

SUNNASSEE, EMMA M., Ph.D. *Translating Theory to Practice: A Multi-Method Study of the Contribution Analysis Evaluation Approach.* (2020)  
Directed by Dr. Ayesha S. Boyce and Dr. John T Willse. 210 pp.

As the evaluation discipline moves away from the “black-box” evaluations, theory-based evaluation approaches such as Contribution Analysis (CA) have gained popularity. This study responded to explicit requests to probe deeper into CA (e.g., Budhwani & McDavid, 2017; Delahais & Toulemonde, 2012; Dybdal et al., 2012; Lemire, 2010) as well as a more general call for a systematic examination of evaluation cases in order to generate “practical knowledge (Schwandt, 2008) about the Contribution Analysis (CA) evaluation approach (Mayne, 2012).

As such, this dissertation employed a multiple-method design of two segments, Phase I consisted a systematic review of the conceptual literature, and the ensuing Phase II called upon a multiple case study of seven empirical CA cases. Taken together, the study design allowed for a systematic inquiry into the theoretical translation and practice of contribution analysis (CA). Specifically, the study investigated how Contribution Analysis is conceptualized by theorists, and how this is understood and translated into practice by evaluation practitioners. In this endeavor, the dissertation identified elements of effective practice by characterizing adaptations, adjustments, and innovations, and identified conditions under which practices may be different (e.g., evaluand, contexts, or evaluators) (Smith, 1993).

The study pursued a richer understanding of the contexts of practice and to probe deeper into the translation of theory to practice using empirical evidence to further develop and improve on the current understandings of CA theory and practice. The

findings from this study contribute to the empirical body on evaluation theory and practice, as well as methodological contribution on conducting research on evaluation. Moreover, findings from this study seek to inform the development of CA contingency theories, which identify conditions under which practices are effective, which are considered to be the strongest types of evaluation approaches as they are buttressed by empirical knowledge of practice (Shadish et al., 1991; Smith, 1993).

The dissertation findings illustrate the complexity of the contexts in which CA evaluations take place, and identified contextual factors related to the program theory, sector of practice, geographic scope, temporal interval, and the effect-object. It seems that for certain contexts and purposes, CA by itself may not be enough as it not amenable for the comparison of causal packages across contexts nor cases. Therefore, approaches like Process Tracing (PT) or Bayesian modeling, or Qualitative Comparative Analysis (QCA) are used to facilitate direct comparisons of effects/influences of impact pathways from case to case. I argue however, that the understated value of Contribution Analysis is in its function as a validation approach, which creates a framework to build robust and sound arguments in support of contribution claims.

*Keywords:* Contribution Analysis, evaluation, contingency theory, evaluation theory, research on evaluation, validation of program theory, program theory

TRANSLATING THEORY TO PRACTICE: A MULTI-METHOD STUDY OF THE  
CONTRIBUTION ANALYSIS EVALUATION APPROACH

by

Emma M. Sunnassee

A Dissertation Submitted to  
the Faculty of The Graduate School at  
The University of North Carolina at Greensboro  
in Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

Greensboro  
2020

Approved by

---

Committee Co-Chair

---

Committee Co-Chair

Dedicated to my parents, Dr. Devdass and Mit Sunnassee,  
and sisters, Melanie and Deesha,  
who have always had more faith in me than I could find in myself.

APPROVAL PAGE

This dissertation, written by Emma M. Sunnassee, has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Co-Chair \_\_\_\_\_  
Ayesha Boyce

Committee Co-Chair \_\_\_\_\_  
John Willse

Committee Members \_\_\_\_\_  
Micheline Chalhoub-Deville

\_\_\_\_\_  
Tiffany Smith

\_\_\_\_\_  
Date of Acceptance by Committee

\_\_\_\_\_  
Date of Final Oral Examination

## ACKNOWLEDGMENTS

This dissertation is the product of many people, and I would like to express my sincere gratitude to the following people:

Dr. Chouinard, who has taught me to think critically and enriched my academic pursuits and professional development. Dr. Chalhoub-Deville, who planted the seed that guided my academic research and ultimately led to this dissertation. Dr. Smith, who has helped me improve the quality of my writing. My committee co-chairs, Dr. Willse, who has persevered with me over the years and helped me jump the last hurdle, and Dr. Boyce, my advisor, mentor, and role model. Her untiring work and guidance helped me both personally and professionally. I am forever indebted to you.

My ERM family, Cherie Avent, Jonathan Rollins, Robyn Pitts, Yan Fu, Aileen Reid, Julianne Zemaitis, Jeremy Acree, Allison Ames, Justin Long, among many others, whose friendships brought both insightful debates and comic relief. My work family in Winston-Salem/Forsyth County Schools, particularly Latandra Vance and Dr. Ward.

My friends and family from around the world, I have always felt your encouragement especially my grandfather who never fails to call every Friday. Though we are far apart, I have always felt your encouragement. In closing, I'd like to thank my future husband, Jonathan Oakley, who has stood by me and offered unconditional love and support over the last five years, and my parents, and sister, who have through their unwavering support, endured sacrifices and been unequivocally committed to my journey since the very beginning. Today would be impossible without the four of you.

## TABLE OF CONTENTS

	Page
LIST OF TABLES .....	x
LIST OF FIGURES .....	xii
CHAPTER	
I. INTRODUCTION .....	1
Statement of the Problem.....	6
Purpose of the Study and Research Questions.....	8
Significance of the Study .....	11
Organization of the Document.....	12
II. REVIEW OF THE LITERATURE .....	13
Chapter Overview .....	13
An Introduction to Evaluation Theory .....	15
Paradigms and Associated Assumptions .....	18
Causality in Evaluation .....	20
Design Approaches to Causation .....	22
Limitations of Method-Driven Approaches .....	29
The Black Box .....	30
Contribution Analysis .....	31
Theoretical-Orientation: Theory-Based Evaluation.....	31
Setting the Evaluation Stage .....	33
Theory of Change (ToC).....	35
Contribution Analysis Steps.....	39
Spheres of Influence and Magnitude of Contribution Claim.....	41
CA as a Validation Framework .....	46
Kane’s Interpretive Argument.....	48
Linking Evaluation Theory and Practice .....	51
Role of Theory in Evaluation.....	52
The Need for Research on Evaluation on CA.....	53
III. METHODOLOGY .....	56
Research Objectives.....	56
Overview of Research Design .....	58

Guiding Worldviews and Philosophies .....	64
Phase I: Systematic Review of the Literature .....	65
Review of Conceptual Literature .....	65
Scoping Review of Empirical Literature .....	66
Intermediary Phase: Development of Instrumentation .....	69
Refining Instruments.....	70
Phase II: Multiple Case Study.....	72
Multiple Case Study Design .....	73
Sampling, Case Selection Rationale, and Inclusion Criteria .....	74
Response Rate.....	77
Data Sources and Data Collection Strategies.....	79
Data Quality: Trustworthiness and Ensuring Rigor .....	88
Credibility .....	89
Transferability.....	91
Dependability.....	92
Confirmability.....	93
Chapter Summary .....	94
IV. FINDINGS.....	96
Research Question 1: What is the Current Theory of the Contribution Analysis Evaluation Approach? .....	96
Overarching Findings.....	97
Finding 1: CA is Currently Entering its Fourth Generation and Becoming Increasingly Refined.....	97
Finding 2: CA is Situated on the “Methods” Branch of the “Evaluation Theory Tree”.....	98
Finding 3: CA is Guided by Realist Ontology and Relativist Epistemology .....	99
Finding 4: Three Key Tenets That Guide CA are: (a) Theory of Change Mechanism, (b) Correspondence between the Sphere of Influence and Magnitude of the Claim Made, and (c) the Plausibility and Credibility of the Contribution Story .....	101
Research Question 2: What Does CA Look Like in Practice?.....	102
Overarching Findings and Themes .....	103
Finding 1: CA Evaluation are Traditionally Impact Evaluations, Examining the Contribution of a Variety of Interventions .....	103
Finding 2: CA Practitioners Tend to be Discipline-specific Researchers, Trained Evaluators, and Had Experience	

Conducting Various Types of Evaluations, across a Diverse Range of Contexts, across the Globe .....	109
Finding 3: Participants Saw Their Roles in the CA Assignments as: Methodological/ Technical (Evaluation) Advisor ( <i>n</i> =3), Critical Analyst ( <i>n</i> =2), Facilitator of Local Change ( <i>n</i> =2), Achiever Working with Program Manager ( <i>n</i> =1), Educator of Clients ( <i>n</i> =1), Knowledge Broker ( <i>n</i> =10), and Resource of Stakeholders ( <i>n</i> =1).....	118
Research Question 2a. How Do Practitioners Implement CA? .....	120
Overarching Findings and Themes .....	120
Finding 1: The ToC is Typically Developed Retrospectively by the Evaluators and Seldom Involved Stakeholders in the Process Other Than to Confirm the Evaluator’s Postulated ToC .....	121
Finding 2: The Interpretation of the ToC was Seldom Conducted with the Stakeholders or Intended Users.....	122
Finding 3: The Postulated ToC is Typically Founded on Scientific Theory .....	123
Finding 4: There are Misunderstandings of Key CA Terminology among Practitioners.....	123
Finding 5: There is a Misunderstanding of When to Use CA—“Analysis of Contribution” Versus Contribution Analysis .....	124
Research Question 2b: What Modifications and Reinforcements are Made in the Design of CA Evaluations? .....	126
Overview of Findings .....	127
Finding 1: The Adaptations Made to the Modification Made on the Evaluation Design Include: Process Tracing and Developmental Evaluation .....	128
Finding 2: The Relevant Explanation Finder (REF), Evidence Table, and Chronological Mapping of Events are Strategies Used to Reinforce the CA Tenets .....	130
Research Question 3: What Conditions and Contextual Factors (e.g., of the Evaluation and Program) Challenge and Facilitate the Implementation of CA in the Field? .....	131
Overview of Findings .....	131
Finding 1: Five Dimensions of Context Were Identified to Have Contributed to the Complexity of the Evaluation: Program Theory Complexity, Sectoral, Geographic, Temporal, and Effect-Object.....	131
Finding 2: Retrospective ToC Require Accessible and Accurate Data Records .....	134

V. DISCUSSION AND CONCLUSIONS.....	136
Study Overview .....	136
Overview of Findings .....	137
Research Question 1: Conceptual Development of Theory.....	137
Research Question 2: CA in Practice .....	137
Research Question 3: Conditions and Contextual Factors that Affect Practice .....	139
Reflections on Key Take-Aways.....	140
Contribution Analysis as an Evaluation Approach .....	140
Methodological Concerns in Research on Evaluation (RoE) .....	140
Improving Evaluation Theory.....	142
Recommendations for Practice .....	144
Contributions of the Study .....	145
Strengths of the Study.....	146
Limitations .....	147
Future Directions .....	149
Concluding Thoughts.....	150
REFERENCES .....	152
APPENDIX A. CONCEPTUAL ARTICLES BY MAYNE.....	177
APPENDIX B. CONCEPTUAL PUBLICATIONS BY OTHER AUTHORS .....	178
APPENDIX C. ALL EMPIRICAL ARTICLES DISCOVERED IN PHASE I .....	180
APPENDIX D. LEGEND OF JOURNAL ACRONYMS.....	185
APPENDIX E. PROSPECTIVE CASES.....	187
APPENDIX F. QUALTRICS QUESTIONNAIRE .....	192
APPENDIX G. RECRUITMENT EMAIL .....	196
APPENDIX H. CONSENT FORM.....	197
APPENDIX I. INTERVIEW PROTOCOL.....	199
APPENDIX J. PRELIMINARY POOL OF INTERVIEW QUESTIONS .....	202
APPENDIX K. CASE EXTRACTION FORM .....	205

APPENDIX L. SOURCES OF EMPIRICAL CA CASES BY YEAR .....	207
APPENDIX M. CA QUALITY MARKERS .....	209

## LIST OF TABLES

	Page
Table 1. Design Approaches, Variants, and Basis for Causal Inference .....	24
Table 2. Three Types of Evaluation Aims and Associated Causal Inferences.....	34
Table 3. Alignment of Key Mechanisms and Steps of CA .....	39
Table 4. Evaluative Inquiry, Magnitude of Contribution, Verification, and Contribution Story .....	43
Table 5. Alignment of IA with CA Mechanisms .....	50
Table 6. Study Timeline .....	62
Table 7. Alignment of Phases to Instruments, Sample, Procedures, and Analyses .....	63
Table 8. Inclusion Parameters for Conceptual Literature Review .....	66
Table 9. Initial Search Parameters for Scoping Review.....	67
Table 10. Summary of Data Sources and Analysis to Research Questions .....	79
Table 11. Meta-Matrix of Cases on Conditions and CA Score.....	88
Table 12. Paradigms Underpinning CA .....	101
Table 13. Summary of Case Profiles.....	105
Table 14. Demographics by Role Assumed in CA Assignment .....	110
Table 15. CA Experience by Role Assumed in CA Assignment.....	113
Table 16. General Evaluation Experience and Training .....	115
Table 17. Characteristics of Typical Evaluation Assignment .....	116
Table 18. Evaluator Role in CA Assignment.....	119
Table 19. Type of ToC Developed, Types of Postulated ToC, and Stakeholder Engagement.....	121

Table 20. Overview of Case by Purpose, Evaluation Approach, and Development of ToC.....	127
Table 21. Type of Complexity Per Case .....	132
Table 22. Conceptual Articles by Mayne Represented in Study .....	177
Table 23. Conceptual Publications by Other Authors (Represented in Study) .....	178
Table 24. Table of All Empirical Articles .....	180
Table 25. Legend of Journal Acronyms .....	185
Table 26. Markers of a “Good” Theory of Change .....	209
Table 27. Spheres of Influence Investigated by Case.....	210
Table 28. Evidence Table .....	210

## LIST OF FIGURES

	Page
Figure 1. Diagram of the “Black-Box,” Illustrating a Lack of Clarity Between Inputs and Output .....	5
Figure 2. Sample Impact Pathway .....	37
Figure 3. A Basic Theory-of-Change .....	38
Figure 4. Direct Control, Direct Influence, and Indirect Influence .....	42
Figure 5. Diagram of the Procedures for Exploratory Sequential Study, Detailing the Systematic Review (Phase I) and Instrument Development (Intermediary), and Multi-Case Study (Phase II) .....	61
Figure 6. Number of Empirical Articles Published Per Year from 2007 to Date (2020) .....	69
Figure 7. Spheres of Influence and Program Components .....	126

## **CHAPTER I**

### **INTRODUCTION**

In its brief 80-year history, the field of evaluation has undergone considerable transformations. For much of its existence, dating from Ralph Tyler’s “Eight-Year Study” in the 1940s (e.g., Alkin, 2013; Donaldson & Lipsey, 2006) and throughout the federally funded large-scale social programs of the 1960s and 1970s (Donaldson & Lipsey, 2006), the discipline as a whole has leaned heavily on values founded in natural sciences (Donaldson & Lipsey, 2006). As such, methods of the 20th century tended to reflect strong positivist roots as studies sought to make generalized causal inferences in order to establish net impact (Donaldson & Lipsey, 2006).

The turn of the 21st century marked a significant shift in the discipline (Donaldson, 2003). While social betterment remains the ultimate desired outcome (Johnson et al., 2009; Mark, Henry, & Julnes, 2000), the function of evaluation has stretched beyond act of “valuing” and making judgments of merit, worth, and significance (Scriven, 1967). Instead, today’s evaluation approaches reflect elevated responsibilities such as inquiring of program effects (Cousins, Whitmore, & Shulha, 2012) and producing knowledge that influences decisions of trivial consequence to those having constitutive effects (e.g., sanctions, policies) (Dahler-Larsen, 2014). More recently, there has been a heightened interest in inquiry designs that can demonstrate clear linkages between the program and the outcomes as by funding agencies compel for

more transparent processes (e.g., see Craig, 2013; Downes, Novicki, & Howard, 2018; Dybdal, Nielsen, & Lemire, 2010; Kotvojs & Shrimpton, 2007; Patton, 2012; Scottish Government Social Research, 2012; Wimbush, Montague, & Mulherin, 2012).

In its attempt to satiate stakeholder and client demands for these new-age responsibilities, the field has become host to a plethora of innovative evaluation approaches and continuous methodological developments. A survey of recent international trends revealed that program effectiveness and cause-and-effect interests are chief motives for commissioning an evaluation (Gates & Dyson, 2017). At the heart of these investigations is the causality inquiry, which seeks to establish attribution of the effects (e.g., program outcomes) to the cause (e.g., intervention) (Dybdal et al. 2010). Attribution is at the core of natural sciences methods (Iverson, 2003), and this interest is as pronounced in the social sciences, including program evaluation (Patton, 2012). In fact, the case has been made that resolving causal questions has been a core mandate of the field since its beginning (Gates & Dyson, 2017; Mayne, 2011).

However, the issue of *how* to best establish causality is central to the discipline and is a thoroughly debated topic. Early social scientific accounts reveal designs borne of the positivist paradigm were most favored. Today, we can see a considerable evolution in inquiry designs to be more considerate of complexities of social phenomena. This perspective holds relatively rigid views of the world and data collection (Fitzpatrick, Sanders, & Worthen, 2011) in that there is one objective reality that is “stable, observation, and measurable” (Merriam, 1998, p. 4). As such, cause-and-effect inquiries followed suit and upheld the positivist values dictating that causality could only be

‘determined’ scientifically and based on empirical observation, systematic experimentation, and quantitative analyses (Iverson, 2003; Patton, 2002).

Today, remnants of the old doctrines still linger in method-driven designs. To adopt a term from Donaldson and Lipsey (2006), method-driven approaches are rooted in the positivist paradigm. These designs tend to include procedures seeking to manipulate a condition to establish the causal claim; for example, randomization to disburse the effects of unwanted variables evenly among groups (Van Melle et al., 2017), or counterfactuals (e.g., control groups) to ascertain what would happen in the absence of the cause or the intervention (White & Phillips, 2012). This inclination towards method-driven designs inspired a powerful movement with quite a bit of momentum (Scriven, 2008) that include “evidence-based decision making” (Donaldson, Christie, & Mark, 2015), “results-based management,” and “What Works Clearinghouse” standards (What Works Clearinghouse, 2002), all of which promote experimental designs (e.g., randomized controlled trials) as the gold-standard in evaluation practice (NIJ, 2010).

While method-driven approaches can support the quest for direct and verifiable causality, it can be a poor fit, particularly in the social sciences (Astbury & Leeuw, 2010; Iverson, 2003). Recent debates have emphasized that these types of designs are difficult to do well in evaluation (Miller, 2010), and are seldom practical, feasible, or appropriate for the evaluation circumstances (Cook, Scriven, Coryn, & Evergreen, 2010; Gates & Dyson, 2017; House, 2001; White, 2010), as context is largely ignored (Scriven, 2008). Be it a large public-sector intervention or an emergent social innovation (Sridharan & Nakaima, 2012), method-driven approaches fail to consider the complexity of the

intervention and the multiplicity of factors that may consequently contribute to the observed impact (e.g., numerous types of activities, feedback loops, emerging outcomes, multi-year time frames) (Koleros & Mayne, 2019). Evaluation is a political and social activity, not merely a technical one (Miller, 2010; Schwandt, 2007). As a result, these approaches are considered by some to be inappropriate and problematic for dealing with multi-faceted programs in complex social settings (Cook et al., 2010; House, 2001; Gates & Dyson, 2017; White, 2010).

Aside from the layers of complexity engendered by the evaluand and practice setting (N. L. Smith, 2015), there is yet another more problematic oversight in method-driven approaches—the people. Michael Patton has famously declared, evaluation is all about the “people, people, people, people, people” (Patton, 2004, p. 291). Method-driven approaches reduce the complexity of people to variables. People interact and react to their environment and are ceaselessly changing and evolving as they learn and adapt. Evaluation is a social practice, yet method-driven approaches attempt to isolate and manipulate human social action the same way that natural objects are studied, and therefore neglect to consider the dynamic nature of people.

Another particularly challenging issue that undermines all investigative efforts of methods-driven designs - that is the black box. As seen in Figure 1, the box is the “space between actual input and expected outcome” (Pederson & Rieper, 2008; as cited in Leeuw, 2012, p. 349). Specifically, this problem refers to the blind acceptance of the observed outcomes without an adequate understanding of how and why they occurred (Astbury & Leeuw, 2010; Befani & Mayne, 2014; Van Melle et al., 2017).

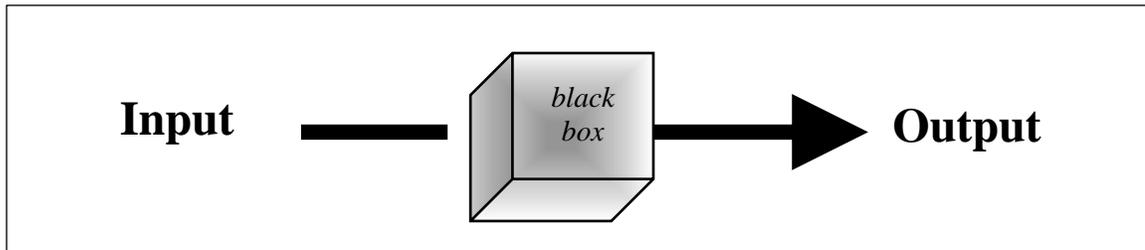


Figure 1. Diagram of the “Black-Box,” Illustrating a Lack of Clarity Between Inputs and Output.

As Astbury and Leeuw (2010) explain, the black box is “viewing social programs primarily in terms of effects, with little attention paid to how these effects are produced” (p. 364). In such a scenario, the observed outcomes could be coincidental or false positives, both of which can lead to misinterpretations and inaccurate conclusions. Opening the black box to understand the *how* and *why*, allows us to describe the impact pathway by mapping theory to program activities and expected outcomes while articulating assumptions and underlying linkages (Van Melle et al., 2017). We begin to understand how our programs are working and how our impacts are achieved. The black box is far too important to ignore because it can unveil the mechanisms by which input leads to output beyond merely “did it work?” Thus, it is imperative that as evaluation architects (Dahler-Larsen, 2015), we ask and expect more of our evaluation designs.

To address these challenges and limitations, principally the black box and overlooking context, evaluators turned to alternative approaches such as theory-based evaluation (TBE), a genre of white-box (or “clear box”) evaluations (Astbury & Leeuw, 2010). One TBE in particular stands out, namely Contribution Analysis (CA), in that it allows us to examine the “how” by using a process of “logical argumentation” (Craig,

2013; Wimbush et al., 2012). It seeks to demystify the links between inputs and outcomes by linking evidence along each step of inference. According to John Mayne, its chief architect, CA aims to “reduce uncertainty about the contribution an intervention is making to observed results through an increased understanding of why results did or did not occur and the roles played by the intervention and the other influencing factors” (Mayne, 2012, p. 271). What makes CA stand out is the added layer of depth not only in its deliberate reflection on the mechanisms through which change is expected to come about but also in its consideration of assumptions and risks that influence the likelihood of said change (Contribution Analysis is elaborated in Chapter II, page 31).

### **Statement of the Problem**

With the continuous demand for rigorous nonexperimental designs and methods to assess causal relationships in social complex settings (Leeuw & Vaessen, 2009; Network of Networks for Impact Evaluation [NONIE], 2008; Rogers, 2009; Stern et al., 2012; White & Phillips, 2012), Contribution Analysis, a theory-based evaluation approach is well-timed. CA presents a persuasive approach to investigate causal relationships (Stame, 2004) while allowing for the unpacking of the black box and being mindful of context. Consequently, CA has gained popularity as funding agencies compel for alternative approaches and transparency (Kane, Levine, Orians, & Reinelt, 2017).

While Contribution Analysis (CA) is promising and currently emerging as an innovative approach in evaluation (Schwandt, 2015), the first decade since Mayne’s initial publication in 2001 is embodied by a few practice-based examinations, with only a handful of examples available mainly in the form of white papers presented at evaluation

conferences (Wimbush & Beeston, 2010). Despite an increase in use in the second decade since this initial publication, there remains little empirical research on the practice of CA in the academic literature. Notwithstanding a few notable contributions such as the special issue dedicated to CA in the journal *Evaluation* (Stern, 2012), CA as an evaluation approach remains relatively underdeveloped and undefined with fundamental traits such as epistemology and ontology still unaddressed (Budhwani & McDavid, 2017; Dybdal et al., 2010).

While Mayne has argued “there is no prescribed way of evaluating a specific intervention,” (as cited in Nkwake, 2015, foreword), he also claims that CA provides a standard analytical framework for evaluation that can be adapted as required (Budhwani & McDavid, 2017; Mayne, 2012). Since its original implementation in performance measurement evaluation, CA has been applied across various settings, from local evaluations to country-wide evaluations (Montague, Young, & Montague, 2003), and has since gathered many proponents across the globe (e.g., Canada, Denmark, the European Union, Fiji, Rwanda, Scotland, and the United States). However, upon closer inspection, each of the cases, as mentioned earlier reflects a portfolio of wide-ranging adaptations of Contribution Analysis (Budhwani & McDavid, 2017).

Notably, each application reflects substantial variation, leading to different understandings and significant adaptations of CA across contexts (Budhwani & McDavid, 2017). This striking variation is reflected in the practice literature (Biggs, Farrell, Lawrence, & Johnson, 2014; Budhwani & McDavid, 2017; Delahais & Toulemonde, 2012; Dybdal et al., 2010; Government of Canada, 2015; Kotvojs &

Shrimpton, 2007; Mentzer, Czerniak, & Struble, 2014; Noltze, Gaisbauer, Schwedersky, & Krapp, 2014; Patton, 2012; Sridharan & Nakaima, 2012; Srivastava & Enriquez, 2013; Wimbush et al., 2012). With so many variations and adaptations, it becomes particularly challenging to identify its unique methodological elements and discern CA from other approaches in practice and (Budhwani & McDavid, 2017; Lemire, 2010), and begs the question, “how would we recognize a methodologically sound CA if it was right in front of us?” (Lemire, 2010, p. 16).

As such, notwithstanding the theoretical pieces (though limited) and the few but diverse CA practice-based cases, there remains a gap in evaluation research. This is problematic and ultimately undermines a core commandment of the field to consider the relationship between theory and practice. As Robin Miller (2010) has stated,

sorting through theories and determining their ultimate feasibility and merit would benefit by close empirical examination of how evaluation theories can be and are applied in practice, whether they consistently and reliably lead to successful evaluation, and under what circumstances ‘good’ evaluations are likely to emerge. (p. 391)

Therefore, an examination of theory and practice is not only warranted but largely overdue.

### **Purpose of the Study and Research Questions**

As a result of the recent boom in Contribution Analysis (CA) evaluations, there is a growing interest within the evaluation community to gain a profound understanding of its use, implications, and consequences across different contexts. Although case examples provide some understanding of this approach in practice, there remains limited empirical

evidence to support a comprehensive understanding. This research responded to explicit requests to probe deeper into CA (e.g., Budhwani & McDavid, 2017; Delahais & Toulemonde, 2012; Dybdal et al., 2010; Lemire, 2010) as well as a more general call for a systematic examination of evaluation cases in order to generate “practical knowledge” about different approaches to evaluation (Schwandt, 2015).

This dissertation sought to examine the practice of the Contribution Analysis (CA) approach to evaluation to inform theory development and improve practice. Specifically, the study researched how CA in practice compares with the conceptualization offered by John Mayne and identified elements of effective practice by characterizing adaptations, adjustments, and innovations, and conditions under which practice may differ (e.g., evaluand, contexts, or evaluators) (Mark, 2008; N. L. Smith, 1993). Though CA has been present in the academic literature for almost two decades, only recently has practice-based accounts surfaced with similar prominence. As such, it is an opportune moment for a thorough and comprehensive investigation of CA in practice. This study addressed the following research questions:

1. What is the current theory of the Contribution Analysis approach to evaluation?
  - a) What are the fundamental tenets that guide CA practice?
2. What does CA look like in practice?
  - a) How do practitioners implement the approach?
  - b) What adaptations, adjustments, or reinforcements are made to the 6-step process?

3. What conditions and contextual factors (e.g., of the evaluation and program) challenge and facilitate the implementation of CA in the field?

This dissertation research explored the theoretical translation and practice of contribution analysis; adaptations and innovations made to CA's distinct 6-step process; and the conditions, influences, and factors that may facilitate or impede CA practice. Accordingly, this dissertation marshaled a sequential research design of two qualitative studies, which together founded a systematic inquiry into the theoretical translation and practice of the Contribution Analysis evaluation approach. First, Phase I launched a generative and emergent inquiry into CA through a systematic review of the literature. This employ conceived a mapping of the conceptual developments of the theory and to identify critical tenets of CA. Next, Phase II drew on the case study methodology (Stake, 1995, 2006; Yin, 2009) to investigate empirical applications of CA mainly through in-depth document analysis of peer-reviewed publications and semi-structured interviews with each case's respective evaluation practitioner. Using multiple sources per study unit allowed to member-check interpretations and ensured that context-sensitive findings were produced. This study provided insight into the conditions, influences, and contextual factors that may affect how CA is practiced.

The two-part research design was complementary to the overall research objectives. The systematic review assisted in the development of a conceptual framework through the identification of key tenets of the Contribution Analysis approach and generated insight into general trends of practice. On the other hand, the multiple case study provided the flexibility needed in exploratory research to ensure the opportunities

to probe deeper into themes identified in the systematic review while also allowing for queries into emerging themes and patterns that materialize.

### **Significance of the Study**

“All evaluation practitioners are nascent evaluation theorists” (Shadish, Cook, & Leviton, 1991, p. 35). As practitioners pursue their craft, they will reflect on their practice, weigh advantages and disadvantages, and learn from their experiences. Nevertheless, making informed choices founded on empirical evidence cannot happen, not because it is not important to use empirical cases to base decisions, but because there is limited research to guide one’s choice (Fitzpatrick et al., 2011). As such, this study sought to address this fundamental oversight and pursue a richer understanding of the contexts of practice and to probe deeper into the translation of theory to practice through the use of empirical evidence to further develop and improve on the current understandings of CA theory and practice.

For practitioners, understanding the extent to which operational practices of contribution analysis reflect its values provides insight into where and how they might improve practice. This research was inspired by Miller and Campbell’s (2006) examination of 47 empowerment evaluation applications, which revealed that just a few evaluators had reliably exhibited all the ten principles of the approach. In turn, Miller and Campbell (2006) were able to make empirically-based recommendations to the theorists to consider a re-assessment of the ten principles as well as elaborate on instructions for practitioners to accurately identify when or how they might be better suited per the project’s size, scope, and aim.

In the same manner, the findings from this study contribute to the empirical body on evaluation practice, as well as identify implications and recommendations for further refinement of CA as an emergent evaluation approach. Moreover, findings from this study inform the development of CA contingency theories (Shadish et al., 1991), which identify conditions under which practices are effective (Mark, 2008; Miller, 2010). Approaches incorporating contingency theories are considered to be the strongest types of evaluation approaches as they are buttressed by empirical knowledge of the practice (Shadish et al., 1991; N. L. Smith, 1993; Vo, 2013). Therefore, findings produced through this research are of value to strengthen the CA's theoretical foundation.

### **Organization of the Document**

This dissertation is organized in five chapters. Chapter I (current chapter) is an introduction to the study, problem statement, purpose of the study and research questions, and significant of the study. Chapter II focuses on the contribution analysis approach to evaluation. it reviews the interest I causality in social science, how CA mitigates the threats posed by traditional method-oriented approaches, and suggests CA as a framework that facilitates the validation process. Chapter III details frameworks pertinent to this study's design and overview of the methods employed to address these questions. Chapter IV presents the findings of the study, and finally, Chapter V discusses the implications of findings as it pertains to the evaluation discipline, and notes on future directions for research.

## **CHAPTER II**

### **REVIEW OF THE LITERATURE**

In Chapter I, I outlined the rationale and purpose of the current study and highlighted the variety in the practice of the Contribution Analysis (CA) approach to evaluation and the need to better understand how the conceptual model translates into practice. I argued for the need to empirically examine the factors and conditions that may impact practice. The chapter serves to provide a comprehensive review of the relevant literature, to elaborate on Contribution Analysis as an evaluation approach, what it purports to be, and how it purports to be applied in situ. I first begin with a brief on the role of theory in the evaluation and paradigms of inquiries. Next, I situate CA amongst other approaches to categorically demonstrate its aim to address causality. Subsequently, a review of the existing literature on the CA approach to evaluation, its technical aspects, how CA helps build validity into evaluation findings. In closing, I present the current need for research on evaluation and argue for an in-depth examination of CA practice.

#### **Chapter Overview**

Before detailing Contribution Analysis and its practice in evaluation, I discuss the tensions that catalyzed the shift towards theory-based evaluation (TBE) approaches such as Contribution Analysis (CA). First, the role of evaluation theory in practice and the key concepts concerning research paradigms is considered to contextualize the discussions offered in this chapter. Next, I discuss the pronounced interest in causality, the classes of

impact evaluation design frequently employed to meet this end and their limitations in social science inquiries. Particularly, I discuss the black box phenomenon, and how theory-based evaluation (TBE) designs help mitigate the fundamental complications couched within it. The text up to this point illustrates the contextual backdrop upon which Contribution Analysis emerged and situates CA in a schema of research designs.

The next section focuses on Contribution Analysis and situates it as a theory-based evaluation (TBE) approach. This is followed by a discussion of the causal questions CA purports to attend and the significance of the theory-of-change in this endeavor, and the elaboration of the CA framework (e.g., signature 6-step process). and the various degrees of contribution claims (as well as corresponding prerequisites) that can be asserted through CA. This segment concludes by presenting the varying magnitudes of contribution claims invoked by CA, and the conditions under which they are possible.

Finally, this chapter comes to a close with a discussion centered on validation, and how CA is inherently a validation framework as it calls for and assists in the systematic examination of the chain of inferences. In that respect, I draw on the corollary from Michael Kane's (2006) interpretive argument to substantiate this claim. Finally, I remind the reader that there is a general ambiguity and vagueness described by practitioners about CA, which suggests more guidance for CA-guided practice is needed, and close by illuminating the urgent need to dissect evaluation practice, particularly in the case of Contribution Analysis.

## **An Introduction to Evaluation Theory**

Program evaluation is defined in numerous ways throughout the academic literature. A synthesis of these various explanations and definitions reveal the act of “valuing” (Scriven, 1967, 1980, 1991, 2003) (e.g., assigning merit, worth, and significance) as the pillar of evaluation (Schwandt, 2015), primarily via a systematic inquiry of an evaluand (e.g., program, performance, process, project, intervention, policy) to produce judgments for program decision-making and knowledge production (e.g., Cousins & Chouinard, 2012; Fitzpatrick et al., 2011). The ultimate desired outcome is social betterment (e.g., see Johnson et al., 2009; Mark et al., 2000).

Although evaluation is a young discipline (approximately 80-year in existence), the field has undergone two significant expansions. The first begun with Ralph Tyler’s “Eight-Year Study,” which is widely accepted to have forged a path for the beginnings of today’s profession and discipline (Alkin, 2013; Donaldson & Lipsey, 2006; Madaus & Stufflebeam, 2000). Unlike traditional research taking place in 1940, the Eight-Year Study represented an example of “implementation research” or what today we call “formative evaluation” (Kridel, 2010). Tyler clearly understood the need for continuous evaluation within the process of creating instruction designed to produce specific outcomes. Furthermore, the study was a model of demographic policy evaluation, where stakeholders were given flexibility in planning for and implementing research so that it would be in line with the needs of the wider organization (Kridel, 2010).

The ensuing period (1960-1970) gave rise to the first major boom in evaluation (Donaldson & Lipsey, 2006) as the federal government funded a plethora of large-scale

social programs in education, income maintenance, housing, health, and criminal justice (Shadish et al., 1991), prompting the widespread institutionalization of the discipline. For the remainder of the 20th century, methods reflected strong positivist roots as design sought to make generalized causal inferences in order to establish the program's net impacts (Donaldson & Lipsey, 2006)

The turn of the 21st century marked the second major boom in evaluation (Donaldson, 2003) and a significant shift in the discipline's priorities as we saw: (a) an emphasis on being more inclusive in the objects of the evaluation (e.g., technology, personnel, proposals, performance), and (b) an emergence of new theories of practice, evaluation methods, and tools, which naturally prompted (c) new developments of general organizing frameworks (Donaldson & Lipsey, 2006) to make sense of these novel theories. For example, Shadish et al.'s (1991) five core necessities of good social program evaluation theory, Stufflebeam's (2001) classification of 22 evaluation models, and the widely popular tree-metaphor (Alkin & Christie, 2004; Mertens & Wilson, 2012) classifying evaluation approaches on the use, methods, values, or social justice branch to name a few.

The various classifications and taxonomies that have been developed over the years underscore the fundamental differences in the guiding values and principles practitioners call upon in conducting sound evaluation, and what is given priority in practice. As many discourses have come and gone, the role of theory in evaluation remains a contentious matter sparking vibrant debates (Donaldson & Lipsey, 2006);

however, most agree its purpose can be boiled down into one of two types, prescriptive models, or descriptive models (Alkin, 2013):

1. Prescriptive models declare what should be (Donaldson & Lipsey, 2006); and use “sets of rules, prescriptions, and prohibitions and guiding frameworks that specify what good or proper evaluation is and how evaluation should be done; such models serve as exemplars” (Alkin, 2013, p. 4-5); and,
2. Descriptive models “characterize what is” (Donaldson & Lipsey, 2006, p. 59) and “offer a set of statements and generalizations which describe, predict, or explain evaluation activities” (Alkin, 2013, p. 5).

In fact, it may be more accurate to perceive these along a continuum rather than separate categories. A descriptive theory is one that has been empirically vetted (Alkin & Ellett, 1985). Thus, a theory that is prescriptive in nature can become descriptive once it has amassed enough evidence to illuminate what evaluation looks like, per different conditions, and types of consequences to result from various approaches (Alkin, 2013; N. L. Smith, 1993).

In support of the research agenda, this explanation is useful to illustrate why research on evaluation practice across a sample of cases is a significant way to contribute to empirical knowledge on the utility of prescriptive evaluation approaches. I will revisit the role of theory at a later point in this chapter (“Linking Evaluation Theory and Practice” on page 51).

## **Paradigms and Associated Assumptions**

Research studies are typically conducted within a narrow set of paradigmatic assumptions that have implications on the type of methods, processes, and conclusions they are expected to produce (Nkwake, 2013). In the quest to establish causation, it is, therefore, imperative to understand the underlying paradigms and make explicit their assumptions to better understand the nature, role, justifications, and nuances of research design choices. The current section reviews the concepts of ontology and epistemology and provides the background for the ensuing discussion on approaches to causation (see “Design Approaches to Causation” on page 22).

While ontology refers to the nature of reality (Nkwake, 2013), epistemology is the study of knowledge, and defines the assumptions underlying the nature of knowledge (Fitzpatrick et al., 2011), and what constitutes knowledge (Mathison, 2005). The ontological choice can be reduced to a dichotomy: an objective reality, in that there is one “truth” independent of the things being studied, which in practice implies that each participant carries the same meaning and understanding of the phenomena (Newman, 1998, as cited in Nkwake, 2013). Alternatively, is reality more fluid and elusive? Thereby subjective as epitomized by multiple realities (Nkwake, 2013). Under the subjective reality, one could never assume that the observed is being interpreted in the same way by all participants. One either “has to accept or reject the notion that there is a single, objective, real-world” (Nkwake, 2013, p. 96, citing Robson, 1993).

However, these underlying notions about reality and knowledge do not occur in a vacuum or isolation. Epistemology and ontology are intertwined: “claims that exist in the

world (reality) imply claims about how that existing reality can be known” (Nkwake, 2013, p. 96, citing Scott & User, 1996). Essentially, epistemology asserts how we “know” reality and defines the kind of knowledge that we allow in our reality, and thus affects how it can be studied (Nkwake, 2013). To further delve into the interlacing of how reality is constructed and how knowledge is defined, I use Guba and Lincoln’s (1989) work to discuss ontological properties of four primary epistemological theories: positivism, realism, critical theory, and constructivism.

Positivism holds at its core the ontological assumption of objective reality (Nkwake, 2013), which is “stable, observable, and measurable” (Merriam, 1998, p. 4). This philosophical stance has rigid views of the world and data collection (Fitzpatrick et al., 2011). Its epistemological assumptions expose an aim to “explain, predict, and ultimately control” that reality (Nkwake, 2013, p. 100) primarily through experimental methods. Like positivism, realism dons an objective reality; however, it assumes it is impossible to capture an untainted version of reality (Healy & Perry, 2000; Nkwake, 2013; Shah & Corley, 2006; Campbell & Wasco, 2000). Realism posits humans as being inherently biased and incapable of separating themselves from their predispositions (e.g., biases, beliefs) to observe reality with an objective lens.

On the other side of the ontological dichotomy are critical theory and constructivism, which assume and perceive reality as primarily subjective and assert a constructed account of reality. Critical theory suggests that there is no objective reality but is instead interpreted through one’s values (e.g., social, political, cultural, economic, ethnic, and gender) (Campbell & Wasco, 2000; Healy & Perry, 2000; Nkwake, 2013). In

other words, knowledge is filtered through various lenses and, therefore, cannot be a representation of “pure fact” (Nkwake, 2013, p. 97). Constructivism edges further than critical theory and assumes that reality is a social construct (Nkwake, 2013). There are many truths, and people’s realities are not merely as they see through their lenses but are shaped by social factors (Nkwake, 2013).

In short, while ontology can be characterized on a binary scale, epistemologies can be positioned along an objective-subjective continuum (Healy & Perry, 2000), positivism lies on the objective end of the continuum, critical theory and constructivism lie on the subjective end of the continuum, and realism seems to lie in between the two (Nkwake, 2013).

### **Causality in Evaluation**

As with all endeavors that seek to further knowledge and understanding, the general interest in causality in the social sciences is no exception. This fact is keenly resounded by Shadish’s (1998) AEA presidential address, which identifies ‘the role of causal inference in evaluation’ is reflected in his shortlist of the ten concepts all evaluators must know. Two decades later, this question is as relevant today as it was then. To a large degree, evaluations tend to focus on results as a measure of success, which Shadish (1998) refers to as “outcome evaluations.” Moreover, despite tremendous advances in the field (e.g., the development of new methodologies, improved access to more sophisticated tools for intricate research designs), properly examining causality remains a challenging problem.

Indeed, the class of evaluations reflected in this dilemma are outcome and impact evaluations which focus on program effects, and hold “outcomes and impacts” (Nkwake, 2013, p. 167), as the object of the evaluation. Impacts are changes to people and their lives and can be viewed as occurring at the individual and personal level (Westhorp, 2014). Outcomes encompass impacts and include other kinds of changes that are above and beyond the individual/personal level. Outcomes can be institutional, community, or organizations such as workers union, governments, and so on (Westhorp, 2014). However, at the heart of impact and outcome evaluations lies the “requirement to link causes and effects and to explain ‘how’ and ‘why’” (Stern et al., 2012).

Establishing causation in social science follows one of two paths, one carrying the unmistakable characteristic of positivist heritage, or one illuminating an alternative understanding of causation (Maxwell, 2012). The positivist interpretation of causation theory adheres closely to an objective reality, where the observer/researcher carries a value-free inquiry (Khakee, 2003). Historically, this stance has been heavily influenced by the scientific method (Maxwell, 2012) and follows a reductionist principle in that “causation is no more than regularity” (Maxwell, 2012, p. 657). In other words, causation is determined by the frequency of occurrence of an outcome.

On the other hand, the alternative understanding views causation from a different angle; causation is perceived to be “generative,” “process,” or “realistic” (Maxwell, 2012, p. 656). This view emerged in response to the positivist theory, specifically in consideration of its inability to consider the complexity of social circumstances. Under the alternative theory to causation, people are presumed to be the actors and agents that

influence change. As Pawson and Tilley (1997) have explained, this take on causality denotes that “it is not programs that make things change, but rather it is people, embedded in their context who, when exposed to programs, do something to activate given mechanisms, and change” (p. 32).

While positivist designs seek to generalize, the impetus undergirding the alternative theory of causation is to understand. For example, consider the following scenario of 100 individuals of similar ability take a test. The method-driven path would seek to infer generalizations from the 100 test results, while the alternative understanding would seek to understand why there were 100 different results. Although there is little agreement across the field on how to best establish causation (Schwandt, 2001), there is a consensus that “qualitative analysis can yield causal explanations rigorously and credibly” (Patton, 2014, p. 1310). While the myriad qualitative approaches have their strengths and weaknesses, my research focused on Contribution Analysis, an approach identified capable of explaining the “why” and “how” (e.g., making causal inferences) by investigating program effects (Stern et al., 2012).

### **Design Approaches to Causation**

In order to situate CA as an evaluation approach intended to establish causal inferences, the following section discusses a taxonomy of design approaches to causality, their associated ontology, and epistemology properties, and how their paradigmatic assumptions derive from and support the nature of causal inferences to create a detailed comparison with other designs seeking to fulfill a similar objective. Stern et al. (2012) investigated design approaches commonly used to support cause-and-effect

investigations, specifically in practicing impact evaluations and identified the following five types: experimental, statistical, theory-based (which includes CA), case-based, and participatory. The taxonomy is summarized in Table 1, along with underlying paradigms, corresponding variants, and basis for causal inference in establishing causality.

Experimental designs and statistical designs reflect positivist epistemological assumptions and are located on the objective end of the spectrum. Experimental designs operate under the criterion of falsifiability, which implies that a claim “must be refutable or falsifiable, and if not, they are merely dogmatic stances” (Mathison, 2005). Experimental designs rely on counterfactual frameworks, illustrating the difference between two identical cases, to make causal claims (Stern et al., 2012).

Randomization is used to disburse the effects of unwanted “variables” evenly among groups (Van Melle et al., 2017). Such an example would use either a counterfactual to attest to what would happen in the absence of the intervention (White & Phillips, 2012), or comparing treatment effects of the control group to the experimental group. Outcomes observed in the presence of the intervention are therefore deemed as evidence of the program’s effects. In other words, the comparison of the outcomes between the two scenarios, without an intervention (established by the counterfactual or control group), compared to the intervention, is used to ascertain that the outcomes observed in the applied setting directly are attributed to the program (Patton, 2012).

Table 1

## Design Approaches, Variants, and Basis for Causal Inference

	<b>Design Approaches</b>	<b>Specific Variants</b>	<b>Basis for Causal Inference</b>	<b>Causality</b>
<b>Positivism</b>	Experimental	RCTs Quasi-experiments, Natural experiments	Counterfactuals: the difference between two otherwise identical cases – the manipulated and the controlled; the co-presence of cause and effects	Attribution
	Statistical	Statistical Modelling Longitudinal Studies Econometrics	Correlation between cause and effect or between variables, the influence of (usually) isolatable multiple causes on a single effect	
<b>Realism</b>	<b>‘Theory-based’</b>	<i>Causal process designs:</i> Theory-of-change, Process tracing, Contribution Analysis, impact pathways  <i>Causal mechanism designs:</i> Realist evaluation, Congruence analysis	Control for ‘confounders’ Identification/confirmation of causal processes or ‘chains’  Supporting factors and mechanisms at work in the context	Contribution
	‘Case-based’ approaches	<i>Interpretative:</i> Naturalistic, Grounded theory, Ethnography  <i>Structured:</i> Configurations, QCA, Within-Case- Analysis, Simulations, and network analysis	Comparison across and within cases of combinations of causal factors  Analytic generalization based on theory	
	Participatory	<i>Normative designs:</i> Participatory or democratic evaluation, Empowerment evaluation  Agency designs: Learning by doing, Policy dialogue, Collaborative Action Research	Validation by participants that their actions and experienced effects are ‘caused’ by program  Adoption, customization, and commitment to a goal	

A more liberal declaration of this opinion is that nonexperimental quantitative designs can also be used to make causal claims (Maxwell, 2004; Shadish, Cook, & Campbell, 2002). These refer to statistical approaches such as correlational studies and structural equation modeling that extrapolate causality by examining the degree of association between the independent (cause) and dependent (effect). Statistical approaches rely on regularity frameworks, linking the frequency of association between cause and effect, as the inference basis for causation claim (Stern et al., 2012). Statistical approaches are not necessarily inept or weak methodologies in social sciences as they are often used in other approaches (such as CA). Experimental and statistical approaches are best when dealing with independent causal factors (Stern et al., 2012). As such, standalone use of statistical approaches is not sufficient, as they do not consider contextualization (Stern et al., 2012). Both experimental and statistical designs are quantitatively driven approaches.

On the other hand, theory-based approaches, case-based approaches, and participatory approaches reflect subjective ontological assumptions: most “theory” and “case” oriented approaches are fundamentally rooted in the realist understanding of the world (Stern et al., 2012), holding the belief that a similar mechanism does not guarantee a common outcome but is instead heavily dependent on context; while constructivist assumptions undergird participatory approaches.

Theory-based evaluation approaches (TBEs) hinge on a program theory (also known as intervention theory) (Rey, Brousselle, & Dedobbeleer, 2016) and consist of two families: causal process designs, and causal mechanisms designs. Causal process designs

aim to “assess the causal processes underlying a program” (Nkwake, 2013, p. 169). CA falls into causal process designs in that it probes the causal mechanism mediating treatment and outcomes (Nkwake, 2013). TBEs are based on well-defined program theory, in that there is a clear definition between the program activities and the cause (or contribution to) outcomes and impacts. TBEs look explicitly at processes and mechanisms for change and includes any evaluation approaches that examine “the assumptions underlying the evaluated intervention’s causal chain from inputs to outcomes and impact” (White, 2009, p. 3). Using this explicit program theory, TBEs endeavor to develop, test, and refine the theory of change (ToC) (Budhwani & McDavid, 2017). Other strains of theory-based evaluations include Realist Evaluation (Pawson, 2006; Pawson & Sridharan, 2009; Pawson & Tilley, 1997) and Developmental Evaluation (Patton, 2010).

Case-based approaches emerged around the turn of the century (Stern et al., 2012). As the name may hint, this set of design approaches means to shift the focus from the narrow fixation on variables to the entire case. Cases may be “policy interventions, institutions, individuals, events, or even countries during a particular period” (Stern et al., 2012, p. 27). Stern et al. (2012) identified two classes of case-based approaches: interpretative and structured.

Despite eschewing causal inquiries, interpretative approaches (e.g., naturalist, grounded, theory, ethnography) contribute in such ways by providing rich understandings of contexts, assisting in “defin[ing] construct validity in terms that make sense to stakeholders on the ground” (Stern et al., 2012, p. 28), and by giving voice to program

beneficiaries both during the formulation of evaluation questions and in the interpretation of findings (Stern et al., 2012). For example, grounded theory studies contribute to an in-depth understanding of context as the researcher makes meaning of the extracted data, rather than fitting the data to a pre-existing framework (Charmaz, 2006). Ethnographic studies strengthen the construct validity as the researcher adopts the perspective of the people being observed and seeks to understand and make meaning of the observations from their perspective (Bryman, 2012).

A newer evolution of case-based approaches are structured approaches (e.g., configurations, Qualitative Comparative Analysis, Within-Case Analysis, network analysis). Structured case-based approaches juxtapose interpretative designs in that they are specifically interested in the causal analysis (Byrne & Ragin, 2009; George & Bennett, 2005; Stern et al., 2012). In fact, their primary interest lies in investigating causality, and in causal analysis to generalize beyond the single case as appropriate (Stern et al., 2012). Case-based studies may test a theory; however, the role of theory is less pronounced in designs that aim to investigate causality. Instead, they more focused on the conditions that are necessary and sufficient to base comparisons of “configurations” of cases and attributes (Byrne & Ragin, 2009; Stern et al., 2012) in Qualitative Comparative Analysis (QCA).

The last class of designs identified by Stern et al. (2012) is participatory approaches. A distinction should be emphasized between the more frequent use of the term “participatory evaluation” and participatory approaches as a design for impact evaluation. The general reference to participatory evaluation refers to a class of

approaches that actively involve program participants and stakeholders in evaluation activities (Cousins & Chouinard, 2012; Gates & Dyson, 2017) to empower the stakeholders, building evaluating capacity, or increasing organizational learning and data-based decision making (Fitzpatrick et al., 2011).

In the context of impact evaluation, however, the term refers to participatory approaches that are used to warrant causal claims (e.g., Participatory Rapid Assessment, Participatory Action Research, and Most Significant Change). These designs emphasize the role of stakeholders as active agents instead of passive recipients (Stern et al., 2012). More specifically, program participants provide the evidence needed to support claims that the program “caused” the behavior changes, as they contribute evidence representative of their experiences and actions, often in the form of participants’ stories (Gates & Dyson, 2017).

Though participatory designs can facilitate the uncovering of unintended consequences, and reveal impacts valued by different stakeholders (Gates & Dyson, 2017), stakeholder buy-in seems to be at a tension. While some stakeholders support participatory approaches in the quest for causality, some authors (Coryn, Schröter, & Hanssen, 2009; Davidson, 2005) have cautioned a different scenario. Instead, some stakeholders do not believe that participatory approaches are “robust enough to make causal claims or that these approaches offer a sufficient degree of certainty” (Gates & Dyson, 2017, p. 40) as this design class lacks the consideration (and elimination) of alternative explanations as well as an independent confirmation or triangulation of participant claims (Leeuw & Vaessen, 2009; NONIE, 2008; Scriven, 2005). Furthermore,

merely inquiring as to whether the intervention produced specific impacts is not enough as stakeholders can potentially manipulate the data to serve specific interests such as the continuation of the intervention (Gates & Dyson, 2017; NONIE, 2008; Rogers, 2009).

### **Limitations of Method-Driven Approaches**

The debate about the ability for qualitative research to establish causality has been questioned for quite some time while the much narrower view that an experiment is necessary to establish a causal link has been traditionally upheld as the gold standard in the research community. The appeal of positivist inquiry designs lies in the production of “defensible causal interpretations” (Campbell & Stanley, 1963) and is facilitated by traditionalist elements such as counterfactuals (e.g., control group, or comparison group). As such, evaluations are commissioned to investigate the relationship between observed outcomes as a direct result of a project, program, or policy. In the end, claims of causality linking the observed outcomes as direct contributions of the intervention are produced by the evaluation.

Because of the positivist stance, objective-oriented approaches have served as the primary framework for cause-and-effect inquiries. However, though well-suited in many cases, traditional method-driven approaches are filled with shortcomings in practice. Experimental and statistical designs tend to be detached from the in situ social complexities. As Weiss (1993) has suggested, “an evaluator who limits his study to program effects conveys the message that other elements in the situation are either unimportant or fixed and unchangeable” (as cited in Nkwake, 2013, pp. 167–168). Method-driven designs can struggle with multiple causalities and inept in capturing

interactions amongst variables or represent irregular, nonlinear paths of influences (Stern et al., 2012). As many intervention designs involve multiple components using multiple pathways of causation to achieve results (Budhwani & McDavid, 2017), experimental designs are not the most appropriate in these ‘complex settings’ (Wimbush et al., 2012).

### **The Black Box**

Another dimension that adds to the complexity of establishing causation in the social sciences, let alone under the positivist paradigm, is the fundamental black box problem. The black box symbolizes the expectation of inputs leading to outcomes without understanding how they function. The ‘black box’ problem has been around since the inception of evaluation with the ‘War on Poverty’ programs in 1964 (Stame, 2004). Preoccupied with the larger social problems at hand, program designers have consistently overlooked “what is expected to happen, the how and why, [and] when input is put in place” (Stame, 2004, p. 58), that is the program theory. Similarly, evaluators followed suit by focusing on “measuring outputs, while attributing the observed difference to the input” (Stame, 2004, p. 58) without much attention to processes and mechanisms of intermediate causal links. To cope with the black box problem, evaluators have developed elegant methods for “measuring the distance between objectives and results” (Stame, 2004, p. 58) instead of turning their attention to the unattended links.

In essence, the sophistication of methods seemed to have emboldened the drawing of the dotted line backward, starting with output and ending with input, without questioning if the series of links between these two points were ever connected. While evaluating the attainment of the program goals is fundamental, it is more important to

understand that “activity A will attain objective B because it is able to influence process C, which affects the occurrence of this objective” (Suchman, 1967, p. 177). In response to the black box problem, theory-based evaluation (TBE) approaches have been proposed to unpack the black box activities. They seek to offer transparency in the dense fog between the obscure links between program inputs and outcomes by focusing on processes and mechanisms that link them and how they contribute to the program's impact.

### **Contribution Analysis**

The following segment renders an overview of Contribution Analysis. First, I present CA's theoretical orientation, categorically as a Theory-Based Evaluation (TBE) against a backdrop of realism (Pawson & Tilley, 1997). Next, I explain the causal questions CA purports to attend, and how the 6-step framework assists in that endeavor. Finally, I elucidate the various spheres of influence, and the corresponding three types of contribution claims tenable through CA. The last two remaining sections within this chapter position CA as a validation framework, and argue for the need for research on evaluation, precisely on Contribution Analysis.

#### **Theoretical-Orientation: Theory-Based Evaluation**

Contribution Analysis is a theory-based evaluation (TBE) approach (Budhwani & McDavid, 2017; Mayne, 1999, 2008, 2012) in that it seeks to move beyond the superficial “what works?” to probing what happens inside the black box to inquire “what works, why, and how?” (Gates & Dyson, 2017, p. 37). This added layer of probity emphasizes the degree to which context is considered in TBE approaches. Sponging from

Pawson and Tilley's realist philosophy, the importance of context in TBEs is inherent in their endeavor to query the triggers and functions that catalyze change to identify what works and for whom by follow the theory of change (ToC) from beginning to end to develop, test, and refine the ToC (Budhwani & McDavid, 2017). As a form of "white box" evaluations, TBEs are considered effective approaches for evaluating complex social interventions (Connell & Kubisch, 1998; Weiss, 1997a, 1997b) in that it seeks to unpack the black box guiding by the principle that "nothing works everywhere or for everyone" (Westhorp, 2014, p. 4).

Generally speaking, however, TBEs have been criticized for concentrating on how the intervention is carried out (implementation/action theory) rather than how the intervention is supposed to work (program theory/results chain) (Dybdal et al., 2010; Weiss, 1997) or using weak developments how and why the intervention will make a difference (theory of change) (Astbury & Leeuw, 2010; Montague, 2019). CA addresses the limitations of traditional methods by shifting the focus from program implementation to an interwoven holistic understanding of the program theory and the necessary actions (e.g., implementation/action theory) for a comprehensive understanding of how the intended change is to occur, theory of change, thus making it an vital evaluation approach for evaluators (Budhwani & McDavid, 2017; Dybdal et al., 2010). CA can also address alternative explanations, which are not necessarily recognized in TBEs (Connolly, 2015; Dybdal et al., 2010).

Indeed, among the plethora of TBE approaches, Contribution Analysis (CA) stands out in that it seeks to build the ToC and subsequently test the ToC. As such, I posit

that CA offers a distinct advantage over other TBEs in that it inherently addresses the possibility of an underdeveloped ToC. In fact, “CA is based on the ToC being examined in detail” (Befani & Mayne, 2014, p. 21). Thus, an explicit and detailed account of the intervention’s ToC is a prerequisite for practice. Furthermore, CA requires that assumptions and risks underlying the ToC are made explicit, further clarifying operational understanding and theoretical underpinnings of an intervention. The dedication and necessity of a detailed ToC, its associated risks, and assumptions, ensure a deeper understanding of the underlying processes that charge input to the outcome.

Unlike method-driven approaches, TBEs do not require a counterfactual argument and do not seek to establish attribution (cause-and-effect). As in the case of CA, TBEs instead build a case of robust evidence in support of an of contribution claim of the degree to which the intervention can be said to have contributed to observed and documented changes, hence the name, Contribution Analysis. This approach is heuristic (Budhwani & McDavid, 2017) in that it allows for the tracing of pathways from inputs to outcomes. It seeks to demystify the links between inputs and outcomes by linking evidence along each step of inference. Briefly put, CA is comprised of a cyclical six-step process through which one can reasonably determine whether the outcomes observed are the result of the intervention’s activities (Mayne, 2012).

### **Setting the Evaluation Stage**

To better understand what CA can investigate, I will first discuss the interests driving stakeholders (e.g., funders, sponsors, managers, beneficiaries), which of those CA purports to attend, and how it facilitates that query via its 6-step process. Table 2 details

three questions driving solicitation of evaluations and the type of causal inference queried by each question (Mayne, 2011).

Table 2

Three Types of Evaluation Aims and Associated Causal Inferences

Evaluation Aims & Inquiries	Cause-and-Effect Inference
1. Were the results as intended?	Attribution
2. If the results were not as intended, why not?	Contribution
3. If the results were as intended, is it reasonable to conclude that the program/intervention played an important contributing factor in attaining those results?	Contribution

Question 1 is usually the focus of most evaluations, where stakeholders are interested in what results can be seen, or what change has transpired (Mayne, 2011). This question has a very narrow cause-and-effect scope and exposes a measurement challenge of measuring intended (as well as unintended) outcomes are not always apparent at first glance. Question 1 is often extrapolated to mean much more than “were the results were as intended” to factual statements ascribing the results to the intervention. The concept of attribution is essential to address as it is a mistake often made in evaluation, in that we often make claims greater than what is supported by the evidence.

Attribution is defined as inferring causality (Mayne, 2011). It refers to the causal relationships, and answer if the “observed outcomes [can] be directly attributed to the program?” (Patton, 2012, p. 364) Attribution, however, ignores the true complexity of the

event by assuming the variables involved are independent of one another and of the system in which they exist (Stern et al., 2012). In other words, by definition, attribution ignores context (Patton, 2012). CA does not seek to answer Question 1. As described previously, methods-based approaches often attempt to answer this line of inquiry but fall short as the black box remains intact (Astbury & Leeuw, 2010), while context is still not considered (Maxwell, 2012).

On the other hand, Questions 2 and 3 are inquiries about contribution and are the questions that CA purports to address. Contribution allows for the consideration of the complexity of the event(s) and is context-sensitive (Mayne, 2008; Patton, 2012). Unlike methods-based approaches, design, context is crucial in evaluations (Greene, 2011). There are multiple interacting independent variables in a dynamic environment with many causal factors potentially contributing to the outcomes. Thus, the evaluation becomes more nuanced: “to what extent and in what ways has the intervention contributed to observed outcomes?” (Patton, 2012, p. 365). To fulfill this function, CA draws on a crucial tenet of theory-based evaluations, the theory of change.

### **Theory of Change (ToC)**

The program (or intervention) theory of change (ToC) is the nucleus of theory-based evaluations and should contain the following elements: (a) impact pathway, and (b) assumptions and risks. This characterization is not consistent throughout the literature as it the case with many notions in the evaluation discipline. However, the operational explications provided in this discussion intend to align with CA’s conceptualization.

The impact pathway reflects the immanent program theory and draws the causal links between the intervention's key steps to articulate how input (activities) will engender impact (Mayne, 2011, 2019; Wimbush et al., 2012). A program theory (or intervention theory, change theory) is a conjecture of what goes on inside the black-box as input is transformed into output (Lipsey, 1987, as cited in Chen, 1990). In other words, the program theory hypothesizes how the intervention will generate the desired change (effect) and what will trigger the anticipated outcomes (Weiss, 1997a, 1997b, p. 46; see also Mayne, 2012); it shows the mechanisms of change (e.g., operations, procedure, interventions) connecting the delivery of programs to the intended outcomes (Weiss, 1997a). The intervention theory elicits the connections from activities to intended and observed outcomes (Funnel & Rogers, 2011; Johnson et al., 2009).

The impact pathway is often modeled using the implementation theory (or action theory). While the program theory reflects the conceptual side of the intervention, the implementation (or action) theory reflects the operational side and explicates how implementation, program design, and mechanisms (e.g., program operation and design, service delivery, logistics) (Montague, 2019; Weiss, 1997) will activate change. Implementation theories are often depicted using theory-of-action models or logic models (e.g., input, activities, and results chain). The results chain focuses on the intervention's products and therefore excludes the inputs and activities and reflects outputs, outcomes, and impacts (Mayne, 2017).

Consider a visual metaphor of a logic model: while the boxes represent the elements of the implementation theory (e.g., input, activities, outputs, outcomes), which

are connected (arrows) in a particular sequence to reflect the program theory (results chain). Taken together, the program theory and implementation theory suggest an impact pathway, “an explanation of how and why a certain type of intervention will make a difference” (Montague, 2019). Figure 2 is of a model impact pathway and illustrates the logical flow of activities, outputs, outcomes, and impacts.

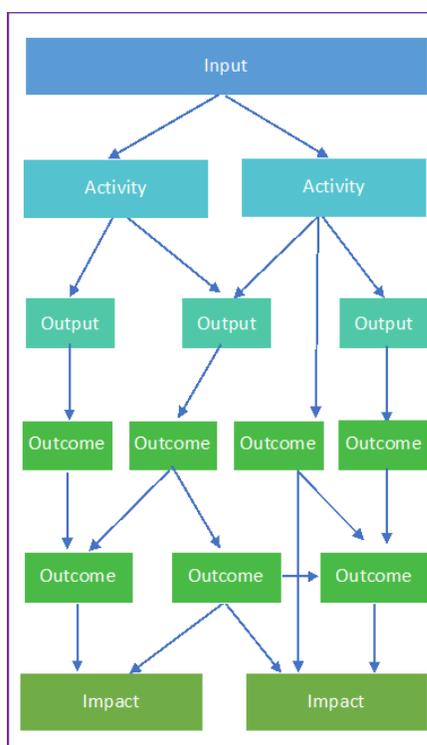


Figure 2. Sample Impact Pathway.

As emphasized throughout this discussion, the impact pathway characterizes each step from program inputs to the intended change using directional arrows, effectively declaring causal links between what is needed (e.g., the precondition) to generate the subsequent outcome (response) (Ton et al., 2019). For that reason, the ToC carries the weight of “specifying and explaining the assumed, hypothesized, or tested causal links”

(Patton, 2002, p. 162). To elaborate, a ToC needs to address the assumptions upon which the causal links are based, the risks to making this assertion, challenges to the assertions, and the context for which is appropriate. Figure 3 depicts a basic theory of change showing the impact pathway and assumptions along with the linkages).

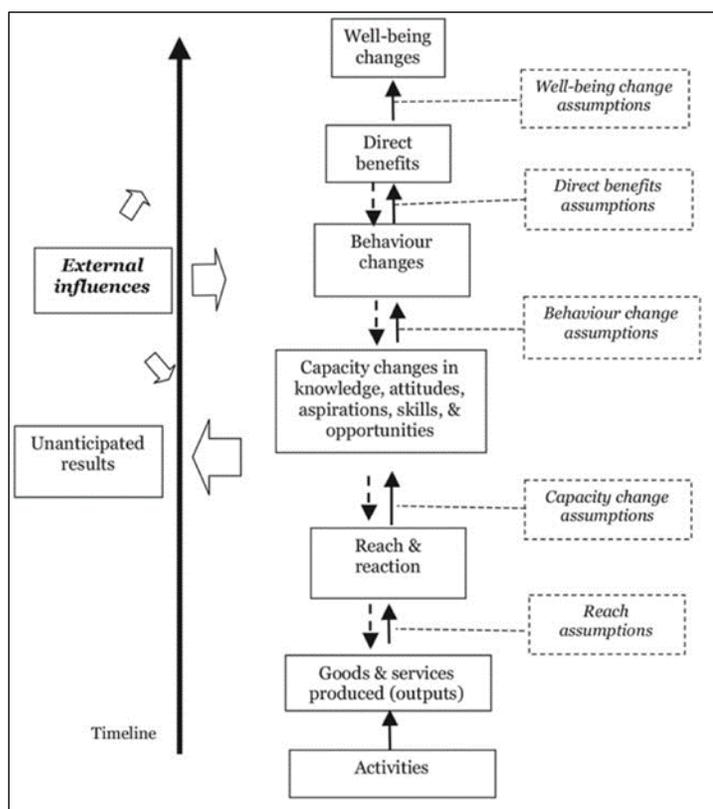


Figure 3. A Basic Theory-of-Change. Adapted from Mayne, 2015b.

A robust ToC needs the deliberation consideration of these elements (assumptions, risks, and context) in examining their role in supporting or obstruct the postulated program theory in producing its expected outcomes. In other words, the ToC describes the causal package (intervention, outputs, assumption) that should be sufficient to engender the expected effect, by explaining how and why the intended impacts are

expected to occur while addressing holes (e.g., risks, threats) in its arguments. To facilitate this quest of a robust ToC, CA draws upon its 6-step process.

### Contribution Analysis Steps

CA consists of six steps that prompt the three fundamental mechanisms of the framework (see Table 3). Together, these six steps assist in building conclusions founded on logical argumentation. Categorically, this process articulates a clear process through which an evaluator can investigate to what end are the observed results a product of the intervention's inputs (e.g., activities), and thereby produce a contribution claim founded on cogent reasoning, which is buttressed by backed by a logical framework or argumentation and its supporting evidence. This process is thought to provide definition and added-valued to theory-based approaches (Delahais & Toulemonde, 2012). As illustrated in the table, CA's six steps stimulate the development of the theory of change, the results chain (also known as the program theory), and finally, the contribution.

Table 3

#### Alignment of Key Mechanisms and Steps of CA

Key Mechanisms of CA	Steps of CA
A. Theory-of-change (ToC)	1. Determine the cause and effect issue to be addressed
	2. Develop a theory of change and risks to its success
	3. Generate evidence in response to the theory of change
B. Results chain (program theory)	4. Assemble the contribution story, and outline the challenges to it
	5. Seek out, test and strengthen(additional) evidence on the causal explanation
C. Contribution Story	6. Revise and strengthen the contribution story
	7. Return to Step 4 if necessary

Specifically, Steps 1 through 3 are meant to develop a robust theory of change. As previously discussed (see “Theory of Change (ToC)” on page 35), the theory of change (ToC) articulates the pathways of contribution; it draws on the program theory to generate the sequence of change leading to the expected outcomes (reflected in Steps 1, 2, and 3).

Steps 4 and 5 seek to develop the results chain (or program theory) by assembling an evidence-base for the claims asserted by the contribution story whereby the results occurred. These steps should explicitly identify the presumptions and conditions necessary to mobilize the intervention’s activities to expected outcomes, roles played by other factors and influences, challenges to the chain of inferences asserted, and aggregate across the various sources of evidence (reflected in Steps 4 and 5) (Wimbush et al., 2012). Primarily, the results chain should be explicitly in expressing the pathways of the intervention’s contribution to the outcomes and is backed by the evidence gathered in the previous steps.

Importantly, Step 5 embodies the strengthening techniques, which consist of deliberate methods to test the performance story and subject it to critical review by “knowledgeable others” which can also help raise the external credibility of the analysis. CA affords a framework to strengthen and critically reviewing the evidence to see how much success (or failure) of a program can be attributed to its focus or other influences – its contribution.

Finally, Step 6 (and 7) are focused on the contribution story or performance story, wherein which the contribution claim asserted is revised and strengthened with additional

evidence as needed. CA embraces a process of refinement as the postulated ToC is examined against evidence from multiple streams of evaluative information (Steps 6 and 7) (Wimbush et al., 2012). The final steps are “where CA adds most value” (Delahais & Toulemonde, 2012, p. 290). This iterative process allows for the construction of claims about the contribution to be examined and re-examined and provides for a rigorous inquiry method (see “CA as a Validation Framework” discussion on page 46).

### **Spheres of Influence and Magnitude of Contribution Claim**

The longer the results-chain, the more difficult it is to empirically verify the linkages in the impact pathway. Therefore, the further the linkages from the input, the less confidence we have in the influence exerted by the intervention on the change (effect). As such, the robustness of a ToC and length of the ToC will allow us to examine the various degrees of the intervention’s influence; namely, its direct control, direct influence, and indirect influence.

Contribution analysis allows for a layered investigation into the spheres over which the intervention has influence, listed in order of the greatest are direct control, direct influence, and indirect influence. Figure 4 illustrates the level of influence of an intervention in conjunction of the increasing role of external factors.

An intervention will have direct control over what is invested and what is produced (inputs to outputs); its direct influence is reflected in the short-term outcomes; while the indirect influence is its effect over a more extensive set of facts (e.g., political, economic, social change). The distance from which the expected outcomes are from the input structures the robustness of the postulated ToC, while the sphere of influence that

can be successfully investigated determines the magnitude of the intervention's contribution claim that can be asserted. Naturally, the magnitude of the contribution claim attempted should correspond with the sphere of influence it is investigating (see Table 4). Mayne (2011) outlined three magnitudes of contribution claims: the minimalist claim, the direct influence claim, and the indirect influence claims.

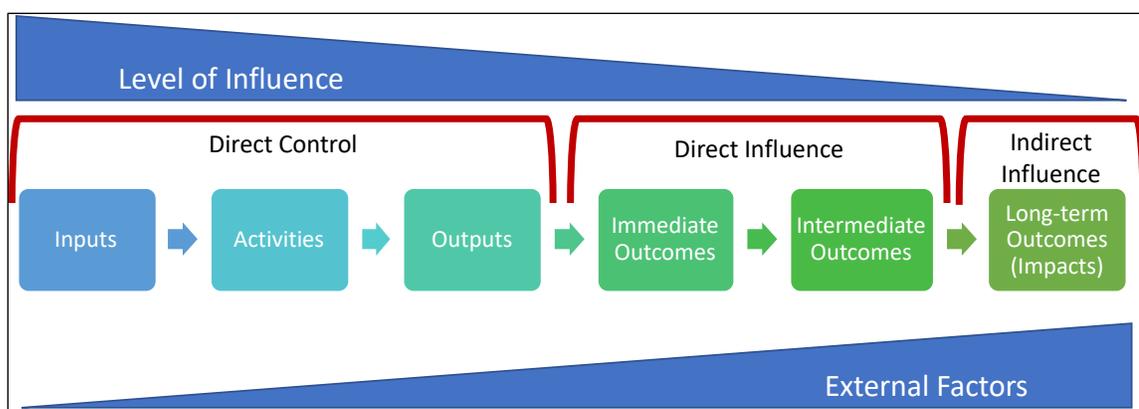


Figure 4. Direct Control, Direct Influence, and Indirect Influence.

Each layer follows a progressive three-level framework, where each layer reflects the types of claims that can be made enabled by the level/type and are as follows (Mayne, 2011): the minimalist level, the direct influence level, and lastly, the indirect influence level of the intervention.

Table 4

## Evaluative Inquiry, Magnitude of Contribution, Verification, and Contribution Story

Evaluative Inquiry	Sphere of Influence	Magnitude of Claim	Verification of Claim			Elements of Contribution Claim
			Results	Causal link assumption	External influences	
Has the intervention (or component) made a difference? (Mayne, 2019b)	Direct control	Minimalist	X			Simple binary statement: outputs were/were not observed.
Has it played a decisive causal role in bringing about change? (Mayne, 2019b)	Direct influence	Direct influence	X	X		Evidence suggests that the intervention was instrumental in creating the expected results in the presence of other influencing factors
How and why has the intervention (or component) made a difference, or not, and for whom? (Mayne, 2019b)	Indirect influence	Indirect influence	X	X	X	Measures intermediate & final outcomes evidence in the ToC in the areas of direct influence

*Note.* Adapted from Mayne (2011).

At a basic level, one must ask whether the expected outcome did indeed occur. However, this line of inquiry is not meaningful about the causal package necessary for the outcome to occur. It does not address the conditions under which the intervention does/or does not work. Instead, for a contribution claim about an intervention to be the most meaningful, it must provide information about “how and in what manner the interventions support factors and efforts brought about or contributed to said change” (Mayne, 2019b, p. 175).

The most straightforward causal inquiry corresponds with the minimalist claim, is just whether the intervention made a difference (Mayne, 2019b) relies on the observation of the expected results, and can be secured without CA. Earlier works by Mayne (see

2011) describe a minimalist claim (direct control sphere) is produced by first postulating a ToC and subsequently verifying that the expected outputs were observed (Mayne, 2011). However, recent communication (Mayne, 2020, personal communication) updated this understanding to stipulate that a ToC is not required to make claims about whether the expected outputs were produced, and therefore do not fall within the scope of CA. In other words, if evaluations are simply focused on whether the expected outputs occurred, there is no need to test a ToC and its underlying assumptions., and therefore does not require a theory of change at all. In other words, the verification of a ToC (and therefore identification of a ToC) nor its assumptions are needed to make claims within the intervention's direct control sphere (e.g., intervention outputs). In essence, evaluations seeking to establish minimalist contribution claims are **black-box** evaluations (e.g., see "The Black Box" discussion on page 30) as making claims of the intervention's direct outputs does not require verifying any underlying program theory (nor assumptions) (J. Mayne, personal communication, 2020).

The direct influence contribution claim can be made when (a) the assumptions about the direct influence links are confirmed (e.g., changes in knowledge, attitudes, and skills of those targeted by the intervention), and (b) other influencing factors are adequate to demonstrate/support that the intervention was influential in producing those direct results in the broader context of other influential factors (Mayne, 2011). The second piece of the direct influence level hinges on providing adequate evidence to demonstrate/support that the intervention was influential in producing those direct results in the broader context of other influential factors (Mayne, 2011). Statements made from a

contribution analysis at the direct level would be based on empirical evidence, verifying assumptions within the direct influence sphere, and the robustness of the postulated theory of change regarding areas of indirect influence.

Lastly, the contribution claim of indirect influence extends the analysis and is conceivably much more challenging (Mayne, 2011). The indirect influence analysis measures the intermediate and final outcomes (some of them) and gathers evidence of the assumptions borne of the indirect influence areas of the theory of change (Mayne, 2011). In other words, this analysis addresses the influence of other factors in consideration of the intermediate and final outcomes. The indirect influence lens would be appropriate in complex settings, where there may be multiple strategies dedicated to the intended outcome. As such, each strategy would have its own impact pathway and thus yield its own contribution claim (Mayne, 2011) that would seek to deliver supporting evidence for these strategies as pieces of the large-scale pathway of change. In this instance, there would be multiple pathways of contribution, and each strategy would generate an individual contribution statement, feeding into the overarching contribution story reflecting a comprehensive and tiered theory of change (Mayne, 2011).

Ultimately the defining factor between the three types of causal stories is founded entirely on the strength and plausibility of the contribution story—the more extensive the sphere of influence sought, the stronger and more credible the supporting evidence by the contribution claim, the more robust and stronger the assumptions.

### **CA as a Validation Framework**

Contribution Analysis (CA) aims to “reduce uncertainty about the contribution an intervention is making to observed results through an increased understanding of why results did or did not occur and the roles played by the intervention and the other influencing factors” (Mayne, 2012, p. 271). It is an evaluation approach that allows for the examination of “how” an intervention contributes to identified outcomes. Specifically, CA uses a process of “logical argumentation” (Craig, 2013; Wimbush et al., 2012) to trace impact pathways, by the linking evidence along each step of inference from inputs to outcomes.

Impact and outcome evaluations require practitioners to estimate the effect of an intervention on one (or more) outcomes of interest (Chen, Donaldson, & Mark, 2011). In order to make value judgments of program effects in impact and outcome evaluations, evaluators make a series of inferences along the theory of change to effectively connect the program inputs (e.g., activity or strategy) to intended outcomes. The evaluation discipline has received some undesirable attention of late as a call for more substantiation of “how evaluators warrant causal claims” (Gates & Dyson, 2017, p. 30) borne from the evaluations linger in obscurity (e.g., black-box evaluations). As such, it begs the question as to how we can improve current practices so that claims about program impact are clearly demonstrated. The discussion that follows proposes the use of CA as a framework to explicitly test the black box and thereby to build validated contribution claims. I present CA in alignment with Michael Kane’s Interpretive Argument, which allows us to “evaluate the rationale, or argument for the claims being made” (Kane, 2006, p. 17). It is

easy to see that CA summons rigor in reasoning and critical thinking (Patton, 2018) in how we form evaluative judgments about an intervention's contribution.

Validity as an argument is a long-standing social science tradition rooted in the interpretivist paradigm (Greene, 2011). Guided by the interpretivist paradigm, evaluations use valuative inference to provide some kind of value judgment (Julnes, 2011). Unlike the positivist-cousin that rests on causal or descriptive inferences to establish an objective truth, "interpretive validity relies on generating inferences that are meaningful, plausible, and of some consequence in the contexts at hand" (Greene, 2011, p. 82). Evaluations make valuative (Julnes, 2011) inferences to come to meaningful judgment, and validity refers to the quality of the inferences (Zumbo & Rupp, 2004). Necessarily, we need to think about "how can we improve current practices to ensure the validity of judgments produces?" In response, I propose Contribution Analysis as a validation framework by discussing its similarities with Michael Kane's (2006) Interpretive Argument (IA). The following section details how Contribution Analysis seeks to approach validation by enabling a logic of argumentation through a comparable process to that of the Interpretive Argument (IA) (Kane, 2006).

The crux of CA lies in the theory of change, through which the contribution story is generated and verified. The contribution story produced by CA rests on an interpretivist argumentation to a) build a nomological network of persuasive inferences and b) balance the evidence and argument presented (Greene, 2001). Indeed, an assumption that is not adequately supported by evidence or even worse, an assumption taken for granted within the ToC, effectively nullifies the entire argument. The warrant is

the evaluator's ultimate responsibility, as "inferences are consequential that we must have confidence that they are warranted" (Greene, 2011, p. 90). In the context of CA, a misconception of how the program theory operates (e.g., illustrated in the postulated ToC) might affect the integrity of any resulting claims derived from CA (Budhwani & McDavid, 2017) and effectively undermine the purpose of the evaluation, and possibly produce negative consequences.

### **Kane's Interpretive Argument**

The theory of validity has been a focus of discourse in academia (see Campbell & Stanley, 1963; Cronbach & Meehl, 1955; Kane, 2001, 2006; Messick, 1989) and can be categorized into two types: a scientific inquiry (e.g., Cronbach, 1988; Messick, 1989), or a rationale argument (Kane, 2006, 2013; Mislavy, Steinberg, & Almond, 2003). While the validity of scientific inquiries rest on the generalizations and/or extrapolations made from the inferences to the real world, the flip side rests on interpretivism, in that it does not seek "the objective truth per se, but rather to provisional, contingent, dynamic understandings about human action in context" (Greene, 2011, p. 84).

Kane's (2006) interpretative argument (IA) is built on the idea of propositions that help connect framing statements to inferences and create the basis for a proposed use/interpretation. Kane (2006, 2013) builds on the work of prominent scholars (e.g., Campbell, 1988; Messick, 1989) and offers additional guidance on the validation process as the development of the argument for or evidence to support, the inferences made [datum] via a two-pronged framework: an interpretive argument (the specification) and the validity argument (the evaluation) (Kane, 2006, 2013; Wallace, 2011). The

interpretive argument can be thought of as a theory (akin to a scientific theory) (Kane, 2006, p. 25), which specifies the network of individual inferences within an argument and the assumptions upon which it depends (Kane, 2006, 2009). Alternatively, the validity argument evaluates the interpretive argument. While the interpretive argument provides a theoretical framework for inferences, the validity argument offers a framework for interpretation and evaluation of the argument (Kane, 2006, 2009).

Kane's IA hosts two stages to the validity argument: development stage and appraisal stage. The interpretive argument is developed and strengthened (development stage) via an iterative process to be subsequently evaluated from a critical stance (appraisal stage) (Kane, 2006). While the interpretive argument cannot be proven, it can be rigorously appraised against the criteria of clarity, coherence, completeness, and plausibility (Kane, 2006, 2009). An accomplished validation argument satisfies in providing conclusive support for the declarations made given that the premises (assumptions) are true.

As validity lies in the supporting arguments (House, 1980), CA brings us a framework to facilitate this process by evaluating the ToC and its chain of inferences to ensure a met objective of social betterment (Furubo & Stame, 2018). Table 5 (page 50) depicts Kane's (2006) IA alignment to approach validity aligns with the CA steps and products (theory of change, results chain, and contribution story).

Table 5

## Alignment of IA with CA Mechanisms

<b>Kane (2006)</b>	<b>Key Mechanisms of CA</b>	<b>Steps of CA</b>
Interpretive argument	A. Theory of Change (ToC)	1. Determine the cause and effect issue to be addressed
		2. Develop a theory of change and risks to its success
	B. Results chain (program theory)	3. Generate evidence in response to the theory of change
Validity argument	C. Contribution Story	4. Assemble the contribution story, and outline the challenges to it
		5. Seek out additional evidence

The complete IA is a journey over several bridges of inferences (Kane, 2006) specifying of input to outcome (interpretive argument), each bridge is linked to the next with evidence supporting each step of inference (validity argument). I propose the interpretive argument in CA is comprised of the ToC and Results Chain, and semblance of Kane's endless pursuit strengthening the argument. The third CA mechanisms align with the Kane's (2006) validity argument and seeks to evaluate the network of inferences.

Evaluation is an essential "methodological toolbox for every other discipline" (Scriven, 2016, p. 27). With transdisciplinary reach, evaluators and evaluations have the potential to influence (Kirkhart, 2013) the far- and wide-reaching impact of their work and products beyond what we can imagine. Thus, we need to be aware of our role in conducting and producing evaluation results. In fact, we also need to acknowledge the evaluator's role, who themselves are instruments and lens through which the argument is

expressed. House stated “data do not assemble and interpret themselves” (House, 2014, p. 12), and promptly reminds us of the importance to be conscious of our own positionality while we reflect on the information in the context of the intervention and give meaning that is most relevant for that current scenario.

### **Linking Evaluation Theory and Practice**

The following section details the connection between evaluation theory and practice, and the dearth of research between the two. I highlight the importance of conducting research on theories, specifically theories serving a prescriptive purpose such as CA. Finally, I conclude with a germinal review of CA in practice.

By and large, evaluation theories are of normative origin and borne from practice rather than theories put into practice (Coryn, Noakes, Westine, & Schröter, 2011). In other words, evaluation is a practitioner-based field, and evaluation theory is derived from practice (Shadish, 1998). However, as N. L. Smith (1993) revealed, “it is not clear what is meant when an evaluator claims to be using a particular theoretical approach” (p. 240), a sensation largely fueled theorists’ presentation of theories in abstract conceptual terms and vagueness in describing how they would be applied in practice (Miller, 2010; N. L. Smith, 1993).

Pawson states that “evaluation can only grow as a science if it learns lessons from investigation to investigation rather than each inquiry emerging freshly out of the egg” (Pawson, 2013, p. 138). Evaluation theory and evaluation practice have a dialectic relationship. Theory informs evaluation practice, and practice informs theory to no end. In actuality, however, there are very few published studies that examine the influence of

theory on practice (Christie, 2003). Under this notion, we are practicing without considering whether theoretical assumptions are so.

Furthermore, many practicing evaluators are not evaluators per se, but rather are professionals in their own field who are conducting evaluations (e.g., Christie, 2003).

This behooves us to continue to examine the unclarified nuances and misalignments of theory in academic literature and practice. Cousins and Earl (1999) counsel:

We need to move beyond the relatively weak connected theoretical musings and anecdotal reports of practice . . . We need to add to the empirical knowledge base through carefully developed and executed studies that have the potential to extend our theories and guide our practices, studies that manifestly strengthen the link between theory and practice. (p. 316)

### **Role of Theory in Evaluation**

In general, evaluation theories guide practitioners on a myriad of decisions, navigating the entire process from design to conducting an evaluation, to disseminating results. Evaluation theories help illuminate the ideological perspectives on evaluation (N. L. Smith, 2007), guide in defining the scope of the evaluation, identify the appropriate role of the evaluator, role of the stakeholders, where control and decision-making power should lie, and the dynamic between the two actors in the evaluation, the depth of involvement (Cousins & Whitmore, 1998; Fetterman, 1994); selecting evaluation questions and pairing with the appropriate methods (e.g., Greene, 2007; Mark et al., 2000; Rossi, Lipsey, & Freeman, 2004); whose informational needs are prioritized by the evaluation (e.g., Abma & Stake, 2001; Greene, 1997; Mark & Shotland, 1985); identifying “when, how, and to whom evaluation findings are disseminated and with what

purpose” (Stufflebeam & Coryn, 2014; e.g., Patton, 2008; Preskill & Torres, 1999). This list is exhaustive but not complete, which highlights the expansive influence of evaluation theory on evaluation practice.

Customarily, evaluation theories are of normative origin and borne from practice rather than theories put into practice (Coryn et al., 2011). In other words, evaluation is a practitioner-based field, and evaluation theory is derived from practice (Shadish, 1998). However, as N. L. Smith (1993) revealed, “it is not clear what is meant when an evaluator claims to be using a particular theoretical approach” (p. 240), a sensation largely fueled theorists’ presentation of theories in abstract conceptual terms and vagueness in describing how they would be applied in practice (Miller, 2010; N. L. Smith, 1993).

### **The Need for Research on Evaluation on CA**

The call for Research on Evaluation (RoE) is well established (Mark, 2001; Miller, 2010; Shadish et al., 1991; N. L. Smith, 1993; Worthen, 2001). Notwithstanding RoE on evaluation use (Cousins & Leithwood, 1986; Johnson et al., 2009; Alkin & Taut, 2003), the answer to the call has yet to surface in the discipline. Mainly, the relationship between evaluation theory and evaluation practice is tenuous (Christie, 2003; Miller, 2010; Shadish & Epstein, 1987), yet the implications of specific evaluation theory for practice have largely not been studied (Miller, 2010).

The prescriptive role of theory in practice is ever-present as they are almost exclusively prescriptive (Akin & Christie, 2008). I argue that Contribution Analysis is one such approach that is prescriptive, and therefore warrants a thorough examination of

practice. As such, the following highlights the role and implications of prescriptive theories in practice and follows with a brief exposition of CA in practice.

As an evaluation framework presently guiding practitioners in their work, Contribution Analysis also reflects a degree of fuzziness. It reflects an underdeveloped evaluation theory with relatively low operational specificity and critical pieces such as epistemology and ontology that remain unaddressed (Budhwani & McDavid, 2017; Dybdal et al., 2010). Though cases provide some understanding of this approach in practice, there remains limited empirical evidence to support a comprehensive understanding. As a field of application, the dialectic theory-practice relationship is fundamental as theory is informed by practice, and which is applied in practice, and further developed by studying practice. An empirical study of CA in practice is needed to inform our understanding of this approach, and theory-based evaluations from a macro lens.

The relationship between theory and practice is too often assumed, and as many in the field have emphasized, research on evaluation is sorely needed (Miller, 2010). Though there are quite a few works focused on Contribution Analysis, and a few published empirical cases, there are even fewer linking the practice and the theory, and none that holistically synthesize the practice literature. As a result, the empirical information on “which approaches to evaluation, implemented how and under what conditions, actually lead to what sort of improvements?” (Mark, 2008, p. 115) is lacking.

Additionally, the abstract nature and explications of theories subject themselves to endless interpretations in practice, which may or may not have been intended by their

theorist (Gates & Dyson, 2017; N. L. Smith, 1993). As such, it is a responsibility of the constituents of the discipline to curtail misuse of theory and investigate degrees of interpretation, clarify theoretical guidelines if needed, and continuously learn from practice. In consideration of the increased requests by funding agencies to use alternative approaches such as CA to investigate casual relationships, an examination of theory and practice is not only warranted but largely overdue.

## **CHAPTER III**

### **METHODOLOGY**

This chapter outlines the methodology for the dissertation study. First, I revisit the research objectives and questions, and present an overview of the dissertation design. Next, I describe the two distinct phases, a systematic review (SR) of the literature and followed by a multiple-case study. The first phase (Phase I) examined the Contribution Analysis (CA) landscape. Mainly, I discuss the systematic review procedures to review the conceptual literature and the empirical literature to elicit the key tenets of Contribution Analysis. The findings from Phase I were used to guide the ensuing study, a multi-case study of empirical CA applications (Phase II). Accordingly, an elaboration of Phase II (multi-case study) is presented wherein which the case selection procedure, data collection methods, data analysis, and the means by which trustworthiness and rigor were protected are detailed. Finally, I end with a discussion on the limitations of the methodologies.

#### **Research Objectives**

Chapter II revealed a covert culture of *laissez-faire* and ambiguity across the discipline, particularly in connecting practice and theory. Moreover, the need to conduct research on evaluation remains a persistent and lingering void that has yet to be sufficiently explored, and Contribution Analysis (CA) is no exception. Despite the surge in CA applications likely instigated by the increase in demand for alternative approaches,

Contribution Analysis is discernably unrefined concerning conceptual development and understanding nuances of practice. Simply put, although Contribution Analysis is currently emerging as a new methodology in evaluation (Schwandt, 2015), the current state of research is inadequate.

As demonstrated in the review of the literature, many salient traits remain undefined, especially given that CA has been theorized far more than it has been explored in practice (Delahais & Toulemonde, 2012), the need to further clarify at both the level of theory and practice is overdue. As such, the primary objectives guiding this study are focused on (1) investigating the theoretical translation and practice of contribution analysis; (2) adaptations and innovations made to CA's distinct six-step process; and (3) the conditions, influences, and factors that may facilitate or impede CA practice. Overall, findings from this research will contribute to a more comprehensive practice-based understanding of CA and illuminate potential challenges, solutions, benefits, and considerations in planning, designing, conducting employing CA-informed impact evaluation. In order to achieve the research objectives, the study asks the following research questions:

- 1) What is the current theory of the Contribution Analysis approach to evaluation?
  - a) What are the fundamental tenets that guide CA practice?
- 2) What does CA look like in practice?
  - a) How do practitioners implement the approach?

- b) What adaptations, adjustments, or reinforcements are made to the 6-step process?
- 3) What conditions and contextual factors (e.g., of the evaluation and program) challenge and facilitate the implementation of CA in the field?

### **Overview of Research Design**

This study followed emergent and exploratory design consisting of two distinct phases, a systematic review of the Contribution Analysis literature (Phase I) and a multiple-case study of seven empirical applications of CA (Phase II). Specifically, Phase I examined (a) the conceptual literature to identify the fundamental tenets of the Contribution Analysis evaluation approach, and (b) the empirical literature base to characterize themes of practice and factors that may influence it, all of which guided the ensuing study. Phase II employed a multiple case (multi-case) study (Stake, 1995, 2006; Yin, 2009) to facilitate a systematic examination of the phenomenon (CA as an evaluation approach). The constructs investigated in the multi-case study were informed by the findings from the systematic review of the literature (Phase I), and consequently helped narrow the research scope to: CA's signature 6-step process, fundamental tenets of Contribution Analysis, the role of stakeholders, and the conditions and contextual factors that may affect practice. Overall, Phase II provided insight into how CA manifests in the field, how the six-step process looks across different contexts, where difficulties lay, which conditions nurture specific barriers and facilitators to implementation, and how practitioners mitigated threats to their practice.

Phase I initiated a fundamentally exploratory and generative inquiry of the extant of the conceptual and empirical literature base to support numerous objectives. Precisely, the systematic review of the conceptual literature provided contextual information on Contribution Analysis as an evaluation approach, the historical developments, and an overdue demarcation and synthesis of the underlying philosophical assumptions (e.g., axiology, ontology, and epistemology). The conceptual literature sample was limited to publications written by John Mayne, and thus this enterprise facilitated in identifying the core tenets of Contribution Analysis as intended by CA's chief architect. On the other hand, a systematic review of the empirical literature intended to investigate the translation of theory into practice. In more detail, this component sought to characterize CA practice (e.g., trends, modification/adaptations, lessons learned) as well as contextual factors that may influence it, which in turn produced propositions that guided the next phase of research. This exercise also assisted in identifying prospective cases for the upcoming cross-case analysis. On the whole, the findings from Phase I informed the directed the focal points investigated in the multiple case study (e.g., CA tenets) (Bryman, 2012; Stake, 1995). In that sense, Phase I served to be suggestive of incipient theories, especially by illuminating contextual factors that may influence practice.

Phase II built on the findings of the systematic review and carried out an in-depth investigation of CA practice via a multi-case study. Specifically, this part of the research sought to examine the translation of theory into practice, as well as the presence and prevalence of the themes identified in Phase I. As such, the multi-case study of seven empirical applications of CA was guided by the CA framework, findings from the

systematic review of the conceptual literature (tenets of CA), and findings from the systematic review of empirical literature (prominent themes of practice, stakeholder engagement, and contextual factors). Cases were identified through Phase I and assessed for their eligibility and relevance to the research objectives (procedures are elaborated in Phase I: Systematic Review of the Literature ). Findings from the multi-case study (Phase II) drew on semi-structured interviews with the evaluation practitioners, and document analysis of the published document was conducted to meet this objective. The use of multiple sources per unit of analysis facilitated member-checking interpretations of the findings and safeguarded the development of context-sensitive findings.

While Phase I provided the conceptual framework that guided the study, notably, it yielded the identification of the core tenets of CA and a set of practice propositions that guided the multiple-case study. Consequently, Phase II reflected a systematic exploration of the phenomenon and empirically informed how CA manifests in practice, where difficulties lay along the six-step process, and how practitioners adapted and/or innovated. It also allowed for multi-case study and a direct probe of cases exemplifying best/worst practices of CA.

Specifically, findings from Phase I contributed to the current extent of the conceptual literature by elaborating and synthesizing on the crucial philosophical assumptions on which it is founded, as well as clarifying the conceptual and operational definitions of the fundamental concepts to CA (see “Theory of Change (ToC)” on page 35; alignment of “Spheres of Influence and Magnitude of Contribution Claim” on page 41, and the validation of the evaluation findings as discussed in “CA as a Validation

Framework” on page 46). Phase II yielded more explicit theory-practice findings based on empirical applications of CA fills out the portrait of the CA practice landscape. This second phase also contributes directly to the Research on Evaluation (RoE) scholarly body by illustrating the translations of an evaluation approach from theory to practice.

Overall, findings from this study are expected to be valuable for practitioners in that they illuminate potential threats (and enablers) to designing, implementing, and practicing evaluation and practice-based solutions. In whole, the research synthesized from this dissertation contributes to the development of contingency theories of the Contribution Analysis approach for outcome and impact evaluations.

Figure 5 illustrates the phases, procedures, and purpose of Phase I and Phase II of the systematic review of the conceptual and empirical literature, leading to the multiple case study of practice cases. Table 6 shows the timeline per study phase and research activity, and Table 7 shows the research phases along with the instrumentation, sample, procedures, and analyses.

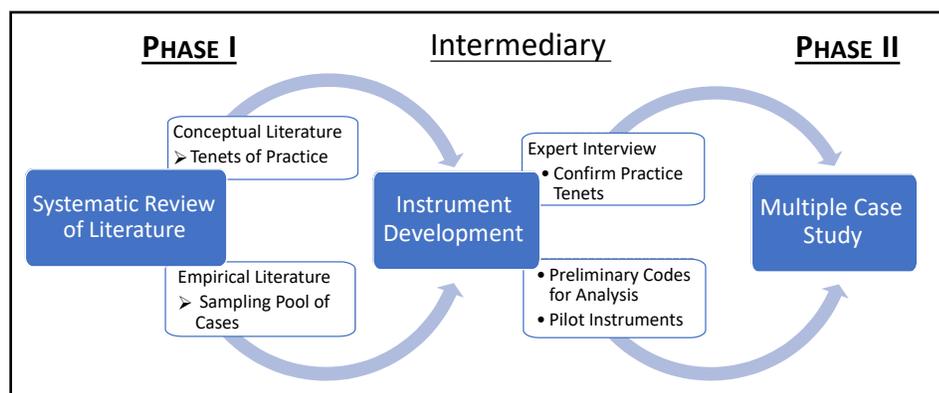


Figure 5. Diagram of the Procedures for Exploratory Sequential Study, Detailing the Systematic Review (Phase I) and Instrument Development (Intermediary), and Multi-Case Study (Phase II).

Table 6

## Study Timeline

	<b>Activity</b>	<b>Product</b>
6/2019	Submit IRB	IRB Approved
8/2019 – 11/2019	<b>Phase I: Systematic Review (SR)</b> Conceptual Literature Empirical Literature	‘Guiding Table’ of cases CA Tenets (Preliminary) Practice Propositions Major themes Identified cases for interviews List of contacts for survey SR extraction data
11/2019 – 1/2020	<b>Intermediary Phase</b> Expert Interview Instrument Development Cognitive Interview Pilot Instruments Finalize Coding	Updated Tenets of CA Updates on instruments (feedback, cognitive interview, pilot study) Data collection organization and code sheet Revised instrumentation
1/2020	<b>Phase II: Mini-Collective Case Studies</b>  Contact participants Email consent, interview questions, and Qualtrics link Conduct interview Summarize & member-check document	Case extraction data (interviews) Interview transcripts Case summaries Within & across theme development Cross-case analyses

Table 7

## Alignment of Phases to Instruments, Sample, Procedures, and Analyses

	<b>Phase I (Systematic Review)</b>	<b>Phase II (Multi-Case Study)</b>
<b>Instrumentation</b>	Document analysis of case publications  Systematic Review  Coding Sheet	Interviews with practitioners  Demographic questionnaire (Qualtrics)
<b>Sample</b>	<i>Conceptual Literature</i> A purposive sample of conceptual publications by John Mayne  <i>Empirical Literature</i> A purposive sample of publications demonstrating empirical applications of evaluations using guided by CA, published after 1999	<i>Publications</i> Most recent publications, with enough detail  <i>CA Practitioners</i> Practitioners with most recent publications  Purposive/snowballing sampling to identify CA practitioners, ranked cases
<b>Procedures</b>	Instruments were tested on articles from the sample pool	Instruments were piloted once IRB approval was received Interview participants were asked to sign consent forms. They were contacted via email to explain research, expected contribution, timeframe, and procedures.
<b>Analyses</b>	Qualitative data were coded by dominant themes and patterns according to the research questions, and summaries will be provided.	Qualitative data were coded by dominant themes and patterns according to the research question, and summaries will be provided.  Descriptive statistics were used to describe information about demographics, and practitioner experience.

### **Guiding Worldviews and Philosophies**

Historically, we have divided the world into two antithetical worldviews of positivism or interpretivism (Creswell, 2009; Merriam, 1998). Positivism holds at its core the ontological assumption of an objective reality (Nkwake, 2013), and carries rigid views in which the world (Fitzpatrick et al., 2011) is seen as “stable, observable, and measurable” (Merriam, 1998, p. 4). On the other hand, interpretivism acknowledges multiple realities that are socially constructed by individuals (Merriam, 1998), and emphasizes understanding the meaning of processes or experience (Khanal, 2014). As such, the two paradigms have been defined by their inherently dissimilar values of theoretical conceptions of reality (ontology) and knowledge (epistemology), impelling how research may be conducted: positivism is steered by deductive reasoning or hypothesis testing to draw conclusions-while interpretivism relies on inductive, hypothesis- or theory generating reasoning (Merriam, 1998).

Greene (2007) describes a paradigm as a worldview, with a set of ascribed philosophical assumptions of ontology and epistemology that circumscribe the “nature of methods, processes, and conclusions” (Nkwake, 2013, p. 98) the study is expected to produce. Constructivism assumes that reality is a social construct (Nkwake, 2013). There are many truths, and people’s realities are not only as they see through their lenses but are shaped by social factors (Nkwake, 2013). Phase I draws heavily on case study methods; specifically, the practitioner interviews will facilitate an investigation into their subjective realities and emphasizes a constructed account of reality (Stake, 1995; Yin, 2009).

Furthermore, the inclusion of multiple cases, thereby, multiple practitioners, will allow for more occasions and perspectives to examine the phenomenon.

### **Phase I: Systematic Review of the Literature**

Phase I consisted of a systematic review of the Contribution Analysis literature. The purpose of the systematic review was two-fold: to identify the tenets of CA, and to amass a corpus of empirical cases to include in the sampling pool for the ensuing multiple case study. A systematic review (SR) differs from a literature review in that it is a scientific study in itself. The process involves collecting and synthesizing all scientific studies on the topic of Contribution Analysis. Systematic reviews (SR) are rigorous and require procedural transparency, which reduces selection bias as criteria for including articles is predetermined, and finally produces an assessment and analyses of studies (Pajo, 2018).

### **Review of Conceptual Literature**

The main objective of the conceptual literature review was to identify crucial tenets of CA and to contribute to the conceptual knowledge base. This segment details the procedures about which this occurred, and the findings produced are described in Chapter IV on page 96.

Table 8 outlines the sampling criteria for the review of the conceptual literature. It details the requirements for works authored by John Mayne and as well as other authors. The search was enacted multiple times over 10 months, wherein new publications did emerge, providing further clarification on CA concepts. The conceptual works on the Contribution Analysis approach drawn upon throughout this dissertation to elucidate the

conceptual theory of CA and tenets of CA are detailed in APPENDIX A (authored by John Mayne), and APPENDIX B (authored by other scholars).

Table 8

Inclusion Parameters for Conceptual Literature Review

Inclusion Criteria	Explanation
<b>Conceptual Literature by John Mayne</b>	<ul style="list-style-type: none"> <li>• A publication describing the Contribution Analysis process, or concepts relevant to the approach (e.g., Theories of Change, evidence in causality)</li> <li>• Gray or Peer-reviewed literature</li> <li>• Timeframe: 1999-2020</li> </ul>
<b>Conceptual Literature by Other Authors</b>	<ul style="list-style-type: none"> <li>• Literature discussing CA from a conceptual perspective</li> <li>• Peer-Reviewed Only</li> <li>• Timeframe: Post-2005</li> </ul>
<b>Type of literature</b>	<ul style="list-style-type: none"> <li>• Narrative, Critical Review, Synthesis of Methodological Guidance</li> </ul>
<b>Evaluation Context</b>	<ul style="list-style-type: none"> <li>• Any</li> </ul>
<b>Period</b>	<ul style="list-style-type: none"> <li>• Publications published in the last 10 years</li> </ul>
<b>Publication status</b>	<ul style="list-style-type: none"> <li>• Peer-reviewed publication in English-Language journal</li> </ul>

### Scoping Review of Empirical Literature

A comprehensive literature review was conducted to identify as many peer-reviewed articles demonstrating the application of contribution analysis. Articles that did not speak of contribution analysis from the author's first-hand perspective were not considered for this study, as the author believes that secondary data sources are prone to higher rates of inaccuracies and/or misrepresentation of the case. Furthermore, although there were evaluation reports depicting contribution analysis applications from various

sources such as government program and policy sites, those reports were not included in this study as they were not published through a peer-review process.

Table 9 outlines the search parameters. The search procedures targeted traditional academic sources (e.g., EBSCOHost, JSTOR, evaluation journals, Google Scholar, major databases), and general search engines using the search terms *contribution analysis*, and *evaluation*. Results were filtered for English-language journal articles published after 1999 (marking Mayne’s first seminal paper on CA). Abstracts were then reviewed for appropriateness, and relevant articles were added to a database. Additionally, the works cited and reference list of each extracted articles (and other relevant sources, e.g., theoretical articles) were reviewed to identify additional prospects. Per these criteria, 89 publications from 2007 to date (January 2020) were identified for consideration for the multiple case study in Phase II. See APPENDIX C for a listing of all empirical articles discovered in Phase I. APPENDIX D contains a legend expanding on journal abbreviations.

Table 9

Initial Search Parameters for Scoping Review

Inclusion Criteria	Explanation
Evaluation Case	Peer-reviewed CA-oriented evaluation from the practitioner (evaluator)’s perspective. This must address the evaluand, evaluation process, and how it was informed by CA.
Evaluation Approach	Contribution Analysis as the primary framework informing the evaluation design and practice.

Table 9

Cont.

Inclusion Criteria		Explanation
Evaluation Context	Any	
Types of Interventions	Any	
Period	Evaluations completed from 1999 to date	
Publication status	Peer-reviewed publication in English-Language journal	

As can be seen from Figure 6, there has been a steady increase in the number of CA publications over the past 13 years. In total, a body of 88 publications was identified from 44 journal publications. The following journals had the most publications of CA cases since 2007: *Evaluation Journal* ( $n=13$ ), *(CJPE)* ( $n=10$ ), *The American Journal of Evaluation* (AJE) ( $n= 5$ ), *Journal of Development Effectiveness* (JDE ) ( $n= 4$ ), *Evaluation and Program Planning* (E&PP ) ( $n=3$ ), and *Research Evaluation* (RE) ( $n=3$ ). The spike seen in 2012 is the special edition on CA in the CJPE. APPENDIX L shows the sources of empirical CA publications by year (page 207). See Figure 6 for the number of publications produced by each journal per year.

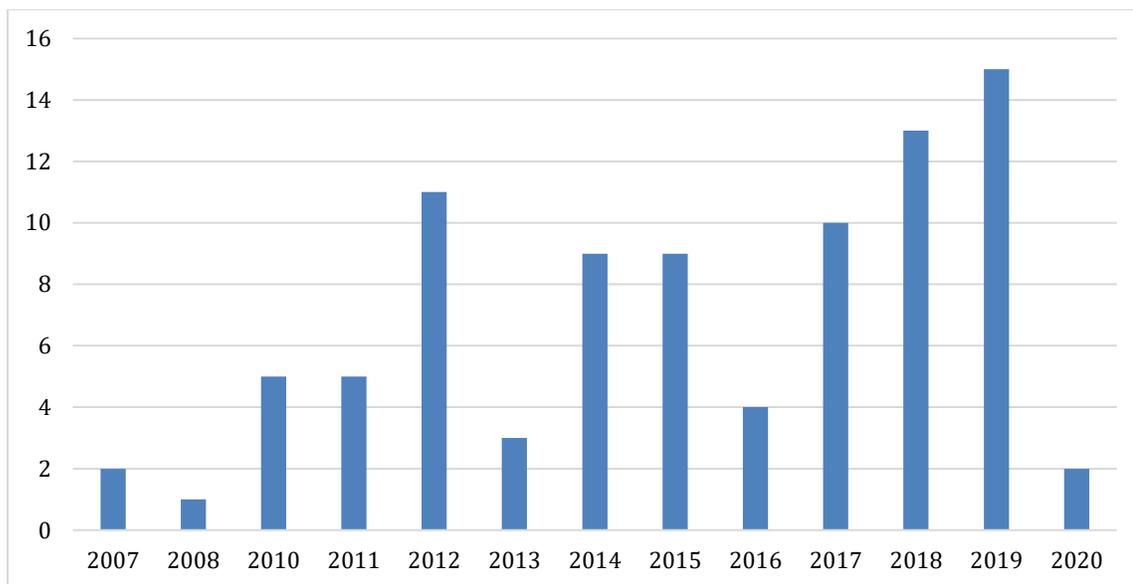


Figure 6. Number of Empirical Articles Published Per Year from 2007 to Date (2020).

### **Intermediary Phase: Development of Instrumentation**

This section discusses the procedures through which the data collection instruments were developed, specifically the CA Scoring Sheet, the Practitioner Interview Protocol, Case Extraction. As discussed, Phase I had the objective of identifying the fundamental tenets of CA. The primary outcome is the draft of CA tenets and values (see Appendix I). According to philosophies of CA (Mayne, 2019; J. Mayne, personal communication, 2020), theory-based evaluation principles (Coryn et al., 2011); and etiological fundamentals (Brousselle & Buregeya, 2018; Gates & Dyson, 2017).

The **CA Tenets** identified in the review of the conceptual literature in Phase I informed the development of the two instruments for Phase II: the interview protocol for Phase II (Appendix E), and CA scoring sheet (Appendix I). The following section describes the data collection instruments for the multi-case study and, specifically, how

they were developed. There were **two** instruments developed: (a) case extraction form, and (b) interview and demographic protocol.

The **interview and demographic protocols** (Appendix I on page 199) were developed from the systematic review of the conceptual literature. This yielded fundamental tenets of CA (Coryn et al., 2011; Gates & Dyson, 2017; Mayne, 2019b). To confirm my understanding of the fundamental values, a consultation with the leading subject-matter CA expert, Dr. John Mayne, ensured that the topics, several focal points, and responses were accurately represented in the interview protocol. On the whole, the feedback from the dissertation, committee, and cognitive interview ensured the quality of the instruments, while feedback from Dr. John Mayne ensured content quality of constructs represented in the instruments.

### **Refining Instruments**

The systematic review of the literature yielded a pool of interview items reflecting a multitude of domains of interest). As I carried on with my dissertation research and engaged in academic discourse on CA, it was clear that there were specific domains that were taking priority in my research, specifically the 6-step process in practice, developing the ToC, and stakeholder engagement. Therefore, the items developed around those domains were drawn upon to create the protocol used in the practitioner interviews.

Once I refined my research questions, my interview protocol and questionnaire were submitted to the dissertation committee for feedback on who is well-versed in both instrument development and the field of evaluation, positioning them as well-qualified to guide on these matters. The feedback received from the dissertation committee members

was incorporated and a cognitive interview (Ryan, Gannon-Slater, & Culbertson, 2012) of the interview was conducted with an ERM student.

At that point, I conducted a cognitive interview (Ryan et al., 2012) of my instruments (interview protocol and Qualtrics questionnaire) with an evaluator practitioner who had some familiarity with CA. This was done to ensure that the items are sound, their underlying constructs are effectively communicated, and overall, the instrument is appropriate for the targeted sample of CA practitioners.

Through the cognitive interview, I was able to probe participants to articulate their thoughts about each question and why they are responding as they are, which will provide valuable insight into how the items are understood (Ryan et al., 2012; Tourangeau, Rips, & Rasinski, 2000), and made modifications as needed to better reflect the desired constructs of interest. Revisions yielded from the cognitive interviews were incorporated in instruments for administration with CA practitioners. The cognitive interview revealed that the demographic questionnaire and interview were too lengthy (45 minutes and 90 minutes, respectively) and thus needed to be reduced to maximize response rates. As such, the revisions focused the instruments on two domains: stakeholder engagement and CA 6-step process. The aim was to have an interview protocol that would take 45 minutes to 1 hour, and for the Qualtrics Questionnaire to take no more than 10 minutes. Finally, I consulted with Dr. Mayne on the instrument constructs, items, and response options that were specific to Contribution Analysis (e.g., CA tenets from Phase I, and the items I developed from these findings to administer to the CA practitioners).

**Researcher log.** Throughout the study, I maintained a researcher log to record events and developments of the study (e.g., data collected, procedures enacted, challenges of the process, etc.). While it served as a technical log, the researcher log was also used to store my reflections, observations, thoughts, ideas, questions, and concerns. This log was not only a source of data but also documented my own developments.

### **Phase II: Multiple Case Study**

To capitalize on the recent surge of practice and conceptual literature on Contribution Analysis, Phase II drew heavily on a multi-case study design (Stake, 1995, 2006) to assist in a holistic investigation of CA. As an ethnographic design (Creswell, 2002), research founded in case study methodology permits for in-depth exploration of a “bounded system” (Merriam, 1998) or a case over time, to produce information-rich in context (Creswell & Maietta, 2002). Specifically, this design will increase the units of study to “illuminate a particular issue” (Creswell, 2002, p. 485) and cultivate a better understanding of the topic, to analyze across context (Baxter & Jack, 2008) and/or generate theory about a broader context (Chmiliar, 2010).

As such, this research design was informed by a multi-case study and is an appropriate step for a multitude of reasons. For one, it allowed for not only explorative investigation of the CA evaluation approach but for more specific theory-to-practice related queries identified by prominent academics (e.g., what worked and how it happens, how contexts differed and what worked across different contexts, how to prevent poor practice, identifying inadequacies and solutions; see N. L. Smith, 1993). Secondly, the inclusion of multiple cases allowed for a more accurate representation of the variety and

range of operating contexts, adaptations, innovations, practices exhibited in CA's practice portfolio discussed in Chapter II.

This research design allowed for a thorough and profound investigation of each case, as well as cross-case analysis of themes across the sample (Bryman, 2012; Yin, 2009). While within-case examinations allowed for the researcher to be attentive to each case and its specific context, comparisons facilitated by the cross-case analyses supported the revelation of concepts that were pertinent to an emerging scheme or ascertain conditions wherein which it will or will not hold (Bryman, 2012; Yin, 2009). Although Phase II is informed by case study design (e.g., methods, procedures, and analyses), I should note that it is not a case study in the technical sense as the qualitative case study methodology was developed to study the experience of real cases operating in real situations" (Stake, 2006, p. 3). In this case, the units of analyses are cases, and I draw on case study strengths to assist an investigation of Contribution Analysis evaluations and events I have not been privy to witnessing in situ.

### **Multiple Case Study Design**

**Defining the case.** Phase II of this dissertation study adopted the multiple case study design (Yin, 2009). A case is the "Contribution Analysis evaluation," is a distinct CA evaluation where there is a richly detailed peer-reviewed publication written from the evaluator/ practitioner's perspective and the practitioner agrees to participate in the phone interview.

While each case was analyzed independently to allow for in-depth examinations (Stake, 2005; Yin, 2009) a cross-case analyses (Miles, Huberman, & Saldaña, 2014)

facilitated a study of themes across the complete sample. The case examinations drew on guidelines from Stake's (2005) multi-case study methodology to prevent cross-contamination between cases.

### **Sampling, Case Selection Rationale, and Inclusion Criteria**

In selected cases, sampling occurred at two levels: (a) the case itself, and (b) data sources within the case (practitioner, in addition to publication identified in Phase I) (Stake, 1995).

Research can employ an array of more than fifteen different purposeful strategies (e.g., extreme or deviant case (outlier) sampling, homogenous sampling, typical case sampling, critical case sampling, snowball or chain sampling, criterion sampling, theory-based sampling, (dis)confirming cases, convenience sampling, as well as a combinations of these types (Patton, 2002). Although random sampling would allow for greater generalization of research findings and control of selection bias, this dissertation study reflects a criterion-sample (Patton, 2002), which is a non-probability purposive sampling technique, deliberately sought cases per specified criteria for numerous reasons.

First, the richness of the information yielded from the cases was crucial in genuinely understanding the phenomenon (CA practice). Therefore, this study intentionally prioritized more recent cases, as it would be more natural for the practitioners to recall more immediate experiences. As such, the criteria used favored conditions wherewith the cases would yield more abundant information and prioritize required that the participate was significantly involved with the evaluation, and prioritized recent application of CA.

Secondly, the sampling strategy aimed to represent the diverse CA landscape in evaluands and contexts, and thus prioritized a pool of diverse cases to that produced a saturated data landscape, and therefore better illuminated the phenomenon (Patton, 2002). This purposeful sampling was enacted to safeguard the quality, depth, and richness of data extracted from the practitioner interviews.

Therefore, the following **steps** were enacted to identify the cases for Phase II. First, the publications were ranked chronologically, from wherein which I selected cases that were rich to allow for a more in-depth study. Next, cases were reviewed against criteria relevant to the following: (a) case, (b) data sources (e.g., practitioner involvement and publication), and (c) evaluation practice.

**Case selection rationale.** To maximize the diverse depiction of CA applications gleaned from the literature, this study sought to sample a diverse representation of CA. Each case chosen for examination varied on context, nature of evaluand (e.g., public health, higher education), and innovations to maximize variety for a complete attempt to glean all contextual factors influencing the CA process (Phase II). As such, reflect a criterion-sample (Patton, 2002), which is a non-probability purposive sampling technique, deliberately sought cases per specified criteria. The impetus driving this strategy lay in the expectation that the inclusion of all cases (meeting the criterion) produced a saturated data landscape, and therefore better illuminated the phenomenon (Patton, 2002).

**Criteria for evaluation practice.** Publications required an explicit narrative of how CA was used in the evaluation. The criteria for inclusion required that the

publication sufficiently demonstrated a focus of CA in the evaluation design and/or practice, and sufficiently described both elements. An ideal case needed to be information-rich (Patton, 2015), and specific constructs of interest to the study, the evaluation design and practice, and reflected on both elements in consideration of CA. This allowed for a more in-depth understanding of the evaluation process to inform my research. Works reflecting fused evaluation frameworks, such as utilization-focused evaluation (UFE) and CA (Patton, 2012), development evaluation and contribution analysis (Michaud-Létourneau, Gayard, & Pelletier, 2019), or combining Process Tracing (PT) and CA (Befani & Mayne, 2014), were included so long that CA was identified to have fundamentally informed evaluation-related decisions. Inclusion of these variations assisted in a better understanding of complementary approaches to CA and how CA practices may be contingent on a combination of approaches.

**Criteria for practitioners.** This study sought to examine evaluation practice; as such, it is imperative that the data sources were able to produce first-hand accounts of the evaluation process. In other words, the practitioner must have been involved in a capacity wherein he/she was engaged in making evaluation-related decisions and thereby reflected on the Contribution Analysis approach to evaluation. Practitioners who were tangentially involved (e.g., conducted interviews, or responsible for specific tasks) may not be able to speak to how CA was considered throughout the entire evaluation process and thus would not be able to inform the research interests of the study. Consequently, priority was given to the first author or individual identified as the CA practitioner to ensure the most knowledgeable individual informs my research. Table 24 (see page 180) displays the

complete corpus of empirical CA studies found through the scoping review in Phase I, from which the cases used in the multi-case study were sampled.

**Criteria for exclusion.** Articles that did not speak of contribution analysis from the author's first-hand perspective were not considered as the study seeks to gain a deeper understanding of CA *in situ*, and therefore seeks first-hand accounts from those directly involved with the evaluation. Furthermore, although the literature search revealed technical evaluation reports from various sources (e.g., government, consulting groups), these were not considered as examining artifacts produced as a consequence of CA evaluation is not within the scope of the study. If additional data (e.g., technical reports, memos) be offered, it was considered so long that it revealed details relevant to the practitioner's interview responses or supports the ongoing investigation.

The seven cases were selected according to the criteria discussed above; however, in a pursuit to maintain the anonymity of the study participants, the seven cases represented in this study are not identified further than the case descriptions and aggregated participant descriptions and demographics.

### **Response Rate**

Of the 22 emails sent to CA practitioners, two emails bounced back and were not delivered, two referred me to the correct individual to interview, and on the whole, seven practitioners ( $n=7$ ) agreed to participate by scheduling an interview and completing the Qualtrics questionnaire. All seven completed all parts of the study. The data collection commenced with the practitioner interviews on January 20, 2020 and concluded on

January 31, 2020. All seven individuals participated in the interviews as well as the Qualtrics questionnaire.

**Demographics.** The Qualtrics questionnaire queried the participants' demographics, practitioner experience, evaluation training, and self-rated expertise. The first item screened participants to ensure that the representative of the targeted sample (e.g., evaluation practitioners with experience with CA evaluation approach). All seven participants responded affirmatively and continued to the main section of the questionnaire, which queried their typical professional role, and sectors of work, the evaluation training and experience, exposure to CA evaluations, and self-rated evaluation and CA expertise. A summary of the demographic characteristics is provided below and is expanded in Chapter IV (see page 96).

Collectively, questionnaire respondents had over 48 years of experience conducting Contribution Analysis evaluations and had conducted over 29 Contribution Analysis assignments. There was a total of five countries represented by the interviews, including France, England, Canada, the United States, Germany, and their work was representative of four continents (Africa, Asia, North America, and Europe).

Of the seven,  $n=5$  were male and  $n=2$  were female;  $n=1$  identified as African, and  $n=6$  identified as European descent;  $n=3$  were over 50 years of age,  $n=3$  were between 31-39 years old, and  $n=1$  was between 31-39 years old. All participants were all highly educated: the self-evaluation practitioners had a master's degree ( $n=1$ ), and a doctoral degree ( $n=1$ ); the researchers all had at least doctorates ( $n=4$ ), and one individual had a post-doctoral ( $n=1$ ).

### Data Sources and Data Collection Strategies

This section elaborates on Phase II, specifically I recount the data that informed the study and how they were collected. Ensuing is a description of data analysis procedures before concluding with strategies to enhance the trustworthiness of this study. The University of North Carolina at Greensboro’s Institutional Review Board approved all instrumentation and protocol (IRB) before use. Table 10 illustrates the data collection strategies and analysis for each research question.

Table 10

Summary of Data Sources and Analysis to Research Questions

Research Question	Data Sources and Collection Strategies	Analysis
RQ1. What is the current theory of the Contribution Analysis approach to evaluation?	Systematic Review (Phase I) Expert Consultation (Dr. Mayne)	Thematic analysis
RQ2. What does CA look like in practice?	Document Analysis (Phase II) Interviews (Phase II) Qualtrics Questionnaire (Phase II)	Thematic analysis Within & Cross-Case Analysis CA Quality Scoring
RQ3. What conditions and contextual factors (e.g., of the evaluation and program) challenge and facilitate the implementation of CA in the field?	Document Analysis (Phase II) Interviews (Phase II) Qualtrics Questionnaire (Phase II)	Thematic analysis Within & Cross-Case Analysis CA Quality Scoring

Once the practitioners confirmed their participation, a document analysis of the publication was made by applying the case extraction form (see Appendix G) to the

publication. Next, a summary of the case was constructed using the case extraction form and the “CA quality.” These were used to guide the interview with the respective practitioners. The document analysis data categories and practitioner interview items mapped onto the primary research questions.

**Document analysis of publication.** First, an examination of empirical applications of CA proxy peer-reviewed publications was conducted to illuminate details relevant to the case (e.g., related to the evaluand, evaluation process, CA framework, modifications, etc.), and to identify additional themes of interest, and areas needing clarification. Although a case extraction form was used, the document analysis was also generative and emergent, where ideas about the direction for deeper inquiry and patterns began to surface as an iterative coding.

**Case extraction form.** This segment was exploratory in nature, a case extraction form (see Appendix K on page 205) was used to systematically mine each publication. This ensured a systematic way of collecting data from the publications so that cross-case comparisons would be possible. The form queried the following categories related to the (a) evaluand (e.g., type of intervention, target beneficiaries, country), (b) CA-specific themes (e.g., rationale for CA, complementary approaches to CA, general CA process and modifications made, factors that facilitated or inhibited the use of CA in practice, lessons learned, suggested recommendations for (future) practice, evidence/examples to support practice behaviors), and (c) themes relevant to evaluation practice. The case extraction forms served as an additional data source about CA and how it informs

evaluation practice in each case. It was also instrumental for the practitioner interviews as it augmented my knowledge as I prepared for the practitioner interviews.

**Interviews with practitioners.** Next, I conducted semi-structured interviews with a leading member of the evaluation team ranging from 50 minutes to 2 hours, with an average of 72 minutes. The interview explored the evaluation process, how CA informed evaluation practice, and invited the evaluators to reflect on the key themes of interest (e.g., CA theory, 6-step process, modifications, barriers, challenges, and facilitators to practice). Findings from the document analysis were used as a reference point throughout the interview as relevant. As appropriate, I probed specific events or details of the particular evaluation.

As all data from the document analysis was secondary, it was particularly important to triangulate (Creswell, 2009; Stake, 2005) my inferences with the case practitioners. As such, the interview provided an opportunity to member-check findings with the practitioner, inquire about unclear themes, and collect additional details and information relevant to this research. These interviews complemented the document analyses by allowing me to capture the complexity of the evaluators' practice experience as it relates to my research questions and ultimately enrich the understanding of what facilitates/hinders CA practice across contexts.

**Recruitment procedures.** One individual from the original evaluation team, who conducted the evaluation, was contacted via email to participate in the interview. Invitations detailed synopsis and purpose of the study, why they were selected (eligibility criteria for participation), specific case (the evaluand) of interest to the interview,

expected role and contribution, and inquire in their interest to participate in the study. If the individual was interested in participating, they were asked to complete a doodle poll to schedule the interview, review the consent and complete the Qualtrics questionnaire attached in the email. Ensuing, a follow-up email was sent to confirm their interview time and expressing my gratitude their support and participation, along with an attachment of the document analysis summary (to member-check my interpretations) and an attachment of the interview questions (which was sampled from the pool of items shown in Appendix F as appropriate for each case).

**Interview procedures.** Empirical studies of evaluation have demonstrated that interviewing with field experts and practitioners is a way to foster dialogue (Christie, 2003). To facilitate this dialogue, a guiding interview protocol of open-ended, descriptive, and interpretive questions ensured systematic data collection without restricting the natural flow of the conversation. The interviews were audio recorded and submitted to Temi, an automated transcription service (Temi, 2019). The transcriptions were checked for accuracy against the audio and corrected before coding. I also took notes throughout the interview and periodically member-check my notes with the participant.

Each interview began with a reminder of the specific evaluation of interest to the study and asked the practitioners to discuss details and examples that are specific to the case. Next, they were asked to confirm their voluntary participation and consent to be audio-recorded verbally so I can maintain the integrity/fidelity of their responses and for transcription purposes. Lastly, they were asked whether they agree to be identified by

name. Since the interview revolved around the specific CA evaluation for specific evaluand (e.g., project, program, intervention, policy) that was identified through publicly available sources (e.g., academic journal or search engine), the data is identifiable with a specific intervention/policy and therefore potentially associated with the practitioner as well. For all cases (as appropriate), the interview transcript was emailed to member-check responses, and/or provide an opportunity to address identifiable information. If there was a concern, any potentially identifiable information that a practitioner did want to be identifiable was reported anonymously if expressed by the practitioner.

***Administration platform.*** From a practical outlook, utilization of an online platform to host my interviews (e.g., Skype, Zoom) were particularly well suited for my research purpose. CA practitioners were expected to exhibit a gamut of geographical locales similar to that of CA's international presence, which could be a potential barrier for participation. However, the use of an online platform facilitated the participation of geographically dispersed individuals. Collecting data through this method further enhanced the cost-effectiveness of this method as travel is not required, while also assisting in recording-keeping of the study's developments (e.g., editing) and responses, and thereby increasing the overall transparency of the research.

**Data analysis.** As is typical of a multi-case design, there are two main objectives for the analysis. First, a detailed case of each CA evaluation (detailed in the Case Profiles presented on page 105) using all data bounded within it to produce a holistic description of the case. The second objective ensue via a cross-case analysis focused on examining

similarities and differences of themes across cases (detailed throughout the themes reported in Chapter IV).

The multi-case study design encourages the researcher to analyze data from earlier interviews before conducting the later ones (Yin, 2009). Therefore, synthesis of data occurred iteratively as new sources of data were added per case (e.g., publication extraction, practitioner interview) which informed the within-case analysis and as complete case synthesis occurred (informing the cross-case analysis). This process was, therefore, iterative and ongoing throughout the study, and allowed for the most thorough and accurate exemplification of each case, and to identify nuance differences between cases.

In all, Phase I procured 88 empirical cases of CA (see APPENDIX C) from which a subset of seven cases are represented in the multi-case study. As aforementioned, maintaining the anonymity of the practitioners is vital in compliance with the ethical values of this study, and as such, the cases are not identified but are described in generalities and attributes, omitting any identifying information.

Each case file consisted of the document analysis of the publication, interview transcript, and practitioner questionnaire. As such, it was imperative to prepare the data before any analysis could be conducted. This next section details the preparation and coding procedures before discussing the within- and cross-case analyses.

**Organizing data and preparing for analysis.** For each case, I first reviewed the research notes from the document analyses, practitioner interviews, and research logs.

This step was to draw out any noteworthy themes or details that emerged during those data collection events before engaging in a deeper analysis.

All cases in the study were organized in a “Guiding Table” (Pajo, 2018) detailing general details (e.g., case ID, publication details, case details, evaluand details), and contextual characteristics. Moreover, themes similar to the following: CA practice trends, describe its use in application, such as the rationale for the approach, benefits ascertained from funders, and benefits ascertained from practitioners, barriers and challenges to practice, types of included stakeholders, level of stakeholder involvement, and whether overall theoretical claims are upheld in practice. The matrix format of the Guiding Table facilitated a preliminary exploratory cross-case analysis focused on identifying patterns, themes, and anomalies. I made a note of any explicit findings for reference in the cross-case analysis.

Next, I constructed a ‘case file’ for each case consisting of all qualitative data gathered (e.g., document analysis data, the interview transcript, and analytic memos), making it suitable for coding. All qualitative data gathered through the open-response items on the questionnaire was coded and analyzed by theme (Creswell, 2009), and summaries of the responses were presented as aligned with research questions. A validity check was conducted by examining the open-ended responses for relevance to ensure that participants understood the underlying construct of the item. Open-ended responses indicating misunderstanding was documented and excluded from synthesis.

The audio files from each interview were transcribed in their entirety through a professional service, and subsequently thematically coded.

**Coding procedures.** First, I developed a coding schema in alignment with primary domains of interest and frameworks (e.g., Mark, 2008) that have informed my research. Ensuing was a series of coding procedures that informed the within and cross-case analyses, and also resulted in the iterative update of the coding schemes as case study research is based on interpretive and emergent research philosophy.

The data extracted underwent thematic analysis. Finally, the document analysis findings and interview findings were synthesized in meta-matrices to display themes across cases to best describe the degree to which cases demonstrate traits identified in each framework. The coded data was organized in a number of cross-case display matrices were constructed to deepen understanding and facilitate the identification of common patterns. This strategy is consistent with many of the analytical techniques described by Miles and Huberman (1994).

**Within-case analysis and cross-cases analysis.** Two sequences of coding activities were conducted: (a) thematic analysis (Creswell, 2006), and (b) pattern matching (Saldaña, 2013) to conduct the within-case and cross-case analyses.

***Within-case analysis.*** For each “case file,” I examined the codes for larger patterns that emerged across all sources. I also identified the information that is missing or implicit. Once the data from each was analyzed, an analytic memo (Miles et al., 2014) was produced to document my reflections and case-specific findings (Stake, 2005) (e.g., barriers/facilitators, contextual factors, CA process).

***Cross-case analysis.*** Once a holistic account of each case was constructed, I conducted a cross-case analysis using pattern matching (Saldaña, 2013), and display tools

(e.g., tables, matrices). This allowed for the examination of data across cases for patterns of variables or themes that transcend the cases. For each case, coding took place through multiple readings of each case file. A detailed summary was constructed for each case according to categories and codes identified in the coding schema. Additionally, any emergent sub-themes of a case that were not captured by the initial schema were added. As “coding is dynamic” (Given, 2008), the coding schema was continuously revised by adding, subtracting, and refining categories and their properties to ensure codes capture the themes and topics being collected as well as others of interest to the study. This step also identified text fragments and quotations which illustrate the research findings. Analytic memos (Miles et al., 2014) were constructed throughout this process. Overall, this supported me in learning the data before ensuing deeper analyses.

**Thematic coding.** First, I used a process of inductive reasoning to generate descriptive themes from the qualitative data acquired through Phase I. Specifically, I used thematic analysis (Bryman, 2012) as it is not tied to any framework and may be applied in a variety of ways. I develop codes and, as a corollary, superordinate themes (Bryman, 2012). The thematic analysis allowed me to code and categorize the data simultaneously to study commonalities, differences, and relationships (Saldaña, 2013). Code frequencies helped identify the prevalence of themes across cases.

**Pattern matching.** Pattern matching is a “second-cycle” coding method (Saldaña, 2013) that was conducted after coding. This involved looking at grouped cases based on one or more codes that characterize similarities, differences, frequencies, sequences, correspondences, or causation observed in the codes (Saldaña, 2013).

**Meta-matrix construction.** Borrowing from Qualitative Comparative Analysis (QCA), a meta-matrix of conditions (e.g., elements of CA tenets) and outcomes were constructed to effectively score the degree to which the cases effectively reflected the Contribution Analysis tenets. (Conditions: 0 or 1 and outcome: Performance of CA). Table 11 is a generic meta-matrix, while Appendix I illustrates the various ways it was used in the synthesis of the data.

Table 11

Meta-Matrix of Cases on Conditions and CA Score

	Conditions				CA Score
	1	2	3	4	
Case A					
Case B					
...					
Case G					

### **Data Quality: Trustworthiness and Ensuring Rigor**

Interpretive research is based on an idealist temperament that social reality is a product of one's mind. Therefore, it is mind-dependent and mind-constructed (J. K. Smith & Heshusius, 1986). For this reason, social inquiry cannot be "value-free" as we cannot separate ourselves from our interpreted reality. On the same token, a researcher cannot be value-free in his/her interpretations and, therefore, must be especially cognizant in projecting their values/bias on someone else's experience.

Given this backdrop, the quality of qualitative research is judged using standards that consider the social aspect of the inquiry as well as the presence of the researcher within the research context. This next section discusses the techniques used to safeguard the quality of the data collected and thereby confidence of inferences borne using Lincoln and Guba's (1985) trustworthiness criteria (credibility, transferability, dependability, and confirmability). It is important to note that the significance each criterion carries is held to be ultimately contingent on the approach and purpose of the study. In other words, the criterion most relevant and indicative of the quality of a study and its findings depends on how it was conducted and what it was trying to achieve. Although many of the strategies presented by Lincoln and Guba (1985) overlap across criteria, they were discussed in a more linear fashion to facilitate a practical discussion.

### **Credibility**

Parallel to internal validity in quantitative research (Lincoln & Guba, 1985), the credibility criteria denote that the claims made should be based on an appropriate amount of data (saturation; Stake, 2005), and that process of analysis and interpretations should be made transparent (Mertens, 2015). In order to enhance the credibility and validity of findings, I used multiple types of triangulation, which is mostly a repetitious process of gathering data followed by criteria review of what is being said (Stake, 2005). This is especially important in consideration of pluralism and the constructivist paradigm guiding this research.

**Multiple modes and methods.** The use of multiple methods to collect my data (e.g., document analysis, interviews) enhanced the validity of my findings (Taber, 2008).

A limitation of using secondary data (e.g., document analysis of publication in Phase I) is that the scope of research is inherently restricted as I was only able to examine what was published for each case. In other words, if details pertinent to my study were omitted, it simply did not inform my research, development, nor analysis, and subsequently, every finding and conclusion that are drawn. As such, the interview with practitioners presented an opportunity for elaboration and/or clarification, as well as to further delve into details lacking in clarity or explications.

The use of document analysis during the interview facilitated a probe-based interview protocol, an effective technique that uses certain materials (e.g., texts, videos, or other artifacts) to evoke comments and interpretations to give focus and scope to the dialogue (Stake, 2005). I would also argue that probe-based interviewing increased the construct validity of responses in this specific study, as using the publications (which are reflections of the evaluation processes) to guide interviews helped center the conversation on the constructs of interest (e.g., evaluation process, CA tenets, stakeholder engagement).

**Member checking.** As detailed throughout the previous sections, data collected underwent multiple points of member checking. Interviewees were provided with selective interview transcripts for feedback and member-checks (Mertens, 2015). Member checking is a process whereby the data collected are taken back to the participants (Creswell, 2009, p. 191). This process offered an opportunity to provide context and alternative explanations (Patton, 2002), and thereby lessened the possibility

of misinterpretations of participants' responses (Maxwell, 2004) as well as enhance the credibility of my interpretation of the data (Lincoln & Guba, 1985).

**Peer debriefing.** I conducted peer debriefing periodically with a peer of my field of study, and who was already knowledgeable on my topic. The conversations helped me check my developments, limitations, and reflect on my findings.

**Researcher log.** While the researcher log was an archive of technical data helped inform data analysis and contemplation upon the study's practical limitations, it was also a routine activity that helped me develop an awareness of my own personal biases through the practice of critical subjectivity. To be critically subjective is to be cognizant of one's values and influences, and how these influences may guide one's practice (Heron & Reason, 1997). As such, it was imperative to document the development of themes and constructs. Engaging in reflection allowed me to see developments over time, identify potential biases, and include these developments when I engaged in peer debriefing sessions, consulted with the lead CA expert, and interacted with practitioners. Documenting my thoughts provided insight into how my understanding of the research and its landscape progressed.

### **Transferability**

Transferability aligns with external validity, the degree to which findings can be generalized to the population represented by the research sample or applied to other situations. As sample sizes are small in qualitative research, findings borne are not intended to be taken out of the research's context. The burden transferability is on the reader to ascertain the degree of similarity between their study site and the one being

described (Merriam, 2006; Mertens, 2015). As such, the onus falls on the researcher to provide enough supporting information to appropriately inform the reader's judgment of the research's transferability.

As previously mentioned, this study was heavily informed by case study methodology, and used strategies from case study methodology to convey degree of transferability for each case, by developing thick descriptions (Merriam, 1988; Stake, 2005) from the following data sources: (a) protocols developed from Stake's (2005) case-study guidelines (e.g., the nature of the case, historical background, physical setting, another context, other cases through which the case is recognized, and the informant through whom the case can be known) (see Appendix F), (b) case extraction forms, (c) analysis of practitioner interviews, and (d) researcher log.

Furthermore, cross-case analysis strengthened the transferability of the findings (Yin, 2009), as I was able to unearth similarities and differences across cases regarding the phenomenon (the practice of CA). Readers were presented with more opportunities to ascertain findings relevant to a greater range of situations.

### **Dependability**

Dependability is the counterpart to reliability in the postpositivist realm (Guba & Lincoln, 1989), affirming that findings are stable over time (Mertens, 2015). Instead of the traditional meaning of whether the study can be replicated, the many implications within this standard refer to having a credible inquiry process so that inferences borne of the study are trustworthy. The current study explicitly detailed the methodological choices and data collection procedures (Guba & Lincoln, 2001), and used extensive

protocols to document each step of the research process. In all, this provided transparency of my research strategy as well as how findings were informed. Lastly, the inclusion of a CA expert further safeguarded that dependability is established.

**The validity of instruments.** To increase the dependability of the data collection instruments, four explicit strategies were used to ensure that the intended constructs were measured. First, I identified the constructs by reviewing the conceptual literature on CA. Next, I consulted with Dr. Mayne on the constructs related to the CA and the ToC. Ensuing was a cognitive interview with an evaluation practitioner who also has licensure in instrument development to ensure that the items were being interpreted as intended. And finally, all instrument protocols were piloted once approved by IRB. In all, these strategies assisted in ensuring the quality of items (e.g., clarity of construct, sequence of items) while getting feedback from the subject-matter expert and a representative of the target population.

### **Confirmability**

The last criterion for qualitative research is confirmability, which parallels objectivity of the postpositivist paradigm (Guba & Lincoln, 1989). While the concept of objectivity denotes that the researcher's influence is minimized, confirmability designates that the data and their interpretation are grounded and can be traced to their source, along with the logic of how the interpretation was constructed (Merriam, 1988). The triangulation of data helped constrict the researcher's bias on the interpretation. The corpus of data (e.g., publications, case summaries, interview transcripts, researcher log) collected throughout this research facilitated a confirmability audit as pieces of data

woven together in interpretation were traceable to its original source, and the logic behind the claims and inferences made were transparent.

### **Chapter Summary**

In this chapter, I described the methodology underlying this dissertation study. I detailed the two-phase multi-method design that guided this qualitative study. The two-phase design enacted by this study allowed for an in-depth examination of the theory and practice of the Contribution Analysis approach to evaluation via a systematic review of the literature and then an in-depth multi-case study to investigate how practice manifests.

First, through Phase I conducted a systematic review of the literature of conceptual articles and empirical articles separately. The review of the conceptual theory of the Contribution Analysis identified fundamental tenets of CA, which informed the interview questions in Phase II. An interview with CA's chief architect provided clarification on these principles as well as clarification of key terms. The review of the empirical literature allowed for a high-level review of trends and themes of practice and produced propositions of practice to examine in Phase II. This phase also yielded the sampling pool of cases for Phase II.

Phase II fulfilled a key objective of the overall study and investigated the CA-practice phenomenon via a multi-case study using secondary data (e.g., journal article) and corroborating it with a primary source (e.g., practitioner interviews). Specifically, this study used the CA six-step framework to guide an in-depth study of how CA manifests in practice, specifically the evaluation process, CA tenets (e.g., development of the ToC), stakeholder engagement. It also provided an opportunity to ascertain whether

the trends, themes, and propositions discovered in Phase I transcended other cases, and to probe the conditions surrounding it. Although a preliminary list of themes was examined, this phase borrowed heavily from the case-study approach and thus was emergent and generative. As such, it also allowed for the emergence of new themes.

## **CHAPTER IV**

### **FINDINGS**

This chapter presents findings from the conceptual publications on Contribution Analysis (Phase I) and multiple-case study (Phase II). To begin, I present an analysis of the current theory of CA by summarizing the historical developments, position on the evaluation tree, ontology and epistemology, and identification of three tenets of practice (RQ1). Next is an examination of the seven CA cases, including a description of each case, settings, and evaluation approach, and details relevant to practitioners experience (RQ2). Following is an analysis of how the Contribution Analysis is practice (RQ2a), and an examination of modifications and reinforcements made to the evaluation process (RQ2b). Finally, a synthesis of the conditions and contextual that impact CA is practiced is presented (RQ3). The roadmap for the findings are: (a) research questions, (b) the data sources, (c) the overarching finding of presented, and (d) an expansion of the key themes.

#### **Research Question 1: What is the Current Theory of the Contribution Analysis Evaluation Approach?**

This research question summarizes the findings obtained through an in-depth review of the conceptual literature of CA, including publications by John Mayne (APPENDIX A), as well as publications by other (APPENDIX B), and an expert consultation with Dr. Mayne, which provided clarification on the fundamental concepts of CA as well as more nuanced operationalization and interpretation of key terms. The analysis of CA's conceptual and empirical developments through 2019, was conducted to

examine the axiological nature of CA to identify how it has developed, what were the key changes, elaborations, and specifications that emerged in each.

### **Overarching Findings**

- Finding 1: CA is currently entering its fourth generation and becoming increasingly refined.
- Finding 2: CA is situated on the “Methods” branch on the Evaluation Theory Tree.
- Finding 3: CA is guided by realist ontology and relativist epistemology.
- Finding 4: Three key tenets that guide CA are: (a) theory of change mechanism, (b) correspondence between the sphere of influence and magnitude of the claim made, and the (c) plausibility and credibility of the contribution story.

### **Finding 1: CA is Currently Entering its Fourth Generation and Becoming Increasingly Refined**

Contribution Analysis (CA) was born out of the need for the consideration of context, conditions of complexity, and multiple interdependent interacting variables (Mayne, 2008, as cited in Patton, 2012). CA, as we know it in the realm of evaluation, stems from financial management and business analysis with a different meaning (Dybdal et al., 2010), with CA’s debut in performance measurement (Dybdal et al., 2010).

Over time, CA has been adopted in various contexts of program evaluation and has since gathered additional proponents in Canada (Dybdal et al., 2010) and the European Union (Delahais & Toulemonde, 2012; Leeuw, 2012; Lemire, Nielsen, &

Dybal, 2012). There have been some recent international applications of this method, particularly in Scotland, notably within the National Health Scotland organization (NHS) (Craig, 2013; Wimbush et al., 2012) and Scottish Government (Scottish Government Social Research, 2012).

Insofar, there have been three generations of CA (1990, 2001, 2011), and there is emerging evidence that Contribution Analysis is currently undergoing a process of reinvention (e.g., Brousselle & Buregeya, 2018; Budhwani & McDavid, 2017; Mayne, 2015). The unremitting discourse surrounding CA attests to the cyclical scrutiny and developments that it has been subject to since its introduction by Mayne in 1999. Each rendition reflects the maturing and refining of CA as an evaluation approach, and the dialectic relationship between evaluation theory and practice. The academic literature reflects a healthy discourse between theory-based evaluations and CA and has contributed to the development of the guiding paradigms of this approach (see Finding 2 and 3 in this section).

### **Finding 2: CA is Situated on the “Methods” Branch of the “Evaluation Theory Tree”**

To situate CA, among other evaluation approaches, I use Alkin and Christie’s (2004) “Evaluation Theory Tree” and propose that CA is positioned on the ‘methods’ branch. The “Evaluation Theory Tree” categorizes evaluation theories based on their primary emphasis on one of four branches, namely, *use, methods, valuing, or social justice* (Alkin & Christie, 2011; Mertens & Wilson, 2012). CA’s iterative 6-step process allows for the constructed contribution claims to be examined and re-examined and provides for a rigorous inquiry method. Although an argument could be made to situate

CA as a ‘use’ approach along with participatory approaches (Alkin & Christie, 2004, p. 15) theorists classify CA as a ‘methods’ oriented approach as it is vested in demonstrating rigor of methodology and transparency in how evaluation findings are established.

Participatory approaches such as Cousin’s pragmatic-participatory evaluation (P-PE), Patton’s utilization-focused evaluation (UFE), and Stufflebeam’s context, input, process, product (CIPP) fall directly in line with the *use* branch as they seek utilization of the evaluation and findings by involving stakeholders in the evaluation and uphold pluralistic perspectives. Focused on creating value in the overall use of the evaluation and its findings and consideration of who will use the information (Alkin & Christie, 2004, p. 14).

Despite its intense use of participants throughout the 6-step process (e.g., Wimbush et al., 2012) and acknowledgment of multiple realities to construct the ToC (aligning with the subjective ontology assumptions), I counter that CA ultimately relies on the iterative testing of the ToC, implying that some perspectives woven into the ToC will not be upheld and will be dismissed in testing (aligning with the realism epistemology). CA assumes a verifiable reality similar to ‘Realist Evaluation’ (Dybdal et al., 2010).

### **Finding 3: CA is Guided by Realist Ontology and Relativist Epistemology**

Though Contribution Analysis has undergone three generations of changes, CA is largely underdeveloped as an approach in that there are still vital theoretical components,

such as epistemology and ontology, that remain unaddressed (Budhwani & McDavid, 2017; Dybdal et al., 2010).

The earlier versions of CA seem align with the successionist model of causation (Mayne, 2012), and focused on establishing causal links in the program theory to rule out rival hypotheses through positivist-based designs (Dybal et al., 2012; Mayne, 1999). More recent clarifications by Mayne (2015) clearly point to a generative model of causality, which embody the intertwined “relationships between context, mechanisms, and outcomes through situationally sensitive theories of change (Befani & Mayne, 2014; Mayne, 2015a; Pawson, 2007)” (Budhwani & McDavid, 2017, p. 16).

Brousselle and Buregeya (2018) have contributed to the development of the paradigms underlying CA to identify a realist ontology, based on judgmental rationality, relative epistemology, and value that is prescriptive. The guiding philosophy underpinning CA is centered on critical realism (Brousselle & Buregeya, 2018). The fourth generation of CA is more aligned with generative causal models and focuses, as CA seeks to assess underlying assumptions and risks behind causal links, and identifies other key competitor, and is situationally sensitive ToC (Befani & Mayne, 2014; Budhwani & McDavid, 2017; Mayne, 2015a; Pawson, 2007). Generative causal models are focused on the causal package and not on the counterfactual to prove causation (see “Design Approaches to Causation” discussion on page 22). Meaning, a causal pathway established through a generative model or framework (e.g., Contribution Analysis) is meant to represent the chain of impacts and contextual factors that are likely to influence the production of the effects (Brousselle & Buregeya, 2018; Mayne, 2012b).

According to Brousselle and Buregeya (2018), the first two steps in the CA serve formative purposes, while Steps 3-6 satisfy a summative inquiry. Furthermore, Step 3 has been identified to be intentional about evaluation use, while steps 4-6 encourage incorporation of evidence to strengthen the program theory.

Table 12

## Paradigms Underpinning CA

Knowledge (Critical Realism)				Use			
Ontology	Rationality	Epistemology	Valuing	CA Steps	Intent	Intended Types	Means to Encourage Use
Realism	Judgmental Rationality	Relativism	Prescriptive	1	Formative	Conceptual Use	Involve different stakeholders
				2			
				3	Summative	Instrumental (Steps 4-6)	Integrative theoretical approach
				4			
				5			
				6			

**Finding 4: Three Key Tenets That Guide CA are: (a) Theory of Change Mechanism, (b) Correspondence between the Sphere of Influence and Magnitude of the Claim Made, and (c) the Plausibility and Credibility of the Contribution Story**

A review of the conceptual literature has identified the key tenets of CA to include: (a) the theory of change mechanism (on page 35), (b) correspondence between the sphere of influence and magnitude of the claim made (see “Spheres of Influence and Magnitude of Contribution Claim” on page 41), and the (c) plausibility and credibility of the contribution story (see “CA as a Validation Framework” on page 46). These concepts have already been covered adequately in this thesis. From both the conceptual and empirical literature, the importance of the theory of change is prominent. The concept of

the theory of change is undoubtedly the underpinning of theory-based evaluations, and CA is no exception (e.g., Mayne, 2012; Chen, 2005; Coryn et al., 2011). The theory of change mechanism has been discussed at length throughout this dissertation (see “Theory of Change (ToC)” on page 35).

The correspondence between the sphere of influence and magnitude of claim is an indicator of the strength and robustness of the linkages in the ToC (the “Spheres of Influence and Magnitude of Contribution Claim” on page 41). Theory-based evaluation approaches provide the opportunity to examine beyond the intervention’s immediate effects (e.g., direct control) to more distal effects like impact. Therefore, I identified the sphere of influence claimed, and the depth of CA investigated to be an indicator of the CA quality. The plausibility and credibility of the contribution story relate back to the validation framework (see “CA as a Validation Framework” on 46). In all, these three tenets are CA (and effectively the quality of the evaluation approach) were used to produce the CA scoring rubric (see Appendix I) by which the seven cases were analyzed.

### **Research Question 2: What Does CA Look Like in Practice?**

The following describes each case represented in the multi-case study, by detailing the evaluand, the purpose of the evaluation, and evaluation approaches used. To be able to specify theories in the literature, it is important to understand who is using the evaluation approach. As such, RQ2 sought to gather information on who the CA practitioners were.

## Overarching Findings and Themes

- Finding 1: CA evaluations are traditionally impact evaluations, examining the contribution of a variety of interventions.
- Finding 2: Participants self-identified as researchers or evaluation practitioners, and had experience conducting various types of evaluations, across a diverse range of contexts, across the globe.
- Finding 3: Participants saw their roles in the CA assignments as: methodological/ technical (evaluation) advisor, critical analyst, facilitator of local change, achiever working with program manager, educator of clients, knowledge broker, and resource of stakeholders.

The following data sources: (a) practitioner interviews, and (b) practitioner questionnaires, (c) interviews were used to gain insight into practitioner attributes. The roadmap for Research Question 2 findings is detailed by the: overarching findings, case profiles of the sample, and themes.

### **Finding 1: CA Evaluation are Traditionally Impact Evaluations, Examining the Contribution of a Variety of Interventions**

- Theme A: Cases reflect mostly impact evaluations ( $n=6$ ).
- Theme B: CA is used in the evaluations to examine the effects of a variety of entities (research organization, health impact assessment, evaluation policy, monetary funds, government policy, professional development program, and advocacy initiative).

Table 13 summarizes the profiles of the seven evaluations represented in the multiple case study. For the purposes of providing enough context while upholding confidentiality of participants identity, descriptive yet generic pseudonyms of the cases are provided.

Table 13

## Summary of Case Profiles

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>
<b>Content areas</b>	Sustainable forest management	Urban revitalizations project	National policy	Financial	Drug and alcohol	Higher education	Child nutrition, Advocacy
<b>Object of evaluation</b>	Research organization	Health impact assessment (HiA)	Evaluation Policy	Funds awards to enterprises	Policy	Professional development program for teachers	Advocacy initiative
<b>Type of evaluation</b>	Impact evaluation	Effectiveness evaluation	Meta-evaluation	Impact evaluation	Policy/strategy	Program evaluation	Advocacy evaluation
<b>Purpose</b>	Impact of research activities on country across many sectors	Impact of HiA on urban revitalization	Impact of policy, quality of evaluations produced	Impact of funds on society; Conditions of effectiveness,	Impact of strategy on reducing harms from drug and alcohol	Effectiveness of intervention on developing knowledge	Impact of advocacy initiative on the policy environment
<b>Evaluation approach</b>	CA and PT	CA only	Meta-evaluation; and CA and PT	CA via comparative case study design	CA only	CA in mixed methods design	DE and CA

**Case A: Sustainable forest management (SFM).** Case A is an impact evaluation of a major research body's potential contribution to the changes observed in the country of operation. The object of the evaluation was to assess the research organization's impact on the observed changes in the country and how. The organization in this case is a large international entity that is focused on research activities on sustainable forest management (SFM) and livelihoods and has since seen the sustainable forest management arena change dramatically, especially regarding general awareness, legal framework, and timber practices. The evaluation used a combination of Process Tracing (PT) with CA to better understand the underlying logic of their casual claim. The strength of evidence was assessed using PT, followed by Bayesian probability logic.

**Case B: Urban revitalization.** Case B is an evaluation of the effect of a health impact assessment (HiA) on urban revitalization (e.g., road infrastructure, parks, green spaces, and residential housing). An HiA is a tool used to estimate the potential impact on the health of non-health-related initiatives. Specifically, it is "a set of procedures, methods, and tools by which a policy, program or project may be judged as to its potential effects on the health of a population, and the distribution of those effects within the population" (European Centre for Health Policy [ECHP], WHO Regional Office for Europe, 1999, p. 4). The impact pathway embodied by this case engages decision-makers in the process of adopting, implementing, and refining the intervention based on recommendations procured from the HiA. CA was conducted using stakeholder interviews, documents observations, and images, CA was able to link the activities implemented in the field from the HiA to health outcomes.

**Case C: Evaluation policy reform.** Case C is a meta-evaluation of a national mandate for reform of public policies at the state level in a European country. The objective to look at the impact of the mandate and evaluate the quality of evaluations completed under the reform (n=80) between 2012-2017. The design was two-phased: first, a meta-evaluation of the 80 evaluations was conducted using interviews, document analysis, administrative decision-making. Next, CA and PT were used to build in the confidence of contribution declared from the evaluation findings.

**Case D: Financial support for large enterprises.** Case D conducted an ex-post policy evaluation of an EU policy across eight member states. The objective of the evaluation was to outline the policy implications by assessing the basis for the implementation of and evidence of effectiveness financial awards to large enterprises. The policy sought effectively influence the behavior of large enterprises through the financial support, to directly affect economic and social benefits directly (e.g., demand for jobs, increased in product levels, etc.), and indirectly (e.g., direct job creation improved local transport infrastructure or local spending of wages by the employee). The evaluation assessed the extent to which resources were used, its effectiveness on the socioeconomic impact of the large enterprises, and the factors contributing to the success or failures of the interventions. The evaluation produced information about the circumstances and conditions under which the offering financial support large enterprises are most meaningful and justified.

**Case E: Drug and alcohol strategy.** Case E reflects an evaluation of a government policy that sought to reduce harms related to drugs and alcohol. The policy

targeted a set of established negative consequences of alcohol and other drug use and was implemented nationwide in a European country. The evaluation queried, “how and to what extent has an implementation of the strategy in the country contributed to reducing drug and alcohol-related harms.” The evaluation sought to investigate the reduction of harms across specific areas (e.g., prevention, provide support for substance misusers, provide support for families). The intervention was modeled after a similar drug and alcohol strategy policy that employed CA in their evaluation. As such, the evaluation was commissioned with a requirement to employ CA. The ToC was retrospectively constructed by the evaluators and then brought to stakeholders for revisions and subsequently submitted to an advisory group. Sources of information for the ToC were the proposal document, the modeled intervention documents,

**Case F: Professional development for teachers.** Case F was an evaluation of the professional development program for the teacher in developing their leadership skills. The evaluation sought to evaluate the degree to which the intervention was effective in developing knowledge in the target sample. The intervention followed a panel design, and the data reflected the repeated measures of the participants (e.g., pre and post-tests). The cause-and-effect was already identified in the project proposal, and therefore was effectively the ToC used in the CA. The practitioner then developed the evaluation plan around the identified program theory in the evaluation proposal.

**Case G: Child nutrition advocacy initiative.** Case G evaluated the impact of an advocacy initiative on policy change over its 9-year term. This evaluation sought to examine the extent to which policy objectives were achieved and identify the key drivers.

This evaluation employed Developmental Evaluation and CA sequentially. DE was used as a framework to engage stakeholders and document the extent to which policy objectives were met in each county, and CA was used to assess the degree to which the policy contributed to the observed changes. The evaluation sought to examine how the activities carried out within the advocacy initiative resulted in the policy environments that were more supportive of nutrition.

**Finding 2: CA Practitioners Tend to be Discipline-specific Researchers, Trained Evaluators, and Had Experience Conducting Various Types of Evaluations, across a Diverse Range of Contexts, across the Globe**

- Theme A: Participants reported experienced with policy evaluation ( $n=5$ ) and program evaluations ( $n=4$ ) the most, followed by needs assessment ( $n=3$ ), performance ( $n=2$ ), process ( $n=2$ ), effectiveness ( $n=1$ ), research project ( $n=1$ ).
- Theme B: All participants were highly educated (minimum of master's degree). Self-identified evaluation practitioners ( $n=2$ ) were formally trained in evaluation, whereas self-identified researchers ( $n=5$ ) tended to be subject-matter experts in each of their respective fields and conducted evaluations to assess interventions within their specific knowledge domain.
- Theme C: The contexts in which the participants practice are diverse: health care ( $n=3$ ), public health ( $n=3$ ), public policy & admin. ( $n=3$ ), regional & urban development ( $n=3$ ).
- Theme D: Participants tend to work in Europe ( $n=3$ ), Asia ( $n=1$ ), Africa ( $n=1$ ), North America ( $n=1$ ), or globally ( $n=1$ ).

- Theme E: Self-identified evaluation practitioners ( $n=2$ ) tended to be more experienced than self-identified researchers in overall evaluation (as measured by number of years) and in CA (total assignments, number of years).
- Theme F: Practitioners published empirically-based peer-reviewed CA literature and five ( $n=5$ ) published CA literature based on secondary data.

Table 14 provides an overview of the evaluation characteristics of the respondents. Participant's self-identified as practitioner/evaluators ( $n=2$ ) or researchers ( $n=5$ ). Practitioners tended to be formally trained in evaluation, whereas researchers were subject-matter experts in each of their respective fields and conducted evaluations to assess interventions within their specific knowledge domain. All participants were highly educated (minimum of master's degree). Participants reported doing policy evaluation ( $n=5$ ) and program evaluations ( $n=4$ ) the most, followed by needs assessment ( $n=3$ ), performance ( $n=2$ ), process ( $n=2$ ), effectiveness ( $n=1$ ), research project ( $n=1$ ).

Table 14

## Demographics by Role Assumed in CA Assignment

Variable	Description	Evaluator/ Practitioner	Researcher/ Consultant
<b>Gender</b>	Female	1	1
	Male	1	4
<b>Race/Ethnicity</b>	African		1
	European/White	2	4
<b>Age Range</b>	31-39		1
	40-49	1	2
	50+	1	2

Table 14

Cont.

Variable	Description	Evaluator/ Practitioner	Researcher/ Consultant
<b>Highest Degree Achieved</b>	Master's	1	
	Doctoral	1	4
	Post-Doctoral		1
<b>Country of Residence</b>	Canada		3
	France	1	
	Germany		1
	Wales		1
	United States	1	
<b>Region of Work</b>	Anywhere		1
	Europe	1	2
	Asia		1
	Africa		1
	North America	1	

The context of practice is diverse: health care ( $n=3$ ), public health ( $n=3$ ), public policy & admin. ( $n=3$ ), regional & urban development ( $n=3$ ). The participants' work settings are academic ( $n=3$ ), or private business/ consulting ( $n=3$ ), and one individual worked in multiple settings (academic and non-profit). Participants worked in Europe ( $n=3$ ), Asia ( $n=1$ ), Africa ( $n=1$ ), North America ( $n=1$ ), or globally ( $n=1$ ).

**Sample demographics and characteristics.** An overall summary of the sample is presented in Table 18. Collectively, questionnaire respondents had over 48 years of experience conducting Contribution Analysis evaluations and had conducted over 29 Contribution Analysis assignments. There was a total of five countries represented by the

interviews, including France, England, Canada, the United States, Germany, and their work was representative of four continents (Africa, Asia, North America, and Europe).

Of the seven, five were male ( $n=5$ ), and two were female ( $n=2$ ) (Table 14); one identified as African ( $n=1$ ), and six identified as European descent ( $n=6$ ); three were over 50 years of age ( $n=3$ ), three were between 40-49 years old ( $n=3$ ), and one was between 31-39 years old ( $n=1$ ). Self-identified evaluation practitioners had a master's degree ( $n=1$ ) and a doctoral degree ( $n=1$ ); researchers all had at least doctorates ( $n=4$ ), and one individual had a post-doctoral ( $n=1$ ).

**CA evaluation experience.** Overall, self-identified evaluation practitioners had more experience (both in CA assignments and total years of experience) than self-identify researchers. Four participants had 4-6 years of CA experience, one had 7-9 years of CA experience, and two had 10 or more years of CA experience. The individuals who identified themselves as practitioners ( $n=2$ ), each had 10 or more years of CA experience and completed more than five CA assignments. Of the five self-identified researchers, three individuals completed three CA assignments, and two completed five or more.

Table 15 summarizes the participants' experience with CA: number of years of experience, number of CA assignments, self-rated expertise, capacities performed in CA assignments, and representation in publications.

Table 15

## CA Experience by Role Assumed in CA Assignment

Variable	Description	Evaluator/ Practitioner	Researcher/ Consultant
Number of years of CA experience	4-6 years		4
	7-9 years		1
	10 or more years	2	
Number of CA Assignments in Role	3		3
	5 or more	2	2
Self-rated level of Expertise in CA	Proficient		1
	Advanced	1	2
	Expert	1	2
Capacities of Participation in CA Assignment	Developing the intervention/research		3
	Designing the evaluation	2	5
	Implementing the evaluation	2	4
	Reporting the results	2	5
	Managing/supervising the evaluation	2	3
Publications (Empirically Based)	Academic (e.g., conference presentations, working paper)	2	5
	Peer-reviewed work (Article, book chapter)	2	5
	Reports & memos	1	3
	Evaluation Reports	1	
	Other	1	
Publications (Secondary-Data)	Developed a research/conceptual framework	1	2
	Academic (e.g., conference presentation, working paper)	0	2
	Evaluation reports/memos	1	1
	Meta-synthesis / Literature review	0	2
	Peer-reviewed work (article, book chapter)	0	2
	No, I have not published	1	1
Felt equipped for CA assignment?	Yes	2	3
	Somewhat		2
	No		

**Sample self-rated CA ability and participation capacities.** Both of self-identified evaluation practitioners and three of the self-identified researchers felt well-equipped to fulfill the requirements in their CA assignment, and two of the self-identified

researchers said they felt “somewhat equipped” to fulfill the requirements of their role in the CA assignment.

The self-identified evaluation practitioners rated their CA ability as advanced ( $n=1$ ), and expert ( $n=1$ ). The self-identified researchers responded as in proficient ( $n=1$ ), advanced ( $n=2$ ), and expert ( $n=2$ ). In terms of which capacities were fulfilled by the participants in the CA evaluation, the two self-identified evaluation practitioners both identified all phases of the evaluation, while the self-identified researchers were almost all involved in developing the intervention ( $n=3$ ) as well as components of the evaluation.

**CA publications.** All participants in the sample were published works that were empirically based: academic ( $n=7$ ), journal article or book chapter ( $n=7$ ), reports and/or memo ( $n=4$ ), evaluation report ( $n=1$ ), and other ( $n=1$ ). Five of seven individuals published literature on ToC based on secondary data: development a research or conceptual framework ( $n=3$ ), academic item ( $n=2$ ), evaluation report/memo ( $n=2$ ), meta-synthesis or literature review ( $n=2$ ), peer-reviewed literature ( $n=2$ ).

**General evaluation experience.** Table 16 displays the multi-case study sample general evaluation experience and evaluation training. The self-identified evaluators had 15-20 years and more than 21 years of evaluation experience. Consequently, their experience reflected in the self-rated expertise of advanced ( $n=1$ ) and expert ( $n=1$ ) and received their evaluation training from on-the-job evaluation training ( $n=2$ ) and through a doctoral degree ( $n=1$ ). The self-identified researchers had 6-10 year of evaluation experience ( $n=3$ ), and 11-15 years ( $n=2$ ), resulting in self-rated evaluation expertise of proficient ( $n=1$ ), advanced ( $n=2$ ), and expert ( $n=2$ ), and receive their evaluating training

in various ways: on-the-job training ( $n=3$ ); doctoral training ( $n=3$ ); informal training (conference, webinars) ( $n=4$ ); training or certification from a professional organization ( $n=3$ ); undergraduate-level course ( $n=2$ ); master's degree in evaluation ( $n=2$ ).

Table 16

## General Evaluation Experience and Training

Variable	Description	Evaluator/ Practitioner	Researcher/ Consultant
Number of years of evaluation experience	6-10 years		3
	11-15		2
	15-20	1	
	More than 21	1	
Self-rated level of expertise in Evaluation	Proficient		1
	Advanced	1	2
	Expert	1	2
Type of Evaluation Training Received	On-the-job training	2	3
	Doctoral Degree	1	3
	Informal training (e.g., conference, webinars)		4
	Training or certification from a professional organization		3
	Undergraduate-level courses		2
	Graduate-level courses		2
	Master's degree		2
	Undergraduate degree		1

**Typical evaluation assignments.** Table 17 illustrates the characteristic of typical evaluation assignments for the sample of practitioners. The self-identified evaluation practitioners typically conducted policy ( $n=1$ ), program ( $n=2$ ), needs-assessments ( $n=1$ ), performance ( $n=1$ ), or process evaluations ( $n=1$ ) varying in scale (combination of large

and small evaluations) ( $n=2$ ). The self-identified researchers typically conducted policy ( $n=4$ ), program ( $n=2$ ), needs-assessments ( $n=2$ ), performance ( $n=1$ ), or process evaluations ( $n=1$ ), effectiveness ( $n=1$ ), and research evaluations ( $n=1$ ), in large ( $n=1$ ), small ( $n=1$ ). Combination of large- and small-scale evaluations ( $n=3$ ).

Table 17

## Characteristics of Typical Evaluation Assignment

Variable	Description	Evaluator/ Practitioner	Researcher/ Consultant
<b>Primary occupation (&gt;80% of the time)</b>	Researcher		4
	External Evaluator	2	
	Other		1
<b>Types of evaluations normally conducted</b>	Policy	1	4
	Program	2	2
	Needs assessment	1	2
	Performance	1	1
	Process	1	1
	Effectiveness		1
	Research Project		1
<b>Scale of Evaluations in Typical Assignment</b>	Combination	2	3
	Large-Scale (ToC, large sample size, large multi-site, multi-organization)		1
	Small-scale (ToC, small sample size, single site, small multi-site)		1
<b>Primary Work Setting</b>	College / University		3
	Private Organization	1	1
	Self-employed	1	
	Other		1
<b>Sectors of Evaluation Work</b>	Health Care		3
	Public Health		3
	Public Policy & Admin.	1	2
	Regional & Urban Development	1	2
	Environmental Initiatives	1	1
	Education (K-12)	2	
	Food & Natural Resources		2
	Higher Education	1	1
	Science, Tech., Engineering, Math (STEM)	1	1
	Social Work	1	1

Table 17

Cont.

Variable	Description	Evaluator/ Practitioner	Researcher/ Consultant
<b>Sectors of Evaluation Work (cont.)</b>	Workforce & Economic Development	1	1
	Agriculture		1
	Health & Human Services		1
	Transport & Mobility	1	
	Community Programs		1
	Nutrition		1

**Primary occupation.** The participants' primary occupations were either as external evaluators ( $n=2$ ), and researchers ( $n=4$ ), and one individual split their time equally between research and academia ( $n=1$ ). In terms of primary work setting, one of the self-identified evaluation practitioners was self-employed, and the other was employed by a private organization. Three of the self-identified researchers worked on a university campus, one individual worked for a private institution, and one split their time equally between college/university setting and working for the federal government.

**Sector of work.** The self-identified evaluation practitioners tend to work in the following sectors: public policy and administration ( $n=1$ ), regional & urban development ( $n=1$ ), environmental initiatives ( $n=1$ ), education (K-12) ( $n=2$ ), higher education ( $n=1$ ), science, technology, engineering, and math (stem) ( $n=1$ ), social work ( $n=1$ ), workforce & economic development ( $n=1$ ), environmental initiatives ( $n=1$ ), education (k-12) ( $n=2$ ), and transport and mobility ( $n=1$ ).

The self-identified researchers tend to work in the sectors: health care ( $n=3$ ), public health ( $n=3$ ), public policy & administration ( $n=2$ ), regional & urban development

(*n*=2), food & natural resources (*n*=2), environmental initiatives (*n*=1), higher education (*n*=1), science, technology, engineering, math (stem) (*n*=1), social work (*n*=1), workforce & economic development (*n*=1), agriculture (*n*=1), health & human services (*n*=1), community programs (*n*=1), and nutrition (*n*=1).

**Finding 3: Participants Saw Their Roles in the CA Assignments as: Methodological/ Technical (Evaluation) Advisor (*n*=3), Critical Analyst (*n*=2), Facilitator of Local Change (*n*=2), Achiever Working with Program Manager (*n*=1), Educator of Clients (*n*=1), Knowledge Broker (*n*=10), and Resource of Stakeholders (*n*=1)**

- Theme A: Self-identified evaluation practitioners would lean towards classifying themselves as technical expertise or methodological experts.
- Theme B: Self-identified researchers saw their roles as subject-matter expertise rather than evaluation or methods experts.

As can be seen from Table 18, the CA practitioners saw their roles as: methodological/ technical (evaluation) advisor (*n*=3), critical analyst (*n*=2), facilitator of local change (*n*=2), achiever working with program manager (*n*=1), educator of clients (*n*=1), knowledge broker (*n*=10), and resource of stakeholders (*n*=1). One individual is not included in this description as they were a commissioner for the specific case. Two additional options were provided “liaisons between stakeholders, benefactors, and donors” and “judge of the program” were not identified by any of the participants. In one case (G), the evaluator found their roles to shift dramatically to becoming an actor in the intervention.

Table 18

## Evaluator Role in CA Assignment

<b>ROLES</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>
Methodological/Technical (evaluation) advisor		X	X	X			
Critical analyst			X		X		
Facilitator of local change						X	X
Achiever working with the program manager							X
Educator of clients			X				
Knowledge broker							X
Resource of stakeholders						X	

There was quite a distinction between self-identified evaluation practitioners and self-identified researchers in how they viewed their roles in the CA assignment. While self-identified evaluation practitioners would lean towards classifying themselves as technical expertise or methodological experts, most of the self-identified researchers emphasized that they provided subject-matter expertise rather than evaluation experts.

Furthermore, one of the interviewees commented:

So, the first thing I would say is that we could not be considered ‘methods expert.’ This was the first time we were using it [ ToC], so you can’t make that claim at all. Yet, I suspect we won the tender though because we had prior exposure to ToC. But that’s not the same as method expert . . . Where we [the evaluators] came in was because we had a very broad set of drug and alcohol research knowledge and experience between us, so if you wanted to label us, we were ‘topic experts.’ And added to that, we were topic exerts with a footprint in the locality.

I would describe the role is as a critical analyst. We are here to make a critical judgment. And we are the tellers, and I think we use this a lot in our report and their story. I think we are the tellers of the performance story.

A unique role (Case G) was in the advocacy initiative evaluation, which employed a Developmental Evaluation followed by CA. The evaluator reported a shift in their role as the evaluation endured from an evaluator to an actor:

I found myself often involved, and so I was not completely an external person. I was one key actor because I became the nutrition advisor of the head of the ministry of health in the department of nutrition.

### **Research Question 2a: How Do Practitioners Implement CA?**

The CA process is described in terms of the theory of change, specifically, how it was developed and verified. The evaluation approaches used in these evaluations are reported in the section on adaptations and reinforcing as often reinforced CA objectives. The following data sources were used to produce these findings: (a) practitioner interviews, and (b) analysis of publication.

### **Overarching Findings and Themes**

- Finding 1: The ToC is typically developed retrospectively by the evaluators.
- Finding 2: The interpretation of the ToC was seldom conducted with the stakeholders or intended users other than to confirm or be a source of data.
- Finding 3: The postulated theory was typical grounded (based on scientific studies) ( $n=3$ ), based on pre-existing program theories ( $n=2$ ) or based on the designer's beliefs.

- Finding 4: There are misunderstanding of key CA terminology amongst practitioners.
- Finding 5: There is a misunderstanding of when to use CA – “Analysis of Contribution” versus CA.

**Finding 1: The ToC is Typically Developed Retrospectively by the Evaluators and Seldom Involved Stakeholders in the Process Other Than to Confirm the Evaluator’s Postulated ToC**

- All but one (G) of the cases developed the theory of change retrospectively and therefore employed evaluation approaches that would facilitate this type of investigation.
- In all cases, the ToC was developed by the evaluator or evaluation team and then submitted for confirmation (if that occurred).

Table 19

Type of ToC Developed, Types of Postulated ToC, and Stakeholder Engagement

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>
<b>ToC</b>	Retro	Retro	Retro	Retro	Retro	Retro	In real time
<b>Type of Postulated ToC</b>	Pre-existing	grounded (based on scientific studies)	Grounded (scientifically based), then tested		Pre-existing	Based on the designer’s beliefs	Grounded, based on new empirical research
<b>Stakeholder involvement</b>	to verify ToC; as data sources	to verify ToC; as data sources	to verify ToC	to verify ToC	to verify ToC		as participants

Note. Retro=Retrospective

Six of the seven cases developed the ToC retrospectively, and one case developed the steps of the ToC as it emerged (Case G). The ToC was seldom built with

stakeholders. Most cases constructed the ToC and then involved stakeholders varying in degrees of engagement (e.g., submission for approval, formal event to negotiate the ToC). In all cases, the stakeholders involved functioned as data sources rather than in a participator form. Case G did, however, also engage stakeholders.

Two cases (E and G) did not have a priori intervention products on which to build the impact pathway. Case E did not have clear intervention outcomes, and G did not have clear identified outputs. For Case E, the evaluation of the drug and alcohol policy strategy, the expected outcomes were not immediately clear. Consequently, the research team effectively reversed engineered the policy to develop the intended ToC:

In a way, we had to try and establish what we thought was the aim of the strategy in cause and effect. This is where it got complex, or really tangled rather than tangible. We realized that, it was trying to do lots and lots of things. Moreover, there was this really interesting moment in that when we [developed] the specific [e.g., links in the causal chain, conditions], and we chose to go and hear from as many stakeholders as we could, partly because we did have them earlier on.

**Finding 2: The Interpretation of the ToC was Seldom Conducted with the Stakeholders or Intended Users**

Instead, once the ToC was interpreted with the collected evidence, the evaluators would then submit it to the intended users/commissioners, and either has a discussion or receive recommendations. In all cases, the stakeholders involved functioned as data sources rather than in collaborators in the evaluation ( $n=7$ ), and in one case (G), they were also engaged in an ( $n=1$ ) participatory manner, wherein which they participated in building the ToC ( $n=1$ ). Almost all cases used stakeholders to verify the ToC ( $n=6$ ) by member-checking or confirming the postulated ToC.

**Finding 3: The Postulated ToC is Typically Founded on Scientific Theory**

Case B established the ToC of the intervention by identified impact pathways from activities to outputs to a sequence of outcomes to impacts and developed ToC for three sub-projects of the revitalization project.

Case C built the ToC based on scientific theory and then submitted it to stakeholders. Stakeholders were involved in determining the steps required for the outcomes to occur; in other words, pieces specific to program implementation.

Case F (Professional Development) did not create a ToC as it used the program theory identified in the program proposal and was not modified. As such, the ToC was based on the program designer's opinions. This case used the ToC to guide implementation and ensure implementation was as proposed and to ensure that the project goals were met.

The child nutrition initiative (Case G) developed the ToC as it emerged, which is impressive given that CA needs to have a ToC and expected result/outcome. However, the nature of the intervention was such that many outcomes (both immediate and middle effects/impact) were not known. As such, as the ToC progressively emerged, the links/assumptions were tested.

**Finding 4: There are Misunderstandings of Key CA Terminology among Practitioners**

The theory of change (ToC) has been called a lot of things: a roadmap, blueprint, theory of action, engine of change, logic model. Beyond the initial conceptualization, there is little consensus on how ToC is defined (Stein & Valters, 2012). The sample of practitioners in this study reflected a similar notion. According to Mayne and the

evaluation scholars, there is a significant difference between a logic model and a theory of change. Whereas a logic model is a linear configuration of the inputs, activities, outputs, outcomes, and impacts; a ToC should delineate the pathways of how the inputs are transformed to the expected outcomes, and more importantly, it should identify assumptions and risks to the demonstrated logic. Though the confusion of the terminology was apparent in both the publications and interviews, almost all participants ( $n=6$ ) did understand and emphasize the significance of assumptions and/or conditions in the postulated ToC. One practitioner had limited understanding of the ToC, assumptions, and verifying the contribution story. This individual, does not engage in academic dialogue (e.g., conferences, etc.):

The “testing of assumptions”, actually, I call data-driven research or something like that. But it sounds like the same thing . . . Well, and let’s face it, people who are professors, they got to get articles published, so they are going to make up some new thing. That’s some old thing, but it has a different name and maybe there’s like 1 little difference in it, you know, so that is always going on.

**Finding 5: There is a Misunderstanding of When to Use CA— “Analysis of Contribution” Versus Contribution Analysis**

All but one case thoroughly articulated how the linkages of the ToC were verified the production of the contribution claim (e.g., external influence, challenges, triangulation). This is in alignment with the evaluation purpose and type, which was to determine the effectiveness of the intervention on improving knowledge and skills. As such, the evaluation considered data that were direct outputs of the intervention in order to declare whether the intervention resulted in significant change in the participants’ knowledge. This also aligns with the role in which the evaluator described for

themselves, which was focused on implementation, and “to make sure that the program goals are met.” Though the intervention did intend for impact beyond the immediate effects, this was beyond the scope of the evaluation.

For one of the cases, it was unclear what level of contribution claim was sought. The interview queried the spheres of influence investigated. The evaluation was reported to be a complicated/complex theory evaluation (interview).

**ES:** Would you say that what you were assessing was up to what, what you have there as the shorter-term term outcomes?

**Participant:** For the article? Yes.

**ES:** And for the contribution analysis results that you have in the article, it is reflecting up to their short-term outcomes?

**Participant:** Well, it is kind of short-medium. Because I am looking at did they achieve the short term [change in knowledge], but then we looked at whether the teachers advanced to more leadership positions.”

Per the definitions provided by John Mayne, Case F is not a CA evaluation but an analysis of contribution and is effectively a black box evaluation.

As discussed in Chapter II, “Spheres of Influence and Magnitude of Contribution Claim” (page 41), investigating direct intervention outputs is a minimalist contribution claim and is effectively a ‘black-box evaluation.’ This evaluation enquiry is fundamentally an “analysis of contribution” and not a Contribution Analysis evaluation as intended by John Mayne.

Specifically, the sphere of influence examined by the evaluation sought to assess under the program's direct control. Referring to Figure 7, the evaluation sought to investigate the effect of the intervention up until the immediate outcomes.

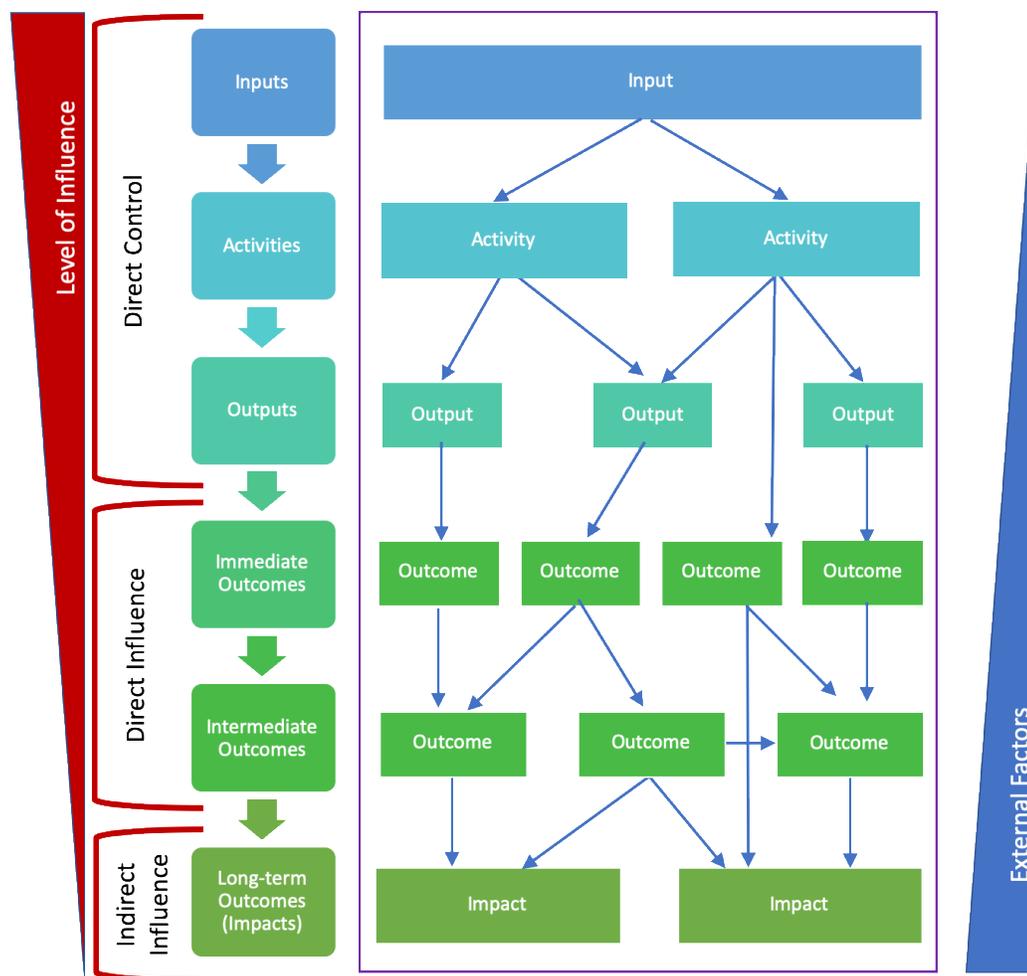


Figure 7. Spheres of Influence and Program Components.

### Research Question 2b: What Modifications and Reinforcements are Made in the Design of CA Evaluations?

- The following data sources were used to produce these findings: (a) practitioner interviews, and (b) analysis of publications. The evaluation

approaches used in these evaluations are reported in the section on adaptations and reinforcing as this often reinforced the CA tenets.

### Overview of Findings

- Finding 1: The adaptations made to the modification made on the evaluation design include: Process Tracing and Developmental Evaluation.
- Finding 2: The Relevant Explanation Finder (REF), evidence table, and Chronological mapping of Events are strategies used to reinforce the CA tenets.

Table 20

Overview of Case by Purpose, Evaluation Approach, and Development of ToC

<b>Case</b>	<b>Purpose</b>	<b>Evaluation Approach</b>	<b>ToC Development</b>
A	Impact of research activities on country across many sectors	CA & then PT	Retrospective/ex-post
B	Impact of HiA on urban revitalization	CA only	Retrospective/ex-post
C	Impact of policy, quality of evaluations produced	meta-evaluation & then CA and PT	Retrospective/ex-post
D	Impact of funds on society; Conditions of effectiveness,	CA via comparative case study design	Retrospective/ex-post
E	Impact of strategy on reducing harms from drug and alcohol	CA only	Retrospective/ex-post
F	Effectiveness of intervention on developing knowledge	mixed method design	Retrospective/ex-post
G	Impact of advocacy initiative on the policy environment	DE & then CA	Emergent (development)

**Finding 1: The Adaptations Made to the Modification Made on the Evaluation Design Include: Process Tracing and Developmental Evaluation**

- Theme A: Two of the cases used CA in combination with Process Tracing.
- Theme A: Most cases employed CA in conjunction with another evaluation research design (meta-evaluation, Development Evaluation).
- Theme A: All but one (G) of the cases developed the theory of change retrospectively and therefore employed evaluation approaches that would facilitate this type of investigation.

All but one case sought to investigate the impact of the intervention, while Case F sought to establish effectiveness. The evaluand types ranged from research organizations, health impact assessment tools, evaluation policies, monetary awards, national government policies, professional development programs, and advocacy initiatives. All but case G developed the theory of change retrospectively and therefore employed evaluation approaches that would facilitate this type of investigation. As such, two cases used solely CA in the evaluation, one case (C) conducted a meta-evaluation on their entire sample before accompanying the analysis with Process Tracing to verify the linkages of the impact pathway using 8 instrumental case and case F conducted a mixed-methods design that was highly reliant on pre and post tests on a sample of 12. As case G aimed to capture and develop the ToC in real time, it reflects a developmental evaluation (DE) aimed to document the intervention's effects and applied CA towards the end of the intervention.

**Theme A: Process tracing.** The use of process tracing has been extensively discussed in the CA literature (Befani & Mayne, 2014).

Adaptions and innovations in operationalizing CA, and I expect more to come! I think the link with process tracing will prove useful. (Mayne, email communication, Jan 28, 2019)

Two cases in the sample (A and C) used process tracing (PT) “to verify the causal claims.” In particular, Case C used CA and PT to test eight cases against 10 empirical tests to validate the developed theory of change. In the sustainable forest management case (A) the ToC was modified over the three phases: an inception phase, which involved the research organization, was focused on building the ToC and define assumption to test, an exploratory phase to refine assumptions and identify areas that would require additional clarification (based on 14 interviews and initial document analysis), and lastly, a deepening phase which producing 3 case studies based on 51 interviews querying each step of the ToC.

**Theme B: Development evaluation.** In the child nutrition initiative (G), CA was coupled with DE. First, DE facilitated the collection of data over an extended period and to examine the postulated ToC and facilitated triangulation through data sources and methods. DE is a good approach to support ToC (in consideration of triangulation of evidence). Accounted for external influences by examining their respective influence on specific elements of the ToC. Using a DE evaluation approach proves useful in strengthening the contribution claims.

What was particular and very good for the evaluation is we documented so much of the innovation that we have a lot of data to also use after to strengthen our contribution claims and enrich all the pieces of analysis that we were doing.

While the cases that are identified as singularly using CA, it should be noted that they all mention conducting rigorous literature reviews to identify factors of influence before engaging in a CA.

**Finding 2: The Relevant Explanation Finder (REF), Evidence Table, and Chronological Mapping of Events are Strategies Used to Reinforce the CA Tenets**

Three specific strategies were identified in the cases that reinforce CA tenets: the Relevant Explanation Framework (REF), an evidence table, and “Chronology of Events.”

**Theme A: Relevant explanation finder (REF) as a reinforcement of CA tenets.** Two cases identified the REF in their CA evaluation. In the HiA evaluation (Case B), the REF framework was used to analyze the postulated ToC, determining degree of influence for the HiA, and the influence of alternative and external factors in consideration of the empirical evidence gathered.

**Theme B: Evidence table as a reinforcement of CA tenets.** The sustainable management forest intervention (Case A) used an evidence table to classify the data into three classes:

1. Not necessary—the intervening bodies contributed, but the evidence is relevant to the causal package of interest in the research evaluation, and the observed changes would be similar without the intervening body.
2. Necessary—the intervening body did contribute, and their contribution was necessary (in conjunction with other factors) to generate the observed changes.
3. Sufficient—that is, the intervening body caused the changes on their own, and no other factors were necessary.

This allowed for the consideration of the degree to which the evidence was part of the causal package for which the effects were observed.

**Theme C: “Chronology of events” as a reinforcement of CA tenets.** Case G attempted to use the REF framework but found it difficult to apply because of the multi-country nature of the evaluation. Consequently, the evaluation adopted a “Chronology of Events” strategy to map the temporality of the events and outcomes observed.

We made a “Chronology of Events,” it is not in CA, but I have continued developing this when I use CA. A “Chronology of Events” is a prospective way of documenting the intervention. For example, for every month we would document what is happening, then we could do after related to different phases of the policy process, “were they moving forward?”

**Research Question 3: What Conditions and Contextual Factors (e.g., of the Evaluation and Program) Challenge and Facilitate the Implementation of CA in the Field?**

**Overview of Findings**

- Finding 1: Five dimensions of context were identified to have contributed to the complexity of the evaluation: program theory complexity, sectoral, geographic, temporal, and effect-object.
- Finding 2: Retrospective ToC require that data sources and archives are accessible and accurate.

**Finding 1: Five Dimensions of Context Were Identified to Have Contributed to the Complexity of the Evaluation: Program Theory Complexity, Sectoral, Geographic, Temporal, and Effect-Object**

Thematic analysis of the data revealed dimensions that seemed to contribute to various degrees of the evaluation’s complexity. The following dimensions were identified

to have contributed to the complexity of the evaluation: program theory complexity ( $n=5$ ), Sector complexity ( $n=4$ ), Geographic complexity ( $n=3$ ), Temporal complexity ( $n=2$ ), and Effects-object complexity ( $n=3$ ).

Table 21

## Type of Complexity Per Case

	A	B	C	D	E	F	G
<b>Program theory</b>	X		X	X	X		X
<b>Sector</b>		X		X	X		X
<b>Geographic</b>	X			X			X
<b>Temporal</b>	X			X			
<b>Effects-object</b>	X			X			X

Note. A: Evaluation complexity= Complicated/complex, Scope=Provincial; B: Evaluation complexity=Complex, Scope=Regional; C: Evaluation complexity=Complicated, Scope=National; D: Evaluation complexity=Complicated, Scope=Multi-country; E: Evaluation complexity=Complex, Scope=National; F: Evaluation complexity=Complicated/complex, Scope=Project; G: Evaluation complexity=Exponentially complex, Scope=Multi-country.

**Program theory complexity** is founded on the intervention itself is based on a complicated theory of change ( $n=5$ ). One practitioner (case C) reported that not fully understanding the mechanisms of the object you are evaluating adds complications, as it is more challenging to map the impact pathway when you do not know what the pivot points are.

**Sectoral complexity** is the context and/or subject addressed by the intervention. An evaluation that embodies sectoral complexity will affect more aspects of society ( $n=4$ ).

Exponential complexity. Because you do it in one country. So, it's complex already because I feel like all public health or nutrition interventions involve a lot

of facets and so it's never simple and complicated. This analogy with simple, complicated, and complex. And then it is multiple exponents that I see.

The policy covered you know, a multiplicity of kind of just issues, in different stages in the cycle of drug and alcohol use and all sorts of things like that. I could argue that everything, but the kitchen sink was inside the strategy.

Another participant explains the context within which the evaluand takes place affects the tangibility of the effects sought to be captured:

One of the difficulties in applying ToC in a not necessarily fully autonomous national policy. What's one of the problems if the, if the bubble itself is not self-contained? I mean, that's why we do ToC.

Other participant reported constraints in conducted impact evaluation of policies:

There is something different when you are applying this to policy evaluation as opposed to single program evaluation. Moreover, somehow, you are not doing things in partnership in the same way., partly because what is up for debate is already fixed. If it is a national policy, it has been fixed by politicians and government. The playing field is already set. There is no negotiation in that broad playing field, there is no negotiation. You know, your negotiation is best is, what evidence you seek, how you interpret that evidence, how are you going find additional evidence we'll do from, but the framework is very fixed.

**Geographical complexity** refers to the scale of the intervention is large (ToC), country-wide, multiple countries, or across a global region. European Union ( $n=3$ ).

**Temporal complexity** is characterized by a prolonged period over which the impact is expected to have occurred ( $n=2$ ). However, as one of the evaluations (G) was conducted in real-time using Development Evaluation, the practitioner reported that an advocacy evaluation tends to be more responsive as the effect it is trying to have is occurring in real-time.

The effects-object complexity by the intended effect in which the change is expected to occur itself abstract and, therefore, difficult to observe (ToC, across multiple sectors or dimensions) (ToC, impact on policy environment) ( $n=3$ ). One practitioner (G) reported a complex evaluation as it examined the points of influence in the policy change cycle. The object (policy cycle) of the intended effects is nebulous and complicated to examine, and as a continuously moving context, it presented an additional challenge for the evaluation.

What we realized is that you cannot say, 'there is a policy change, or there is not.' So, the complexity of this evaluation was to be able to assess the kind of movement or influence throughout the policy cycle and the stages within it. Otherwise, there is not much you can say. It's so much more than that.

### **Finding 2: Retrospective ToC Require Accessible and Accurate Data Records**

Evaluations that sought to establish distal effects of a multi-year policy or strategy relied more on secondary data (reports, logs, memos) to verify the ToC as it was more challenging to track stakeholders who may have participated in the intervention as more time passed. However, it was often seen ( $n = 4$ ; A, C, D, E) that the evaluators did attempt to trace/verify the assumption made by the ToC by tracking down program participants or implementers. In two cases (A and C), the ToC was rigorously verified using process tracing. It should be noted that those two cases have a common evaluator, and therefore the similarity of methods should not be considered a coincidence but intentional.

Case D reports the difficult in synthesizing the change mechanisms across multiple units of analysis.

It was to understand how these interventions were designed. So, we basically started bottom-up in each of the eight countries and programs to have a look at the interventions. What was the input? What was the project about, or what kind of projects were funded? What was the output? What was the expected outcome and impact? We did that for a total of 40 - 50 ToCs. And that was driven by a set of bottom-up work.

## **CHAPTER V**

### **DISCUSSION AND CONCLUSIONS**

#### **Study Overview**

This qualitative study explored the theoretical translation and practice of contribution analysis, the adaptations and innovations made to CA's distinct 6-step process, and the conditions, influences, and factors that may facilitate or impede CA practice. Following a sequential exploratory multi-method design, I first investigated the conceptual theory of CA as well as empirical practice by reviewing the extent of the literature. Conceptual literature informed the development of the CA tenets as well as the constructs contained within the data collection instruments, while the empirical literature identified preliminary themes of practice and procured a pool of empirical cases for consideration in the next study. Next, I conducted a multi-case study that focused on investigating CA in practice to characterize adaptations and innovations made to CA's distinct 6-step process; the conditions, influences, and factors that facilitate or impede CA practice. This study sought to address the following research questions:

1. What is the current theory of the Contribution Analysis approach to evaluation?
  - a) What are the tenets and values of CA?
2. What does CA look like in practice?
  - a) How do practitioners implement the approach?

- b) What adaptations, adjustments, or reinforcements are made to the 6-step process?
- 3. What conditions and contextual factors (e.g., of the evaluation and program) challenge and facilitate the implementation of CA in the field?

### **Overview of Findings**

The following gives a sweeping overview of the (a) dissertation findings and themes by research questions (RQ1, RQ2, RQ2a, RQ2b, RQ3), (b) an argument for the role of Contribution Analysis amongst other evaluation approaches, and lastly (c) methodological concerns regarding current practices of Research on Evaluation (RoE).

#### **Research Question 1: Conceptual Development of Theory**

CA has gone through a considerable evolution since its emergence in 1999. Despite the emergence of a fourth generation of the approach, the conceptual developments may have outpaced practice. The literature on the application of CA in various contexts is still very limited, leaving practitioners lacking accessible and relevant content for their specific practice (e.g., policy advocacy initiatives) (Michaud-Létourneau et al., 2019). This study identified a very similar breadth of practice-based applications in terms of contexts and evaluand function.

#### **Research Question 2: CA in Practice**

As was extensively discussed in Chapter II, there is a significant degree of variation in how CA is conceptualized and operationalized in practice. These findings of the dissertation study support the synthesis in Chapter II.

**Breadth in context.** The findings highlight the diversity of applications and, therefore, contexts in which CA is applied. The evaluand types ranged from research organizations, health impact assessment tools, evaluation policies, monetary awards, national government policies, professional development programs, and advocacy initiatives. The content areas were wide-ranging from sustainable forest management to urban revitalizations, from financial to drug and alcohol reduction and prevention policy, and from higher education programs to maternal and child nutrition.

**Methods in practice.** Every evaluation needs to be tailored respective of its evaluand, purpose, and context; as Mayne himself has emphasized: “there is no prescribed way of evaluating specific intervention” (as cited in Nkwake, 2015, foreword). Contribution Analysis (CA) is a standard analytical framework that can be adapted as required (Budhwani & McDavid, 2017; Mayne, 2012). CA has not been linked to any specific evaluation method. Instead, Mayne (2001, 2012) strongly advocates for the combination of a range of methods of both quantitative and qualitative methods. Some have come as far as to say that CA is methodologically neutral as it can be “readily combined with other methods” (Kane et al. 2017, p. 15). Theory Based Evaluations are ‘neutral’ and “utilize all methods that might be suitable, without privileging any one of them, and without depending on them” (Stame, 2004, p. 63). The discussion has evolved from mixing inquiry methods to mixing CA with other evaluation approaches. For instance, Patton discusses CA in the context of Utilization Focused Evaluation (UFE) and supplies an example (cite). Befani and Mayne (2014) discuss the hypothetical combination of CA and Process Tracing (PT).

The cases reflected in Phase II illustrate a variety of methods used to satisfy the inquiries, which a variety of combinations (e.g., meta-evaluation, process tracing, developmental evaluation; mixed methods). While the cases that are identified as singularly using CA, it should be noted that all mentioned conducting rigorous literature reviews to identify factors of influence prior to engaging in a CA. The findings from this study are consistent with the literature, as CA is primarily seen in ex-post evaluations (Dybdal et al., 2010), and smaller numbers for formative purposes, though a developmental approach and CA have otherwise not been observed (e.g. case G). The research designs (case study designs, time series, mixed methods), and data collection methods observed in the cases are also consistent with the literature (e.g., desk reviews, rigorous literature review, interviews, observations, and surveys).

### **Research Question 3: Conditions and Contextual Factors that Affect Practice**

The dissertation findings illustrate the complexity of the contexts in which evaluations take place. The study identified contextual factors related to the program theory, sector of practice, geographic scope, temporal interval, and the effect-object. Furthermore, the way in which the theory of change (ToCs are built (e.g., retrospective) are ultimately dependent on factors that are completed outside of the evaluator and intervention's control as the amount of data and quality of data would have already taken place.

### **Reflections on Key Take-Aways**

The following section is a reflection on the key-take away from this dissertation study, specifically I contemplate whether CA is sufficient on its own as an evaluation approach, and discuss methodological concerns when conducting research on evaluation.

#### **Contribution Analysis as an Evaluation Approach**

The cases included in this study used CA in conjunction with another evaluation approach/analytical strategy. It seems that for certain contexts and purposes, CA by itself may not be sufficient as it is not amenable for the comparison of causal packages across contexts nor cases. Therefore, approaches like Process Tracing (PT) or Bayesian modeling, or Qualitative Comparative Analysis (QCA) are used to facilitate direct comparisons of effects/influences of impact pathways from case to case. In other words, CA is sufficient for building a reasonable argument as to whether the intervention has contributed to observed effects, but it may not be sufficient for categorically identifying how much more effective an intervention is on a case by case basis.

#### **Methodological Concerns in Research on Evaluation (RoE)**

As canvassed throughout this document, the CA evaluation literature is rife with remarks on the lack of examples to exhibit just how the theory of the CA works in the field. Contrary to the prevailing discourse, however, the present application revealed a plethora of empirical CA applications that have been increasing over the past decade. A significant discovery from this study is that the practice of CA appears to be increasing and may be more widespread than is often supposed. In total, 84 publications meeting strict criteria for inclusion were identified in the published literature during the period

2007 to early 2020. Interestingly, over half (55%;  $n=44$ ) of all CA evaluation cases were in non-evaluation journals. This suggests that CA is practiced by discipline-specific researchers in specific contexts and settings rather than evaluation.

Consequently, this dissertation has pinpointed a crucial weakness in conventional research on evaluation (RoE) methodologies. More specifically, the search protocol employed in this study expanded the scope of the literature review to go beyond the purview of evaluation practice journals, to include international, discipline-specific, online publications (e.g., *Australian Centre for International Agricultural Research*; *Drugs: Education, Prevention and Policy*; *Journal of Applied Gerontology*; *Maternal & Child Nutrition*). By expanding the scope of publications considered (Phase I), the search was able to identify an unprecedented number of CA cases, many of which would not be identified under traditional inclusion criteria (e.g., publications within evaluation journals) as they located peripherally of the traditional evaluation practitioner and scholar research scope. Case and point, as an illustration of the seven practitioners, interviewed, five did not primarily identify as evaluation practitioners. Consequently, the locales of their publications were justly in journals that were relevant to their disciplines and not in evaluation journals.

This discovery indicates that when conducting RoE, specifically when investigating evaluation practice, one must be mindful of the degree to which the approach is also desirable or applicable outside the immediate purview of evaluation discipline. This finding is in line with the transdisciplinary nature of evaluation. As aforementioned, evaluation is the methodological toolbox for every other discipline

(Scriven, 2016). As such, we need to reflect that in our work and reflect the multi- and transdisciplinary nature of the discipline.

### **Improving Evaluation Theory**

Studying evaluation practice can help us understand the conditions under which a specific approach or design would not work, “failure is often explained by factor outside the evaluation” (Furubo & Stame, 2018, p. 4). A driving force of this dissertation was to identify conditions upon which CA as an evaluation approach or design is not ideal or simply would not work. The conditions identified as “not ideal” are those in which compromise the principles fundamental to CA. These conditions were recognized using the CA tenets and quality markers. Ultimately, these conditions were identified in cases where the compromise on the tenets undermined the purpose of using CA (e.g., validation process) or a theory-based evaluation approach. As such, this dissertation research sought to identify these dimensions under which conditions of “malpractice” are high and to identify which conditions need further exploration or explanation.

This second phase also contributes directly to the Research on Evaluation (RoE) scholarly body by illustrating the translations of an evaluation approach from theory to practice. Specifically, this study elucidates how the theory is understood in the field at large, what key pillars are missing in ensuring uniform understanding, how practitioners understand the concepts, and finally how these practitioners apply these concepts from their given understandings. The study showed where there are gaps in practitioner understanding of CA and provides insight into where theorists may consider illuminating,

operationalizing, or providing conditions of which specific CA steps and pillars are relevant.

A key finding of this study was that CA practitioners may not be formally trained in evaluations. Findings from this research illustrated that five of seven CA evaluators self-identify as discipline-specific researchers, who conduct evaluations as an auxiliary part of their responsibilities. Many studies across the field have demonstrated the nebulousness of evaluation terminology. Therefore, we cannot expect that individuals from other disciplines will have enough clarity through self-directed learning to be able to navigate through the confusion and nuances of the evaluation concepts when evaluators themselves struggle with the conceptual understanding.

A prime example of the lack of consensus of terminology is the theory of change. The ToC has been called a lot of things: a roadmap, blueprint, theory of action, engine of change, logic model. Beyond the initial conceptualization, there is little consensus on how ToC is defined (Stein & Valters, 2012). The sample of practitioners in this study also reflected a lack of agreement in interpretation and use of theory of change.

According to Mayne and other evaluation scholars, there is a significant distinction between a logic model and a theory of change. Whereas a logic model is a linear configuration of the inputs, activities, outputs, outcomes, and impacts; a ToC should delineate the pathways of how the inputs are transformed to the expected outcomes, and more importantly, ToCs should identify assumptions and risks to the demonstrated logic.

Though the confusion of the terminology was apparent in both the publications and interviews, almost all participants ( $n=6$ ) did understand and emphasize the significance of assumptions and/or conditions in the postulated ToC. One practitioner had limited understanding of the ToC, assumptions, and verifying the contribution story.

### **Recommendations for Practice**

#### **Develop corresponding guidelines and expectations for theory and practice.**

While the implementation of Contribution Analysis was predominately satisfactory in the sample represented in this dissertation study, the meaning of key terms and concepts was occasionally the source of confusion and misunderstandings. To minimize the confusion and provide greater conceptual clarity for practitioners, theorists should provide detailed operational definitions of terminology and concepts, describe how they relate, and the implications for practice. Ideally, nuances between context would be identified, as well as potential pitfalls and misuse to prevent poor evaluation practice. In the same vein, practitioners should echo this habit and report terms of reference (ToR), for both the evaluation results and the evaluation process on how the findings came to be.

Furthermore, as terminology is understood differently depending on training and context of application, a good practice is to have dialogue with evaluation team and stakeholders to define the operational meanings of key CA and theory-based evaluation concepts. This will not only ensure that everyone has a uniform understanding, but that expectations of what is required to fulfill an inquiry is recognized.

Moreover, practitioners should include implications for practice, limitations of CA, and conditions under which the generative causation logic is threatened and the

evaluation cannot be expected to draw valid conclusions. In other words, practitioners need to be mindful of the operating context, resources, evaluation/evaluand attributes, and working timeline, and delineate the degree to which CA can inform on the intervention's contribution to the observed effects. Acknowledging limitations and threats to the evaluation design upfront will safeguard against potentially drawing invalid and unsubstantiated conclusions.

**Simple contribution stories are more effective.** Recommendations that resounded amongst many practitioners through Phase II was related to the presentation of the theory of change and contribution story to stakeholders. Multiple practitioners reported unsuccessful delivery of evaluation findings if the theory of change and more importantly, contribution story was elaborate and detailed. Stakeholders were more receptive to simple and concise contributions stories whereas a longer and more detailed contribution story were often not fully digested and/or comprehended, and therefore presented a barrier to the utility of the evaluation findings.

### **Contributions of the Study**

This dissertation study has two key contributions. First and foremost, it contributes to the empirical literature and evaluation practice and specific domain of theory-based evaluation and Contribution Analysis. As such, the findings from this study contribute to the empirical body on evaluation practice, as well as identify implications and recommendations for further refinement of CA as an evaluation approach. Moreover, findings from this study inform the development of CA contingency theories, which identify conditions under which practices are effective. Approaches incorporating

contingency theories are considered to be the strongest types of evaluation approaches as they are buttressed by empirical knowledge of the practice (Shadish et al., 1991; N. L. Smith, 1993). Therefore, findings produced aim to be of value to strengthen the CA's theoretical foundation. In all, this dissertation illuminates how theory can best be translated into practice. For practitioners, understanding the extent to which operational practices reflect CA's values and tenets providing insight on where and how they might improve practice. It also informs theorists on how they might sharpen or modify the theory or address its limitations.

Secondly, this dissertation contributed to the methodology of research on evaluation. As detailed in chapter III, the methodological process enacted by this study was appropriate to the research objectives of investigating the dialectical nature between theory and practice. It allowed for an investigation of CA on a conceptual level, to validate the representation of the theories' construct in research (e.g., consultation with expert), and subsequently apply the vetted constructs against practice to better to understand how the theoretical notions translate into real-work tangibles.

### **Strengths of the Study**

As detailed throughout this chapter, this dissertation study endured rigorous qualitative methods that called upon multiple channels of information to triangulate evidence for both Phase I (Systematic Review) and Phase II (Multiple-Case Study). The use of multiple data collection methods (i.e., literature review, interviews, questionnaire) enhanced the validity of the research findings (Taber, 2008). Interviewees were provided with selective interview transcripts for feedback and verification to "member check" the

collected data. Furthermore, the inclusion of cognitive interviews and subject matter experts (both in content and survey construction) safeguarded against invalid inferences being drawn from this research. This process lessened the possibility of the researcher's misinterpretations of participants' answers (Maxwell, 2005) as well as enhance the credibility of the researcher's interpretation of the data (Lincoln & Guba, 1985).

The nature of naturalistic inquiry is such that data collection and preliminary analysis occur concurrently and continuously inform the research landscape. Each additional case contributed to new developments, either enhancing or diverging earlier findings. Because I did not seek to build a generalizable theory across all cases, I therefore was not seeking to confirm nor refute findings with evidence from each ensuing case. Instead, my pursuit was to expand the evidence amassed to reflect the diversity of cases and practice to best inform the development of contingency theories.

### **Limitations**

While some interesting findings are represented in this study, the findings should be interpreted with caution for numerous reasons. First, the cases in this study represented a small portion ( $n=7$ ) of the total number of empirical cases identified ( $N=88$ ). Secondly, in consideration of the nature of the sampling and inclusion criteria, applications that were not peer-reviewed were excluded. Consequently, the current sample may differ in unknown ways from the grey literature, conference presentations, and evaluation reports.

Finally, the retrospective nature of the analysis depended on the quality and availability of data. Although the triangulation information across sources mitigated the

potential of inaccurate interpretations, the nature of the study relied on publication and practitioner's recollection. If information was not reported in the publication and the practitioner was not able to recall (or recalled incorrectly) then this information was not represented or inaccurately represented.

**Retrospective cases and recall bias.** While the interview allowed for an in-depth examination of CA practice for each case and provided insights into the complexity of the intervention and its ToC, the evaluation process, and CA specific process from a (the) primary source, an overarching limitation remains the use of retrospective assessments of events. Interviews required that the practitioners make retrospective assessments of contextual findings using their memory (and notes, case publication) to discuss the evaluation process. From a temporal sense, the more recent a case, the more accurate the extracted information is expected to have been. Consequently, this study was also as reliant on data sources that were not necessarily intended for the purposes of the study (e.g., case publications written for different purposes) (Morton, 2015). Taken together, the reliance on retrospective cases may result in incomplete and/or potentially inaccurate data and preclude any comparative analysis over time except to reflect how it was historically recorded or recalled.

In conducting case studies, a researcher may unknowingly allude to his/her opinions. I am not an expert on the topic of Contribution Analysis, but it has been my primary research interest since 2016. As such, I have undertaken numerous research endeavors focused on CA, exploring its relationship to evaluation use, complementary aspects to collaborative approaches to evaluation, overcoming the causality dilemma, and

how CA can be used to build validity into evaluation findings. There are many precautions that should be taken in interpreting the findings from this research. In the small representation of the cases to the methodological limitations of case study research, findings are not intended to be taken out of the research's context.

### **Future Directions**

I argue for greater precision in the use of terminology and concepts implicated in the CA evaluation process; development of guidelines for identifying the level of attribution that is feasible given the attributes of the evaluand, evaluation, resources, timeline, context, etc. This will allow for more robust curation of evidence to determine the degree to which the interventions contribution to the observed effects, and also then produce some sort of structure for the meta-evaluation of CA practice, and provide a more substantial basis for improving the theory of practice. For factors implicated in implementation processes; development of guidelines for selecting a research design and study plans that account for practical constructs and allow for the study of mechanisms; psychometrically strong and pragmatic measures of mechanisms; and more robust curation of evidence for knowledge transfer and use.

As the findings have shown, there are numerous potential combinations of methods and evaluation approaches with which CA can align and is an incredibly attractive prospect. Since CA is entirely reliant on the quality of evidence, future research would recommend investigating what kinds of evidence or knowledge might be most effective in supporting CA evaluations, which conditions lend themselves well to CA as

dominant evaluation theory, and as a secondary strategy, how CA might enhance the evidence produced.

### **Concluding Thoughts**

This study responded to explicit requests to probe deeper into CA (e.g., Budhwani & McDavid, 2017; Delahais & Toulemonde, 2012; Dybdal et al., 2010; Lemire, 2010) as well as a more general call for a systematic examination of evaluation cases in order to generate “practical knowledge” about different approaches to evaluation (Schwandt, 2015). Therefore, the purpose of this dissertation was to examine the practice of contribution analysis, to inform theory development, and improve practice. Specifically, the study investigated how CA in practice compares with the conceptualization offered by John Mayne, identified elements of effective practice by characterizing adaptations, adjustments, and innovations, and identified conditions under which practices may be different (e.g., evaluand, contexts, or evaluators; N. L. Smith, 1993). Though CA has been present in the academic literature for almost two decades, only recently have practice-based accounts surfaced with similar prominence. As such, it is an opportune moment for a thorough and comprehensive review of the CA practice.

This study addressed a critical oversight and pursued a richer understanding of the contexts of practice and probed deeper into the translation of theory-to-practice using empirical evidence to develop and improve on the current understandings of CA theory. Findings from this study inform practitioners of potential threats to evaluation design, implementation, and practice.

Despite the varied degrees of applications and contexts in which Contribution Analysis is used, this thesis has identified that CA is still a promising theory of evaluation practice primarily because of its attention to the theory of change. The findings from this study strongly suggest that CA's role as a guiding evaluation framework or strategy is advantageous as it ensures a systematic investigation of the production of the program theory assumptions, linkages, and eventual impact. John Mayne and other scholars have made enormous contributions to the development of CA. However, as this dissertation discovered, CA is transdisciplinary, and the reality. As such, it is my hope that research on CA theory will continue, and that we maximize the dialectic relationship between theory and practice and for continuous development of strong theoretical base and empirical base.

## REFERENCES

- Abma, T. A., & Stake, R. E. (2001). Stake's responsive evaluation: Core ideas and evolution. *New Directions for Evaluation*, 92, 7–22.
- Alkin, M. C., & Taut, S. (2003). Unbundling evaluation use. *Studies in Educational Evaluation* 29(2003), 1-12.
- Alkin, M.C. (Ed.). (2013). *Evaluation Roots: A Wider Perspective of Theorists' Views and Influences* (2nd edition). Thousand Oaks, CA: Sage Publications.
- Alkin, M. C. (2003). Evaluation theory and practice: Insights and new directions. In C. A. Christie (Ed.), *New directions for evaluation*, No. 97. The practice-theory relationship in evaluation (pp. 81–89). San Francisco: Jossey-Bass.
- Alkin, M. C., & Christie, C. A. (2004). An evaluation theory tree revisited. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 381-392). Thousand Oaks: Sage.
- Alkin, M. C., and Ellett, F. S., Jr. "Evaluation Models and Their Development." In H. J. Walberg and G. D. Haertel (eds.), *International Encyclopedia of Education: Research and Studies*. New York: Pergamon, 1985.
- Astbury, B., & Leeuw, F. L. (2010). Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation*, 31 (3), 363-381.
- Bamberger, M., Rugh, J., & Mabry, L. (2006). *Real world evaluation: Working under budget, time, data and political constraints*. Thousand Oaks, CA: Sage.

- Bannister, J., & O'Sullivan, A. (2013). Knowledge mobilization and the civic academy: The nature of evidence, the roles of narrative and the potential of contribution analysis. *Contemporary Social Science*, 8(3), 249–262.  
doi:10.1080/21582041.2012.751497
- Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, 13(4), 544-559.
- Befani, B., & Mayne, J. (2014). Process tracing and contribution analysis: A combined approach to generative causal inference for impact evaluation. *Institute of Development Studies Bulletin*, 45(6), 17–36. <http://doi.org/10.1111/1759-5436.12110>
- Better Evaluation. (2013). *Contribution Analysis*. Retrieved from [http://betterevaluation.org/plan/approach/contribution\\_analysis](http://betterevaluation.org/plan/approach/contribution_analysis)
- Betts, J., & Wood, B. (2012). The Paris declaration evaluation process and methods. *The Canadian Journal of Program Evaluation*, 27(3), 69-n/a.
- Bickman, L., & Rog, D. (1998). *Handbook of applied social research methods*. Sage Publications.
- Biggs, J. S., Farrell, L., Lawrence, G., & Johnson, J. K. (2014). A practical example of contribution analysis to a public health intervention. *Evaluation*, 20(2), 214–229.  
<http://doi.org/10.1177/1356389014527527>
- Brousselle, A., & Buregeya, J.-M. (2018). Theory-based evaluations: Framing the existence of a new theory in evaluation and the rise of the 5th generation. *Evaluation*, 24(2), 153–168. doi:10.1177/1356389018765487

- Budhwani, S., & McDavid, J. C. (2017). Contribution analysis: Theoretical and practical challenges and prospects for evaluators. *Canadian Journal of Program Evaluation*, 32(1), 1–24. doi:10.3138/cjpe.31121
- Bryman, A. (2012). *Social research methods*. Oxford university press.
- Byrne, D., & Ragin, C. C. (2009). *The Sage handbook of case-based methods*. Sage Publications.
- Campbell, D. T. (1988) “Degrees of Freedom” and the Case Study’, in D. T. Campbell (1988) *Methodology and Epistemology for the Social Sciences: Selected Papers*. Thousand Oaks, CA: Sage.
- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. *Handbook of research on teaching*. Chicago, IL: Rand McNally.
- Campbell, R., & Wasco, S. M. (2000). Feminist approaches to social science: Epistemological and methodological tenets. *American journal of community psychology*, 28(6), 773-791.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative research*. Sage Publications Ltd, London.
- Chen, H. T. (1990). *Theory-driven evaluation*. Newbury, CA: Sage.
- Chen, H. T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Thousand Oaks, CA: Sage.
- Chmiliar, L. (2010). Multiple-case designs. *Encyclopedia of Case Study Research*, 1-2.
- Christie, C. A. (2003). What guides evaluation practice? A study of how evaluation practice maps onto evaluation theory. In C. A. Christie (Ed.), *The practice- theory*

relationship in evaluation (pp. 7-36). *New directions for evaluation*, No. 97, San Francisco, CA: Jossey-Bass.

Connell, J. P., & Kubisch, A. C. (1998). Applying a theory of change approach to the evaluation of comprehensive community initiatives: Progress, prospects, and problems. In K. Fulbright-Anderson et al. (Eds.), *New approaches to evaluating community initiatives Vol. II: Theory, measurement, and analysis*. Washington, DC: The Aspen Institute.

Connolly, J. (2015). *Contribution Analysis' as an Innovative Approach to Policy Planning, Implementation and Evaluation in Partnership Contexts*. Paper at presented at the Social Innovation and Research Conference, University of Fudan, Shanghai, China, May 2015.

Connolly, J. (2016). Contribution analysis as an approach to enable public managers to demonstrate public value: The Scottish context. *International Journal of Public Sector Management*, 29(7), 690-707. <http://doi.org/10.1108/IJPSM-12-2015-0225>

Cook, T., Scriven, M., Coryn, C., & Evergreen, S. (2010). Contemporary thinking about causation in evaluation: A dialogue with Tom Cook and Michael Scriven. *American Journal of Evaluation*, 31, 105-117.

Coryn, C., Noakes, L., Westine, C., & Schröter, D. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American journal of Evaluation*, 32(2), 199-226.

Coryn, C. L. S., Schröter, D. C., & Hanssen, C. E. (2009). Adding a time-series design element to the success case method to improve methodological rigor: An

- application for nonprofit program evaluation. *American Journal of Evaluation*, 30, 80-92.
- Cousins, J. B. & Chouinard, J. A. (2012). *Participatory evaluation up close: An integration of research-based knowledge*. Greenwich, CT: Information Age Publishing.
- Cousins, J. B., & Earl, L. M. (1999). When the boat gets missed: Response to M. F. Smith. *American Journal of Evaluation*, 20(2), 309–317.
- Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of educational research*, 56(3), 331-36.
- Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, 80, 5-23.
- Cousins, J. B., Whitmore, E., & Shulha, L. (2013). Arguments for a common set of principles for collaborative inquiry in evaluation. *American Journal of Evaluation*, 34(1), 7–22. <http://dx.doi.org/10.1177/1098214012464037>
- Craig, N. (2013). Seeing the wood and the trees: using outcomes frameworks to inform planning, monitoring and evaluation in public health. *Journal of Public Health*, 35(3), 467–74.
- Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative approaches to research*. Upper Saddle River, NJ: Merrill/Pearson Education.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage.

- Creswell, J. W., & Maietta, R. C. (2002). Qualitative research. *Handbook of research design and social measurement*, 6(1), 143-184.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer (Ed.), *Test validity*, 3–17. Hillsdale, NJ: Erlbaum.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*. 52, 281-302.
- Dahler-Larsen, P. (2014). Constitutive effects of performance indicators: Getting beyond unintended consequences. *Public Management Review*, 16(7), 969–986.
- Dahler-Larsen, P. (2015). The evaluation society: Critique, contestability, and skepticism. *Spazio Filosofico*, 13, 21-36.
- Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.
- Delahais, T., & Toulemonde, J. (2012). Applying contribution analysis: Lessons from five years of practice. *Evaluation*, 18(3), 281–293.  
<http://doi.org/10.1177/1356389012450810>
- Donaldson, S. I. (2003). Theory-driven program evaluation in the new millennium. In S. I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 109-141). Mahwah, NJ: Erlbaum.
- Donaldson, S. I., & Lipsey, M. W. (2006). Roles for theory in contemporary evaluation practice: Developing practical knowledge. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The SAGE Handbook of Evaluation: Policies, Programs, and Practices* (pp. 57-75). SAGE Publications Ltd.

- Donaldson, S. I., Christie, C. A., & Mark, M. M. (Eds.). (2009). *What counts as credible evidence in applied research and evaluation practice?* Thousand Oaks, CA: Sage.
- Downes, A., Novicki, E., & Howard, J. (2018). Using the contribution analysis approach to evaluate science impact: A case study of the national institute for occupational safety and health. *American Journal of Evaluation*.  
doi:10.1177/1098214018767046
- Dybdal, L., Nielsen, S. B., Lemire, S. (2010). Contribution analysis applied: Reflections of scope and methodology. *Canadian Journal of Program Evaluation*, 25(2), 29–57.
- European Centre for Health Policy (ECHP), WHO Regional Office for Europe (1999). Gothenburg consensus paper. *Health impact assessment: Main concepts and suggested approach*. Brussels, Belgium.
- Fetterman, D. (1994). Empowerment evaluation. *Evaluation Practice*, 15, 1-15.
- Fitzpatrick, J. L., Sanders, J. R. & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines*. Upper Saddle River, NJ: Pearson Education.
- Funnell, S., & Rogers, P. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco, CA: Jossey-Bass.
- Furubo, J. E., & Stame, N. (Eds.). (2018). *The Evaluation Enterprise: A Critical View*. Routledge.

- Gates, E., & Dyson, L. (2017). Implications of the changing conversation about causality for evaluators. *American Journal of Evaluation*, 38(1), 29–46.  
doi:10.1177/1098214016644068
- George, A. L., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. MIT Press. Chicago .
- Given, L. M. (Ed.). (2008). *The Sage encyclopedia of qualitative research methods*. Sage publications.
- GIZ Denmark. (2015). Capturing results using contribution analysis. Retrieved from <http://www.giz.de/en/downloads/giz2015-en-capturing-results.pdf>
- Government of Canada. (2012). *Theory-based approaches to evaluation: Concepts and practice*. Retrieved from <http://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/theory-based-approaches-evaluation-concepts-practices.html#tocAck>
- Government of Canada. (2015). *Evaluation of the 2009 Policy on Evaluation*. Retrieved from <http://www.tbs-sct.gc.ca/hgw-cgf/oversight-surveillance/ae-ve/cee/orp/2015/e09poe-epse09-eng.asp>
- Government of Scotland. (2011). *Social Sciences Methods Series. Guide 6: Contribution analysis*. Retrieved from <http://www.gov.scot/resource/doc/175356/0116687.pdf>
- Greene, J.C. (2001). Evaluation extrapolations. *American Journal of Evaluation*, 22(3), 397- 402.
- Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, CA: Jossey-Bass.

- Greene, J. C. (2011). The construct (ion) of validity as argument. *New Directions for Evaluation*, 130, 81-91.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Thousand Oaks, CA: Sage.
- Guba, E. G., & Lincoln, Y. S. (2001). *Guidelines and checklist for constructivist evaluation*. Retrieved from [http://www.wmich.edu/evalctr/archive\\_checklists/constructivisteval.pdf](http://www.wmich.edu/evalctr/archive_checklists/constructivisteval.pdf)
- Healy, M., & Perry, C. (2000). Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualitative Market Research: An International Journal*, 3(3), 118-126.
- Heron, J., & Reason, P. (1997). A Participatory Inquiry Paradigm. *Qualitative Inquiry*, 3(3), 274-294.
- House, E. R. (1980). *Evaluating with validity*. Beverly Hills: Sage.
- House, E. R. (2001). Unfinished Business: Causes and Values. *American Journal of Evaluation*, 22(3), 309–315. <http://doi.org/10.1177/109821400102200304>
- House, E. R. (2014). Origins of the ideas in Evaluating with Validity. In J.C. Griffith & B. Montrosse-Moorhead (Eds.), *Revisiting truth, beauty, and justice: Evaluating with validity in the 21st century*. *New Directions for Evaluation*, 142, 9–15.
- Iverson, A. (2003). Attribution and Aid Evaluation in international Development: A Literature Review. *Evaluation Unit International Development Research Centre*. Retrieved from [http://web.idrc.ca/ev\\_en.php?ID=32055\\_201&ID2=DO\\_TOPIC](http://web.idrc.ca/ev_en.php?ID=32055_201&ID2=DO_TOPIC)

- Johnson, K., Greenesid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use a review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30(3), 377-410.
- Julnes, G. (2011). Reframing validity in research and evaluation: A multidimensional, systematic model of valid inference. *New Directions for Evaluation*, 2011(130), 55–67. <https://doi.org/10/ddpbxn>
- Kane, M. T. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement*, 38(4), 319–342. <https://doi.org/10/dsxrtc>
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.), Westport, CT: Praeger Publishers.
- Kane M. T. (2009). Validating the interpretations and uses of test scores. In: Lissitz RW (ed.) *The Concept of Validity: Revisions, New Directions and Applications*. Charlotte, NC: Information Age Publishing, Inc, 39–64.
- Kane, R., Levine, C., Orians, C., & Reinelt, C. (2017). Contribution analysis in policy work: Assessing advocacy’s influence. Centre for Evaluation Innovation. Retrieved from [https://www.evaluationinnovation.org/wp-content/uploads/2017/11/Contribution- Analysis\\_0.pdf](https://www.evaluationinnovation.org/wp-content/uploads/2017/11/Contribution- Analysis_0.pdf)
- Khakee, A. (2003). The emerging gap between evaluation research and practice. *Evaluation*, 9(3), 340-352.
- Khanal, R. (2014). Concerns and challenges of data integration from objective post-positivist approach and a subjective non-positivist interpretive approach and their

validity/credibility issues. *Journal of the Institute of Engineering*, 9(1), 115-129.

doi:10.3126/jie.v9i1.10677

Kirkhart, K. E. (2013, April). *Repositioning validity*. Paper presented at the Inaugural Culturally Responsive Evaluation & Assessment (CREA) Conference, Chicago, IL.

Koleros, A., & Mayne, J. (2019). Using Actor-Based Theories of Change to Conduct Robust Contribution Analysis in Complex Settings. *Canadian Journal of Program Evaluation*, 33(3). <https://doi.org/10/gf247h>

Kotvojs, F., & Shrimpton, B. (2007). Contribution analysis: A new approach to evaluation in international development. *Evaluation Journal of Australasia*, 7(1), 27–35.

Kramer, J., & Kaszap, A. (2017). Theory-based impact evaluation in practice: Key findings and policy learnings from the ex-post evaluation on cohesion policy support to large enterprises. *European Structural and Investment Funds Journal*, 5(2), 120-133.

Kridel, C. (Ed.) (2010). *Encyclopedia of curriculum studies*. Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781412958806

Leeuw, F. L. (2012). Linking theory-based evaluation and contribution analysis: Three problems and a few solutions. *Evaluation*, 18(3), 348-363.

Leeuw, F., & Vaessen, J. (2009). *World Bank: Impact evaluations and development*. NONIE Guidance on Impact Evaluation. Retrieved from [http://siteresources.worldbank.org/EXT/OED/Resources/nonie\\_guidance.pdf](http://siteresources.worldbank.org/EXT/OED/Resources/nonie_guidance.pdf)

- Lemire, S. (2010). *Contribution analysis: The promising new approach to causal claims*. Retrieved from <http://www.alnap.org/resource/7431>
- Lemire, S., Nielsen, S. B., & Dybdal, L. (2012). Making contribution analysis work: A practical framework for handling influencing factors and alternative explanations. *Evaluation, 18*(3), 294–309. <http://doi.org/10.1177/1356389012450654>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Madaus, G., & Stufflebeam, D. (2000). Program evaluation: a historical overview. In D. L. Stufflebeam, C. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models* (pp. 3–18). Boston, MA: Kluwer Academic.
- Mark, M. M. (2001). Evaluation's future: Furor, futile, or fertile? *American Journal of Evaluation, 22*, 457-479.
- Mark, M. M. (2008). Building a better evidence base for evaluation theory. *Fundamental issues in evaluation, 111-134*.
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA: Jossey Bass.
- Mark, M. M., & Shotland, R. L. (1985). Stakeholder-based evaluation and value judgments. *Evaluation Review, 9*, 605-626.
- Marks, D. F. & Sykes, C. M. (2003). Synthesizing evidence: Systematic reviews, meta-analyses and preference analyses. Marks, D. and Yardley, L. (Eds.), *Research methods for clinical and health psychology* (pp. 185-209) London: Sage Publications Ltd.

- Mathison, S. (2005). *Encyclopedia of evaluation*. Thousand Oaks, CA: SAGE Publications.
- Mayne, J. (1999). Addressing attribution through contribution analysis: Using performance measures sensibility [Discussion paper]. Retrieved from Office of the Auditor General of Canada  
[http://www.oagbvg.gc.ca/internet/docs/99dp1\\_e.pdf](http://www.oagbvg.gc.ca/internet/docs/99dp1_e.pdf)
- Mayne, J. (2001). Addressing attribution through contribution analysis: Using performance measures sensibly. *Canadian Journal of Program Evaluation*, 16(1), 1–24.
- Mayne, J. (2008). *Contribution analysis: An approach to exploring cause and effect*. ILAC Brief 16. Retrieved from  
[http://betterevaluation.org/resources/guides/contribution\\_analysis/ilac\\_brief](http://betterevaluation.org/resources/guides/contribution_analysis/ilac_brief)
- Mayne, J. (2011). Contribution analysis: Addressing cause and effect. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution and beyond* (p. 53-96). New Brunswick, NJ: Transaction Publishers.
- Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, 18(3), 270–280.  
<http://doi.org/10.1177/1356389012451663>
- Mayne, J. (2015a). Foreword. In A. M. Nkwake, *Credibility, Validity, and Assumptions in Program Evaluation Methodology* (p 6-9). Springer.
- Mayne, J. (2015b). *Contribution analysis: Making causal claims in the face of complexity*. Retrieved from

<http://www.torontoevaluation.ca/evaluatingcomplexity/pdf/workshops/15-07-17/John%20Mayne-Presentation%20.pdf>.

Mayne, J. (2015c). Useful theory of change models. *Canadian Journal of Program Evaluation*, 30(2), 119–142. <http://doi.org/10.3138/cjpe.230>

Mayne, J. (2017). Theory of change analysis: Building robust theories of change. *Canadian Journal of Program Evaluation*, 32(2), 155–173. <https://doi.org/10.3138/cjpe.31122>

Mayne, J., & Stern, E. (2013). *Impact evaluation of natural resource management research programs: A broader view*. ACIAR Impact Assessment Series Report No. 84, Australian Centre for International Agricultural Research, Canberra.

Mayne, J. (2019a). Assessing the relative importance of causal factors. CDI Practice Paper 21. Brighton, England: IDS. Retrieved from <https://opendocs.ids.ac.uk/opendocs/handle/123456789/14647>

Mayne, J. (2019b). Revisiting Contribution Analysis. *Canadian Journal of Program Evaluation*, 34(2).

Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, 33(2), 3–11.

Maxwell, J. A. (2012). The importance of qualitative research for causal explanation in education. *Qualitative Inquiry*, 18(8), 655-661.

Mentzer, G. A., Czerniak, C. M., & Struble, J. L. (2014). Utilizing program theory and contribution analysis to evaluate the development of science teacher leaders.

*Studies in Educational Evaluation*, 42, 100–108.

<http://doi.org/10.1016/j.stueduc.2014.03.003>

- Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. San Francisco, CA: Jossey-Bass.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco, CA: Jossey-Bass.
- Mertens, D. M. (2015). *Research and Evaluation in Education and Psychology*, Sage, Thousand Oaks CA 2015.
- Mertens, D. M., & Wilson, A. T. (2012). *Program evaluation theory and practice: A comprehensive guide*. Guilford Press.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*, 3rd ed. (pp. 13-103). New York: Macmillan.
- Michaud-Létourneau, I., Gayard, M., & Pelletier, D. (2019). Strengthening advocacy and policy change for infant and young child feeding. *Maternal & Child Nutrition*, 15. <http://doi.org/10.1111/mcn.12749>
- Miles, M., & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.
- 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. Retrieved from <http://doi.org/10.1177/1098214006291015>
- Mislevy, R. J., Steinberg, L. S., & Almond, R. A. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-67.
- Moreau, K. A., & Eady, K. (2015). Connecting medical education to patient outcomes: The promise of contribution analysis. *Medical Teacher*, 37(11), 1060. <http://dx.doi.org/10.3109/0142159X.2015.1060307>
- Montague, S. (2019). Does your implementation fit your theory of change? *Canadian Journal of Program Evaluation*, 33(3). doi:10/gf4jng
- Montague, S., Young, G., & Montague, C. (2003). Using circles to tell the performance story. *Canadian Government Executive*, 2, 12-16. Retrieved from <http://pmn.net/library/usingcirclestotelltheperformancestory.htm>
- Network of Networks on Impact Evaluation Subgroup 2. (2008). *NONIE impact evaluation guidance*. Retrieved from [http://www.worldbank.org/ieg/nonie/docs/NONIE\\_SG2.pdf](http://www.worldbank.org/ieg/nonie/docs/NONIE_SG2.pdf)
- Nkwake, A. M. (2013). *Working with Assumptions in International Development Program Evaluation: With a Foreword by Michael Bamberger*. Springer Science & Business Media.
- Nkwake, A. M. (2015). *Credibility, Validity, and Assumptions in Program Evaluation Methodology*. With a Foreword by John Mayne, Springer.

- Noltze, M., Gaisbauer, F., Schwedersky, T., & Krapp, S. (2014). Contribution analysis as an evaluation strategy in the context of a sector-wide approach: Performance-based health financing in Rwanda. *African Evaluation Journal*, 2(1), 8 pages. doi:10.4102/aej.v2i1.81
- Nour, K., Dutilly-Simard, S., Brousselle, A., Smits, P., Buregeya, J. M., Loslier, J., & Dennis, J. L. (2016). *Health Research Policy and Systems*, 14.
- Obodai, J., Adjei, P. O. W., Hamenoo, S. V. Q., & Abaitey, A. K. A. (2018). Towards household food security in Ghana: Assessment of Ghana's expanded forest plantation programme in Asante Akim south district. *GeoJournal*, 83(2), 365–380. doi:10.1007/s10708-017-9776-9
- Pajo, B. (2018). *Introduction to research methods: A hands-on approach*. Sage.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Patton, M. Q. (2004). The roots of utilization-focused evaluation. In M. C. Alkin (Ed.), *Evaluation roots tracing theorists' views and influences* (pp. 276-292). Thousand Oaks, CA: Sage Publications, Inc.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: SAGE.
- Patton M. Q. (2010) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford Press.
- Patton, M. Q. (2012). A utilization-focused approach to contribution analysis. *Evaluation*, 18(3), 364–377. <http://doi.org/10.1177/1356389012449523>

- Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q. (2018). Evaluation Science. *American Journal of Evaluation*, 39(2), 183-200.
- Pawson, R. D. (2007). Causality of Pawson, R. D. (2007). *Causality for beginners: First draft*. Retrieved from [eprints.ncrm.ac.uk/245/1/Causality\\_for\\_Beginners\\_Dec\\_07.doc](https://eprints.ncrm.ac.uk/245/1/Causality_for_Beginners_Dec_07.doc)
- Pawson, R. D. (2013). *The Science of Evaluation: A Realist Manifesto*. London: SAGE.
- Pawson, R. D. and Sridharan, S. (2009) Theory-driven evaluation of public health programmes. In: Killoran A and Kelly M (eds) *Evidence-Based Public Health: Effectiveness and Efficiency*. Oxford: Oxford University Press, 43–61.
- Pawson, R. D., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Preskill, H., & Torres, R. T. (1999a). Building capacity for organizational learning through evaluative inquiry. *Evaluation*, 5, 42-60.
- Qualtrics Labs, Inc. (2018). Provo, Utah: Qualtrics Labs, Inc. Retrieved from <https://www.qualtrics.com/blog/citing-qualtrics/>
- Rey, L., Brousselle, A., & Dedobbeleer, N. (2011). Logic Analysis: Testing Program Theory to Better Evaluate Complex Interventions. *Canadian Journal of Program Evaluation*, 26(3), 29.
- Rogers, P. J. (2009). Matching impact evaluation design to the nature of the intervention and the purpose of the evaluation. *Journal of development effectiveness*, 1(3), 217-226.

- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed). Thousand Oaks, CA: SAGE.
- Ryan, K., Gannon-Slater, N., & Culbertson, M. J. (2012). Improving survey methods with cognitive interviews in small-and medium-scale evaluations. *American Journal of Evaluation*, 33(3), 414-430.
- Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). London, England: Sage.
- Schwandt, T. (2007). On the importance of revisiting the study of ethics in evaluation. In *Dilemmas of engagement: Evaluation and the new public management* (pp. 117-127). Emerald Group Publishing Limited.
- Schwandt, T. (2015). Evaluation foundations revisited: Cultivating a life of the mind for practice. Stanford University Press. Schwartz, R., & Pais, G. (2012). Challenges and approaches to evaluating comprehensive complex tobacco control strategies. *The Canadian Journal of Program Evaluation*, 24(3), 1.
- Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Curriculum evaluation. AERA Monograph Series on Curriculum Evaluation* (vol. 1). Chicago, IL: Rand McNally.
- Scriven, M. (1980). *The logic of evaluation*. Inverness, CA: Edgepress.
- Scriven, M. (1991). *Evaluation thesaurus*. Thousand Oaks, CA: Sage.
- Scriven, M. (2003). Evaluation in the new millennium: The transdisciplinary vision. In S. I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems: Vision for the new millennium* (pp. 19-42). Mahwah, NJ: Lawrence Erlbaum.

- Scriven, M. (2005). Causation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 44-48). Thousand Oaks, CA: Sage.
- Scriven, M. (2008). A summative evaluation of RCT methodology: & an alternative approach to causal research. *Journal of multidisciplinary evaluation*, 5(9), 11-24.
- Scriven, M. (2016). Roadblocks to recognition and revolution. *American Journal of Evaluation*, 37(1), 27-44.
- Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19(1), 1-19.
- Schwartz, R., Forss, K., & Marra, M. (Eds.). *Evaluating the Complex: Attribution, Contribution, and Beyond*. 187-204. New Brunswick, NJ: Transaction Publishers, 2011.
- Searle, M., & Donnelly, C. (2017). Optimizing Use in the Field of Program Evaluation by Integrating Learning from the Knowledge Field. *Canadian Journal of Program Evaluation*, 31(3).
- Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19(1), 1-19.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: SAGE.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin.

- 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.
- Shah, S. K., & Corley, K. G. (2006). Building better theory by bridging the quantitative–qualitative divide. *Journal of management studies, 43*(8), 1821-1835.
- Smith, N. L. (1993). Improving evaluation theory through the empirical study of evaluation practice. *Evaluation Practice, 14*(3), 237-242.
- Smith, N. L. (2007). Empowerment evaluation as ideology. *American Journal of Evaluation, 28*, 169-178.
- Smith, N. L. (2015). Using action design research to research and develop evaluation practice. *New Directions for Evaluation, 2015*(148), 57-72.
- Smith, J. K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational inquirers. *Educational researcher, 15*(1), 4-12.
- Sridharan, S., & Nakaima, A. (2012). Towards an evidence-base of theory-driven evaluations: Some questions for proponents of theory-driven evaluation. *Evaluation, 18*(3), 378–395. <http://doi.org/10.1177/1356389012453289>
- Srivastava, R., & Enriquez, M. (2013). Just Communities’ Institute for Equity in Education – Contribution Analysis: A pilot study. Retrieved from [http://www.racialequitytools.org/resourcefiles/JCs\\_IEE\\_Contribution\\_Analysis\\_-\\_Full\\_Report\\_-\\_2013.pdf](http://www.racialequitytools.org/resourcefiles/JCs_IEE_Contribution_Analysis_-_Full_Report_-_2013.pdf).
- Stake, R. E. (1995). The art of case study research. Thousand Oaks, CA: Sage.

- Stake, R. E. (2004). *Standards-based and responsive evaluation*. Thousand Oaks, CA: Sage.
- Stake, R. E. (2006). *Multiple case study analysis*. New York, NY: Guilford Press.
- Stame, N. (2004). Theory-based evaluation and types of complexity. *Evaluation*, 10(1), 58–76.
- Stein, D., & Valters, C. (2012). *Understanding Theory of Change in International Development*. The Justice and Security Research Programme (JSRP). London: United Kingdom.
- Stern, E. (Ed.). (2012). Contribution Analysis [Special issue]. *Evaluation*, 18(3), 269-399.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). Broadening the range of designs and methods for impact evaluations. *Report of a study commissioned by the Department for International Development, London, UK*.
- Stocks-Rankin, C. R. (2014). *Reflective literature review of contribution analysis*. Retrieved from <http://blogs.iriss.org.uk/contribution/files/2015/06/Reflective-Literature-Review-of-Contribution-Analysis-Stocks-Rankin-2014.pdf>
- Stufflebeam, D. L. (2001). Evaluation models. *New Directions for Evaluation*, 89, 1-106.
- Stufflebeam, D. L., & Coryn, C. L. (2014). *Evaluation theory, models, and applications* (Vol. 50). John Wiley & Sons.
- Suchman, E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs*. New York, NY: Russell Sage Foundation.

- Taber, K. S. (2008). Of models, mermaids, and methods: The role of analytical pluralism in understanding student learning in science. *Science education in the 21st century*, 69-106.
- Tangata Whenua Community & Voluntary Sector Research Centre. (2015). *Contribution analysis*. Retrieved from <http://whatworks.org.nz/methods-tools-and-techniques/contribution-analysis>
- Tem. (2019). *Tem* [Transcription software]. Retrieved from <https://www.tem.com/>
- Ton, G. (2017). Contribution analysis of a Bolivian innovation grant fund: Mixing methods to verify relevance, efficiency and effectiveness. *Journal of Development Effectiveness*, 9(1), 120–143. <http://doi.org/10.1080/19439342.2016.1231702>
- Ton, G., Mayne, J., Delahais, T., Morell, J., Befani, B., Apgar, M., & O'Flynn, P. (2019). *Contribution Analysis and Estimating the Size of Effects: Can We Reconcile the Possible with the Impossible?* CDI Practice Paper 20, Brighton: IDS
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.
- U.S. National Institute of Justice. (2010). *NIJ tests and evaluates programs, practices and equipment*. Retrieved from [www.nij.gov/nij/about/testing-evaluation.htm](http://www.nij.gov/nij/about/testing-evaluation.htm)
- Van Melle, E., Gruppen, L., Holmboe, E. S., Flynn, L., Oandasan, I., & Frank, J. R. (2017). Using Contribution Analysis to Evaluate Competency-Based Medical Education Programs: It's All About Rigor in Thinking. *Academic medicine*, 92(6), 752-758.

- Vo, A.T. (2013). Visualizing context through theory deconstruction: A content analysis of three bodies of evaluation theory literature. *Evaluation and Program Planning*, 38, 44–52.
- Wallace, T. L. (2011). An argument-based approach to validity in evaluation. *Evaluation*, 17(3), 233-246.
- Weiss, C. H. (1997a). How can theory-based evaluation make greater headway? *Evaluation Review*, 21(4), 501–524. <http://doi.org/10.1177/0193841X9702100405>
- Weiss, C. H. (1997b). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation*, 76, 41–55. <http://doi.org/10.1002/ev.1086>
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies*. Upper Saddle River, NJ: Prentice Hall.
- Westhorp, G. (2014). Realist impact evaluation: an introduction. *London: Overseas Development Institute*, 1-12.
- What Works Clearinghouse. (2002). Retrieved from <http://ies.ed.gov/ncee/wwc>.
- White, H. (2009). Theory-based impact evaluation: principles and practice. *Journal of Development Effectiveness*, 1(3), 271–284.
- White, H. (2010). A contribution to current debates in impact evaluation. *Evaluation* 16 (2), 153–64.
- White, H., & Phillips, D. (2012). *Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework*. Working Paper 15, International Initiative for Impact Evaluation (3ie). Retrieved from [http://www.3ieimpact.org/3ie\\_working\\_papers.html](http://www.3ieimpact.org/3ie_working_papers.html)

Wilson-Grau, R., & Britt, H. (2012). Outcome harvesting. Cairo: Ford Foundation.

Retrieved from [http://www.managingforimpact.org/sites/default/files/resource/outome\\_harvesting\\_brief\\_final\\_2012-05-2-1.pdf](http://www.managingforimpact.org/sites/default/files/resource/outome_harvesting_brief_final_2012-05-2-1.pdf)

Wimbush, E., & Beeston, C. (2010). Contribution analysis: What is it and what does it offer impact evaluation? *The Evaluator*, Spring, 19–24.

Wimbush, E., Montague, S., & Mulherin, T. (2012). Applications of contribution analysis to outcome planning and impact evaluation. *Evaluation*, 18(3), 310–329.

doi:10.1177/1356389012452052

Worthen, B. R. (2001). Whither evaluation? That all depends. *American Journal of Evaluation*, 22, 409–418.

Yin, R. K. (2009). *Case study research: Design and methods*. Los Angeles, CA: Sage.

Zumbo, B. D., & Rupp, A. A. (2004). Responsible modeling of measurement data for appropriate inferences. *The SAGE handbook of quantitative methodology for the social sciences*, 73.

## APPENDIX A

### CONCEPTUAL ARTICLES BY MAYNE

Table 22

Conceptual Articles by Mayne Represented in Study

Type	Year	Author	Title	Publication
JA	2019	Mayne	Revisiting Contribution Analysis	<i>CJPE</i>
JA	2019	Mayne	Assessing the Relative Importance of Causal Factors	<i>CDI For Development impact</i>
manuscript	2019	Mayne	Developing Useful ToCs REV3	--
manuscript	2019	Mayne	A Brief on Contribution Analysis Principles and Concepts	--
JA	2017	Mayne	Theory of Change Analysis: Building Robust Theories of Change	<i>CJPE</i>
book	2013	Mayne & Stern	Impact evaluation of natural resource management research programs: a broader view	
JA	2012	Mayne	Contribution analysis: Coming of age?	<i>Eval.</i>
JA	2012	Mayne	Making Causal Claims	<i>ILAC Brief</i>
JA	2012	Stern, Stame, Mayne, Forss, Davies, & Befani	Broadening the range of designs and methods for impact evaluations	<i>DFID Working Paper</i>
BS	2011	Mayne	Contribution Analysis: Addressing Cause and Effect	<i>Evaluating the Complex: (Book)</i>
JA	2008	Mayne	Building an evaluative culture for effective evaluation and results management	<i>ILAC Brief</i>
JA	2008	Mayne	Contribution analysis: An approach to exploring cause and effect	<i>ILAC Brief</i>
JA	2006	Mayne & Rist	Studies are not Enough: The Necessary Transformation of Evaluation	<i>CJPE</i>
JA	2001	Mayne	Addressing attribution through contribution analysis: using performance measures sensibly	<i>CJPE</i>

## APPENDIX B

### CONCEPTUAL PUBLICATIONS BY OTHER AUTHORS

Table 23

Conceptual Publications by Other Authors (Represented in Study)

Type	Year	Author	Title	Publication
JA	2019	Whynot, Lemire, & Montague	How We Model Matters: A Manifesto for the Next Generation of Program Theorizing	<i>CJPE</i>
JA	2019	Montague,	Does Your Implementation Fit Your Theory of Change?	<i>CJPE</i>
JA	2019	Koleros & Mayne	Using Actor-Based Theories of Change to Conduct Robust Contribution Analysis in Complex Settings	<i>CJPE</i>
JA	2019	Ton, Mayne, Delahais, Morell, Befani, Agpar, & O'Flynn	Contribution Analysis and Estimating the Size of Effects: Can We Reconcile the Possible with the Impossible?	<i>CDI Practice Paper</i>
JA	2018	Brousselle & Buregeya	Theory-based evaluations: Framing the existence of a new theory in evaluation and the rise of the 5th generation	<i>Eval.</i>
JA	2017	Budhwani & McDavid	Contribution Analysis: Theoretical and Practical Challenges and Prospects for Evaluators	<i>CJPE</i>
JA	2017	Gates & Dyson	Implications of the Changing Conversation About Causality for Evaluators	<i>AJE</i>
JA	2014	Befani & Mayne	Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation	<i>IDS Bulletin</i>
JA	2014	Biggs, Farrell, Lawrence, & Johnson	A practical example of Contribution Analysis to a public health intervention	<i>Eval.</i>
manuscript	2014	Stocks-Rankin	Reflective Literature Review of Contribution Analysis	--
JA	2013	Bannister & O'Sullivan	Knowledge mobilisation and the civic academy: the nature of evidence, the roles of narrative and the potential of contribution analysis	<i>Cont.SS</i>
manuscript	2013	Mayne & Stern	Impact evaluation of natural resource management research programs: a broader view	<i>ACIJAR</i>

Table 23

Cont.

Type	Year	Author	Title	Publication
JA	2012	Delahais & Toulemonde	Applying contribution analysis: Lessons from five years of practice	<i>Eval.</i>
JA	2012	Leeuw	Linking theory-based evaluation and contribution analysis: Three problems and a few solutions	<i>Eval.</i>
JA	2012	Lemire, Nielsen, & Dybdal	Making contribution analysis work: A practical framework for handling influencing factors and alternative explanations	<i>Eval.</i>
JA	2012	Patton	A utilization-focused approach to contribution analysis	<i>Eval.</i>
JA	2012	Stern, Stame, Mayne; Forss, Davies, & Befani	Broadening the range of designs and methods for impact evaluations	DFID Working Paper
JA	2011	Coryn, Noakes, Westine, & Schröter	A systematic review of theory-driven evaluation practice from 1990 to 2009	<i>AJE</i>
JA	2011	Dybdal, Nielsen, & Lemire	Contribution Analysis Applied: Reflections on Scope and Methodology	<i>CJPE</i>
JA	2011	Wimbush	Implementing an outcomes approach to public management and accountability in the UK—are we learning the lessons?	<i>PM&amp;M</i>
JA	2010	Astbury & Leeuw	Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation	<i>AJE</i>
manuscript	2010	Lemire	Contribution Analysis: The promising new approach to causal claims	--
JA	2010	Wimbush & Beeston	Contribution Analysis -What is it and what does it offer impact evaluation?	<i>The Eval.</i>
JA	2009	Eirich & Morrison	Guide 6 - Contribution analysis	<i>SSMS</i>
book	2009	Leeuw & Vaessen	Impact evaluations and development: NONIE guidance on impact evaluation (Network of Networks on Impact Evaluation).	World Bank [NONIE]
manuscript	2003	Iverson	Attribution and Aid Evaluation in international Development: A Literature Review	<i>IDRC</i>

## APPENDIX C

### ALL EMPIRICAL ARTICLES DISCOVERED IN PHASE I

Table 24

Table of All Empirical Articles

Type	Year	Authors	Title	Journal	Language (If other than English)
JA	2007	Douthwaite, Alvarez, Cook, Davies, George, Howell, MacKay, & Rubiano	Participatory impact pathways analysis: A practical application of program theory in research-for-development	<i>CJPE</i>	
JA	2007	Kotvojs & Shrimpton	Contribution analysis: A new approach to evaluation in international development	<i>EJA</i>	
JA	2008	Patton	Advocacy Impact Evaluation	<i>JMDE</i>	
JA	2010	Armytage	Judicial reform in Asia: case study of AusAID's experience in Papua New Guinea: 2003–2007	<i>JDE</i>	
JA	2010	Graham & Mackinnon	Grasping the thistle: The role of alcohol brief interventions in Scottish alcohol policy	<i>D&amp;AR</i>	
JA	2010	Rotem, Zinovieff, & Goubarev	A framework for evaluating the impact of the United Nations fellowship programme	<i>HRH</i>	
JA	2010	Wimbush	Debate: Accountability for outcomes— international lessons	<i>PM&amp;M</i>	
JA	2010	Wimbush & Beeston	Contribution Analysis -What is it and what does it offer impact evaluation?	<i>The Eval.</i>	
BS	2011	Schwartz & Garcia	Intervention Path Contribution Analysis (IPCA) for Complex Strategy Evaluation: Evaluating the Smoke-Free Ontario Strategy		
BS	2011	Toulemonde, Carpenter, & Raffier	Coping with the Evaluability Barrier: Poverty Impact of European Support at Country Level		
JA	2011	Dybdal, Nielsen, & Lemire	Contribution Analysis Applied: Reflections on Scope and Methodology	<i>CJPE</i>	
JA	2011	Heyward, Cannon, & Sarjono	Implementing school-based management in Indonesia: impact and lessons learned	<i>JDE</i>	
JA	2011	Wimbush	Implementing an outcomes approach to public management and accountability in the UK—are we learning the lessons?	<i>PM&amp;M</i>	
JA	2012	Betts & Wood	The Paris Declaration Evaluation Process and Methods	<i>CJPE</i>	
JA	2012	Dabelstein & Kliest	Preparing, Governing, and Managing The Paris Declaration Evaluation	<i>CJPE</i>	
JA	2012	Delahais & Toulemonde	Applying contribution analysis: Lessons from five years of practice	<i>Eval.</i>	
JA	2012	Lemire, Nielsen, & Dybdal	Making contribution analysis work: A practical framework for handling influencing factors and alternative explanations	<i>Eval.</i>	
JA	2012	Montague & Lamers-Bellio	Advocacy Evaluation Theory as a Tool for Strategic Conversation: A 25-year Review of Tobacco Control Advocacy at the Canadian Cancer Society	<i>CJPE</i>	
JA	2012	Patton	Meta-Evaluation - Evaluating The evaluation of the Paris Declaration	<i>CJPE</i>	

Table 24

Cont.

Type	Year	Authors	Title	Journal	Language (If other than English)
JA	2012	Patton	A utilization-focused approach to contribution analysis	<i>Eval.</i>	
JA	2012	Schwartz & Pais	Challenges and Approaches to Evaluating Comprehensive Complex Tobacco Control Strategies	<i>CJPE</i>	
JA	2012	Sridharan & Nakaima	Towards an evidence base of theory-driven evaluations: Some questions for proponents of theory-driven evaluation	<i>Eval.</i>	
JA	2012	Vaessen & Raimondo	Making sense of impact: A methodological framework for assessing the impact of prizes	<i>Eval.</i>	
JA	2012	Wimbush, Montague, & Mulherin	Applications of contribution analysis to outcome planning and impact evaluation	<i>Eval.</i>	
JA	2013	Bannister & O'Sullivan	Knowledge mobilisation and the civic academy: the nature of evidence, the roles of narrative and the potential of contribution analysis	<i>Cont.SS</i>	
JA	2013	Smith, Wilkinson, & Gallagher	'It's what gets through people's radars isn't it': relationships in social work practice and knowledge exchange	<i>Cont.SS</i>	
thesis	2013	Buckley	Indigenous Firm Performance in a Small Late Developing State : A Case-study of the Role and Contribution of Public Venture Capital in Ireland		
JA	2014	Bauman, King, & Nutbeam	Rethinking the evaluation and measurement of health in all policies	<i>HPI</i>	
JA	2014	Befani & Mayne	Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation	<i>IDS</i>	
JA	2014	Biggs, Farrell, Lawrence, & Johnson	A practical example of Contribution Analysis to a public health intervention	<i>Eval.</i>	
JA	2014	Brandon, Smith, Ofir, & Noordeloos	Monitoring and Evaluation of African Women in Agricultural Research and Development (AWARD)	<i>AJE</i>	
JA	2014	Lloyd & Villanger	Assessing aid impacts revisited: results measurement in Norwegian aid	<i>JDE</i>	
JA	2014	Mentzer, Czerniak, & Struble	Utilizing program theory and contribution analysis to evaluate the development of science teacher leaders	<i>SEE</i>	
JA	2014	Nakrošis	Theory-based evaluation of capacity-building interventions	<i>Eval.</i>	
JA	2014	Noltze, Gaisbauer, Schwedersky, & Krapp	Contribution analysis as an evaluation strategy in the context of a sector-wide approach: Performance-based health financing in Rwanda	<i>AEJ</i>	
JA	2014	Ton, Vellema, & Ge	The Triviality of Measuring Ultimate Outcomes: Acknowledging the Span of Direct Influence	<i>IDS</i>	
JA	2015	Buckley	Using Sequential Mixed Methods in Enterprise Policy Instrument Evaluation: The Pragmatic Design Choice?	<i>Eu Conf. on RMBMS</i>	
JA	2015	Cousins, Svensson, Szijarto, Pinsent, Andrew, & Sylvestre	Assessing the Practice Impact of Research on Evaluation: Assessing the Practice Impact of Research on Evaluation	<i>NDE</i>	
JA	2015	Dauphinee	The role of theory-based outcome frameworks in program evaluation: Considering the case of contribution analysis	<i>MT</i>	

Table 24

Cont.

Type	Year	Authors	Title	Journal	Language (if other than English)
JA	2015	Joly, Gaunand, Colinet, Larédo, Lemarié, & Matt	ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization	<i>RE</i>	
JA	2015	Mayne & Johnson	Using theories of change in the CGIAR Research Program on Agriculture for Nutrition and Health	<i>Eval.</i>	
JA	2015	McDermott, Johnson, Kadiyala, Kennedy, & Wyatt	Agricultural research for nutrition outcomes – rethinking the agenda	<i>FS</i>	
JA	2015	Moreau & Eady	Connecting medical education to patient outcomes: The promise of contribution analysis	<i>MT</i>	
JA	2015	Morton	Progressing research impact assessment: A ‘contributions’ approach	<i>RE</i>	
JA	2015	Schmitt & Beach	The contribution of process tracing to theory-based evaluations of complex aid instruments	<i>Eval.</i>	
JA	2016	Buckley	Using Contribution Analysis to evaluate small & medium enterprise support policy	<i>Eval.</i>	
JA	2016	Celermajer & Saul	Preventing Torture in Nepal: A Public Health and Human Rights Intervention	<i>JBI</i>	
JA	2016	Connolly	Contribution analysis as an approach to enable public managers to demonstrate public value: The Scottish context	<i>IJPSM</i>	
JA	2016	Nour, Dutilly-Simard, Brousselle, Smits, Buregeya, Loslier, Denis, & Jean-Louis	Evaluation of the effects of health impact assessment practice at the local level in Monteregie	<i>HRP&amp;S</i>	
JA	2017	Belcher, Suryadarma, & Halimanjaya	Evaluating policy-relevant research: lessons from a series of theory-based outcomes assessments	<i>Palgrave Comm.</i>	
JA	2017	Budhwani & McDavid	Contribution Analysis: Theoretical and Practical Challenges and Prospects for Evaluators	<i>CJPE</i>	
JA	2017	Buregeya, Brousselle, Nour, & Loignon	Translation (word) - Comment évaluer les effets des évaluations d’impact sur la santé : le potentiel de l’analyse de contribution	<i>CJPE</i>	French
JA	2017	Delahais & Toulemonde	Making rigorous causal claims in a real-life context: Has research contributed to sustainable forest management?	<i>Eval.</i>	
JA	2017	Hersey & Adams	Using contribution analysis to assess the influence of farm link programs in the U.S.	<i>JAFSCD</i>	
JA	2017	Kramer & Kaszap	Theory-based Impact Evaluation in Practice: Key Findings and Policy Learnings from the Ex-post Evaluation on Cohesion Policy Support to Large Enterprises	<i>ESIF</i>	
JA	2017	Nour, Lafontaine, Mariève, Brousselle, Smits, Buregeya, Loslier, & Denis	L’analyse de contribution pour évaluer l’impact de la démarche ÉIS sur les processus décisionnels municipaux : un choix méthodologique intéressant ?	<i>GHP</i>	
JA	2017	Ofek	Evaluating social exclusion interventions in university-community partnerships	<i>E&amp;PP</i>	
JA	2017	Ton	Contribution analysis of a Bolivian innovation grant fund: mixing methods to verify relevance, efficiency, and effectiveness	<i>JDE</i>	

Table 24

Cont.

Type	Year	Authors	Title	Journal	Language (if other than English)
JA	2017	Van Melle, Gruppen, Holmboe, Flynn, Oandasan, & Frank	Using Contribution Analysis to Evaluate Competency-Based Medical Education Programs: It's All About Rigor in Thinking	<i>AM</i>	
JA	2018	Adlbrecht, Bartholomeyczik, & Maye	Entwicklung einer Programmtheorie als Grundlage für die Evaluation einer Demenzstation: Eine Möglichkeit der theoretischen Fundierung einer komplexen Intervention	<i>Pflege</i>	German
JA	2018	Belcher & Palenberg	Outcomes and Impacts of Development Interventions: Toward Conceptual Clarity	<i>AJE</i>	
JA	2018	Björkquist & Hansen	Reducing service barriers to people with dual diagnosis in Norway	<i>Cog. SS</i>	
JA	2018	Faure, Barret, Blundo-Canto, Dabat, Devaux-Spatarakis, Le Guerroué, Marquié, Mathé, Temple, Toillier, Triomphe, & Hainzelin	How different agricultural research models contribute to impacts: Evidence from 13 case studies in developing countries	<i>AS</i>	
JA	2018	Koleros, Mulkerne, Oldenbeuving, & Steine	The Actor-Based Change Framework: A Pragmatic Approach to Developing Program Theory for Interventions in Complex Systems	<i>AJE</i>	
JA	2018	Lawless, Baum, Delany-Crowe, MacDougall, Williams, McDermott, & Eyk	Developing a Framework for a Program Theory-Based Approach to Evaluating Policy Processes and Outcomes: Health in All Policies in South Australia	<i>IJHPM</i>	
JA	2018	Leiber	Impact evaluation of quality management in higher education: a contribution to sustainable quality development in knowledge societies	<i>EJHE</i>	
JA	2018	Maag, Alexander, Kase, & Hoffmann	Indicators for measuring the contributions of individual knowledge brokers	<i>ES&amp;P</i>	
JA	2018	Morton, Wilson, Inglis, Ritchie, & Wales	Developing a framework to evaluate knowledge into action interventions	<i>BMC-HSR</i>	
JA	2018	Obodai, Adjei, Hamenoo, & Abaitey	Towards household food security in Ghana: assessment of Ghana's expanded forest plantation programme in Asante Akim South District	<i>GeoJournal</i>	
JA	2018	Riley, Kernaghan, Stockton, Montague, Yessis, & Willis	Using contribution analysis to evaluate the impacts of research on policy: Getting to 'good enough'	<i>RE</i>	
JA	2018	Terrapon-Pfaff, Gröne, Dienst, & Willington	Impact pathways of small-scale energy projects in the global south – Findings from a systematic evaluation	<i>R&amp;SER</i>	
JA	2019	Baum, Delany-Crowe, MacDougall, van Eyk, Lawless, Williams, & Marmot	To what extent can the activities of the South Australian Health in All Policies initiative be linked to population health outcomes using a program theory-based evaluation?	<i>BMC-PH</i>	
JA	2019	Delahais & Lacouette-Fougère	Try again. Fail again. Fail better. Analysis of the contribution of 65 evaluations to the modernisation of public action in France	<i>Eval.</i>	

Table 24

Cont.

Type	Year	Authors	Title	Journal	Language (if other than English)
JA	2019	Dewar, Barrie, Sharp, & Meyer	Implementation of a Complex Intervention to Support Leadership Development in Nursing Homes: A Multimethod Participatory Study	<i>JAG</i>	
JA	2019	Dinh, Worth, Haire, & Hong	Confucian Evaluation: Reframing Contribution Analysis Using a Confucian Lens	<i>AJE</i>	
JA	2019	Downes, Novicki, & Howard	Using the Contribution Analysis Approach to Evaluate Science Impact: A Case Study of the National Institute for Occupational Safety and Health	<i>AJE</i>	
JA	2019	Gadda, Harris, Tisdall, & Millership	'Making children's rights real': lessons from policy networks and Contribution Analysis	<i>IJHR</i>	
JA	2019	Koleros & Mayne	Using Actor-Based Theories Of Change to Conduct Robust Contribution Analysis in Complex Settings	<i>CJPE</i>	
JA	2019	Livingston, Madoc-Jones, & Perkins	The potential of contribution analysis to alcohol and drug policy strategy evaluation: an applied example from Wales	<i>DEP&amp;P</i>	
JA	2019	Michaud-Létourneau, Gayard, & Pelletier	Translating the International Code of Marketing of Breast-milk Substitutes into national measures in nine countries	<i>M&amp;CN</i>	
JA	2019	Michaud-Létourneau, Gayard, & Pelletier	Contribution of the Alive & Thrive-UNICEF advocacy efforts to improve infant and young child feeding policies in Southeast Asia	<i>M&amp;CN</i>	
JA	2019	Schumacher, Dornoff, Carraccio, Busari, van der Vleuten, Kinnear, Kelleher, Sall, Warm, Martini, & Holmboe	The Power of Contribution and Attribution in Assessing Educational Outcomes for Individuals, Teams, and Programs:	<i>AM</i>	
JA	2019	Ton, Mayne, Delahais, Morell, Befani, Agpar, & O'Flynn	Contribution Analysis and Estimating the Size of Effects: Can We Reconcile the Possible with the Impossible?	<i>CDI-PP</i>	
JA	2020	Buregeya, Loignon, & Brousselle	Contribution analysis to analyze the effects of the health impact assessment at the local level: A case of urban revitalization	<i>E&amp;PP</i>	

## APPENDIX D

## LEGEND OF JOURNAL ACRONYMS

Table 25

## Legend of Journal Acronyms

Abbreviation	Journal	Evaluation Journal
<b>ACFIAR</b>	<i>Australian Centre for International Agricultural Research</i>	
<b>AEJ</b>	<i>African Evaluation Journal.</i>	X
<b>Ag. Systems</b>	<i>Agricultural Systems</i>	
<b>AJE</b>	<i>American Journal of Evaluation</i>	X
<b>AM</b>	<i>Academic Medicine</i>	
<b>BMC-HSR</b>	<i>BMC Health Services Research</i>	
<b>BMC PH</b>	<i>BMC Public Health</i>	
<b>CDI PP</b>	<i>CDI Practice Paper</i>	
<b>CJPE</b>	<i>Canadian Journal of Program Evaluation</i>	X
<b>Cog. SS</b>	<i>Cogent Social Sciences</i>	
<b>Cont. SS</b>	<i>Contemporary Social Science</i>	
<b>D&amp;AR</b>	<i>Drug and Alcohol Review</i>	
<b>DEP&amp;P</b>	<i>Drugs: Education, Prevention and Policy</i>	
<b>E&amp;PP</b>	<i>Evaluation and Program Planning</i>	X
<b>EJA</b>	<i>Evaluation Journal of Australasia</i>	X
<b>EJHE</b>	<i>European Journal of Higher Education</i>	
<b>EStIF</b>	<i>European Structural and Investment Funds Journal</i>	
<b>ES&amp;P</b>	<i>Environmental Science &amp; Policy</i>	
<b>Eu Conf. on RMBMS</b>	<i>European Conference on Research Methodology for Business and Management Studies</i>	
<b>Eval</b>	<i>Evaluation</i>	X
<b>FS</b>	<i>Food Security</i>	
<b>BHP</b>	<i>Global Health Promotion</i>	
<b>HER</b>	<i>Health Education Research</i>	
<b>HPI</b>	<i>Health Promotion International</i>	
<b>HRP&amp;S</b>	<i>Health Research Policy &amp; Systems</i>	
<b>HRH</b>	<i>Human Resources for Health</i>	
<b>IDS</b>	<i>IDS Bulletin</i>	X
<b>IJHPM</b>	<i>International Journal of Health Policy and Management</i>	
<b>IJHR</b>	<i>International Journal of Human Rights</i>	
<b>JAFCSD</b>	<i>Journal of Agriculture, Food Systems, and Community Development</i>	

Table 25

Cont.

<b>Abbreviation</b>	<b>Journal</b>	<b>Evaluation Journal</b>
<b>JAG</b>	<i>Journal of Applied Gerontology</i>	
<b>JBI</b>	<i>Journal of Bioethical Inquiry</i>	
<b>JDE</b>	<i>Journal of Development Effectiveness</i>	
<b>JMDE</b>	<i>Journal of Multidisciplinary Evaluation</i>	X
<b>M&amp;CN</b>	<i>Maternal &amp; Child Nutrition</i>	
<b>MT</b>	<i>Medical Teacher</i>	
<b>NDE</b>	<i>New Directions for Evaluation</i>	X
<b>PM&amp;M</b>	<i>Public Money &amp; Management</i>	
<b>R&amp;SER</b>	<i>Renewable and Sustainable Energy Reviews</i>	
<b>RE</b>	<i>Research Evaluation</i>	X
<b>SEE</b>	<i>Studies in Educational Evaluation</i>	X
<b>SSMS</b>	<i>Social Science Methods Series</i>	
<b>The Eval.</b>	<i>The Evaluator</i>	X

**APPENDIX E**  
**PROSPECTIVE CASES**

- Bannister, J., & O’Sullivan, A. (2013). Knowledge mobilisation and the civic academy: The nature of evidence, the roles of narrative and the potential of contribution analysis. *Contemporary Social Science*, 8(3), 249–262.  
<http://doi.org/10.1080/21582041.2012.751497>
- Befani, B., & Mayne, J. (2014). Process tracing and contribution analysis: A combined approach to generative causal inference for impact evaluation. *Institute of Development Studies Bulletin*, 45(6), 17–36. <http://doi.org/10.1111/1759-5436.12110>
- Betts, J., & Wood, B. (2012). The Paris declaration evaluation process and methods. *The Canadian Journal of Program Evaluation*, 27(3), 69-n/a.
- Biggs, J. S., Farrell, L., Lawrence, G., & Johnson, J. K. (2014). A practical example of contribution analysis to a public health intervention. *Evaluation*, 20(2), 214–229.  
<http://doi.org/10.1177/1356389014527527>
- Connolly, J. (2015). *Contribution Analysis’ as an Innovative Approach to Policy Planning, Implementation and Evaluation in Partnership Contexts*. Paper at presented at the Social Innovation and Research Conference, University of Fudan, Shanghai, China, May 2015.
- Connolly, J. (2016). Contribution analysis as an approach to enable public managers to demonstrate public value: The Scottish context. *International Journal of Public Sector Management*, 29(7), 690-707. <http://doi.org/10.1108/IJPSM-12-2015-0225>

- Downes, A., Novicki, E., & Howard, J. (2018). Using the Contribution Analysis Approach to Evaluate Science Impact: A Case Study of the National Institute for Occupational Safety and Health. *American Journal of Evaluation*.  
<https://doi.org/10.1177/1098214018767046>.
- Dybdal, L., Nielsen, S. B., Lemire, S., & Ramboll Management Consulting. (2011). Contribution analysis applied: Reflections of scope and methodology. *Canadian Journal of Program Evaluation*, 25(2), 29–57.
- GIZ Denmark. (2015). Capturing results using contribution analysis. Retrieved from <http://www.giz.de/en/downloads/giz2015-en-capturing-results.pdf>
- Hersey, A., & Adams, M. (2017). Using contribution analysis to assess the influence of farm link programs in the U.S. *Journal of Agriculture, Food Systems, and Community Development*, 7(3), 83.
- Kramer, J., & Kaszap, A. (2017). Theory-based impact evaluation in practice: Key findings and policy learnings from the ex-post evaluation on cohesion policy support to large enterprises. *European Structural and Investment Funds Journal*, 5(2), 120-133.
- Kotvojs, F. (2006) *Contribution Analysis – A New Approach to Evaluation in International Development*. Australian Evaluation Society Conference, Darwin Aus, 4-7 Sept 2006.
- Mentzer, G. A., Czerniak, C. M., & Struble, J. L. (2014). Utilizing program theory and contribution analysis to evaluate the development of science teacher leaders. *Studies in Educational Evaluation*, 42, 100–108.

- Michaud-Létourneau, I., Gayard, M., & Pelletier, D. (2019). Strengthening advocacy and policy change for infant and young child feeding. *Maternal & Child Nutrition*, 15.
- Montague, S. and Results Management at the Canadian Cancer Society.  
<http://www.pmn.net/library>
- Monnier, E. & Barakat, N. (2008) *Designing a Contribution Analysis for Assessing Expected Impacts*. European Evaluation Society Conference, Lisbon, 1-3 October 2008.
- Morton, S. (2015). Progressing research impact assessment: A ‘contributions’ approach. *Research Evaluation*.
- Noltze, M., Gaisbauer, F., Schwedersky, T., & Krapp, S. (2014). Contribution analysis as an evaluation strategy in the context of a sector-wide approach: Performance-based health financing in Rwanda. *African Evaluation Journal*, 2(1), 8 pages.  
<http://doi:10.4102/aej.v2i1.81>
- Nour, K., Dutilly-Simard, S., Brousselle, A., Smits, P., Buregeya, J.M., Loslier, J., & Dennis, J.L. (2016). *Health Research Policy and Systems*, 14.
- Obodai, J., Adjei, P.O.W., Hamenoo, S.V.Q., Abaitey, A.K.A. (2018). Towards household food security in Ghana: Assessment of Ghana’s expanded forest plantation programme in Asante Akim south district. *GeoJournal*, 83(2), 365-380.  
<http://dx.doi.org/10.1007/s10708-017-9776-9>
- OECD. (2005). Paris declaration on aid effectiveness. Retrieved from Organisation for Economic Co-operation and Development website:

<https://legalinstruments.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=141&Lang=en>

Rotem, A., Zinovieff, M. A., & Goubarev, A. (2010). A framework for evaluating the impact of the united nations fellowship programmes. *Human Resources for Health*, 8, 7. doi: <http://dx.doi.org/10.1186/1478-4491-8-7>

Schwartz, R. & Garcia, J. (2011). Intervention Path Contribution Analysis (IPCA) for Complex Strategy Evaluation: Evaluating the Smoke-Free Ontario Strategy. In K. Forss, M. Mara, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution and beyond* (pp. 199-220). New Brunswick, NJ: Transaction

Schwartz, R., & Garcia, J. (2011). Intervention Path Contribution Analysis (IPCA) for complex strategy evaluation: evaluating the smoke-free Ontario strategy. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the complex. Attribution, contribution, and beyond*. New Brunswick, USA: Transaction Publishers.

Schwartz, R., & Pais, G. (2012). Challenges and approaches to evaluating comprehensive complex tobacco control strategies. *The Canadian Journal of Program Evaluation*, 24(3), 1.

Sridharan, S., & Nakaima, A., A. (2012). Towards an evidence base of theory-driven evaluations: Some questions for proponents of theory-driven evaluation. *Evaluation*, 18(3), 378–395.

Srivastava, R., & Enriquez, M. (2013). Just Communities' Institute for Equity in Education – Contribution Analysis: A pilot study. Retrieved from

[http://www.racialequitytools.org/resourcefiles/JCs\\_IEE\\_Contribution\\_Analysis\\_-\\_Full\\_Report\\_-\\_2013.pdf](http://www.racialequitytools.org/resourcefiles/JCs_IEE_Contribution_Analysis_-_Full_Report_-_2013.pdf)

- Tangata Whenua Community & Voluntary Sector Research Centre. (2015). *Contribution analysis*. Retrieved from <http://whatworks.org.nz/methods-tools-and-techniques/contribution-analysis>
- Ton, G. (2017). Contribution analysis of a Bolivian innovation grant fund: Mixing methods to verify relevance, efficiency and effectiveness. *Journal of Development Effectiveness*, 9(1), 120–143. <http://doi.org/10.1080/19439342.2016.1231702>
- Toulemonde, J. et al (2010) Contribution of EC support to poverty reduction in Tanzania. In: Schwartz, R., Forss, K. and Marra, M. (eds) op cit.
- Toulemonde, J., Carpenter, D., & Raffier, L. (2011). Coping with the evaluability barrier: Poverty impact of European support at country level. In K. Forss, M. Mara, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution and beyond* (pp. 123–144). New Brunswick, NJ: Transaction.
- Wimbush, E. (2009) *What is Contribution Analysis and what does it offer impact evaluation?* UK Evaluation Society Conference, London, 24-25 Nov 2009.
- Wimbush, E., Montague, S., & Mulherin, T. (2012). Applications of contribution analysis to outcome planning and impact evaluation. *Evaluation*, 18(3), 310–329. <http://doi.org/10.1177/1356389012452052>

## APPENDIX F

### QUALTRICS QUESTIONNAIRE

#### CA Practitioner Questionnaire: Consent, Practitioner Experience, & Demographics

##### Consent and Screening

Q1.1 Thank you for your interest in my study. Your participation will include:

- 1) questionnaire about your background & overall CA experiences (current activity)
- 2) one audio-taped interview (45-60 minutes) (to occur at a later time)

It should take around 10 minutes to complete. Please review the informed consent document attached in the invitation to begin the survey.

With full knowledge of all foregoing, I agree, of my own free will, to participate in this study.	<input type="radio"/> Yes	<input type="radio"/> No
I agree to have my interview audio recorded.	<input type="radio"/> Yes	<input type="radio"/> No
I agree to the use of anonymous quotations in any thesis or publication that comes of this research.	<input type="radio"/> Yes	<input type="radio"/> No
I agree to participate in this research & understand the above statement about the research. I hereby grant the researcher permission to use my responses provided in this survey for research & publication purposes.	<input type="radio"/> Yes	<input type="radio"/> No

Q2.1. Contribution Analysis (CA) is a theory-based evaluation approach that seeks to probe “what works, why, and how” by building and testing the theory-of-change (ToC) via a seven-step framework (Mayne, 2001). Which statement best describes your experience with Contribution Analysis within the past 10 years?

- I **have applied experience** with CA (that is, any participation in an evaluation guided by CA).
- I have **conducted research** on CA, but I **do not have** applied experience.
- None of these.

##### CA Experience

Q3.1. This section will ask about your **general experience with Contribution Analysis evaluations**.

Q3.2 Which statement best describes your **role/experience** with the Contribution Analysis evaluation approach?

- |   |   |
|---|---|
| <input type="radio"/> Evaluation Practitioner | <input type="radio"/> Program Implementer/Staff     |
| <input type="radio"/> Consultant              | <input type="radio"/> Policy-Maker                  |
| <input type="radio"/> Researcher              | <input type="radio"/> Other (please specify): _____ |

Q3.3. In which **capacities** did you participate (by yourself or in a team) in your role as a(an)  $\$(Q3.2\_Response)$  in the CA evaluation(s)? (Select all that apply)

- |   |  |
|---|--|
| <input type="checkbox"/> developing the intervention/research | <input type="checkbox"/> managing/supervising the evaluation |
| <input type="checkbox"/> designing the evaluation             | <input type="checkbox"/> Other (please specify): _____       |
| <input type="checkbox"/> implementing the evaluation          | <input checked="" type="checkbox"/> None of these            |
| <input type="checkbox"/> reporting evaluation results         |  |

Q3.4. **How many CA** evaluations have you participated in your role as a(an)  $\$(Q3.2\_Response)$ ?

- |                         |                                 |
|-------------------------|---------------------------------|
| <input type="radio"/> 1 | <input type="radio"/> 4         |
| <input type="radio"/> 2 | <input type="radio"/> 5 or more |
| <input type="radio"/> 3 |                                 |

Q3.5. How many **years of CA evaluation experience** do you have in your role as a(an)  $\$(Q3.2\_Response)$ ?

- |                                   |                                  |
|-----------------------------------|----------------------------------|
| <input type="radio"/> Less than 1 | <input type="radio"/> 7-9        |
| <input type="radio"/> 1-2         | <input type="radio"/> 10 or more |
| <input type="radio"/> 4-6         |                                  |



Q5.3. In which **sectors** do you typically do most of your evaluation work? (Select all that apply)

- |  |   |   |
|--|---|---|
| <input type="checkbox"/> Agriculture               | <input type="checkbox"/> International & Cross-Cultural           | <input type="checkbox"/> Social Work                      |
| <input type="checkbox"/> Disaster & Emergency Mgmt | <input type="checkbox"/> Medicine                                 | <input type="checkbox"/> Regional & Urban Development     |
| <input type="checkbox"/> Environmental Initiatives | <input type="checkbox"/> Organization Behavior                    | <input type="checkbox"/> Transport & Mobility             |
| <input type="checkbox"/> Education (K-12)          | <input type="checkbox"/> Public Health                            | <input type="checkbox"/> Workforce & Economic Development |
| <input type="checkbox"/> Food & Natural Resources  | <input type="checkbox"/> Public Policy & Admin.                   | <input type="checkbox"/> Other (please specify): _____    |
| <input type="checkbox"/> Health Care               | <input type="checkbox"/> Science, Tech., Engineering, Math (STEM) |   |
| <input type="checkbox"/> Higher Education          |   |   |
| <input type="checkbox"/> Health & Human Services   |   |   |

Q5.4. Which of the following describes your **primary work setting**? (Select one)

- |  |  |
|--|--|
| <input type="radio"/> College / University     | <input type="radio"/> Local/Regional Agency                                    |
| <input type="radio"/> Private Organization     | <input type="radio"/> Local Non-profit organization                            |
| <input type="radio"/> School System (Pre/K-12) | <input type="radio"/> International Non-profit Organization (National, Global) |
| <input type="radio"/> Self-employed            | <input type="radio"/> Other (please specify): _____                            |
| <input type="radio"/> Federal Agency           |  |

Q5.5. What **size** of evaluations do you typically work on?

- Small-scale (e.g., small sample size, single site, small multi-site)
- Large-scale (e.g., large sample size, large multi-site, multi-organization)
- A combination of large-scale & small scale

Q5.6. Which statement best characterizes your **typical position** in your evaluation practice?

- embedded (e.g., also entrusted with design & delivery of intervention)
- internal (e.g., to organization but not entrusted with design & delivery of intervention)
- external (to organization)
- consultant (e.g., fragmented engagement w/ evaluation)
- other (please specify): \_\_\_\_\_

### Demographic Section

Q6.1. What's the **highest degree** that you have achieved?

- |  |   |
|--|---|
| <input type="radio"/> Bachelor's                         | <input type="radio"/> Doctoral (e.g., PhD)          |
| <input type="radio"/> Master's                           | <input type="radio"/> Other (please specify): _____ |
| <input type="radio"/> Professional Degree (e.g., MD, JD) |   |

Q6.2 In which **field** did you receive your highest degree?

- |   |  |
|---|--|
| <input type="radio"/> Arts/Humanities                 | <input type="radio"/> Psychology                           |
| <input type="radio"/> Business & Economics            | <input type="radio"/> Science (Natural, Physical)          |
| <input type="radio"/> Education                       | <input type="radio"/> Social work                          |
| <input type="radio"/> Environmental Studies           | <input type="radio"/> Sociology                            |
| <input type="radio"/> Evaluation                      | <input type="radio"/> Quantitative Psychology / Statistics |
| <input type="radio"/> Health (Public, Medicine)       | <input type="radio"/> Other (please specify): _____        |
| <input type="radio"/> Political Science/Public Policy |  |

Q6.3. What is your **primary occupation** (>80% of your time)?

- |  |   |
|--|---|
| <input type="radio"/> Internal evaluator           | <input type="radio"/> Researcher                    |
| <input type="radio"/> External evaluator           | <input type="radio"/> Professor/educator            |
| <input type="radio"/> A mix of interval & external | <input type="radio"/> Other (please describe) _____ |

Q6.4. How many **years** have you **conducted evaluations**?

- |  |  |
|--|--|
| <input type="radio"/> Less than a year | <input type="radio"/> 11 to 15 years     |
| <input type="radio"/> 1 to 5 years     | <input type="radio"/> 15 to 20 years     |
| <input type="radio"/> 6 to 10 years    | <input type="radio"/> More than 21 years |

Q6.5. How would you rate your **evaluation expertise**?

- |                                  |                                |
|----------------------------------|--------------------------------|
| <input type="radio"/> Novice     | <input type="radio"/> Advanced |
| <input type="radio"/> Proficient | <input type="radio"/> Expert   |

Q6.6. What is your country of residence?

- |                                     |   |
|-------------------------------------|---|
| <input type="radio"/> United States | <input type="radio"/> France                        |
| <input type="radio"/> Canada        | <input type="radio"/> Other (please specify): _____ |

Q6.7. In which **region** do you tend to **work**?

- |                                     |                                       |  |
|-------------------------------------|---------------------------------------|--|
| <input type="radio"/> North America | <input type="radio"/> Central America | <input type="radio"/> Oceania                      |
| <input type="radio"/> Africa        | <input type="radio"/> Eastern Europe  | <input type="radio"/> South America                |
| <input type="radio"/> Asia          | <input type="radio"/> Middle East     | <input type="radio"/> Other (please specify) _____ |
| <input type="radio"/> Carribbean    | <input type="radio"/> Oceania         |  |

Q6.8. What is your gender identity?

- |                              |  |
|------------------------------|--|
| <input type="radio"/> Male   | <input type="radio"/> Self-identify other  |
| <input type="radio"/> Female | <input type="radio"/> Prefer not to answer |

Q6.9 How old are you?

- |                                |                             |
|--------------------------------|-----------------------------|
| <input type="radio"/> Under 30 | <input type="radio"/> 40-49 |
| <input type="radio"/> 31-39    | <input type="radio"/> 50+   |

Q6.1.0 What is your racial/ethnic background? (Select all that apply)

- |   |  |
|---|--|
| <input type="checkbox"/> African  | <input type="checkbox"/> Latino                              |
| <input type="checkbox"/> African American or Black                          | <input type="checkbox"/> Middle Eastern European             |
| <input type="checkbox"/> American Indian, Native American, or Alaska Native | <input type="checkbox"/> Native Hawaiian or Pacific Islander |
| <input type="checkbox"/> Asian/East Indian/Indian                           | <input type="checkbox"/> Prefer not to answer                |
| <input type="checkbox"/> Caribbean Islander                                 | <input type="checkbox"/> Other (specify) _____               |
| <input type="checkbox"/> European American, White                           |  |

End Survey

Thank you very much for completing the questionnaire! Your time & effort are very much appreciated! You are welcome to leave any suggestions, comments, or add anything additional here!

---

**APPENDIX G**  
**RECRUITMENT EMAIL**

Dear *{name}*,

I would like to invite you to participate in a dissertation study focused on examining the Contribution Analysis (CA) approach to evaluation. Specifically, this study seeks to gain insight into practice and investigate barriers and facilitators to implementation. You have been selected for this study because of your visible involvement and notable contributions to the peer-review literature, and have been identified as a practitioner with professional experience practicing the Contribution Analysis approach to evaluation.

Specifically, I will ask about your experience for the case reflected in the following article:  
*{citation}*

Your participation will consist of participating in one audio-taped telephone interview (about 45-60 minutes) wherein you will answer and respond to prepared questions concerning your experiences with Contribution Analysis evaluation, as well as a short demographic questionnaire. You may be asked to review a transcription of the interview to verify the accuracy of your response.

If you agree to participate, please complete this doodle *{Doodle Link}* to schedule the telephone interview & you may access the short demographic questionnaire here *{Qualtrics link}*. I've also included the interview questions and consent form (for your records) if you wish to participate.

Thank you very much for your consideration of this research study. If you do not wish to participate in this study, you may respond to this email requesting no further contact.

Sincerely,  
Emma M. Sunnassee  
PhD Candidate, Dept. of Educational Research Methodology  
The University of North Carolina at Greensboro

Attachment:  Consent Form
------------------------------------

Attachment:  Interview Questions
---

## APPENDIX H

### CONSENT FORM

#### INFORMATION CONSENT LETTER FOR PRACTITIONER INTERVIEW

**Project Title:** An Investigation of Contextual Factors, Facilitators, and Barriers to the Contribution Analysis Evaluation Approach

**Principal Investigator:** Emma M. Sunnassee

**Dissertation Advisors:** Drs. Ayesha Boyce and John T. Willse

**Institution/Department:** University of North Carolina-Greensboro, Educational Research and Methodology

#### **Introduction**

This study offers an opportunity for you, as an evaluator, to share your experience in practicing the Contribution Analysis approach to evaluation (Mayne, 2008, 2011, 2015, 2017, 2019) and reflection on your professional practice. Your responses will help achieve a better understanding of the context of practice as it relates to CA, and how specific contextual factors impact practice. The study results will directly contribute to the general knowledge base of evaluation practice and the advancement of the Contribution Analysis evaluation approach.

#### **What are some general things you should know about research studies?**

You are being asked to take part in a research study. Your participation is voluntary. Details about this study are discussed in this consent form.

#### **What is the study about?**

The purpose of this study is to examine the Contribution Analysis (CA) approach in evaluation. Specifically, this study seeks to better understand the translation of theory into practice by querying the 7-steps described by John Mayne and identify barriers and facilitators encountered during this process via an interview with CA evaluation practitioners regarding their experiences. The contents will be used for the purpose of preparing and submitting a Ph.D. (Education) dissertation thesis at the University of North Carolina at Greensboro.

#### **Why are you asking me?**

You were selected to participate because you have been identified as an evaluator practitioner with professional experience practicing the Contribution Analysis approach to evaluation. More specifically, I would like to include you in my study because of your interest in, understanding of, and experience using CA.

#### **What will you ask me to do if I agree to be in the study?**

Your participation in the study will consist of one audio-taped telephone interview (about 45 minutes) wherein you will respond to prepared questions about your experience with Contribution Analysis. You may be asked to review a transcription of the interview to verify the accuracy of your response. Onboarding for the study will also include a brief demographic and evaluation experience questionnaire via Qualtrics.

#### **Is there any audio/video recording?**

The interview will be audio recorded using a digital recorder.

#### **How will you keep my information confidential?**

All information obtained in this study is strictly confidential unless disclosure is required by law. Digital recordings of the telephone interviews, as well as other data collected, will be kept in a secure manner and conserved for up to three years, at which time all materials will be destroyed.

#### **Will my de-identified data be used in future studies?**

Research findings may be used in written publications, such as journals or abstracts. Your de-identified data will be kept indefinitely and may be used for future research without your additional consent.

#### **Notes on Internet Research:**

Absolute confidentiality of data provided through the Internet cannot be guaranteed due to the limited protections of Internet access. Please be sure to exit your browser once completed with the Qualtrics questionnaire.

The commercial survey tool, Qualtrics, will be used in this study. All survey information will be retained and hosted on the researcher's Qualtrics (third party) account. Your confidentiality will be maintained to the degree permitted by the technology used. Specifically, no guarantees can be made regarding the interception of data sent via the Internet by any third parties.

**What are the risks to me?**

The risks involved in this study include the potential that confidentiality is not guaranteed since information is being gathered through an online questionnaire format that must contain an identifying email so that the participants' responses after each round can be reported back to the participant. In order to maintain confidentiality for the data gathered through the questionnaires, they will be labeled with your email. These email addresses will be erased from the temporary account within 30 days of the final presentation of the research.

Other than the time you spend on this project there are no known or foreseeable risks involved with this study, the Institutional Review Board at the University of North Carolina at Greensboro has determined that participation in this study poses minimal risk to participants.

**Are there any benefits to society as a result of me taking part in this research?**

Findings from this study may inform Contribution Analysis practitioners on how they may plan for and improve their practice, as well as the broader academic body of evaluation research.

**Are there any benefits to me for taking part in this research study?**

There are no direct benefits to participants in this study.

**Will I get paid for being in the study? Will it cost me anything?**

There are no costs to you or payments made for participating in this study.

**What if I want to leave the study?**

You have the right to refuse to participate or to withdraw at any time, without penalty. If you do withdraw, it will not affect you in any way. If you choose to withdraw, you may request that any of your data that has been collected be destroyed unless it is in a de-identifiable state. The investigators also have the right to stop your participation at any time. This could be because you have had an unexpected reaction, or have failed to follow instructions, or because the entire study has been stopped.

**What about new information/changes in the study?**

If significant new information relating to the study becomes available which may relate to your willingness to continue to participate, this information will be provided to you.

If you have questions, want more information or have suggestions, please contact the researcher: Emma M. Sunnassee ((336) 848-0426; [emsunnas@uncg.edu](mailto:emsunnas@uncg.edu)), or dissertation advisors (Dr. Ayesha Boyce at [Ayesha.boyce@uncg.edu](mailto:Ayesha.boyce@uncg.edu); Dr. John Willse at (336) 446-9428 or [jtwillse@uncg.edu](mailto:jtwillse@uncg.edu)).

If you have any concerns about your rights, how you are being treated, concerns or complaints about this project or benefits or risks associated with being in this study please contact the Office of Research Integrity at UNCG toll-free at (855)-251-2351.

**Voluntary Consent by Participant:**

By signing this consent form/completing this survey/activity (used for an IRB-approved waiver of signature) you are agreeing that you read, or it has been read to you, and you fully understand the contents of this document and are openly willing consent to take part in this study. All your questions concerning this study have been answered. By signing this form, you are agreeing that you are 18 years of age or older and are agreeing to participate, in this study described to you by Emma M. Sunnassee.

Signature: \_\_\_\_\_ Date: \_\_\_\_\_

## APPENDIX I

### INTERVIEW PROTOCOL

#### CONTRIBUTION ANALYSIS PRACTITIONERS INTERVIEW QUESTIONS

##### REVIEW CONTEXT & EVALUAND DETAILS

- 1) Can you briefly describe the evaluand, its objectives, & your role?
  - a) When did you join the evaluation; was the intervention already underway?
  - b) Were you part of an evaluation team?
  - c) Who else was involved the evaluation?
- 2) What the motivation to use CA for this evaluation?
- 3) How would you describe the complexity of the intervention (e.g., simple, complicated, complex)?

##### STAKEHOLDER INVOLVEMENT

- 4) Which stakeholder groups, if any, were involved in the evaluation (e.g., intended beneficiaries/recipients, program developers, program managers, implementers, frontline workers)?
 

SH #1: \_\_\_\_\_ SH #2: \_\_\_\_\_

SH #3: \_\_\_\_\_ SH #4: \_\_\_\_\_
- 5) Rate each of the stakeholder groups on the extent to which they participated in the evaluation tasks from 1 (never) to 5 (always).

Evaluation Phase	SH #1	SH #2	SH #3	SH #4
Program description				
Developing evaluation scope & questions				
Designing the evaluation				
Instrument development				
Collecting data & evidence				
Data processing & analysis				
Interpretation				
Reporting, sharing & discussing findings				

- 6) Generally speaking, who made decisions about the planning, design or conduct of the evaluation?
 

1 (Stakeholders)    2    3    4    5 (Evaluators)

##### EVALUATOR ROLE

- 7) How would you describe your role in this evaluation? (e.g., Achiever working with program manager, Educator of clients, Facilitator of local change, Judge of program, Liaison between stakeholders, benefactors, & donors, Knowledge broker, Methodological expert, Part of the program team, Resource for stakeholders, Shepherd for public good)
- 8) Is this consistent with other evaluations (where in which you did not use CA)?

### CONTRIBUTION ANALYSIS: 7-STEP PROCESS

This section seeks to understand how CA in practice. Specifically, I am interested in understanding how each step unfolded, special concerns you had to consider (e.g. planning ahead, capacity building to ensure data access), mechanisms that assisted in fulfilling each objectives, barriers/challenges that you encountered & how they can be mitigated/addressed, & what, if any, adaptations were made. **Use the CA reference table provided on the next page.**

An important piece of CA is the theory of change. Can you walk me through the process of developing the theory of change? What were some barriers encountered? What mechanisms assisted in meeting its objective?

#### 9) Theory-of-Change

- a) Step 1:
- b) Step 2
  - i) Sources for postulated theory of change:
- c) Step 3:
  - i) Sources of evidence:
- d) Overall, what were the biggest challenges/barriers you encountered in developing the ToC?
- e) Lessons learned/considerations for future applications regarding developing the ToC?

#### 10) Results-chain

- a) Step 4
- b) Step 5
- c) Challenges/barriers In developing the results-chain?
- d) Lessons learned/considerations for future application?

#### 11) Contribution Story

- a) Step 6
- b) Step 7 (if applicable)
- c) Challenges/barriers in developing the contribution story?
- d) Lessons learned/considerations for future application?

*This concludes my questions for you. Do you have anything else you'd like to share about CA or the study? Thank you very much for, your participation in this research & this interview!*

Table 1. Reference Table for CA-Specific Items

• Mechanisms	• CA Steps
A. Theory of Change (ToC)	Step 1: Set out the cause-effect issue to be addressed <ul style="list-style-type: none"> <li>• Acknowledge the attribution problem.</li> <li>• Scope the problem: determine the specific cause-effect question being addressed; determine the level of confidence needed in answering the question</li> <li>• Explore the contribution expectations</li> <li>• Determine the other key influencing factors</li> <li>• Assess the plausibility of expected contribution given the intervention size &amp; reach</li> </ul>
	Step 2: Develop the postulated theory of change & risks to it <ul style="list-style-type: none"> <li>• Set out the postulated theory of change of the intervention, including: identify the roles of the other influencing factors, identify the risks to the assumptions &amp; links in the theory of change, &amp; determine how contested is the postulated theory of change</li> </ul>
	Step 3: Gather the existing evidence on the theory of change <ul style="list-style-type: none"> <li>• Assess the strengths &amp; weaknesses of the links in the theory of change</li> <li>• Gather the evidence that exists from previous measurement, past evaluations, &amp; relevant research (1) for the observed results, (2) for each of the links in the results chain, &amp; (3) for the other influencing factors</li> </ul>
	
B. Results chain (program theory)	Step 4: Assemble & assess the contribution story, & challenges to it <ul style="list-style-type: none"> <li>• Set out the contribution story</li> <li>• Assess the strengths &amp; weaknesses in the postulated theory of change in light of the available evidence, &amp; the relevance of the other influencing factors</li> <li>• If needed, refine or update the theory of change</li> </ul>
	Step 5: Seek out additional evidence <ul style="list-style-type: none"> <li>• Determine what kind of additional evidence is needed to enhance the credibility of the contribution story.</li> <li>• Gather new evidence</li> </ul>
	
C. Contribution Story	Step 6: Revise & strengthen the contribution story <ul style="list-style-type: none"> <li>• Build the more credible contribution story</li> <li>• Reassess its strengths &amp; weaknesses</li> <li>• Revisit Step 4</li> </ul>
	Step 7: In complex settings, assemble & assess the complex contribution story <ul style="list-style-type: none"> <li>• Develop the contribution story for each additional sub-theory of change (using Steps 2-6)</li> <li>• Develop the contribution story for any general theory of change (using Steps 2-6)</li> </ul>

## APPENDIX J

### PRELIMINARY POOL OF INTERVIEW QUESTIONS

#### General Questions

##### **The Evaluand**

1. Can you describe the evaluand (program being evaluated)?
  - a. Program objective details:
  - b. Timeline of the program (intervention, expected outcomes, long-term outcomes)
2. How would you describe the context of the program?
3. Can you describe the stakeholder groups involved in the program (organization, program beneficiaries, etc.)?

##### **The Evaluation**

4. What is the purpose of this evaluation?
5. What is the timeline allotted to this evaluation?
  - a. How does it align with the program's timeline?
6. How would you describe the complexity (e.g., cross-cultural evaluation, the complexity of intervention and ToC, conditions of funding, etc.) of this evaluation?

##### **Stakeholder Involvement**

7. Can you describe the stakeholder groups involved in the evaluation (e.g., funders/sponsors, program beneficiaries, program managers)?
8. Overall, how receptive are the stakeholders in partaking in the evaluation process?
  - a. What barriers (e.g., time, location, scheduling) have you encountered regarding stakeholder engagement?
  - b. What has facilitated stakeholder engagement?
9. Have you noticed any changes in stakeholders throughout their involvement with the evaluation (e.g., increase in knowledge of program theory, attitudes/beliefs, ownership of the program, long-term planning)?

##### **Evaluator Role**

10. How would you describe your role in this evaluation?
  - a. Is this consistent with other evaluations (wherein which you did not use CA)?
11. How involved is this evaluation for an evaluator (e.g., big workload, quick turnaround deadlines for deliverables and information, high frequency of communication)?
  - a. In your opinion, would it require a team of evaluators?
12. How would you describe your relationship with the stakeholders?
  - a. How does it compare to other evaluations (those you did not use CA)?

### Contribution Analysis

13. What was the motivation to use CA for this evaluation?
14. What factors of the evaluand influenced/directed towards a CA approach?
15. What factors of the evaluation purpose influenced/directed towards CA as an evaluation approach?
16. How would you describe the evaluation approach (e.g., a fusion of CA with other evaluation approaches, Results-Based Management + CA, Utilization-Focused Evaluation + CA)?
17. How was CA introduced (pitched) to the evaluation's commissioners?
  - a. Did you encounter any barriers?
18. How was CA introduced to the stakeholders identified to participate in the evaluation?
  - a. How did they receive it?
  - b. What were some barriers you encountered in introducing CA to the stakeholders?
  - c. How would you describe their level of understanding of CA (e.g., distinctions between contribution vs. attribution, types of questions CA cannot answer)?

### CA 6-Step Process

Key Mechanisms of CA	Steps of CA
A. Theory-of-change (ToC)	1. Determine the cause and effect issue to be addressed 2. Develop a theory of change and risks to its success 3. Generate evidence in response to the theory of change
B. Results chain (program theory)	4. Assemble the contribution story, and outline the challenges to it 5. Seek out additional evidence
C. Contribution Story	6. Revise and strengthen the contribution story *Return to Step 4 if necessary

19. Of the six outlined steps by Mayne, can you walk me through your experience and activities that took place as the evaluation progressed?
20. An important piece of CA is the theory of change. Can you walk me through the process of developing the theory of change?
  - a. What types of challenges did you encounter in this process?
  - b. What facilitated/enhanced the development of the ToC?
  - c. Lessons learned/considerations for future application regarding developing the ToC?
21. In general, were you able to follow the 6-steps as outlined?
  - a. If not, why?
  - b. What steps did you modify, in any? Did you expect to have to make modifications?
  - c. What were some barriers you encountered in following Mayne's 6-step process?

- d. What factors facilitated the 6-step process?

### **Level of CA**

22. Level of CA Mayne identifies three levels of contribution analysis investigation (minimal, indirect influence, direct influence). How would you describe the level of CA for this evaluation?

### **Knowledge Management/Learning**

23. Are there any mechanisms (formal or informal) that have formed to stakeholders are new knowledge and/or facilitate learning (e.g., WhatsApp group chat, shared Google Drive folder, etc.)?

### **Pro/Cons**

24. What has CA facilitated/enhanced in terms of the evaluation and its purpose?  
 25. What has CA facilitated/enhanced in terms of the evaluand?  
 26. What barriers have you encountered applying CA? How do you think those could be addressed/prevented?  
 27. Have you witnessed any unexpected developments (because of the application of CA)?  
 28. What additional uses can you see CA enhancing/facilitating?  
 29. What advice do you have for evaluators seeking to apply this method?  
 a. Lessons learned?  
 b. What would you have done differently?

### **Theory versus Practice**

30. What has aligned with what is theorized about CA versus your experience in its application?  
 31. What has not aligned with what is theorized about CA versus your experience in its application?  
 a. How can CA be practiced more effectively?  
 b. Is an emphasis on XX impeding the overall purpose?  
 c. Did the conditions of the evaluand and overall context facilitate the evaluation?

**APPENDIX K**  
**CASE EXTRACTION FORM**

<b>Categories</b>	<b>Codes</b>
<b>About the evaluand:</b>	
The types of settings of Intervention	
The target population of the program:	
Level of complexity of intervention (ToC):	
Rationale for CA	
<b>Evaluation:</b>	
Type:	performance evaluation, evaluation capacity, evaluation practice, program evaluation, impact evaluation
Scale of evaluation	
Scope of Intervention	
Purpose of the evaluation	
The methodologies and research strategies employed:	
Complementary approaches to CA:	
General facilitators	
General barriers.:	
Stakeholder involvement	
<b>About CA:</b>	
General CA process	
Development of Theory-of-change	
Adherence to CA tenets:	
Lessons learned:	
Benefits and challenges:	
Facilitators of the approach:	
Barriers of the approach:	
Plausibility of ToC	
Depth of CA - level:	
<b>Six-step Process:</b>	
Step 1	
....	
Step 6	
<b>Contextual Characteristics:</b>	
Country of evaluation	
Evaluation Team:	
# of team members	

---

**Demographic variables (of Practitioner)**

---

Education

Profession

Ethnicity

Gender

Evaluation experience (# of years)

Expertise

Evaluation training (type of degree)

Geographical location

---

## APPENDIX L

### SOURCES OF EMPIRICAL CA CASES BY YEAR

	2007	2008	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Grand Total
	2	1	5	5	11	3	9	9	4	10	13	15	2	84
<i>Eval.</i>					6		2	2	1	1		1		13
<i>CJPE</i>	1			1	5					2		1		10
<i>AJE</i>							1				2	2		5
<i>M&amp;CN</i>												5		5
<i>JDE</i>			1	1			1			1				4
<i>E&amp;PP</i>										1			2	3
<i>RE</i>								2			1			3
<i>Cont.SS</i>						2								2
<i>IDS</i>							2							2
<i>IJHPM</i>											2			2
<i>MT</i>								2						2
<i>PM&amp;M</i>			1	1										2
<i>AEJ</i>							1							1
<i>AM</i>										1		1		2
<i>AS</i>											1			1
<i>BMC-HSR</i>											1			1
<i>BMC-PH</i>												1		1
<i>CDI-PP</i>												1		1
<i>Cog. SS</i>											1			1
<i>D&amp;AR</i>			1											1
<i>DEP&amp;P</i>												1		1
<i>EJA</i>	1													1
<i>EJHE</i>											1			1
<i>ES&amp;P</i>											1			1
<i>ESiF</i>										1				1
<i>Eu Conf. on RMBMS</i>								1						1
<i>FS</i>								1						1
<i>GeoJournal</i>											1			1

	2007	2008	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Grand Total
<i>GHP</i>										1				1
<i>HPI</i>							1							1
<i>HRH</i>			1											1
<i>HRP&amp;S</i>									1					1
<i>IJHR</i>												1		1
<i>IJPSM</i>									1					1
<i>JAFSCD</i>										1				1
<i>JAG</i>												1		1
<i>JBI</i>									1					1
<i>JMDE</i>		1												1
<i>NDE</i>								1						1
<i>Palgrave Comm.</i>										1				1
<i>Pflege</i>											1			1
<i>R&amp;SER</i>											1			1
<i>SEE</i>							1							1

**APPENDIX M**  
**CA QUALITY MARKERS**

**Quality Markers of CA**

A “Good” Theory-of-Change \*(from Mayne, 2012a, p. 273) includes:

Table 26

Markers of a “Good” Theory of Change

ToC Elements	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
A results (causal) chain showing the basic logic of the intervention							
The underlying assumptions behind the links in the results chain							
The risks to each link occurring							
Identification of unintended effects							
Identification of possible alternative rival explanations							
Total Identified							
Remarks							

*Note.* Source: Mayne, 2012a, p. 273

Table 27

## Spheres of Influence Investigated by Case

<b>Spheres of Influence</b>	<b>Case 1</b>	<b>Case 2</b>	<b>Case 3</b>	<b>Case 4</b>	<b>Case 5</b>	<b>Case 6</b>	<b>Case 7</b>
Direct control							
Direct influence							
Indirect influence							
Remarks							

The evidence analysis table was adapted from the REF framework to evaluate the quality of evidence used to verify the ToC for the cases in the multiple-case study.

Table 28

## Evidence Table

<b>REF Component</b>	<b>Explanation</b>
1. Description	
2. Type	<ul style="list-style-type: none"> <li>- Direct rival</li> <li>- Commingled rival</li> <li>- Implementation rival (factors)</li> </ul>
3. Level (of external influencing factors)	<ul style="list-style-type: none"> <li>- Individual</li> <li>- Interpersonal</li> <li>- Institutional</li> <li>- Infra-structural</li> </ul>
4. Identifiers	Specific data patterns indicating the presence of rivals
5. Degree of influence	<ol style="list-style-type: none"> <li>1. Certainty</li> <li>2. Robustness</li> <li>3. Range</li> <li>4. Prevalence</li> <li>5. Generalized specificity</li> </ol>
6. Implications	

*Note.* Adapted from REF Framework.