely on more than others.*

- a. Yes, I would say I have a preferred evaluation approach
- b. No, I do not have a preferred evaluation approach
- c. I don't know enough about evaluation approaches to say

21. [If Q13 YES] **Free of resource and contextual constraints**, what is your preferred evaluation approach(s) or theoretical orientation(s)? And for what reasons do you prefer this approach?

- a. [textbox]

22. [If Q13 YES] **When planning an evaluation with contextual and resource constraints**, what evaluation approach(s) or theoretical orientation(s) do you most utilize in your evaluation practice? And for what reasons do you usually use this evaluation approach?

- a. [textbox]

23. The second phase of this research intends to look at evaluators' decision making in practice. Would you be interested in participating in an hour long follow up interview? Participants will receive a \$25 gift card as a token of appreciation.

- a. Yes
- b. No

24. [If Q22 YES] Thank you! Please provide your name and email address for follow up communication. *This information will be kept seprate from your responses and will only be used for research purposes.*
- a. First Name
 - b. Last Name
 - c. Email Address

Appendix B

CES Member Sample Demographics Compared to Population Demographics

CES Member Sample Demographics Compared to Population Demographics

Characteristic	CES Member Sample (<i>n</i> = 313)	CES Member Population (<i>N</i> = 634)
Gender		
Female	70.3%	69.9%
Male	27.8%	28.4%
Another gender	0.3%	0.3%
Population Group		
White	67.7%	69.2%
Prefer not to answer	6.4%	7.1%
Black	6.1%	5.4%
Asian	5.4%	4.4%
Mixed	3.8%	3.3%
Chinese	2.6%	1.6%
Other	2.6%	3.2%
Arab	1.6%	1.6%
Filipino	1.3%	1.3%
No selection made	1.3%	0.8%
Aboriginal (First Nations, Metis, Inuit)	0.6%	0.8%
Korean	0.3%	0.2%
Latin American	0.3%	1.3%
Educational Degrees Received		
Bachelors	47.0%	46.1%
Master's	70.6%	70.7%
Doctorate	20.4%	21.5%
Major Job Responsibility		
Evaluation practice	31.9%	31.4%
Consulting	17.3%	16.9%
Management/administration	15.3%	15.3%
Research	10.9%	11.4%
Policy of Program development	9.9%	11.2%
Other	4.8%	5.0%
Student	3.5%	3.3%
Teaching/Training	3.5%	3.3%
Prefer not to answer	1.9%	1.6%
No selection made	1.0%	0.6%
Type of Employer		
Other	17.3%	18.5%
Provincial gov't/agency	17.3%	15.9%

Characteristic	CES Member Sample (<i>n</i> = 313)	CES Member Population (<i>N</i> = 634)
Self-employed	16.9%	17.7%
Federal gov't/agency	13.7%	13.1%
Private business	10.5%	10.3%
University	8.9%	10.1%
Student	4.5%	5.0%
Foundation	3.5%	2.5%
Municipal gov't/agency	2.2%	3.5%
College	1.9%	0.9%
Prefer not to answer	1.3%	0.8%
Elementary/secondary school system	0.6%	0.6%
No selection made	0.6%	0.3%
Retired	0.6%	0.8%

Appendix C
Interview Protocol

Interview Protocol

Interview Preface

Do you have any questions for me about the consent form I sent via email?

Would it be alright if I recorded our discussion today?

In this study, I want to know more about the influencing factors behind evaluators' decisions in practice.

I'd like to start off with some general questions and then move into a more in-depth discussion of an evaluation you've conducted.

Questions

Opening

1. How did you first become an evaluator?
 - a. What is your educational background?
 - b. How many years have you been a practicing evaluator?
 - c. In what fields do you conduct evaluations?
2. What motivated you to be an evaluator?
3. What do you see as the role of an evaluator?

Case Example – Influential Factors in Decision Making

4. Could you tell me about an evaluation you're particularly proud of?
 - a. How did you become involved as the evaluator?
 - b. What was the purpose of the evaluation? To look at the process, make improvements, or determine outcomes or impact?
 - c. What approach or methods did you use?
 - i. Why did you choose these approaches/methods?
 - ii. What information did you draw on to make this decision?
 - iii. Did you consider any alternative approaches?
 - iv. Were stakeholders involved in the process of choosing these methods?
 - d. How did you involve stakeholders in other aspects of the evaluation process, if you involved them at all?
 - i. *Are there other factors that influenced your thinking on stakeholder involvement?*

- e. Did you identify specific criteria during the evaluation? (By criteria, I mean the indicators or dimensions of the program you measured.)
 - i. Who was involved with choosing these criteria?
 - ii. What was the motivation for choosing these criteria?
- f. Did you identify specific standards or benchmarks during the evaluation? (By standards, I mean a comparison point to determine whether the evaluation findings should be considered good or bad.)
 - i. Who was involved with choosing these standards?
 - ii. What was the motivation for choosing these standards?

I want to ask about the argument or reasoning underlying your evaluation. By which I mean, the logic that connects the data you chose to collect to the conclusions you made.

- g. Why do you think stakeholders trust your evaluative conclusions? (What made them buy-in to the findings? Or what made them find the conclusions valid?)
 - i. Evaluative arguments can be constructed in ways that make them stronger or weaker. How strong would you say this evaluative argument was?
 - ii. Are there any other factors that influenced your thinking on building your evaluative argument?
- h. How did you report your findings?
 - i. What led to your decision to use this reporting method?
 - ii. What were the important factors that influenced your decision to report findings like that?
- i. I originally asked you to talk about an evaluation you were particularly proud of. Why did you choose this one?
- j. In what ways is this example typical of your evaluation practice?
 - i. In what ways is it importantly different?

Direct Report of Theory in Practice

One of the concepts I am particularly interested in is how evaluators use evaluation theory in their practice. *By evaluation-specific theory, models, or approaches, I am referring to a coherent set of principles that form a framework for guiding the study and practice of evaluation. For example, some major theories of evaluation include Utilization Focused Evaluation, Democratic evaluation, and the logic of evaluation.*

5. Would you say that evaluation theory influenced this evaluation?

6. IF YES, FORMAL TRAINING

- a. Which theories do you rely on the most in your practice?
 - i. In what ways do you use these theories?
- b. What do you see as the role of evaluation theory in your evaluative thinking?
- c. How does learning about multiple evaluation theories/approaches inform:
 - i. Your ability to be responsive to contextual factors?

- ii. The values you incorporate into an evaluation?
- iii. How you create an evaluative argument?
- iv. Ability to think critically about evaluation?
- v. Ability to think creatively about evaluation?

7. IF NO FORMAL TRAINING

- a. Do you think evaluation theory has influenced your evaluative thinking?
 - b. What is the most important guide to decision making in your practice?
8. What other ideas, experiences, or learnings have importantly influenced your thinking or practice as an evaluator? Please tell me about 1 or 2 of them.
9. Is there something else you think I should know to better understand your evaluative thinking or evaluation practice?

Closing

10. Is there something that occurred to you during this interview that you might not have thought about before this interview?
11. Is there anything you would like to ask me about my research?

That's all the questions I have. Thank you again for taking the time out of your day to talk with me.

Appendix D
HSIRB Approval

WESTERN MICHIGAN UNIVERSITY



Institutional Review Board
FWA00007042
IRB00000254

Date: September 20, 2018

To: Chris Coryn, Principal Investigator
Lyssa Wilson Becho, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair

Re: IRB Project Number 18-09-14

This letter will serve as confirmation that your research project titled "An Exploratory Investigation into the Role of Evaluation Theory in Evaluative Thinking" has been **approved** under the **expedited** category of review by the Western Michigan University Institutional Review Board (IRB). The conditions and duration of this approval are specified in the policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note: This research may **only** be conducted exactly in the form it was approved. You must seek specific board approval for any changes to this project (e.g., ***you must request a post-approval change to enroll subjects beyond the number stated in your application under "Number of subjects you want to complete the study"***). Failure to obtain approval for changes will result in a protocol deviation. In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the IRB for consultation.

Reapproval of the project is required if it extends beyond the termination date stated below.

The Board wishes you success in the pursuit of your research goals.

Approval Termination:

September 19, 2019

Office of the Vice President for Research
Research Compliance Office
1903 W. Michigan Ave., Kalamazoo, MI 49008-5456
PHONE: (269) 387-8293 FAX: (269) 387-8276
WEBSITE: wmich.edu/research/compliance/hsirb

CAMPUS SITE: 251 W. Walwood Hall