Fall 2020

The Use of Graphic Conceptual Models in Theory-Driven Evaluation

Susana Marianne Bonis
Claremont Graduate University

The Use of Graphic Conceptual Models in Theory-Driven Evaluation

By

Susana Marianne Bonis

Claremont Graduate University

2020

APPROVAL OF THE REVIEW COMMITTEE

This dissertation has been duly read, reviewed, and critiqued by the Committee listed below, which hereby approves the manuscript of Susana Marianne Bonis as fulfilling the scope and quality requirements for meriting the degree of Doctor of Philosophy in Psychology.

Dr. Stewart I. Donaldson, Chair
Distinguished University Professor,
Claremont Graduate University

Dr. Michelle C. Bligh,
Professor and Dean,
School of Social Science, Policy, and Evaluation,
Claremont Graduate University

Dr. Michael Quinn Patton,
Professor of Evaluation Practice,
Claremont Graduate University

Dr. Huey T. Chen, Professor and Director,
Center for Evaluation and Applied Research,
Mercer University

ABSTRACT

The Use of Graphic Conceptual Models in Theory-Driven Evaluation

By

Susana Marianne Bonis

Claremont Graduate University

2020

Theory-driven evaluation has been adopted by numerous philanthropic organizations and government agencies across the world. Many evaluators also have embraced elements of theory-driven evaluation, regardless of their approach to evaluation. In theory-driven evaluation, the beliefs or assumptions behind an intervention are made explicit and used to guide the evaluation. These may be based on the perspectives and experiences of stakeholders, social science theory, observation, and previous evaluation and research. A graphic conceptual model, often in the form of a logic model or theory of change, is commonly developed to show how an intervention is expected to lead to desired outcomes. While there have been various conceptual, methodological, and theoretical writings about theory-driven evaluation since it came to prominence in the early 1990s, empirical research on aspects of theory-driven evaluation are more recent.

This study took a comprehensive approach to understanding the development and use of graphic conceptual models in theory-driven evaluation practice. It also investigated issues that have been raised in relation to the use of models, such as the need to adapt them in different cultural contexts and more complex environments, and the possible benefits of identifying program archetypes. The study involved two parts: a content analysis of 116 published

evaluations in peer-reviewed journals focused on public health, education, and evaluation and a survey of 141 evaluators. A distinction of this study from other research on theory-driven evaluation is its deeper examination of the graphic conceptual models included in published articles on evaluation studies.

Findings reveal several encouraging practices in the design, development, and use of graphic conceptual models in evaluation. Graphic conceptual models were developed using multiple sources, including theory and research, and they often depicted detailed paths of mediation and moderation. While most models reviewed were linear, there was indication that evaluators are exploring alternative ways of describing programs whose path of change is less predictable and describing programs in different cultural contexts. Co-design of models through a participatory process was common and resulted in shared understanding of how an intervention is expected to lead to change. In many instances, program theory was used to develop evaluation questions and the evaluations measured constructs and relationships outlined in the model. Finally, this study took a step towards understanding possible archetypes. While programs may focus on distinct issues and serve different people, there may be similarities across design and implementation characteristics and impact pathways. Archetypes could serve as a heuristic device for practitioners and help build knowledge about programs in a systematic manner.

The results indicate that theory-driven evaluation is practiced across various disciplines and cultural contexts in ways that adhere to many of its core principles. The content analysis involved a larger sample than previous empirical research on theory-driven evaluation, allowing for confirmation and contradiction of earlier findings that could add to the knowledge base and point to directions for future research. This study also gathered information on emerging issues and approaches proposed to strengthen theory-driven evaluation that have not been studied

empirically on a large scale. The findings around use of graphic conceptual models in complex interventions and in interventions carried out in different cultural contexts, as well as on program archetypes, could serve as part of the foundation for further inquiry. Overall, this study contributes to a better understanding of how evaluators develop and use graphic conceptual models in evaluation practice and suggests possible future directions for research on theory-driven evaluation practice.

Key words: theory-driven evaluation, graphic conceptual models, program theory, program evaluation, research on evaluation

DEDICATION

I dedicate this dissertation to my parents, Dr. Joseph A. Bonis and Dr. Susana Marta Topor Bonis, and grandparents, Dr. *György* Bajkay Topor and Teresa S. Topor; they never had the chance to know this was a path that I would take, but they influenced and empowered me by their example, love, and support when I was a child. I also dedicate it to my two daughters, Maritza and Mirka, who motivated and inspired me to persevere on this journey, and who brought me joy along the way.

ACKNOWLEDGMENTS

I am deeply grateful to have reached this moment. My path to this goal was long, not always linear, and influenced by many factors sometimes in my control and sometimes not. I never lost sight of the goal, in large part because of my faith and because I had such a remarkable circle of support.

I begin by acknowledging and thanking the chair of my committee, Dr. Stewart I. Donaldson. As someone for whom theory-driven evaluation resonated from the earliest days of her career, it was an exceptional opportunity for me to learn from a scholar who has contributed so significantly to the development and understanding of this evaluation theory. I appreciate his guidance, support, and patience.

I also am truly thankful to the other members of my committee. Dr. Michelle C. Bligh supported me when I pursued my co-concentration in organizational behavior and evaluation and applied research. She remained on my committee even once my dissertation focused on evaluation. I appreciate her sincere encouragement and thoughtful questions to help me improve my work. Dr. Michael Quinn Patton joined my committee late in the process, yet through his insightful questions and comments I learned so very much and was able to strengthen the narrative of my dissertation. Finally, it was a great privilege to have Dr. Huey T. Chen, one of the founders of theory-driven evaluation, serve as my external examiner. Overall, I am appreciative of the opportunity to have received constructive feedback from such a distinguished committee.

In addition to my formal academic committee, I am immensely grateful to Dr. Jorja Leap for always believing in me and supporting me. I could not have completed this journey without her. I also give profound thanks to Dr. Leslie Fierro, Dr. Dreolin Fleischer, Dr. Todd Franke, and

Robert Fierro for their generosity with their time and knowledge as I worked through different aspects of this study. They were willing to listen, review, share ideas, and offer encouragement whenever I asked. I also wish to express appreciation to Tiffany McBride and Devin Larsen for spending several months coding published evaluation studies with me. Their intelligence, commitment, and good nature made the process an enjoyable and productive one. From that experience, I also made a wonderful new friend and colleague in Tiffany. I must acknowledge her separately as well. Tiffany remained a thought partner, editor, and advocate throughout the process; her support was pivotal. I also wish to acknowledge Ray Ramirez and Dr. Christina Christie, without whom I would not even have started this path; they encouraged my early interest in and study of evaluation.

I was blessed with many other family members and friends without whom I could not have completed this journey. Monica Chu, my cousin who is more like a sister, stood by me the whole way. She, along with a few others, knows the great challenges that I had to face while trying to complete my doctoral program and this dissertation. We supported each other and grew together. I could reach this point only because she was with me. My dear friend Miguel Suaste played a similar role. He didn't let me give up and helped me think of strategies to keep going. I am also grateful to Dr. Sylvia Lang, Dr. Saba Medhane, Anabella Bastida, Vanessa Aramayo, Pamela Agustin, Joel Perez, Ismael Castro, James Griffin, and Natasha Morisawa for motivating me and reminding me that I *could* do this.

Then, I treasure the support of my daughter Maritza, who walking the grounds of Claremont with me when she was about eleven years old told me that I *had* to do this. Her words deeply touched my heart. My choice to complete my doctoral studies was mine, but at that moment I realized that I was setting an example for her and her sister. I had to model

perseverance, working towards a goal even as challenges abounded. I also wanted them to see that with a good education a person can stand strongly on their own. I started my doctoral program before Martiza knew what that is; today she talks about possibly working towards her Ph.D. one day. My daughter Mirka was born while I was in the program. Her whole life she has seen me working on some project tied to it. There were parts of my study that were done sitting in the car, on a bleacher, or in a chair by a baseball game or track meet. Maritza and Mirka's bright smiles and sweet natures made it easier to juggle being a mom, student, and working professional.

Finally, I am grateful to many talented and dedicated colleagues. I have been blessed with a very meaningful career since I moved to Los Angeles from Maryland. The incredible nonprofit organizations, foundations, staff, and community members that I have worked with inspired me to continue with my studies. It is my intention to become a better partner in building organizations and programs for social justice and equity.

Completing this dissertation is one step in that direction. One chapter may close, but our learning never ends. It is with humility, gratitude, and optimism that I move on to my next chapter and offer this document as a contribution to my field and future research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE
INTRODUCTION

Individuals, organizations, and governments across the world have responded to social challenges by designing programs and policies intended to improve well-being. Donaldson (2003) defines social programming as "organized efforts to train, educate, and/or change human behavior to achieve participant and social betterment" (110). In the United States, programs designed to address social problems such as unemployment, crime, and unequal access to medical care proliferated during the Kennedy and Johnson eras (Shadish, Cook, and Leviton, 1991; Rossi, Freeman, and Lipsey, 1999). During this time as well, aid to foreign countries to promote social and economic development saw a transformation with the passage of the Foreign Assistance Act of 1961. The proliferation of psychological, educational, and behavioral interventions to address individual and social problems gave rise to the need to systematically determine their efficacy and effectiveness. Program evaluation was a response to this need.

The field of program evaluation has burgeoned over the last four decades. Evidence of this growth includes the increase in approaches to evaluation (Alkin, 2004; Shadish, Cook, and Leviton, 1991) and the increase in professional associations to support evaluation research and practice, and to set guidelines for the profession. While there were only five such associations in 1990, there are now over 200 associations worldwide (Donaldson, 2020). The American Evaluation Association alone has approximately 8,000 members (Donaldson, 2020).

Social programs as a solution to social problems hold promise. Lipsey and Wilson (1993) found positive findings for social programs after conducting a meta-analysis of 302 meta-analyses of over 10,000 studies. They contended, however, that the questions of interest in program evaluation should focus less on whether or not a treatment or intervention works, but on

how it works and how it can be improved. This recommendation is reinforced by an earlier meta-analysis by Lipsey (1988), which discovered that many studies often reported comparisons on outcomes between treatments and control conditions, with little attention to program theory or mediating and moderating factors. Lipsey described the unknown space between a program's input and output as the "black box." Too many evaluations had focused on measuring outputs while attributing the observed difference to the input. Theory-driven evaluation gained popularity in the 1990s as a way to open the black box (Gargani, 2003; Stame, 2004; Coryn et al., 2011).

Overview of Theory-Driven Evaluation

In theory-driven evaluation, the beliefs or assumptions behind a program are made explicit. These assumptions may be based on the perspectives and experiences of stakeholders, social science theory, observation, and previous evaluation and research (Donaldson, 2001, 2007; Leeuw and Donaldson, 2015; Weiss, 1997). Leeuw and Donaldson (2015) propose two typologies of theory in the evaluation field. The first typology involves theories derived from the practice of stakeholders and evaluators and their sets of underlying assumptions about an intervention. The second one draws on theories based on research from the social, behavioral, and policy sciences that offer insight about mechanisms and contexts underlying policies and programs. Ideally, evaluation is based on both types of theories.

Relationships between program activities and outcomes, and between proximal and distal outcomes, are delineated in theory-driven evaluation. Several authors differentiate between these sets of relationships using diverse terms. For example, Weiss (1997b) distinguishes between implementation theory, which describes the program's inputs, activities, and outputs, and programmatic theory, which describes the outcome chain. Chen (1990) refers to the distinction

as the action model and the change model, while Donaldson (2007) uses the terms program process theory and program impact theory.

The various terms, however, all draw attention to the importance of focusing on the mechanisms that make change happen. “The mechanism of change is not the program activities per se, but the response that the activities generate,” (Weiss, 1997b, p. 46). The attention to indirect or mediator relationships as opposed to direct effects is a hallmark of theory-driven evaluation and is a means to make more visible the transformations taking place within the “black box” criticized by Lipsey. Also important are moderator relationships that influence program results, such as participant characteristics and context. These mediator and moderator relationships comprise the program theory and are often depicted as a graphic conceptual model. With the model developed, evaluation questions can be identified and answered (Donaldson, 2007). Chen (2005) distinguishes four types of theory-driven evaluation, depending on which part of the conceptual model the evaluation is focused: theory-driven process evaluation, intervening mechanism evaluation, moderating mechanism evaluation, and integrative process/outcome evaluation. Creating models of program theory are beneficial not only in program evaluation but also in program development (Donaldson, 2001). Models of program theory can generate discussion around important questions for program success, such as whether or not it is reasonable to expect certain changes to occur based on past experience and research findings and given current resources, and whether or not the intervention is of sufficient dosage to lead to the expected changes.

Proponents of theory-driven evaluation cite several benefits for evaluators, program practitioners, funders, and policy makers (Weiss, 1997a; Donaldson, 2007). Evaluators are able to better focus their study on key questions, offer information on interim markers of progress,

and provide better explanations for why a program is more or less successful. Practitioners involved in developing program theory at the start of their program may have a stronger shared understanding of their program, thereby potentially improving implementation. The subsequent evaluation may provide them with information on which mechanisms of change worked and which didn't, how and why. This information could be used to improve local programming. It also could benefit others seeking to replicate a similar program elsewhere and could help generate new theories for social betterment.

Chen and Rossi (1983, 1987) were early advocates of including a strong theoretical component in evaluation, especially good social science theory. Rossi made theory-driven evaluation one of three core components of his approach to evaluation; the other two components being tailored evaluation and comprehensive evaluation. Theory-driven evaluation permits the option of either tailoring an evaluation to focus on particular aspects of program theory, or conducting a comprehensive evaluation studying the entire program theory. Rossi, as an evaluator whose approach integrates the work of several other evaluators (Shadish, Cook, and Leviton, 1991) and as the author of one of the most popular textbooks on program evaluation, *Evaluation: A Systematic Approach,* had much influence on the field of program evaluation. Chen also did much to advance theory-driven evaluation by writing one of the first textbooks focused on the approach, *Theory-Driven Evaluations* (1990). Weiss' 1995 article *"Nothing as practical as a good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families"* has become a classic in the field of evaluation and is one of its most influential articles. Weiss emphasized the importance of basing evaluations on solid theories of change that underlie interventions. Weiss brought rigor to theory-driven evaluation, encouraging stakeholders not only to have conversations on what they are trying to do and why,

but also to base their theories of change on social science knowledge and research. Individuals involved with a program may have different ideas about how a program should work. Multiple theories of change must be brought to light so a consensus can be reached on which deserve to be tested. Pawson and Tilley (1997) emphasized context and mechanisms as essential components of program theory. In their realist approach, it is not programs that make things change, but people embedded in their context who, when exposed to programs, do something to activate given mechanisms, and change (pp. 32-34). Donaldson (2007) proposed the term "program-theory driven evaluation science" to better capture the essence of this approach to evaluation, which while placing program theory at the center of the evaluation does not do so at the expense of rigorous and systematic empirical evidence. To make the approach more accessible to evaluators, Donaldson recommends a concise three-step model for conducting an evaluation: developing program impact theory, formulating and prioritizing evaluation questions, and answering evaluation questions. Leeuw and Donaldson (2015), as noted earlier, suggested two typologies of theory in evaluation: 1) "theories of policy makers, stakeholders and evaluators underlying their professional work in making policies and doing evaluations" (p. 468) and 2) "scientific theories capable of contextualizing and explaining the consequences of policies, programs and evaluators' actions" (p. 470). The proposed typologies are offered to reduce confusion around the use of theory in evaluation. Leeuw and Donaldson also put forward promising approaches to combine different aspects of theory in evaluation, including theory knitting and theory layering. In theory knitting, previous theories are integrated into a single higher order theory. In theory layering, mechanisms of change are analyzed by different theories but viewed as part of a nested system with "upward causation".

There have been numerous conceptual, methodological, and theoretical writings about theory-driven evaluation that have introduced many terms, such as theory-based evaluation, theory of change, program logic, logic model, logical frameworks, outcomes hierarchies, intervening mechanism evaluation (Chen, 2005), realist or realistic evaluation (Pawson and Tilley, 1997), program theory-driven evaluation science (Donaldson, 2007), and others. While there may be slight differences in theory-oriented approaches, and in the terms used to describe them, the following are similarities which unite them (Stame, 2004):

1. Evaluation is based on an account of what may happen, as understood by stakeholders. Values are taken into account.
2. Some type of graphical conceptual model and/or narrative depicting program theory is developed.
3. Programs are considered in their context.
4. All methods are suitable. Program theory is developed first, and the evaluation is built around it.
5. There is a commitment to looking for causality (internal validity), while allowing for comparisons across different situations.

Theory-driven evaluation, in one form or another, has been adopted by many philanthropic organizations and government agencies, including the W.K. Kellogg Foundation (1998, 2000), the United Way of America (1996), the Centers for Disease Control and Prevention (Milstein, Wetterhall, and CDC Working Group, 2000), and the World Bank (Carvalho and White, 2004). The Government Accountability Office (2009) has even recommended theory-driven evaluation as one possible alternative to randomized controlled trials or randomized experiments. Many evaluators also have embraced elements of theory-

driven evaluation, regardless of their approach to evaluation. Crano (2003) observed that several papers presented at a symposium on evaluation dealt directly or indirectly with theory-driven evaluation. Donaldson (2007) points out that theory-driven evaluation is both a separate theory of evaluation practice and one whose elements have been incorporated into other theories of evaluation practice.

Overview of this Study

This study looks at the types of graphic conceptual models used in different cultural and programmatic contexts and substantive fields, and at the factors that influence the effectiveness of models in developing shared understanding about a program among stakeholders and in planning an evaluation. As research on evaluation, the study could be considered a practice component study (Henry and Mark, 2003), aiming to provide information about what works well in the development process of graphic conceptual models, and what the outcomes are of that process. It also may contribute to the knowledge base of "exemplary evaluations" (Donaldson, 2020), highlighting successful applications of theory-driven evaluation, enabling factors, and opportunities for improvement.

Understanding how graphic conceptual models could be better developed and used is important for the field of evaluation. Astbury and Leeuw (2010) pointed out that the significance and importance of mechanisms—so fundamental to theory-driven evaluation—does not seem to be understood by many evaluators. Frequently in graphic conceptual models, the focus is on describing the way a program fits together, often in a simple sequence of inputs, activities, outputs, and outcomes. There is less of an attempt to explain how the program works, with whom, and under what circumstances. In 1997, Weiss expressed concern that many evaluators developed program theory but then did not use it to guide the evaluation. Ten years later, Rogers

(2007) observed that while there was some improvement in the use of models to guide evaluation, the ways they were used were often simplistic. Evaluations frequently only consisted "of gathering evidence about each of the components in the model and answering the question "Did this happen?" about each one" (p. 65). Such an approach does not use the full potential of theory-driven evaluation. The findings of Coryn et al. (2011) corroborate Rogers' observation. "In many of the cases reviewed, the explication of a program theory was not perceptibly used in any meaningful way for conceptualizing, designing, or executing the evaluation reported and easily could have been accomplished using an alternative evaluation approach (e.g., goal-based or objectives-oriented)" (p.15). It is hoped that the findings of this study could help to improve the use of graphic conceptual models so that more of the potential of theory-driven evaluation could be realized.

CHAPTER TWO

REVIEW OF THE RELEVANT LITERATURE

Developing graphic conceptual models of program theory is a fundamental part of theory-driven evaluation. A graphic conceptual model in evaluation is a representation of elements and inter-relationships of an intervention. The evaluation literature includes descriptions of various types of graphic conceptual models. There have also been critiques and suggestions for improvement from various evaluation theorists around the development and use of graphic conceptual models.

Types of Graphic Conceptual Models

Funnell and Rogers (2011) highlight three approaches to describing program theory in a graphic conceptual model: outcomes chain logic models, pipeline logic models, and realist matrices. The outcomes chain logic model shows a sequence of results leading to the ultimate outcomes or impacts of interest. The pipeline logic model represents an intervention as a linear process, where inputs are entered at one end and impact appears at the other end.

While Funnell and Rogers do not use the terms program process theory and impact theory as used by Rossi et al. (1999), Chen (2005), and Donaldson (2007) to discuss outcomes chain logic models and pipeline logic models, the terms could apply. The program process theory describes the assumptions and expectations about how the program is supposed to operate. It combines the program's organizational plan (the personnel, structures, processes, resources, and general organization needed to support a program) and the service utilization plan (the steps to be taken in order to deliver the intervention to the target population). Impact theory, on the other hand, is a causal theory focusing on connecting a program to proximal and distal outcomes.

Funnell and Rogers' outcome chain logic models could be said to depict impact theory, while the pipeline logic models could be said to depict elements of the process theory and impact theory.

Examples of linear models depicting process theory and impact theory are included in Appendices A and B. While the most common linear models may involve a series of shapes and arrows presented linearly, there can be other ways to demonstrate linear relationships, as in the Logical Framework (Rosenberg, 1970) commonly used in international evaluation and Bennett's hierarchy (Rockwell and Bennet, 2004) developed for educational programs.

In the logic of realist evaluation, causal outcomes are seen as following from mechanisms acting in particular contexts. The emphasis placed on context, and how this interacts with causal mechanisms, is a unique feature of realist evaluation. Realist evaluation involves the development of context, mechanism, and outcome (CMO) configurations describing how an intervention is expected to work for different groups of people and the contextual conditions needed to trigger mechanisms, the causal process resulting in particular outcomes. Often, CMO configurations are also presented in a model which Funnell and Rogers call a "realist matrix." Examples of realist matrices may be found in Appendix C.

The model types discussed thus far have been linear, but other models have been developed to address some of the perceived shortcomings of linear models. One concern with linear models is their ability to address complicated and complex programs. Patton (2008) makes the following distinctions among simple, complicated, and complex programs. In simple programs the inter-relationships between elements are tight and centralized, and there is a simple linear cause and effect. In complicated programs, the relationships are looser but still clustered around a central core. Cause and effect, however, is non-linear. There also are fewer "knowns," but relationships are able to be modeled and predicted. In complex programs, there is no central

core for inter-relationships between elements; the center is loosely connected to a network. Cause and effect are difficult to understand and are not predictable. Lastly, the situation is understandable only in retrospect. Interventions may be complex as a result of their design or as a result of the environment in which they are implemented, or both.

To create graphic conceptual models for more complicated and complex programs, such as multi-site interventions, community initiatives, systems-level change, and programs taking a social determinants of health perspective, some evaluators have adopted principles of the social ecological model influenced by Brofenbrenner's Ecological Systems Theory (1994) and Stokol's Social Ecology Model of Health Promotion (1992, 1996). In the ecological model, problems are addressed at multiple levels and highlight the interaction and integration of biological, behavioral, environmental, and social determinants and the influence of organizations (e.g., work place and schools), other persons (e.g., family, friends, and peers), and public policies which together help individuals achieve certain outcomes. Sample socio-ecological models may be seen in Appendix D.

Multi-dimensional models are another way to demonstrate interactions at multiple levels (examples are presented in Appendix E). The "triple helix" is one such model (Etzkowitz, 2002); it is a spiral that captures multiple reciprocal relationships at different points in the process of an intervention. The first dimension of the triple helix model is internal transformation of each of the helices. The second dimension is the influence of one helix upon the other. The third dimension is the creation of a new overlay of trilateral networks and organizations from the interaction among the three helices.

A shift to network models as an alternative method to represent complexity has also been suggested, particularly for development aid programs (Davis, 2003), and system innovation

projects (Van Mierlo, Arkesteign, and Leeuwis, 2010). The argument in favor of network analysis is that it is about social relationships, which is at the core of many development aid programs and system innovation projects. Also, network models of change can incorporate mutual and circular processes of influence, as well as simple linear processes of change. In work involving systems change, the process of change has a long-time horizon and includes many interrelated actors and factors. The network perspective places greater emphasis on identifiable actors and the structure of relationships between them. This perspective also keeps attention on the problem analysis in addition to what is planned and then achieved, which is sometimes lost in approaches like the Logical Framework. Keeping the problematic in the model encourages comparison between the problem analysis and the plan, and between the plan and what is achieved. This can help distinguish failures caused by poor implementation versus failures caused by design-related issues. There are a wide range of methods for visualizing network structures. Two examples are included in Appendix F.

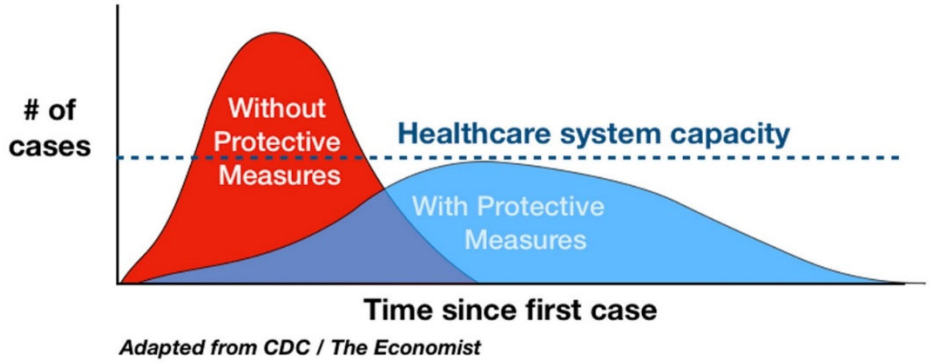
Another concern with linear models is that they may be limiting to certain groups. For example, Johnston (2002) commented that Western evaluation logic models are linear and interested in isolated domains, such as indicators or factors, while in the Ojibwe communities she works with, knowledge is holistic and the focus is on how spheres overlap to produce growth. LaFrance (2004) echoes this sentiment, explaining that traditional logic modeling formats might be too sequential and narrative-driven, and therefore not appropriate to capture the connections between program activities and underlying assumptions in Indian Country (Indian tribes and Alaskan Natives in North and South America, and the Pacific). Frazier-Anderson, Hood, and Hopson (2011) found a similar challenge when using a traditional logic model to describe the African American Culturally Responsive Evaluation System (ACESAS). Instead, they

transformed the traditional linear logic model with boxes and arrows into a Sankofa bird, an African American symbol representing the concept of reaching back into the past to gather strength for the future. The visual metaphor demonstrated how Culturally Responsive Evaluation fulfills the principles of Sankofa in its evaluation practice: looking back from a cultural and sociopolitical perspective in order to move forward (p. 362). In choosing this alternate model, they recognized that the African way of thinking is not necessarily linear and that a visual with symbolic significance for the community in which it will be used may make a greater impression on the community.

Visual metaphors are useful not only in diverse cultural settings. Keene (2011) and colleagues integrated web 2.0, graphic design, and data visualization with the linear logic model to create the "fuzzy logic model" for the Oregon Paint Stewardship Pilot Program (Figure G1 in Appendix G). The fuzzy logic model aims to improve the capacity of logic models to navigate non-linearity, feedback loops, and other concepts of complexity, and to expand the access and use of the evaluation process. Examples of models built around a visual metaphor are shared in Appendix G.

The power of graphic representations to convey a theory has been demonstrated recently with the "flatten the curve" graph (Figure 1). It provides a concise, straightforward story about the importance of slowing the rate of an epidemic through mitigation actions that even non-specialists can quickly understand.

Figure 1. The "Flatten the Curve" Graph



First introduced by the CDC in 2007, the graph represents two outcome scenarios: a pandemic where no interventions are taken (depicted by a curve with a steep peak) and a pandemic where containment, suppression, and mitigation measures are implemented (the flatter curve). Of the two curves presented, the flatter, lower one is the more desirable. With appropriate interventions, there is a decrease or delay in the peak of an epidemic wave so that it does not strain or exceed the capacity of healthcare systems. Over the last decade, various versions of the model were created. The "flatten the curve" graph reappeared in 2020 with the onset of the COVID-19 pandemic. This time, it received an addition of a dotted line by Harris (2020) to represent hospital capacity. The revised image, first introduced on Twitter, exploded across social and

other media and since has become a popular visualization tool to explain the benefits of community mitigation actions like social distancing and wearing masks.

Program Archetypes

In the last few years, there has been a suggestion among some theorists and researchers that program archetypes be identified. A program archetype is a core example of a type of program and includes the essential elements of a program for categorization. Pawson (2006) has noted that all programs are associated with some theory or theories and that while programs may be unique, there can only be a limited number of program theories. As a result, Pawson suggests that knowledge can be gained about the circumstances and success of program theories that can be of value to programs. Funnell and Rogers (2011) were the first evaluation researchers to propose program archetypes that share a similar logic or approach. These include programs that expect to achieve change by providing people with information, by motivating people through incentives or sanctions, by working case by case, by taking a community approach to change, and by offering a service. The archetypes can be used as a heuristic device or building blocks to facilitate the development of program theories that may then be depicted in a graphic conceptual model.

The advisory, public information, and education program archetype refers to programs that try to modify behavior through influencing attitudes, knowledge, and skills. The archetype makes two assumptions. First, behaviors targeted by the intervention are mainly the result of people knowing or not knowing something or having or not having the right attitudes and interests. Second, the information given will affect their knowledge, attitudes, and ways they behave. The archetype goes beyond a simple transmission of information but may use processes like demonstrations, advocacy, counseling, and facilitated learning processes. Examples include

training or education programs. The social science theories frequently associated with this archetype include diffusion theory, theory of reasoned action, and theory of planned change.

These social science theories also connect to the carrot and stick (incentives and sanctions) program archetypes. These archetypes refer to programs that try to influence behavior through use of incentives to promote positive behaviors or use of penalties and threats to discourage undesirable ones. Both are motivational programs that require an effective communication strategy for people to know about the rules guiding rewards and sanctions and to believe they will be applied. While the carrot and stick archetypes could be viewed as separate ones, they are often used in conjunction with one another.

In the case management program archetype, cases are bound together in the program by the nature of the problem that the program is trying to address. The program works with each case (e.g., individual, family, organization, or community) in a personalized manner, offering different services and treatments leading to individualized outcomes. There is a recognition that that there may be different factors and processes needed for behavior change. Two social science theories underlying case management program archetype include empowerment theory and the transtheoretical model (stages of change).

In the community capacity building program archetype, programs move to enhance community capacity to address certain challenges affecting the community or to seize opportunities. Community capacity building may be a cyclical process, with new capacity building on existing capacity. Common approaches for community capacity building programs include audits of formal skills and assets, community consultations, and facilitated strategic planning. The community capacity building archetype relates to empowerment theory.

In the direct service delivery program archetype, end results are achieved simply by having members of the target audience use the product or service. These programs do not try to change behavior except to the extent to which they try to encourage members of the target group to use or not use the program's service or to use it in a particular way.

Empirical Research on Theory-Driven Evaluation

While a theoretical base has been established for theory-driven evaluation and while its practice has proliferated, little empirical research has been done on the approach. This has been the case with several approaches to evaluation, leading Henry and Mark (2003) to call for greater research on evaluation to understand how it is being practiced, why, by whom, and to what effect (p. 69). Henry and Mark's agenda for evaluation research includes six types: research on evaluation outcomes, comparative research on evaluation practice, meta-evaluation, analog studies, practice components studies, and evaluation of technical assistance and teaching. Three recent empirical studies on theory-driven evaluation have added to the knowledge base on the practice of theory-driven evaluation.

Coryn et al. (2011) reviewed 45 cases of theory-driven evaluations from a twenty-year period (1990 to 2009). Cases were drawn from books or articles in evaluation-related journals and substantive journals in disciplinary areas where cases of theory-driven evaluations could be found. To determine how closely theory-driven evaluation practices align with key elements of theory-driven evaluation, they first identified fundamental principles of theory-driven evaluation developed through a systematic analysis of major theoretical writings on the approach. Five principles, and related sub-principles, were developed; the five principles are as follows (p. 205):

1. Theory-driven evaluations/evaluators should formulate a plausible program theory.
2. Theory-driven evaluations/evaluators should formulate and prioritize evaluation questions around a program theory.
3. Program theory should be used to guide planning, design, and execution of the evaluation under consideration of relevant contingencies.
4. Theory-driven evaluations/evaluators should measure constructs postulated in program theory.
5. Theory-driven evaluations/evaluators should identify breakdowns, side effects, determine program effectiveness (or efficacy), and explain cause-and-effect associations between theoretical constructs.

The findings of the study draw attention to some important trends in theory-driven evaluation practice that merit further investigation. For example, existing theory and research was the most common source for theory formulation, used in 41 of 45 cases (91%). The program theory articulated, however, was less frequently used to guide the evaluation. Program theory was used to develop evaluation questions in 76% of instances, and to prioritize evaluation questions in 22%. When program theory did guide the evaluation, process and outcome constructs identified in the program theory were both measured in nearly 50% of the articles (45% and 49%, respectively); contextual constructs were measured in 35%. True to the intent of theory-driven evaluation, though, many of the case studies attempted to describe (82%) and explain (67%) cause-and-effect associations between theoretical constructs, looking closely at mediators and moderators to more fully explain simple main causal effects.

Munter, Cobb, and Shekell (2016) undertook a study along similar lines to that of Coryn et al. Their investigation looked at the extent to which program theory was articulated and used

in evaluation studies of mathematics programs included in the U.S. Department of Education's What Works Clearinghouse. They found that while 27 of the 37 reports (73%) identified a guiding theory for programs, only six mapped the program's causal chain—including mediating and moderating variables linking the program's action with its intended outcome. And even among the six, attention to moderators was considered minimal. The models in the remaining 21 reports described primarily how a particular program was implemented, presenting resources, inputs, outputs, and so forth. Program theory often was not developed using social science theory or previous research and evaluation. Nineteen percent drew broadly from research literature while 30% provided a brief description of the program with no reference to the literature.
In reviewing use of program theory in the evaluation, the authors discovered that in 65% of the reports, research questions were limited to whether a program led to differences in outcomes. The percentage was similar for analyses conducted in the evaluation reports, with 62% describing a change in outcomes. In identifying construct measures, only 3 of the 14 studies that used multiple measures for mathematics achievements connected the choice of outcomes to program theory. Based on their assessment, Munter, Cobb, and Shekell concluded that few evaluations articulated program theory and used it to inform all phases of an evaluation.

Torres, Hopson, and Casey (2013) conducted a study of what logic model use looks like in practice. Logic models are one of the most popular versions of graphic conceptual models utilized by practioners (McLaughlin and Jordan, 1999; Gargani, 2003). Torres, Hopson, and Casey's investigation included case studies of educational reform projects funded by the National Science Foundation, in-depth interviews, and a forum where experts responded to findings. The study examined circumstances of use and provided a conceptual framework for examining and improving those circumstances. Six benefits of logic model use were found; these

include idea formation/organization, communication tool, internalized understanding, guiding framework, decision making, and shared understanding. The conceptual framework explored who is involved in the logic model development process, when and how the process is carried out, in what context the process takes place, and what type of model is developed.

Stakeholder engagement was critical in obtaining benefits from logic model use. The range of stakeholders varied, from program/organization staff at all levels, direct program beneficiaries, members of the broader community, partner organizations, and evaluators. Several factors influenced stakeholder engagement, including motivation/capacity of individuals to promote model development and/or use, prior knowledge of and experience with models, the degree of difference in perspectives among stakeholders, availability of stakeholders for involvement, and evaluator cultural competency and facilitation skills. The modeling process itself worked best when it was inclusive and honored and built upon the history and culture of stakeholders; challenged underlying, deeper assumptions; was flexible and well-resourced, and maintained the visibility and accessibility of the model to stakeholders over time. Contextual factors were also identified that positively influence logic model use; many of these are ones that can help promote evaluation use in general, such as the favorability of the organization's leadership, structure, processes, and culture towards learning and evaluation. Characteristics of models created were described, as opposed to discussing specific types of models.

Connection of this Study to Previous Research

The studies by Coryn et al. (2011), Torres, Hopson, and Casey (2013), and Munter, Cobb, and Shekell (2016) are important advances in understanding theory-driven evaluation and serve as potential launching points for further study. Like Coryn et al., the current research project looks at how core principles in theory-driven evaluation are applied in published

evaluation studies. The current project, however, more deeply examines graphic conceptual models: their development process, design, and use in evaluation. Then, the focus on the model development process ties this inquiry to the work of Torres, Hopson, and Casey and is another contribution to the evaluation field. Coryn et al. point out that despite nearly two decades since theory-driven evaluation became prominent, "documenting and recounting how the approach is enacted, procedures and analytic frameworks, and the subsequent uses of evaluation results is surprisingly low (p. 216)". This investigation also probes how graphic conceptual models are adapted based on complexity or cultural context, as well as what program archetypes could be helpful to the field. In this way, the study also explores empirically some critiques and opportunities for improvement proposed in relation to theory-driven evaluation (Johnston, 2002; LaFrance 2004; Patton, 2008; Frazier-Anderson, Hood, and Hopson, 2011; and Funnell and Rogers, 2011. The research project includes two components: a content analysis of published evaluation studies and a survey of evaluators.

This study involved two parts. In Part One, a content analysis was conducted of articles in peer-reviewed journals that described a program evaluation and that included a graphic conceptual model. The content analysis was guided by the following questions:

1. How do evaluators use graphic conceptual models in studies that they publish?
    a. How frequently are graphic conceptual models included in published evaluation studies?
    b. To what extent are core principles of theory-driven evaluation applied in the graphic conceptual models?
    c. What types of graphic conceptual models are most frequently used?

d. Do the model types vary by substantive field; by cultural context; and/or by program ecology (e.g., simple or complex)?
e. How are the graphic conceptual models used in the evaluation?

2. What are the most common program archetypes?

Part Two of this study involved an online survey designed to gather information about how evaluators develop and use graphic conceptual models. The survey was completed by evaluators, the majority of whom belong to the American Evaluation Association. The research questions from Part One also guided Part Two. Additional research questions were included that focused on the model development process, on using the process to promote shared understanding about a program among stakeholders, and on conditions affecting the use of models in evaluation. The survey format facilitated inquiry around these questions while the content analysis did not. In this way the survey was a means by which to gather additional information on the development and use of graphic conceptual models in evaluation.

CHAPTER THREE
CONTENT ANALYSIS OF PUBLISHED EVALUATION STUDIES

Research Questions and Hypotheses

As noted in Chapter One, the research questions that drove the content analysis included the following:

1. How do evaluators use graphic conceptual models in studies that they publish?
    a. How frequently are graphic conceptual models included in published evaluation studies?
    b. To what extent are core principles of theory-driven evaluation applied in the graphic conceptual models?
    c. What types of graphic conceptual models are most frequently used?
    d. Do the model types vary by substantive field; by cultural context; and/or by program ecology (e.g., simple or complex)?
    e. How are graphic conceptual models used in the evaluation?
2. What are the most common program archetypes?

For question 1a, the expectation was that a small number of articles would serve as case examples of theory-driven evaluation using a graphic conceptual model. In their search for articles and book chapters, Coryn et al. (2011) systematically reviewed databases in the social sciences between January 1990 and December 2009. They initially identified 205 articles and chapters for possible inclusion in their study and determined that only 45 fully met their criteria. Their finding indicated that in the twenty-years since the publication of Chen's *Theory-Driven Evaluations* (1990), an average of 10 articles and book chapters directly related to theory-driven

evaluation were published per year. The search parameters for the study presented in this paper were less extensive than that of Coryn et al: only articles including graphic conceptual models in a set number of peer-reviewed journals over a thirteen-year period. Given the scope of this study and the experience of Coryn et al., it was anticipated that the number of articles focused on theory-driven evaluation with a graphic conceptual model in the specified journals would be small.

Question 1b focuses on application of core principles of theory-driven evaluation, particularly as related to the development of graphic conceptual models. Three aspects of model development and the models themselves were of interest: the extent to which social science theory and research or evidence influenced model design and the level of detail described in the model in the form of mediators and moderators. It was hypothesized that graphic conceptual models in referred journals would be based more frequently on social science theory and research/previous evaluations than on stakeholder theory or program observation; that most graphic conceptual models would describe only one mediator relationship (e.g., program-mediator-outcome); and that moderator relationships would be infrequently described in graphic conceptual models.

The first hypothesis for question 1b was based on the similarity of this sample to that of Coryn et al.; both include articles from peer-reviewed journals. In their sample, Coryn et al., found that 91% of articles and chapters included graphic conceptual models based on social science theory. The second and third hypotheses built on the findings of Rogers et al. (2000) and of Coryn et al. (2011). Rogers et al. noted that at its simplest, program theory shows a single intermediate outcome by which a program achieves its ultimate outcome. More complex program theories show a series of intermediate outcomes. The authors observed that even the

inclusion of one mediating variable would be an improvement over some current practice. A decade later, Coryn et al., in their review of 45 articles describing theory-driven evaluations, discovered the following: in 82% of the cases, the evaluation described cause-and-effect associations between theoretical constructs; in 53% of cases, the evaluation explained differences in direction and/or strength of relationship between program and outcomes; and in 67% of cases, the evaluation explained the extent to which one construct accounts for or mediates the relationship between other constructs. These findings indicate that mediator and moderator relationships were explored in the evaluation. While the results are encouraging because such relationships are at the core of theory-driven evaluation, another discovery raises the question of when these critical relationships were identified. In only 51% of cases was program theory used to design, plan, and conduct an evaluation. This left open the possibility that mediating and moderating relationships were developed separately from the models of program theory.

Questions 1c and 1d center on the type of graphic conceptual models most frequently developed, and whether these vary based on circumstances. It was hypothesized that linear models would be the most common and that substantive field, cultural context, or program ecology would not strongly influence the type of model developed. The majority of guides around development of logic models or other graphic conceptual models take a linear approach. There are many such guides created by philanthropic foundations and government agencies, in both the United States and other countries. Some popular ones include the Innovation Network's *Logic Model Workbook* (2010), the CDC Evaluation Research Team's *Logic Model Basics* (2008), the European Commission's *Methodological Guidance for Evaluation* (2006), the World Bank's *The Logframe Handbook* (2005), the W.K. Kellogg Foundation's *Logic Model Development Guide* (2004), and the University of Wisconsin's *Enhancing Program Performance*

with Logic Models (2003). While it is true that in recent years there has been criticism of the value of linear approaches in certain contexts (as outlined in the literature review), it was not expected that alternative approaches had entered mainstream practice.

For question 1e it was hypothesized that implementation and effectiveness evaluation would predominate. Coryn et al. (2011) noted that in their sample several evaluations focused only on the presence or absence of elements in the model—asking more descriptive rather than evaluative questions. Rogers et al. (2000) observed a similar pattern. Then, as mentioned above, Coryn et al. found that 82% of evaluations described cause-and-effect associations between theoretical constructs while only 53% explained differences in direction and/or strength of the relationship between program and outcomes and 67% percent explained the extent to which one construct accounts for/mediates the relationship between other constructs.

A hypothesis was not developed for question two. Given the limited empirical information related to program archetypes, an inductive approach was taken to address it. Each coder wrote a brief description of the program and outcome pathway described in the model. Two coders then reviewed the narratives to identify potential archetypes.

Methods

Content analysis is unique in that it has both a quantitative (Krippendorff, 2004; Neuendorf, 2002) and a qualitative methodology (Patton, 2015; Berg, 2001), and it can be used in deductive and inductive ways. Quantitative content analysis is deductive, intended to test hypotheses or address questions generated from theories or previous empirical research. Qualitative content analysis is mainly inductive, grounding the investigation of topics and themes, and inferences drawn from them, in the data.

Quantitative content analysis has been a method used often in recent years to study evaluation practice. For example, the two empirical studies of theory-driven evaluation described earlier used the method, with Coryn et al. (2011) investigating 45 cases drawn from books or articles and Munter, Cobb, and Shekell (2016) reviewing 37 evaluation reports submitted to the U.S. Department of Education. Jacobson, Azzam, and Baez (2013) conducted a content analysis of 32 articles to determine the level of inclusion of people with intellectual, developmental, and psychiatric disabilities in the evaluation of programs aimed to serve them. Miller and Campbell (2006) examined 47 case examples of empowerment evaluation published over a decade. Christie and Fleischer (2010) performed a content analysis on 117 evaluation studies to determine the designs and data collection methods reportedly used in evaluation practice in light of federal guidelines enacted prior to 2004.

This investigation followed the approach of these studies, as well as the process of content analysis research outlined by Neuendorf (2002): (a) theoretical and conceptual backing, (b) conceptualization decisions, (c) operationalization measures, (d) coding schemes, (e) sampling, (f) training and initial reliability, (g) coding, (h) final reliability, and (i) tabulation and reporting. There are limitations to this approach, however, which are recognized in the section "Strengths and Limitations."

Journal sample. Peer-reviewed journals focused on program evaluation and journals in the fields of education and public health were the focus of this investigation. Criteria used for journal selection were as follows: (a) the journal's mission is to advance the field of evaluation, or the journal's focus is on education or public health, and (b) the journal is likely to include evaluation studies from various parts of the world. The initial proposal for this research called for the inclusion of journals on organizational development. This was not possible because the

primary peer-reviewed journals focused on organizational development did not meet criteria "b." These journals included very few evaluation studies. The paucity may be because organizational development is a relatively young field. It was more probable to find evaluations of organizational development interventions in an evaluation-focused journal than in a journal on organizational development. Table 1 lists the journals included in the content analysis.

Table 1. *Journals Included in Study*

Journal Title	Start Date	Issues per Year
American Journal of Evaluation	1998	4
Journal of Multidisciplinary Evaluation	2004	2
Evaluation and Program Planning	1978	4
Evaluation: The International Journal of Theory, Research and Practice	1995	4
Canadian Journal of Program Evaluation	1986	3
Evaluation Journal of Australasia	2001	2
Studies in Educational Evaluation	1975	1
Educational Evaluation and Policy Analysis	1979	4
American Journal of Public Health	1971	12
Health Promotion Practice	2000	6
Preventing Chronic Disease	2004	4

Article sample. The journals were searched for evaluation studies published between 2003 and 2016. By 2003, several guides to the development of graphic conceptual models were available, including Practical Concepts' guide to the logical framework (1979), the United Way's guide to outcome measurement (1997), the W.K. Kellogg Foundation's guide to logic model development (2003), and the University of Wisconsin Extension's guide to logic models (Taylor-Powell, Jones, and Henert, 2003). Several scholarly publications on theory-driven evaluation were also accessible, including Chen's *Theory-Driven Evaluations* (1990) and *Practical Program Evaluation* (2005), which elaborated upon the concepts of action theory and impact theory. Publications on the use of graphic conceptual models in evaluation have continued.

Recent contributions include Donaldson's *Program Theory-Driven Evaluation Science* (2007) and Funnell and Rogers' *Purposeful Program Theory* (2011). The numerous resources available on graphic conceptual models between 2003 and 2016 could be taken as indication of interest in their use. Thus, it was anticipated that there could be a high likelihood that evaluation studies during this period may include some type of graphic conceptual model.

Every journal publication between 2003 and 2016 was reviewed individually to identify articles that described a program evaluation and included a graphic conceptual model as part of the evaluation. First the table of contents was skimmed and then article abstracts were read. Articles describing a program evaluation were reviewed to determine if they included a graphic conceptual model. During the collection of articles for the sample, a record was kept of the number of articles in the journal, the number of articles on program evaluation, and the number of articles on program evaluation that included a graphic conceptual model. Editorials, commentary, announcements, and related pieces were not included in the total number of articles in a journal.

The review of the eleven journals resulted in 141 articles that described a program evaluation and included a graphic conceptual model. A distinction was not made between efficacy and effectiveness evaluations. Ten articles were used in training with the coding team and were not included in the final sample. Five were removed from the sample before coding because upon closer examination, it was apparent that the articles, while describing a program evaluation and including a graphic conceptual model, did not include enough information for coding to be meaningful. Ten articles included the context-mechanism-outcome (CMO) configuration used as the main structure for realist analysis. While realist evaluation is considered a part of theory-driven evaluation, the CMO tables in the articles could not be easily

analysed using the coding protocol developed for the study; there was ambiguity in the way that authors seemed to interpret terms like "context" and "mechanism." During training with the coders, it became clear that the CMO tables may need to be reviewed separately. Furthermore, several of the articles that presented a CMO configuration did not explain in detail the context of the program and/or the evaluation developed in response to the CMO configuration. This too, then, made it difficult to code these articles using the coding guide. As a result, the ten articles with a CMO configuration were removed from the sample. The final number of articles included in the analysis was 116.

Coding protocol. The coding guide included 50 codes at the start of the coding process: nineteen related to program context, five to program theory, twelve to the graphic conceptual model, thirteen to the evaluation, and one identified the article. The context codes were meant to assist with determining whether graphic conceptual models vary by substantive field; by cultural context; and/or by program ecology (e.g., simple, complicated, or complex). The codes tied to program theory and to graphic conceptual models were based on core principles put forth by prominent theorists focused on theory-driven evaluation. The protocol was shared for review by nine practicing evaluators. Their feedback was incorporated in revisions of the instrument. Table 2 lists the variables that were coded. The complete coding protocol is included Appendix G. A coding schema was developed for each variable, along with definitions for each item in the schema.

Table 2. *Variables Coded*

Type	Variable
---	Identification of article
Context	Publication year
Context	Evaluation team
Context	Evaluator role
Context	Evaluator geographic area
Context	Evaluand
Context	Evaluand
Context	Geographic location of evaluand
Context	Target population of the evaluand
Context	General substantive field
Context	Specific substantive focus area
Context	Number of sites
Context	Area where evaluand implemented
Context	Single or multiple organizations
Context	Number of organizations involved
Context	Interdisciplinary / intersectoral collaboration
Context	Disciplines / sectors involved
Context	Author describes project as “complex”
Context	Stage of program
Context	Primary commissioner of the evaluation
Context	Level of change

Program theory	Program theory narrative
Program theory	Details of theory formulation
Program theory	Overview of theory formulation
Program theory	Social science theory
Program theory	Archetypes (archetypical outcome pathways)
Model	Number of causal strands
Model	Proportion of impact
Model	Presence of mediators
Model	Length of mediator/outcome chain
Model	Pathway of causation
Model	Moderator included
Model	Type of moderator
Model	Explanation of path moderators
Model	Description of the intervention process
Model	Design of model
Model	Identified as logic model
Model	Other identification of model
Evaluation	Evaluation purpose (with supporting text)
Evaluation	Evaluation approach
Evaluation	Evaluation questions stated
Evaluation	Evaluation questions tied to program theory
Evaluation	Focus of the evaluation
Evaluation	Evaluation design

Evaluation	Measure of implementation fidelity
Evaluation	Theory-guided construct measurement 1 – assessment of constructs identified in program theory
Evaluation	Theory-guided construct measurement 2 – types of constructs assessed
Evaluation	Focus of the analysis
Evaluation	Mediator analysis
Evaluation	Moderator analysis

Originally, six codes were identified to categorize design of model: linear model describing process theory, linear model describing impact theory, linear model describing process theory and impact theory, model based on realist evaluation, social ecological model, multi-dimensional model, and model built on a visual metaphor. When the decision was made to remove the ten articles on realist evaluation from the sample, this code was dropped from the protocol. Then, even though network models have been increasingly suggested as an option to consider, it was expected that they would appear rarely in the sample. A specific code for network model was not included; if such a model would appear, it was decided that it would be captured under the category of "other model" and then described.

Coders. Three coders were involved in the project. They went through a week of training, coding ten articles together, developing a shared understanding of definitions, and making final refinements to the coding guide. The ten practice articles were not included in the final sample. The 116 in the sample were divided across three teams: coders A and B, A and C, and B and C. Each coding team read and coded an article simultaneously, either in person or via

Google Hangouts. Coders individually reviewed the article and placed their codes and notes in the spreadsheet. The coders would then discuss their answers. When there was a discrepancy, they would deliberate the issues and reach an agreement by consensus for a final code. The initial individual responses were used in determining the inter-rater reliability among coders. The final code was used in determining the findings of the content analysis.

Analysis

The coding guide for the content analysis was composed primarily of deductive codes based on concepts from the literature on theory-driven evaluation. Analysis of archetypes initially was done using an inductive approach, but when the codes revealed matched closely existing codes in the literature, a second review was done using pre-determined codes. The deductive codes resulted in quantitative data that were analyzed in SPSS and in some cases Excel. The primary analysis conducted on the quantitative data yielded descriptive statistics, such as measures of central tendency (i.e., mode, mean, and median) and measures of variability (i.e., average deviation, variance, and standard deviation).

Results from the Content Analysis

Characteristics of the sample of articles.

The sample comprised 116 articles describing a program evaluation and including a graphic conceptual model. Over 62.9% of evaluations focused on a public health intervention and 23.3% on an education intervention. About sixty-six percent (66.4%) of the interventions evaluated were based in the United States; between 5% and 10% were based in Australia, New Zealand, Africa, or Canada; and less than 5% were based in Europe, Latin America, Asia, or the Middle East. Sixty-two percent (62.1%) of programs were implemented by multiple organizations; 53.4% involved interdisciplinary or intersectoral collaborations; and 72.4% were implemented in multiple sites. Interventions targeted change at various levels, with individual

change being the most frequent (in 94.8% of interventions), followed by organizational change (in 56.0%), interpersonal change (in 55.2%), change at the community level (in 35.5%), and lastly by policy change (in 19.0%). The majority targeted two levels of change. Seventy-five percent (75.5%) of evaluations were undertaken to assess merit and worth of the intervention, 37.9% were done to determine ways to improve a program, and 34.5% aimed to add knowledge to the field.

Frequency of graphic conceptual models in published evaluation studies.

This study drew on articles from eleven journals between 2003 and 2016: six focused specifically on evaluation and comprised evaluations of interventions in different substantive areas; two focused on educational evaluation; and three focused on public health, including evaluation of public health programs. In this thirteen-year period, the eleven journals published 1,578 articles describing a specific program evaluation. Among these articles, 207 (13.1%) included a graphic conceptual model and appeared to follow tenets of theory-driven evaluation.

Articles were sought that included a graphic conceptual model outlining program theory. While most interventions focused on implementation in community settings, it is possible that some efficacy evaluations were included along with effectiveness ones. Articles did not necessarily identify theory-driven evaluation as their evaluation approach. In fact, the majority (61.2%) did not specify the evaluation approach taken. Among the 38.8% of articles that did indicate the evaluation approach used, 19.0% indicated theory-driven evaluation. The discrepancy between the percentage of graphic conceptual models included in evaluation studies and the percentage of specific mention of theory-driven evaluation as the approach guiding the evaluation may be due to two reasons. First, theory-driven evaluation is compatible with other theories of evaluation practice and for this reason might not be identified separately. Second,

naming the evaluation approach used is not yet commonplace among many evaluators. Table 3 shows the distribution of articles with a graphic conceptual model across the eleven journals reviewed.

Table 3. *Articles on Program Evaluation with a Graphic Conceptual Models Between 2003 and 2016*

Journal	Number of Articles On Program Evaluation	Number of Articles with a Graphic Conceptual Model	Percentage of Articles with a Graphic Conceptual Model
American Journal of Evaluation	61	17	27.9%
The Canadian Journal of Program Evaluation	27	7	25.9%
Evaluation and Program Planning	248	53	21.4%
Evaluation: The International Journal of Theory, Research and Practice	51	28	54.9%
Evaluation Journal of Australasia	42	7	16.6%
Journal of Multidisciplinary Evaluation	26	5	19.2%
Educational Evaluation and Policy Analysis	111	11	9.9%
Studies in Educational Evaluation	64	9	14.1%
American Journal of Public Health	528	28	5.3%
Health Promotion Practice	160	20	12.5%
Prevention of Chronic Disease	259	22	8.5%
Total	1,578	207	13.1%

From the period of 2003 to 2016, a yearly average of 16 articles with a graphic conceptual model were published in journals focused on program evaluation and journals in the fields of education and public health that include evaluation studies. This is six more articles per year than Coryn et al. (2011) noted in the time frame of 1990 to 2009. The percentage of

published articles presenting evaluations using elements of theory-driven evaluation, however, remains small given the total number of articles on program evaluation.

Extent that core principles of theory-driven evaluation are applied in graphic conceptual models.

Although 207 articles were identified with a graphic conceptual model, only 116 had sufficient detail to be included in the content analysis. The principles of theory-driven evaluation connected to graphic conceptual models that were the focus of the review included sources used for developing program theory; a specific and detailed representation of the path of mediation; and identification of path moderators.

Sources for program theory. In theory-driven evaluation, the program theory may be developed using multiple sources, including social science theory, evidence base (e.g., previous evaluations and programs or practices considered models or exemplary), stakeholder theory, program observation, and document review. In the published articles, social science theory was used in 38 cases (32.8%) to develop program theory. One or more specific social science theories were referenced in each case. Within the 38 articles that mentioned a social science theory, social cognitive theory was referenced most frequently (31.6%). Ecological systems theory was mentioned in 13.1% of articles, the health belief model in 10.5%, the theory of reasoned action/planned behavior in 7.9%, and the transtheoretical model of behavior change in 5.3%. Other theories cited appeared only once.

In eighty-eight cases (75.9%), program theory was based on results from research and evaluation or on programs or practices considered models or exemplary (in the coding protocol this source of program theory was labeled as “evidence”). It was often the case that program theory was built on both social science theory and evidence. In several instances, theoretical and

empirical sources for program theory were complemented by sources closer to the intervention. In 51 instances (44.0%), stakeholder theory informed the program theory. Less frequently used were observation in twelve articles (10.3%) and document review in thirteen (11.2%).

Description of path of mediation. Mediators are critical in explaining the effects of an intervention. The amount of change in a desired outcome that a program can produce is influenced by the strength of the relationships that exist between mediators and outcomes. It is important that program activities aim at the right targets (mediators). A graphic conceptual model can be used to explain the expected path of mediation of an intervention.

Among the case examples, mediators were included in the graphic conceptual model in 97 instances (83.6%). On only 19 occasions (16.4%) was a "black box" effect depicted, with activities directly linked to outcomes and no path of mediation outlined. In models where the path was depicted, there were mediator chains of varying lengths. The length of the chain was of interest because this could be taken to indicate a focus on details of how a program is expected to lead to results. Table 4 presents the length of the mediator chains in the graphic conceptual models in the study. (The longest mediator chain in each model was considered for this code.)

Table 4. *Length of Mediator Chains in Graphic Conceptual Models (GCM)*

Number of Mediators*	Number of GCMs	Percent of GCMs
0	19	16.4%
1	27	23.3%
2	28	24.1%
3	17	14.7%
>3	18	15.5%

Note: In seven cases, it was not possible to determine the number of mediators in the chain because of how the model was designed.

Donaldson (2001) describes the mediator in program development as a variable that is affected by the program, which in turn affects an outcome of interest. Figure 2 shows a graphic conceptual model with one mediator. The model describes the theory underlying a mass media campaign using billboards, newspapers, radio, and poster advertisements to promote walking and local community-sponsored wellness initiatives. The relationship between campaign exposure and behavior regarding walking and wellness activities is mediated by pro-walking beliefs.

Figure 2. Model of a Walking and Wellness Campaign (Wray, Jupka, and Ludwig-Bell, 2005)

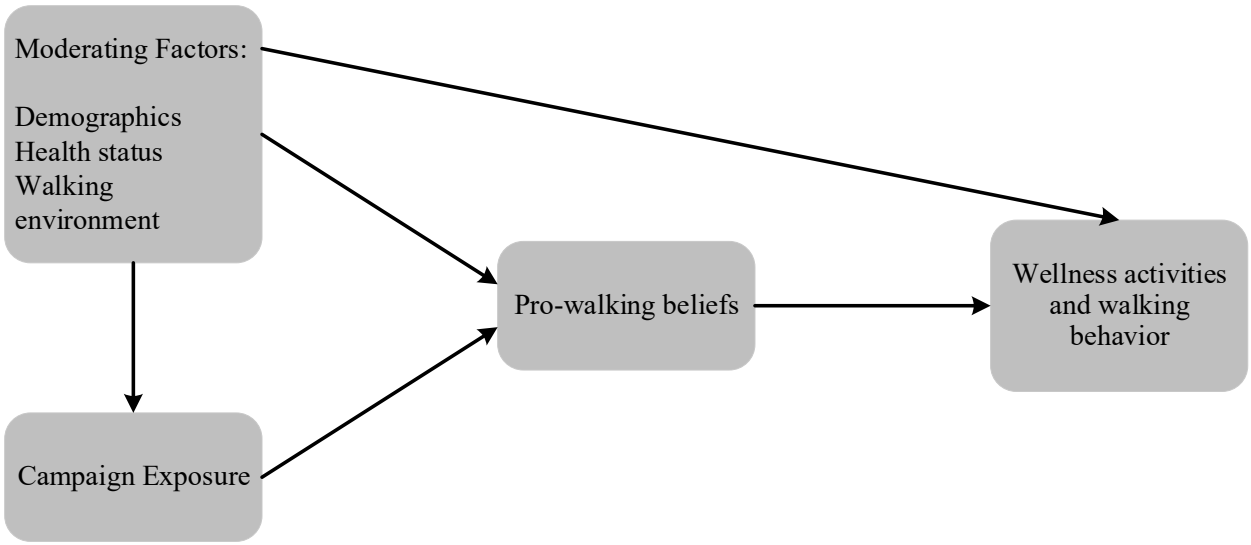


Figure 3 presents the program theory for a health education program that includes culturally-tailored curricula taught by community health workers to improve knowledge and heart healthy behaviors among diverse racial and ethnic groups. It serves as an example of a graphic conceptual model with multiple mediators. In serial mediation, there are two or more mediators, with one of the mediators being the cause of the other mediator. For instance, in one causal chain in this model, there are two mediators between the program and the outcome of decreased risk of heart disease: increased heart health knowledge, which influences increased heart health behaviors.

Figure 3. Model of a Heart Health Education Program (Hurtado et al., 2014)

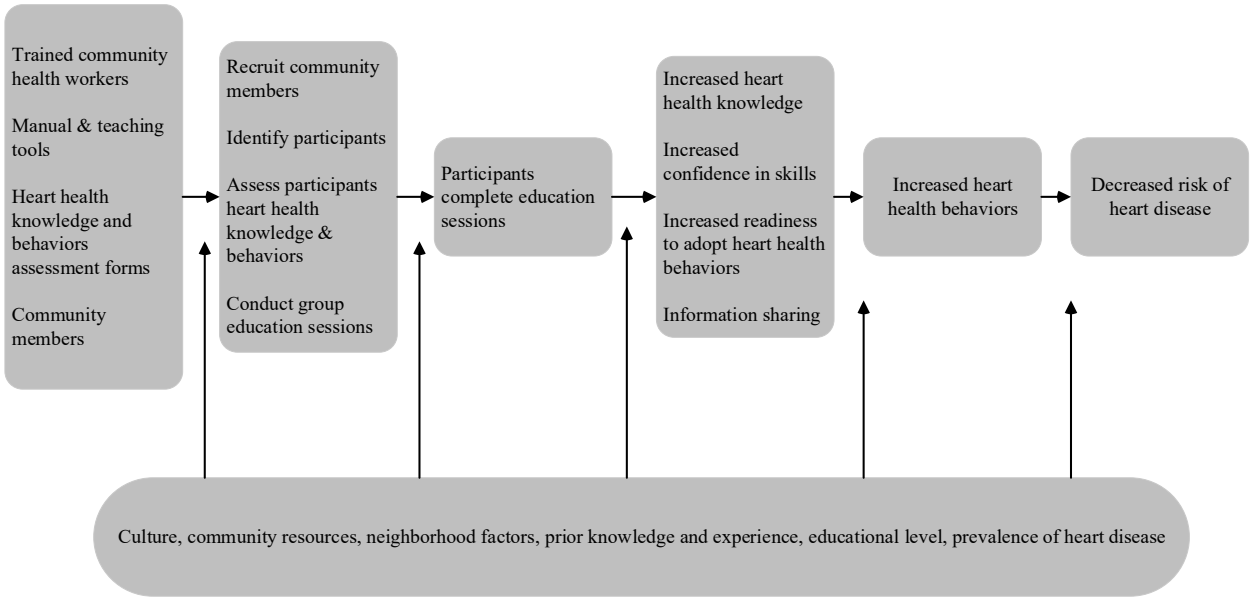
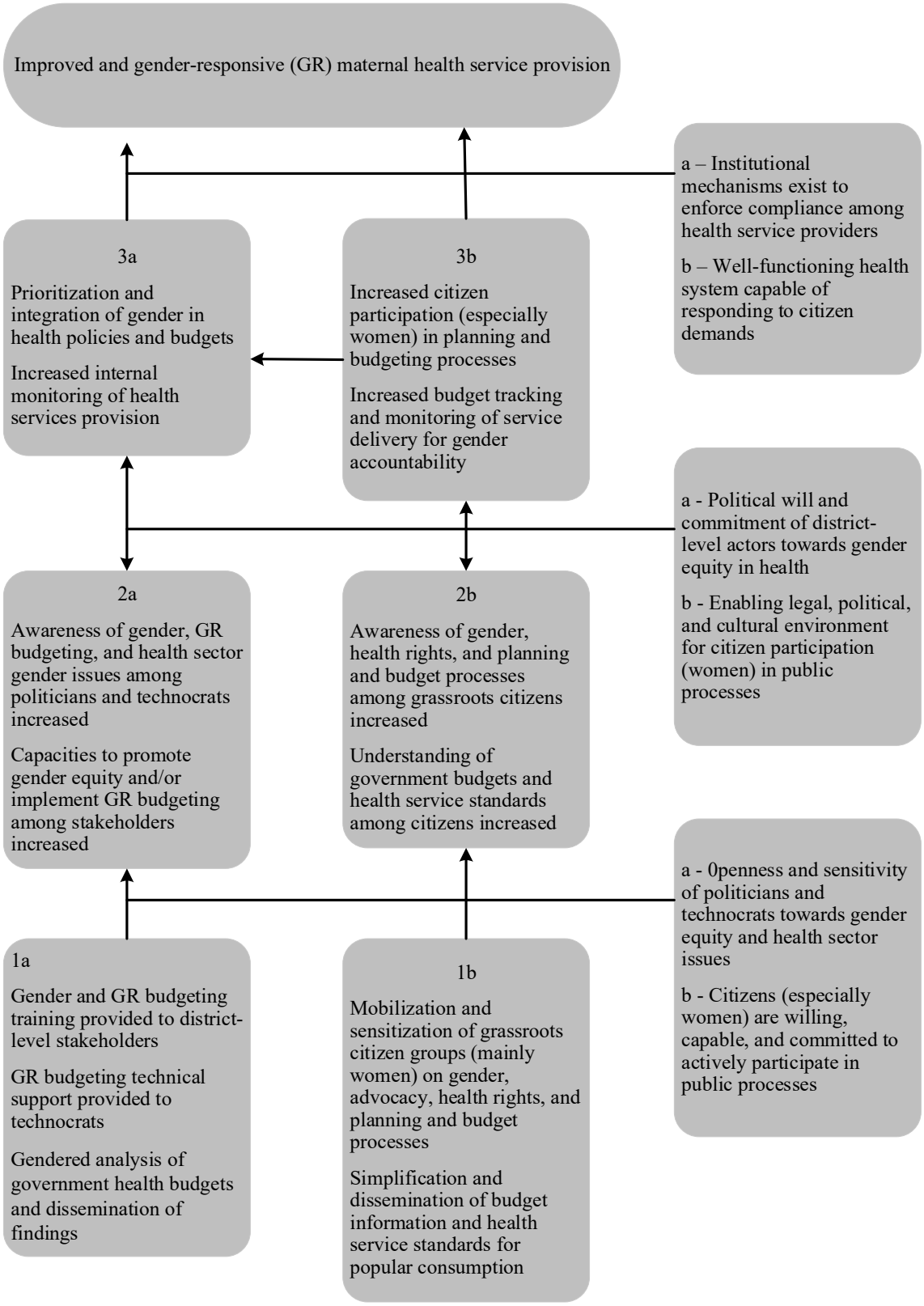


Figure 4 offers another example of serial mediation in a slightly more involved intervention. The model presents the program theory of a local-level, civil society-led gender-responsive budgeting initiative for maternal health. The causal chains depicted follow distinct paths for two target groups: government district-level stakeholders and grassroots citizen groups, particularly those composed by women. There are both unidirectional and bidirectional relationships in the model. A few examples of serial mediation include the following:

- Example of two mediators: Program activities (budget training, sensitizing and mobilizing of citizens) are expected to lead to increased awareness of gender, health rights, and planning and budget processes among grassroots citizen groups (mediator 1), which then increase citizen participation in planning and budget processes (mediator 2), which then could result in improved and gender-responsive maternal health service provision (outcome).

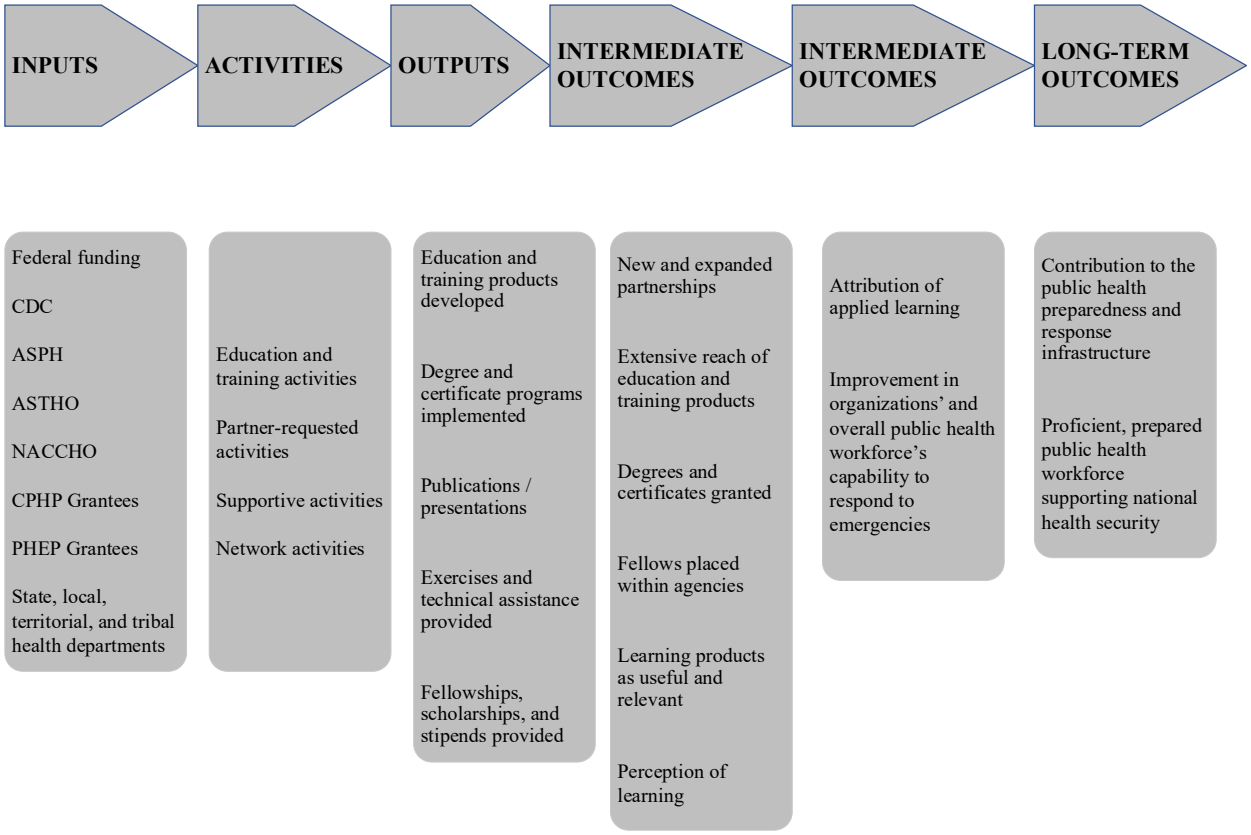
- Example of three mediators: Program activities (budget training, sensitizing and mobilizing of citizens) are expected to lead to increased awareness of gender, health rights, and planning and budget processes among grassroots citizen groups (mediator 1), which then increase citizen participation in planning and budget processes (mediator 2), which then influence district-level stakeholders to prioritize and integrate gender in health policies and budgets (mediator 3), which could result in improved and gender-responsive maternal health service provision (outcome).
- Example of more than three mediators: Program activities (budget training, sensitizing and mobilizing of citizens) are expected to lead to increased awareness of gender, health rights, and planning and budget processes among grassroots citizen groups (mediator 1), which then increase citizen participation in planning and budget processes (mediator 2), which then strengthens awareness further (mediator 3), which then encourages even greater participation (mediator 4), which then influences district-level stakeholders to prioritize and integrate gender in health policies and budgets (mediator 5), which finally could result in improved and gender-responsive maternal health service provision (outcome).

Figure 4. Model of a Gender-Responsive Budgeting Initiative in Maternal Health (Bamanyaki and Holvoet, 2016)



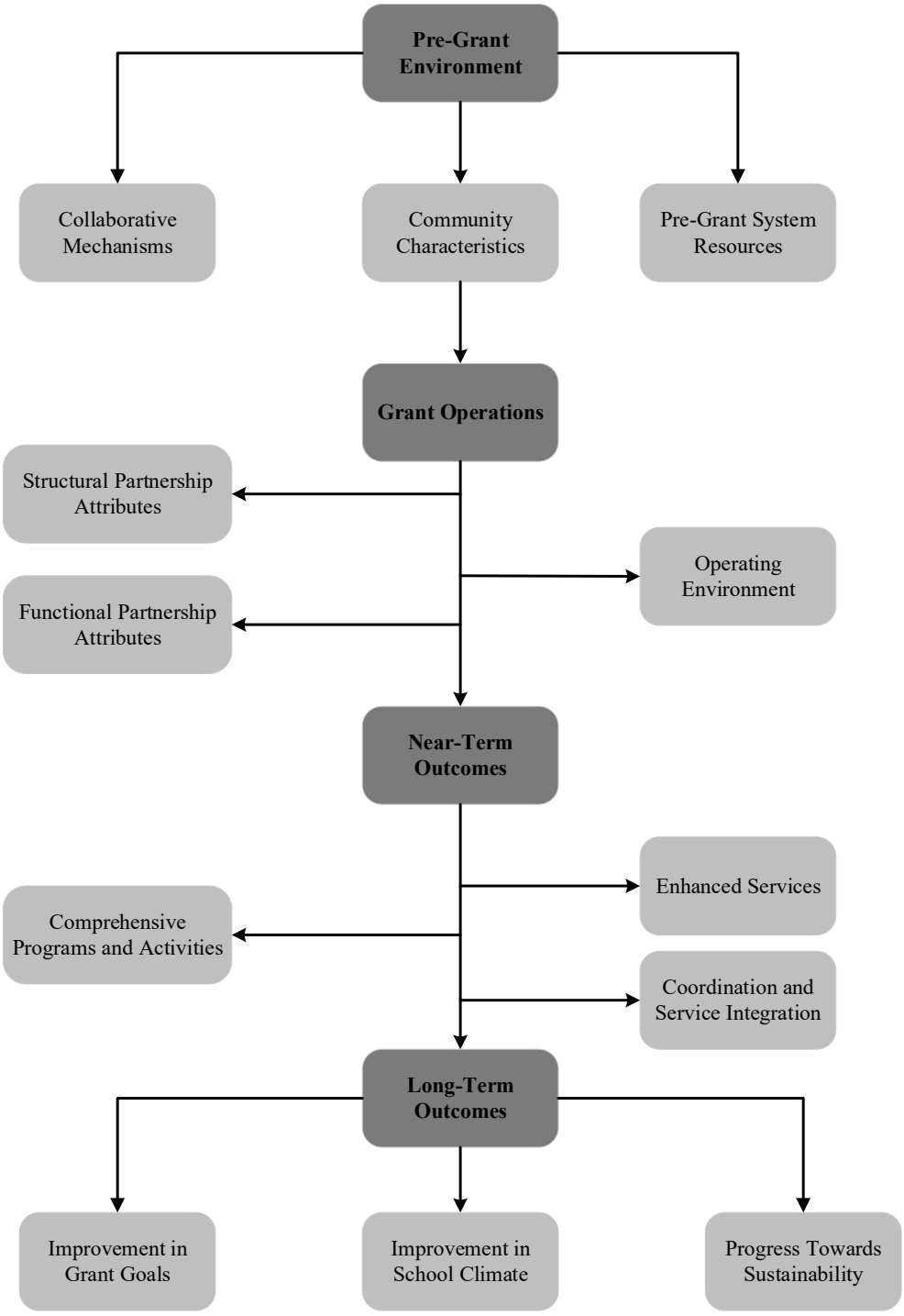
In many graphic conceptual models, the path of mediation is explained using arrows or other means (e.g. headings or the order in which items are presented). For instance, in Figure 4, arrows, numerical and alphabetical headings, and the consistent ordering of related concepts were used to demarcate different mediation paths. How specifically the path of mediation is explained is an important characteristic of graphic conceptual models so that viewers of a model can better understand the mechanisms anticipated to produce change. When the path is only generally presented or not depicted at all, it may be difficult to develop a shared understanding of how a program is expected to result in certain outcomes. Figure 5 presents the graphic conceptual model for a program that brought together various academic institutions to train health agencies in terrorism preparedness and emergency response. While the model presents aspects of the program theory, stronger writing or design steps could help make connections on the path of mediation more clear.

Figure 5. Model of a Terrorism Preparedness and Emergency Response Program (Sobelson and Young, 2013)



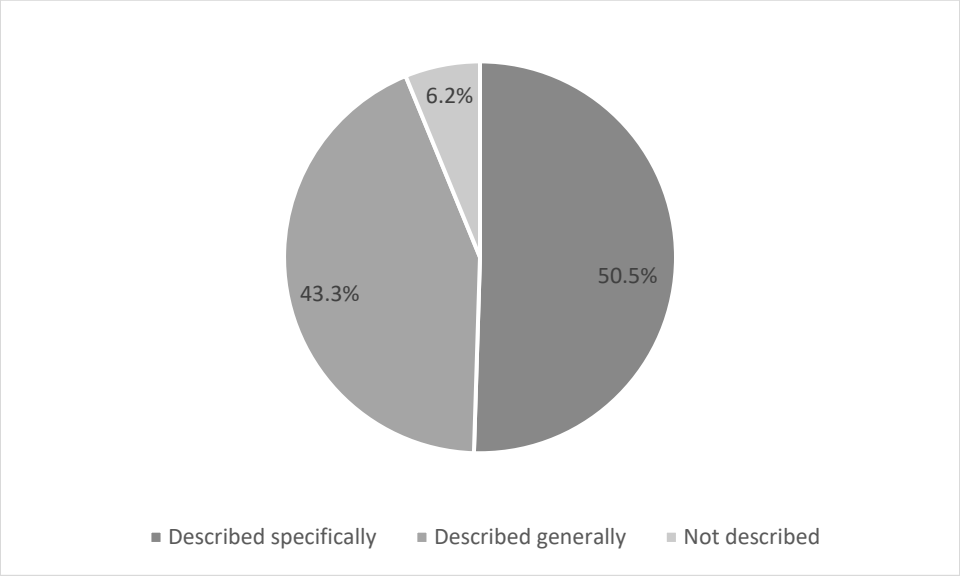
While the more tabular format used in Figure 5 was somewhat difficult to follow, this is not to say that tables are less effective than using arrows. There can be instances when arrows themselves are not helpful in explaining connections among program components, as is the case in Figure 6. The program theory depicted is for an initiative designed to foster integrated systems that create safe and respectful school climates and, consequently, promote the mental health of students and prevent violence and substance abuse. Even though arrows are used, it is still difficult to ascertain the specific paths of mediation.

Figure 6. Model of an Initiative Fostering Integrated Systems for Safe and Respectful School Climates (Rollison et al., 2012)



Figures 5 and 6 provide examples of a "general" description of the path of mediation—meaning that the path was not able to be easily or fully followed. The examples in Figures 2, 3, and 4 offered examples of models with more "specific" paths. This distinction was considered important because graphic conceptual models are expressions of how a program is understood to potentially effect change. Clear explanations of the relationship between different mediators facilitate understanding of how a program is expected to function. As noted in Figure 7, among the 97 graphic conceptual models in the study that included mediators, the path of mediation was specifically outlined in 49 instances (50.5%), generally presented in 42 cases (43.3%), and not delineated 6 times (6.2%).

Figure 7. Extent That Path of Mediation is Described in Detail



Identification of path moderators. In program development, moderators may affect the direction or strength of the relationships between the program and mediator or mediator and outcome (Donaldson, 2001). When conceptualizing how a program is supposed to work, it can

be beneficial to identify significant moderators. These relationships may be helpful for understanding program effects, or the lack of effects.

Several of the illustrations shared above offered examples of moderators. For instance, in the mass media campaign to promote participation in wellness activities and walking (Figure 2), moderators include demographic characteristics, pre-existing health status, and the walking environment. These moderators are expected to influence exposure to the campaign itself, as well as the relationship between the campaign and the mediator of pro-walking beliefs and between pro-walking beliefs and behavior change. The specific relationships affected by the moderators are clearly depicted with discrete arrows. Figure 4 also provides an example of moderators specifically called out in a model. There are several moderators in this model, such as the willingness and readiness of citizens to participate in public processes. This moderates the relationship between the program and increased awareness of gender, advocacy, health rights, planning and budgeting and increased understanding of government health budgets and service standards. As was the case in Figure 2, the moderation is specifically outlined in the model. While moderators were presented in 106 (91.3%) articles, they appeared in only 50 (43.1%) graphic conceptual models. Among these 43.1% of models, relationships moderated were shown specifically in 76.0% cases. In the remaining 56 (48.3%) articles, moderators were introduced in the narrative of the article only. The most frequently mentioned moderators are presented in Table 5.

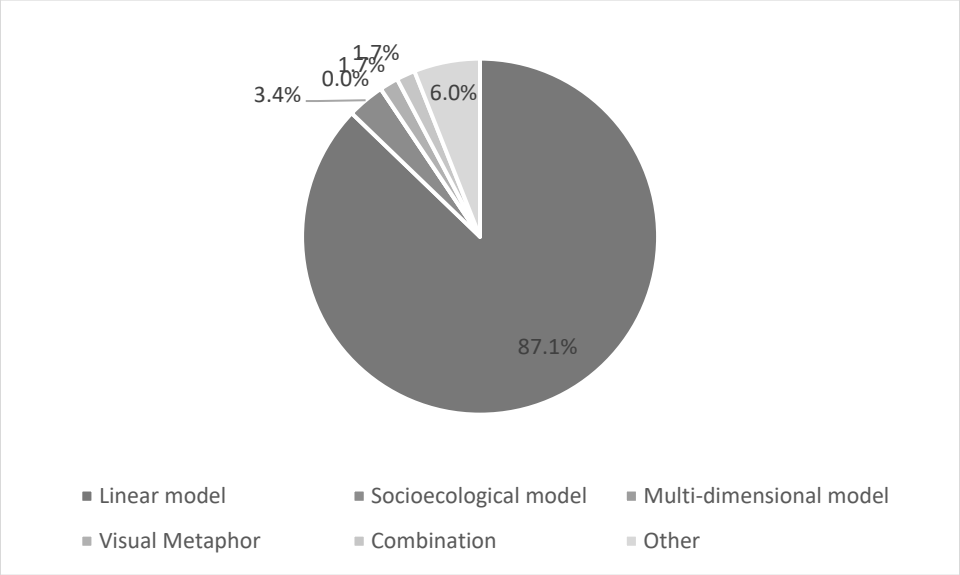
Table 5. *Most Common Moderators Included in Graphic Conceptual Models (GCMs)*

Moderator Type	Number of GCMs	Percent of GCMs
Participant Characteristics	36	72.0%
External Factors	39	78.0%
Setting	16	32.0%
Provider Characteristics	11	22.0%
Dosage of Intervention	4	8.0%
Assumptions	6	12.0%
Attendance	1	2.0%
Other	3	6.0%

Type of graphic conceptual models most frequently used, and variations in model type.

The coding protocol included the following model types: linear model, socioecological model, multi-dimensional model, and visual metaphor. The most frequently used model was a linear model, appearing in 101 cases (87.1%). In this sample of 116 published articles, it did not appear that factors like cultural context, substantive focus area, or program complexity made necessary a non-linear model. Linear models dominated; there were only a limited number of socioecological models and models based on a visual metaphor, and no multi-dimensional models (noted in Figure 8).

Figure 8. Most Common Types of Graphic Conceptual Models Used



In the majority of linear models (57 models, 49.1%), both the process and impact theory were presented. In 34 models (48.6%), the process theory included inputs and activities, and in 17 models (24.3.%), the process theory comprised inputs, activities, and outputs. These categories frequently appear in logic models, which are one specific type of linear graphic conceptual model. Indeed, logic models were frequently referenced in the articles. In fifty articles (49.5%), the graphic conceptual model included was referred to as a logic model.

While graphic conceptual models in theory-driven evaluation are expected to be able to tell the story of a program's functioning on their own, they can be complemented by a narrative description of the program theory. The narrative ideally goes beyond only describing program components and expected outcomes as presented in the model, but also discusses how inputs are expected to lead to outcomes—discusses the anticipated mechanisms of change and/or relationships between mediators and moderators. In the sample, such a narrative was included in 72 articles (62.1%).

Use of graphic conceptual models to guide evaluation.

In theory-driven evaluation, there are several expectations around the use of a graphic conceptual model to guide inquiry. Among these are the following: Evaluation questions should be designed around the program theory represented in the graphic conceptual model; assessment of fidelity of implementation is strongly recommended to help determine if poor implementation or some aspect of the intervention is the reason for certain findings; constructs identified in the program theory should be among the ones measured; and there should be a commitment to explain cause-and-effect associations between theoretical constructs. Table 6 describes aspects of the evaluations carried out for the interventions presented in Figures 1 through 5; it is included to illustrate the type of information looked at during the coding. The extent to which the models included in the articles reviewed guided the evaluation are described in the sections following the table.

Table 6. *Evaluation Design of Interventions in Figures 1 Through 5*

Intervention	Evaluation Questions	Fidelity of Implementation	Measurement of Constructs	Data Collection Method and Analysis
Campaign to promote walking (Figure 2)	Evaluation questions were not stated explicitly.	Frequency and duration of exposure were assessed.	Process, outcome, and context constructs in the model were assessed in the evaluation.	A quasi-experimental study was conducted. Data was collected via a survey. Mediators and moderators were tested statistically.
Heart health education program	Evaluation questions were	Quality of delivery was assessed.	Process and outcome constructs in the	A single group pre-post test was conducted. The

(Figure 3)	not stated explicitly.		model were assessed in the evaluation.	influence of moderators like demographic variables on achieving proximal outcomes (e.g., knowledge, confidence) was investigated in the analysis. The mediating effect of proximal outcomes on more distal ones (e.g., health behavior) were not measured.
Gender responsive budgeting initiative (Figure 4)	Evaluations questions were stated explicitly and tied to the program's underlying logic or theoretical foundations.	Duration and frequency of exposure and quality of delivery were assessed.	Process, outcome, and context constructs in the model were assessed in the evaluation.	Process tracing was used to investigate the theory presented in the model. Evidence was gathered to support relationships (e.g., prior research, interviews, participation logs). Empirical evaluation of evidence was conducted using Bayesian logic to confirm or disconfirm the presence of the causal mechanism linking the intervention to observed changes.

Terrorism preparedness and emergency response program (Figure 5)	Evaluations questions were stated explicitly and tied to the program's underlying logic or theoretical foundations.	Fidelity of implementation was not assessed.	Process, outcome, and context constructs in the model were assessed in the evaluation.	While various constructs were measured, the relationships among them were not.
Initiative fostering integrated systems for safe and respectful school climates (Figure 6)	Evaluations questions were stated explicitly and tied to the program's underlying logic or theoretical foundations.	Fidelity of implementation was not assessed.	Process, outcome, and context constructs in the model were assessed in the evaluation.	Data from surveys, site visits, interviews, and focus groups across grantees were collected. Qualitative data supplemented quantitative data. Some narrative-based qualitative observations were converted into cross-site matrices with ordinal values. Relationships in the program theory model were examined though statistical analysis.

Evaluation questions tied to program theory. Evaluation questions were stated explicitly in 44 studies (37.9%). When the evaluation questions were stated explicitly, they were tied to program theory in 41 cases (93.2%). In 67 articles (57.8%), the evaluation questions could be surmised. In the absence of specific questions, however, the coders did not map the supposed questions back to the program theory.

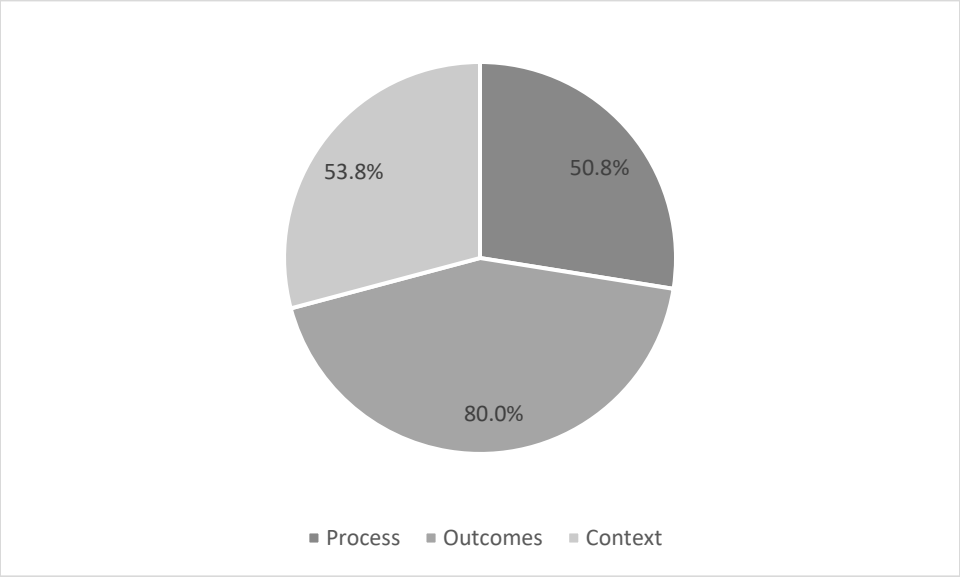
Assessment of fidelity of implementation. Fidelity of implementation was assessed in 69 evaluation studies (59.5%). In most instances, only one dimension of fidelity of implementation was measured. Table 7 identifies the frequency with which dimensions of fidelity were evaluated. The dimensions used were drawn from James Bell Associates (2009). Program adherence refers to the extent to which program components are delivered as outlined in the model. Exposure is the amount of program delivered in relation to the amount prescribed by the program model. This can include the frequency and duration (e.g. dosage) of sessions. Quality of delivery reflects the manner in which a program is delivered. It may include provider knowledge, preparedness, and delivery style. Participant responsiveness refers to the way in which participants react to or engage in a program, such as their level of interest or engagement.

Table 7. *Frequency of Dimensions of Fidelity of Implementation Measured*

Dimensions	Number	Percent
Adherence	32	46.4%
Exposure: Frequency	36	52.2%
Exposure: Dosage	15	21.7%
Delivery Quality	49	71.0%%
Participant Response	34	49.3%

Measurement of constructs in program theory. In 112 articles (96.6%), constructs identified in the graphic conceptual model were measured. In 47 of these (40.5%), the evaluation measured constructs in the graphic conceptual model as well as additional constructs that had not been included in the model. It was not explained why these additional constructs were not included in the model. In the 65 articles (56.0%) where constructs only from the graphic conceptual model were measured, the constructs most frequently measured were related to outcomes (90.0%) and context (53.8%), as presented in Figure 9.

Figure 9. Constructs from Graphic Conceptual Models Measured in the Evaluation



In the 47 evaluation studies that measured constructs beyond those in the model, the breakdown of constructs followed a similar pattern: 87.2% measured constructs related to outcomes, 76.6% measured constructs related to context, and 48.9% measured constructs related to process. The larger percentage of contextual constructs measured in this group of studies may be due to the fact that moderators in some cases were described in the text but not included in the model.

Explanation of associations between theoretical constructs.

The majority of evaluation studies in the sample looked closely at cause-effect associations between theoretical constructs. Coders used three means of classification: 1) the evaluation only measured the extent to which outcomes were attained (presentation of effects without discussion of cause); 2) there was an acknowledgment and discussion of relationships but the relationships were not tested (description of cause-effect associations); and 3) the relationships among mediators or moderators were tested (explanation of cause-effect

associations). As noted in Table 8, 42.2% of studies tried to explain cause-effect associations and 20.7% to describe them; only 20.7% engaged solely in measuring outcomes.

Table 8. *Focus of Outcome Analysis*

Focus	Number	Percent
Presentation of effects without discussion of cause	24	20.7%
Description of cause-effect associations	24	20.7%
Explanation of cause-effect associations	49	42.2%

Note: In nineteen cases, outcome analysis was not part of the evaluation.

Both quantitative and qualitative approaches were used to describe and make inferences about causal relationships. Among the 73 studies that either described or explained cause-effect associations, 41.1% used a non-experimental design, 38.4% used a quasi-experimental design, and 20.5% used an experimental one. In the 49 studies that tested relationships, the extent to which one construct accounts for or mediates the relationship between other constructs was assessed in 32 cases (65.3%). In 40 instances (81.6%), methods were used to test the extent to which one construct moderates the relationship between other constructs.

Program archetypes.

Each member of each coding team wrote a brief description of the program and the outcome pathway described in the graphic conceptual model during the initial coding of articles. Since each coding team was comprised of two individuals, there were two descriptions for each article. Two coders first used an inductive approach to identify program archetypes, but when those codes matched closely the codes put forward by Funnell and Rogers (2011), a second review was done using those pre-determined codes. (The third coder was no longer available to participate in this part of the project.)

The graphic conceptual models in this study offered the opportunity to assess whether the archetypes proposed by Funnell and Rogers align well with commonly developed graphic conceptual models and to identify any needed refinements. Of the 116 graphic conceptual models in the sample of articles, only 104 provided enough detailed information to code for possible archetypes. Table 9 presents the frequency with which these archetypes were found in the graphic conceptual models.

Table 9. *Program Archetypes Observed in the Study Sample*

Archetype	Number of GCMs	Percent of GCMs
Case Management	5	4.8%
Direct Service	11	10.6%
"Carrots and Sticks"	2	1.9%
Advisory, Public Information, and Education	60	57.7%
Network Theory	3	2.9%
Combinations		
Case Management and Advisory, Public Information, and Education	6	5.8%
Incentive and Advisory, Public Information, and Education	4	3.8%
Community Capacity Building and Network Theory	13	12.5%

In the sample, the majority of graphic conceptual models were based on the advisory, public information, and education archetype. Among the 73 graphic conceptual models of this kind, direct education/training opportunities were the most frequently used intervention, appearing in 58 models (79.5%). Information dissemination by community leaders or mass media was the next most common intervention under this archetype (15.1% of models).

The archetypes of advisory, public information, and education; case management; "carrots and sticks," and direct service could be applied much as Funnell and Rogers described.

Network theory is an archetype that Funnell and Rogers suggested would be useful to develop further. The findings of this study underscore that position. Several graphic conceptual models described organizing and mobilizing at the community level, by either individuals or institutions, that could not be captured fully by the current description of the community capacity building archetype of Funnell and Rogers. Organizing and mobilizing for community change could be a component of the network theory archetype, or it could be included in a revised version of the community capacity building archetype.

CHAPTER FOUR

SURVEY OF EVALUATORS

Research Questions and Hypotheses

Phase Two of this study expanded inquiry about the use of graphic conceptual models to additional evaluators through a survey. Several research questions from the content analysis carried over to the survey. These were complemented by questions focused on the model development process, on fostering a shared understanding about program theory among stakeholders, and on conditions affecting the use of models in evaluation. The research questions included the following:

1. How do evaluators describe their application of theory-based evaluation? In particular,
    a. To what extent are core principles of theory-driven evaluation applied in graphic conceptual models?
    b. What types of graphic conceptual models do they use most frequently given substantive field, program complexity, and cultural context?
    c. What is the process used for developing graphic conceptual models and to what extent does the process foster a shared understanding about a program among stakeholders?
    d. How are graphic conceptual models used in the evaluation?
2. What conditions must be in place for the graphic conceptual model development process and product to be useful for developing shared understanding among stakeholders and for designing an evaluation (e.g., evaluator knowledge and skills, political considerations)?

The hypotheses for questions 1a and 1d were the same as in the content analysis: linear models were expected to dominate and the majority of evaluations were expected to focus on description rather than explanation. In developing the models (1b), it was expected that stakeholder theory may be more commonly used than social science theory or research or evidence. Questions 1c, 2a, and 2b were more exploratory in nature so specific hypotheses were not developed for them.

Methods

Questionnaire. The survey questions were informed by the coding protocol used in the content analysis. Additional questions were formulated to better understand the model development process, the creation of shared understanding around program theory, and conditions that influence the design and use of program theory in evaluation. The majority of the survey included closed questions with response categories. There were a few open-ended questions that focused on stakeholders engaged in the model development process, facilitation techniques for creating models, modifications to models based on cultural context, and barriers to developing and using program theory in evaluation. The survey did provide a definition of theory-driven evaluation and an example of a graphic conceptual model in the event that participants did not use those specific terms but do follow principles of the approach. The survey was revised based on feedback from five people who completed draft versions. The final survey is included in Appendix H. The survey was administered online through Qualtrics and remained open for 3 to 4 weeks. Individuals invited to participate received two reminders to take the survey.

Participants and recruitment. Target respondents were individuals over 18 years of age who have conducted evaluations of social betterment programs or policies. Individuals invited to

participate included members of the American Evaluation Association and evaluators identified through snowball sampling. The target was to recruit 150 respondents. Participants were recruited by email. Three $100 Amazon gift cards were offered as incentives, for those participants who wished to enter a lottery done at the close of the survey.

Analysis

The survey included both closed-ended and open-ended questions. The analysis process was similar to that done in the content analysis. After the data was transferred to SPSS, the quantitative analysis again involved descriptive statistics, including crosstabs. Open-ended questions resulted in text that was exported from Qualtrics into Word and then themes were induced from the data through open coding of the text (Bernard and Ryan, 2010). Responses were examined for repetition of content. Once organized, the data were reviewed closely to identify sub-themes and to select key identifying quotations that characterized a particular theme or sub-theme.

Results from the Survey

Characteristics of survey respondents.

One hundred forty-one individuals fully completed the survey. (Twelve individuals did not complete the survey because they indicated that they never use theory-driven evaluation; their responses are not included in these findings.) There was a nearly equal distribution of evaluators with a Master's Degree (48.6%) and those with a Ph.D. (47.9%). The most common means by which respondents received training in evaluation was through a degree program (68.1%), through professional development (63.1%), and through on-the-job training (57.4%); most individuals engaged in more than one training opportunity. Over eighty-seven percent of respondents (87.2%) conduct evaluations in the U.S.; only between 3.5% and 15.7% reported that they carry out evaluations in other parts of the world. The primary fields in which

participants reported that they "always" or "often" conduct evaluations include K-12 education (43.2%), health (41.9%), community development (23.4%), higher education (20.6%), and workforce training and development (20.5%). Less than 10% of respondents evaluate programs connected to agriculture, the arts, criminal justice, economic development, or housing. Fifty-six of the 141 respondents (39.8%) specialize in carrying out evaluation in particular fields while 85 (60.2%) conduct evaluations in various ones.

Extent that core principles of theory-driven evaluation are applied in graphic conceptual models.

The survey questions related to core principles of theory-driven evaluation in graphic conceptual models focused on aspects of models similar to those that were the emphasis in the content analysis: sources for developing program theory, mediators, and moderators.

Sources for program theory. For each possible source of program theory on the survey, over 50% of respondents reported using the source "often" or "always" when designing program theories. Table 10 shows the frequency with which these sources were reported to be utilized. Document review and previous research and evaluation were marked as the sources most used.

Table 10. *Sources for Program Theory Used "Often" or "Always" by Evaluators Surveyed*

Source	Number	Percent
Social Science Theory	80	56.7%
Research and Evaluation	103	73.0%
Stakeholder Theory	73	51.7%
Program Observation	88	62.4%
Document Review	113	80.1%

Note: The scale choices included "always," "often," "sometimes," "rarely," and "never."

Description of path of mediation. The survey did not ask participants to submit a sample graphic conceptual model. In the content analysis, the models provided useful information about the inclusion of mediators. In the survey, responses to two questions provide some insight into the role of mediators in graphic conceptual models. When asked whether or not evaluators discuss connections among model components with stakeholders, 72.3% indicated that they did so "often" or "always." While for a large percentage of evaluators discussion of mediators may have been a focus when designing program theory, studying the path of mediation in the evaluation itself was less of a focus, with only 25.5% of evaluators reporting that they "often" do so.

Identification of path moderators. Among respondents, 36.9% marked that they "often" or "always" include moderators in their graphic conceptual model. The moderator types most frequently used are noted in Table 11.

Table 11. *Moderators Most Commonly Included in Graphic Conceptual Models Developed by Survey Respondents*

Moderator Type	Number of Respondents	Percent of Respondents
Participant Characteristics	74	60.7%
External Factors	89	73.0%
Setting	75	61.5%
Provider Characteristics	51	41.8%
Dosage	79	64.8%
Assumptions	80	65.6%
Attendance	56	45.9%

Type of graphic conceptual models most frequently used, and variations in model type.

The graphic conceptual models most frequently developed by participating evaluators were linear models, using either a table or diagram with shapes and arrows. Table 12 indicates the frequency with which certain model types were developed.

Table 12. *Type of Graphic Conceptual Models Developed "Often" or "Always" by Evaluators Surveyed*

Model Type	Number of Respondents	Percent of Respondents
Table	88	62.4%
Diagram with shapes and arrows	112	80.1%
Model built on a visual metaphor (e.g., tree)	18	12.8.%
Multidimensional model (e.g., cube)	3	2.2%

Note: The scale choices included "always," "often," "sometimes," "rarely," and "never."

Respondents also had the opportunity to note other types of models or representations that they create in their evaluation practice. These included circular and ecological models, space-time modeling, and non-graphical representations like skits and scenarios. The percentage of evaluators using these models, however, was very small (less than 3% of respondents).

Graphic conceptual models most frequently used in evaluations of different substantive areas.

Respondents in the survey indicated the substantive areas in which they most frequently conduct evaluations. Table 13 indicates they type of model most frequently developed for an evaluation in different substantive fields. In almost all areas, more linear models, either as a table

or a diagram with shapes, dominated. In many instances, diagrams with shapes were preferred slightly over tables. Visual metaphors were most common in education-related fields such as K-12 education and workforce training and development.

Table 13. *Type of Graphic Conceptual Models Developed "Often" or "Always" by Substantive Areas (Percent)*

Substantive Area	n	Table	Diagram with Shapes	Visual Metaphor	Multidimensional Model
Agriculture	38	21.1%	23.7%	5.3%	0
Arts	48	8.3%	6.3%	0	0
Community Development	103	22.3%	26.2%	6.8%	0
Criminal Justice	54	3.7%	5.6%	1.9%	0
Economic Development	59	20.3%	22.16%	5.1%	0
Education	109	34.9%	44.0%	9.2%	0.9%
Health	108	35.2%	44.4%	6.5%	2.8%
Higher Education	89	20.2%	23.6%	5.6%	1.1%
Housing	55	12.7%	12.7%	3.6%	0
Workforce Training and Development	64	25.0%	26.0%	9.4%	0

Note: The scale choices included "always," "often," "sometimes," "rarely," and "never."

Graphic conceptual models most frequently used in interventions where the causal pathway is adaptive or emergent.

For interventions where the causal pathway is adaptive or emergent, respondents were more likely to create models based on a visual metaphor, multidimensional models, or other models. Graphic conceptual models in the form of a table or diagram using shapes were more commonly used for interventions where the path to reach outcomes is more evident, e.g., knowledge change resulting from a well-tested and implemented training program. Table 14

demonstrates the different types of models used by the 141 respondents for interventions with more "known" paths and those with more emergent or adaptive ones.

Table 14. *Models Most Frequently Created for Known and Emerging or Adaptive Paths*

Model Type	Known Path	Emergent or Adaptive Path
Table	18.4%	15.6%
Diagram with shapes and arrows	73.8%	53.9%
Model built on a visual metaphor (e.g., tree)	1.4%	13.5%
Multidimensional model (e.g., cube)	---	2.1%
Other	6.4%	14.9%

Graphic conceptual models and cultural context.

Seventy-three evaluators described modifications that they often make to graphic conceptual models in different cultural contexts. Some individuals shared more than one strategy used, resulting in 105 open-ended responses. These were coded into seven categories: simplify complexity of the model, reduce the linearity of the model, modify the language used in the model, use culturally-relevant metaphors, use culturally-relevant and visually appealing images, make culture explicit in the model, and provide an alternative to the model. These categories are described below.

The first category is to simplify the complexity of a model. Respondents accomplished this by leaving certain details out of the model or breaking the model into parts. In a few instances, additional information would be included with the intention of making the model more understandable.

The second category is to reduce the implication of linearity or causality. This approach was important in some settings where such a worldview is less common. In such cases, respondents used shapes that could reflect more cyclical or holistic thinking styles. The circle was the shape most commonly mentioned among evaluators in the sample. Circles used as part of ecological models could demonstrate a more holistic view of behavior, drawing on individual and environmental determinants.

The third category is to modify the language used. The most frequent adjustments to language included simplifying the language, making culturally-sensitive word choices, using more positive and empowering language, and direct translation. Of these four approaches, the one that may need more explanation is the use of more positive and empowering language. Respondents indicated doing this mainly in the description of moderators, which sometimes may focus on deficits in a community as opposed to assets or more positive attributes.

The fourth category is to use culturally-relevant metaphors. The metaphors to which different audiences could better relate may vary. In this sample, metaphors tied to trees or a growing plant were the most common. Other examples of metaphors used included rivers, recipes, houses, and maps. The program theory was then depicted in a model or drawing based on the metaphor.

The fifth category is to use culturally-appropriate and visually appealing images. This differs slightly from models based on a culturally-relevant metaphor. In models based on a metaphor, the entire format of the model may evoke the metaphor. Models using culturally-relevant and visually appealing images make use of photos or pictures in the model itself, which may still be a more linear model.

The sixth and seventh categories occurred less frequently. In the sixth, cultural factors could be emphasized as an input, assumption, or environmental factor in a model. The seventh category is to provide an alternative to the model. This could include writing out the model as a story or narrative. It could also involve setting up activities to let people describe situations orally. These oral narratives could then be recorded and later demonstrated visually for other audiences, if needed. Table 15 describes how frequently respondents indicated their use of one of the seven approaches to make graphic conceptual models more accessible to stakeholders from different cultures.

Table 15. *Most Frequent Approaches Used to Make Graphic Conceptual Model to More Accessible in Different Cultural Contexts*

Approach	Number	Percent
Simplify complexity of model	15	14.3%
Reduce linearity of model	15	14.3%
Modify language	28	26.7%
Use culturally-relevant metaphors	13	12.4%
Use culturally-relevant images and colors	24	22.9%
Make culture explicit in model	5	4.7%
Provide an alternative to model	5	4.7%

Process used to develop graphic conceptual models and foster shared understanding.

Among the 141 evaluators who completed the survey, 29.8% most often are called to develop a program theory for established programs that have room for few modifications, 34.0% for programs that are being tested and can still be modified, and 26.2% during the design phase

of the program (9.9% usually evaluate programs that already have a program theory developed and that cannot be modified). To develop models, evaluators work with multiple groups of stakeholders—program staff, executive leadership, program beneficiaries, and funders are the ones most often consulted. Evaluators were asked to list the stakeholders that they most frequently engage to develop models of program theory as a response to an open-ended question. Table 16 describes the breakdown of stakeholder groups with whom 139 evaluators most frequently interact to develop models (two individuals provided a general response to the question that could not be coded).

Table 16. *Number and Percent of Evaluators Who Engage with Distinct Stakeholders to Develop Graphic Conceptual Models*

Stakeholders	Number	Percent
Board of Directors	13	9.4%
Executive Leaders and Administrative Staff (e.g., Executive Director, Development Director)	53	38.1%
Program Staff	132	95.0%
Program Beneficiaries	132	95.0%
Family of Beneficiaries	14	10.1%
Partner Organizations	19	13.7%
Community Members	15	11.5%
Funders	23	16.5%
Government Employees (e.g., Policy Makers, Department / Ministry Leaders)	9	6.5%
Academics or Content Experts	14	10.1%
Business and Industry Leaders	5	3.6%
Teachers and Principals of schools benefitting from program implemented on site	7	5.0%

In the open-ended responses, some evaluators shared challenges that they have had with convening representatives from multiple stakeholder groups to engage in developing a program theory. The two most common were time constraints, e.g., facing a grant deadline or limited availability of different stakeholders, and resistance from senior leadership and program staff to access other stakeholders. (The survey included specific questions related to challenges in implementing theory-driven evaluation; the responses to these questions will be shared in a subsequent section.)

All 141 survey respondents wrote a description of the process that they use for developing graphic conceptual models. These were analyzed for themes. The process for developing graphic conceptual models can be very contextualized. Two approaches at a macro level apparent among responses include 1) the evaluator creating a visualization from documents or artifacts related to the program that are then checked or revised with stakeholders through face-to-face and/or virtual meetings and 2) working directly with stakeholders to create a visualization for the program theory. When co-creating a model, respondents indicated that the number of meetings could range from one to six meetings, with meetings two hours in length or a half day in length being the most common. The majority indicated that they engaged in the second "macro approach": co-designing graphic conceptual models with stakeholders. Several themes emerged related to this co-design process in the open-ended responses; these are depicted in Table 17.

Table 17. *Themes for the Process of Creating Graphic Conceptual Models*

Theme	Exemplary Quote
Conduct preliminary research before meeting with stakeholders to understand how individuals articulate program theory and how it appears to be enacted. Research can include document review, one-on-one interviews, and a survey of stakeholders	[I] utilize a "pre-search" questionnaire to interview program faculty, staff, and administrators about program operations and how they perceive things working or not working as intended. [I] utilize program document review (always) and research examples of similar programs (if available) as other sources of information.
Select participants carefully; include individuals with knowledge of the program and who can support others if there is concern or resistance to the process	I try to get the right people in the room. It's best to get those who have been working on the programs directly and have the most insight and experience. It's great when everyone is involved because some people may have relevant experience and knowledge that others are not aware of. I try to include individuals who are perceived positively among the groups so they can function as persuaders.
Keep meetings small; if a large number of individuals must participate, consider incorporating small group activities	Coming in from the outside, I've also divided program staff (12+ from different units) into smaller groups to work on specific sections of a program theory or logic model that relate most closely to their daily work and then have them present back to the group.
Start the conversation gently, focusing on what people do, what they hope to accomplish, and how they think what they do relates to the end goal; it is important to remember that not everyone is familiar or comfortable with models of their program or reflecting on their work in this way	I have found it counter-productive to explicitly say that we are constructing a theory when there is none in writing, so we have conversations that vary a lot in length and depth depending on the availability of staff, which tends to be minimal! We discuss what the short- and medium-term outcomes are expected to be, what needs to happen for these to be realized, [and] what the possible barriers or facilitating factors might be.

Use interactive activities to prompt thinking about program elements and their relationships	Common activities we do include appreciative inquiry about the program; exercises that help stakeholders/practitioners think about what program success looks like; hands on activities using notecards; and role-playing activities about award acceptance speeches, etc.
Share the model on a screen, whiteboard, or flipchart so it can be developed or revised collectively	I have done a half day discussion with stakeholders based on the logframe. I put [the logframe] on the screen, asked the team to put each item of the logframe on the board (colored tags) with the links between them and asked them to question it and think beyond the logframe. I took notes in a Flip chart and proposed a program's theory based on the discussion.
Facilitate discussion of steps in the process of implementing the program and link them with expected outcomes	Depending on the group I'm convening, questions usually center around their perceived outcomes and purpose of the program, and then a dialogue about the inputs and resources which have been brought to bear in the initiative. We then have a discussion about how those inputs and resources are theorized to lead to the outcomes.
Bring social science theory into the conversation about program theory	I typically find that stakeholders have an implicit rather than explicit program theory. Therefore, I often start with asking them about what they do and why they do it, working from what they know explicitly to their implicit theory. I will then, sometimes, bring in theoretical or research literature to help broaden their view of what they are trying to do or capable of doing in the program context.

While Table 17 presents themes in the process of creating graphic conceptual models among respondents, Table 18 lists the topics that are frequently discussed during this process. In

this case, evaluators responded to a closed question with response categories tied to topics important in theory-driven evaluation to facilitate articulation of program theory and analysis of connections among components of the theory. One hundred-forty-one evaluators indicated to what extent they usually tend to discuss these topics with stakeholders when developing graphic conceptual models.

Table 18. *Topics Discussed to a "Great" or "Very Great Extent During the Process for Creating Graphic Conceptual Models*

Topics	Number	Percent
Beliefs or assumptions that underlie the program	83	58.9%
How social science theory or previous research and evaluation influences choice of strategies and desired outcomes	36	25.5%
If identified resources are sufficient to implement strategies to desired level	80	56.7%
If sufficient numbers of people are being served to expect influence on the desired outcomes	67	47.5%
Strength of connection between identified strategies and outcomes	102	72.3%
If duration and sequence of chosen strategies are sufficient to accomplish desired outcomes	85	60.3%
External factors that may influence desired outcomes	84	59.6%

Note: The five-point scale included "very great extent," "great extent," "moderate extent," "small extent," and "not at all."

An expectation in theory-driven evaluation is that by facilitating conversations using approaches like those listed in Table 17 and around topics like the ones listed in Table 18, shared understanding around program theory will be developed among key stakeholders. Seventy-nine (56.1%) of survey respondents considered that stakeholders involved in the process to explicate

program theory developed a shared understanding of how the program is expected to work to a "great extent" or "very great extent".

Use of graphic conceptual models to guide evaluation.

The survey focused on the design of evaluation questions tied to program theory represented in the graphic conceptual model; on the measurement of constructs identified in the program theory; and explanation of cause-and-effect associations between theoretical constructs.

Among the 141 respondents, program theory was used by 114 evaluators (80.8%) "often" or "always" to develop evaluation questions. Table 19 presents the focus of the evaluations usually conducted by respondents. There was little difference in the percentage of evaluators that focused on process and outcome measures, with the majority of evaluators focused on both. Moderators and the path of mediation received less attention. Although a large number of evaluators indicated that program theory guides the design of their evaluation, a smaller number noted that they report findings in relation to program theory (90 out of 141, 63.9%).

Table 19. *Focus of Evaluations "Often" or "Always" Conducted by Respondents*

Focus	Number of Respondents	Percent of Respondents
Process and Implementation	125	88.7%
Outcomes and Impact	132	93.7%
Context / Moderators	53	37.5%
Relationships in the Outcomes Chain	49	34.7%

Note: The scale choices included "always," "often," "sometimes," "rarely," and "never."

Conditions that may affect development and use of graphic conceptual models.

To identify conditions that may affect development and use of models, respondents completed a closed-ended question and an open-ended one. The closed-ended question listed several possible barriers to implementing theory-driven evaluation, including political dynamics tied to the organization, funder, and program itself, as well as evaluator knowledge and skills. Table 20 demonstrates the extent to which respondents felt that these barriers affected their implementation of theory-driven evaluation. The largest barriers were political factors, including time, money, and attitude of program staff.

Table 20. *Barriers to Implementing Theory-Driven Evaluation to a "Very Large Extent" or to a "Large Extent"*

Barrier	Number of Respondents	Percent of Respondents
Program/Organization is not interested in critically examining the program theory—they believe the program is fine as it is.	55	39.0%
Program/Organization only wants to know about final outcomes--little interest in relationships among outcomes along the causal chain or in variables that could affect outcomes.	49	34.8%
Funder is only interested in knowing about final outcomes—little interest in relationships among outcomes along the causal chain or in variables that could affect outcomes.	46	32.6%
Evaluator content knowledge of the field in which the program is based.	7	5.0%
Evaluator technical knowledge about theory-based evaluation and/or research methods.	6	4.3%

Time allocated to conduct a theory-driven evaluation.	63	44.6%
Financial resources allocated to conduct a theory-driven evaluation.	69	49.0%

Note: The scale choices included "to a very large extent," "to a large extent," "to a moderate extent," "to a small extent," and "not at all."

Eighty-five respondents chose to complete the open-ended question, which asked if evaluators encountered any additional barriers to implementing theory-driven evaluation. Twenty-six of these responses emphasized that the listed barriers in the previous closed-ended question captured well their own experiences. The remaining fifty-nine answers offered new sub-themes to barriers related to organizations, programs, and evaluators. These are outlined in Table 21.

Table 21. *Additional Barriers to Theory-Driven Evaluation*

Theme	Exemplary Quote
<u>Barriers Related to Organization Staff</u>	
Staff evaluation capacity	I think our biggest challenge is getting frontline staff members and their immediate supervisors (who typically come from the frontline) to think beyond activities to theory. This is new discipline for them and most are unfamiliar with it and a minority are resistant and view it as a waste of time.
Staff do not see relevance or value of developing models of program theory	Many local programs that I work with are focused on "doing" and sometimes see conceptual models and theories as "nice-to-think-about," but not germane to their everyday work.

Staff resistance to examining program logic because of belief that the program is fine as it is or that it is “too special” to be explained by existing theory	Stakeholders who are carrying on a legacy model (meaning carrying on the work of a charismatic program founder) may be reluctant to closely examine…what they are doing. Staff who have risen through the organization to positions of leadership with little experience in other programs may have difficulty thinking about the program in any way other than the way they have experienced it—and this also applies to people who are now staff but were originally served as clients by the program.
Staff distrust due to past negative experience with developing graphic conceptual models	I think many practitioners have had a lot of experience with 'evaluators' coming in and telling them how their program should work or applying theory in a way that doesn't feel true to the program—so there is a lot of distrust. Evaluators need to recognize that they are not programmatic experts. The practitioners are always the 'experts' in their program. The evaluator needs to use facilitation techniques that help the practitioners come to a better understanding of the relationships between process, outcome, and theory that drive program success (or lack of) and help some practitioners appreciate why that understanding is critical to program success and illustration of outcomes. Evaluators and practitioners must be partners for it to work—the old adage of 'translate research to practice' is completely outdated. It should be more of a two-way street in order for theory driven evaluation to really take hold and improve programs.

Barriers Related to the Evaluator	
Evaluator awareness or access to social science theory	I currently work in a university at the moment, but I imagine that if I didn't, gaining access to relevant social science theories would be an enormous barrier.
Evaluator skills and tools for model development.	My graphics skills are deficient.
Barrier Related to the Intervention	
Amount of emergence or innovation in a program	No social theory exists, or intervention is too complicated to reduce it to a program logic.

CHAPTER FIVE

DISCUSSION

At its core, theory-driven evaluation involves creating a model to show how an intervention leads to outcomes and using the model to guide the evaluation. There are those who believe there are weaknesses to this approach (Scriven, 1997; Stufflebeam, 2001), and others who see both its potential and limitations (Pawson and Tilley, 1997; Rogers et al., 2000; Donaldson, 2003; Leeuw and Donaldson, 2015; Patton, 2008; Frazier-Anderson, Hood, and Hopson, 2011). This investigation was undertaken in the spirit of contributing to a better understanding of how theory-driven evaluation is practiced by some evaluators. The study comprises a content analysis of evaluation studies published in peer-reviewed journals and a survey of evaluators. Prior to this inquiry, there have been only a limited number of empirical studies on theory-driven evaluation (Coryn et al., 2011; Torres, Hopson, and Casey, 2013; and Munter, Cobb, and Shekell, 2016). Of these, the research by Coryn et al. most influenced the design of the present investigation. To facilitate discussion of findings, similarities and differences between the two studies are presented in Table 22.

Table 22. *Comparison of Coryn et al. and Bonis Studies*

	Similarities	Differences or Contradictions
Focus	Both Coryn et al. and Bonis looked at how core principles in theory-driven evaluation are applied in published evaluation studies.	Bonis looked more deeply at graphic conceptual models in published evaluation studies: their development process, design, and use in evaluation. Bonis also investigated how models are adapted based on complexity or cultural context, as well as what program archetypes could be helpful to the field.

		Following the content analysis of published evaluation studies, Bonis also investigated many of the points referenced above through a survey of evaluators.
Method	Both reviewed evaluation studies in evaluation-related journals and substantive journals in disciplinary fields	Coryn's sample for content analysis covered a nineteen year period (1990 to 2009) and included book chapters in addition to journal articles. The final sample included 45 articles. One criteria for inclusion was that studies identify theory-driven evaluation as the driving approach. Bonis' sample for the content analysis covered a thirteen-year period (2003 – 2016) and focused only on journal articles. The sample was larger (116) because the inclusion criteria did not require theory-driven evaluation to be identified as the driving approach. A key consideration was the use of a graphic conceptual model to guide the evaluation study. Bonis' study also involved a survey of evaluators and their use of graphic conceptual models in evaluation (n=141).
Key Findings of the Content Analysis		
Frequency of Published Studies of Theory-Driven Evaluation	The reviews of both Coryn et al. and Bonis indicate that the number of studies published	

	that use elements of theory-driven evaluation remains relatively small (10 per year in Coryn et al's sample and 16 per year in Bonis' sample).	
Source for Developing Program Theory	The majority of program theories presented in published articles were based on theory and research. Stakeholder theory was the second most commonly used source.	Bonis differentiated between social science theory and previous research or evaluation. Previous research or evaluation was used more often in developing models than a specific social science theory.
Program Theory Guided Evaluation Questions	The use of evaluation questions tied to program theory to guide the evaluation was high in both studies.	Bonis found that in cases where evaluation questions were explicitly stated, the questions were used in 93% of cases to drive the evaluation. Coryn et al. found this to be true in 75% of studies.
Measurement of Constructs Presented in the Program Theory	Both Coryn et al. and Bonis found that evaluation studies measured constructs identified in the program theory. Process constructs were measured at about the same frequency (in 45% of the studies reviewed by Coryn et al. and in 50% of studies reviewed by Bonis).	Bonis found more evaluation studies that measured constructs related to outcomes and context presented in the program theory than did Coryn et al. Eighty percent (80%) of articles reviewed by Bonis included measurement of constructs related to outcomes and 54% included measurement of constructs related to context. In contrast, Coryn et al. found constructs measuring outcomes outlined in the program theory in 49% of articles and constructs measuring context in 36% of articles.

Investigation of Mediators and Moderators	Coryn et al. and Bonis both found that in approximately 65% of studies seeking to explain a cause-effect association, mediator relationships were investigated.	Bonis found more studies that intended to explain a cause-effect association to investigate moderators than did Coryn el al. (81% vs. 53%).

Findings from this study will be discussed in four sections: 1) "Frequency of the Use of Graphic Conceptual Models in Evaluation," 2) "Design of Models to Describe Program Theory," 3) "Development Process of Models to Describe Program Theory, and 4) "Use of Program Theory in Evaluation." When appropriate, results will be discussed in relation to the study of Coryn et al. (2011). This is the case in sections one and four; sections two and three represent additional contributions of this study.

Frequency of the Use of Graphic Conceptual Models in Evaluation

Over a thirteen-year period, an average of 22.2% of articles describing a program evaluation in five major peer-reviewed journals focused on evaluation used a graphic conceptual model. The five evaluation journals included the *American Journal of Evaluation, Evaluation and Program Planning, the Canadian Journal of Program Evaluation, Journal of Multidisciplinary Evaluation*, and *Evaluation Journal of Australasia*. Overall, this percentage is relatively small given the length of the time period. The finding concurs with that of Coryn et al. In only one evaluation journal, *Evaluation,* was there a relatively large percentage of articles describing a program evaluation that included a graphic conceptual model; the percentage was 54.9%. (It must be noted that the aim of this journal's editorial board is to make the journal

"theory-led" and a content search by Leeuw and Donaldson (2015) provide support for this orientation.)

While overall fewer studies may be published that use theory-driven evaluation, the survey results seem to indicate that evaluators often are inclined to use program theory to guide their evaluations. Among survey respondents, 30.5% reported that they "always" use program theory in their evaluations, 34.0% said that they "often" do, and 26.2% indicated that they "sometimes" do so.

Design of Models to Describe Program Theory

Sources of program theory.

Results from both the content analysis and the survey about the sources for designing program theory align with the findings of Coryn et al. (2011). Social science theory and existing research and evaluation were the primary sources used for developing the program theory depicted in graphic conceptual models. In the sample of articles reviewed by Coryn et al., 91% of models were developed from existing theory and research. In this study, 75.9% of models in the articles were based on existing research and evaluation and 32.8% were based on a specific social science theory. Stakeholder theory was the second most commonly used source for developing the models, with this being true in 49% of the articles reviewed by Coryn et al. and in 44.0% of articles reviewed here. Survey results followed a similar pattern. Social science theory, research and evaluation, and stakeholder theory contribute in different, worthwhile ways to the program theory. The prevalence of their use is encouraging for theory-driven evaluation. Most definitions of program theory indicate that it should include a meaningful description of how the program is expected to work by stakeholders. Donaldson (2007) adds that "it is highly desirable

if program theory is rooted in, or at least consistent with, behavioral or social science theory or prior research."

Structure of the graphic conceptual models.

One significant distinction of this study from that of Coryn et al. (2011) is its deeper examination of the graphic conceptual models included in published articles on evaluation studies. Another difference is the survey of evaluators, which allowed for inquiry around the development and design of models. As a result, the findings discussed in this section are not reviewed in relation to the study of Coryn et al.

Paths of mediators and moderators: The identification of paths of mediation and moderation in program theory was of interest in this study because they are central to theory-driven evaluation. Before discussing findings, two assumptions of this study are recognized. First, it was assumed that the graphic conceptual models presented in the evaluation articles represented closely the thoughts of intervention designers and evaluators of how the intervention was expected to result in change. It is possible, however, that some authors included a graphic conceptual model in the article without that intention, so the models might not tell the whole story of an intervention. Second, the mediating and moderating relationships were determined based on the construction of the model. Assessment was not made of the quality or strength of relationships described in the model (the benefit of doing so in future investigation will be discussed below).

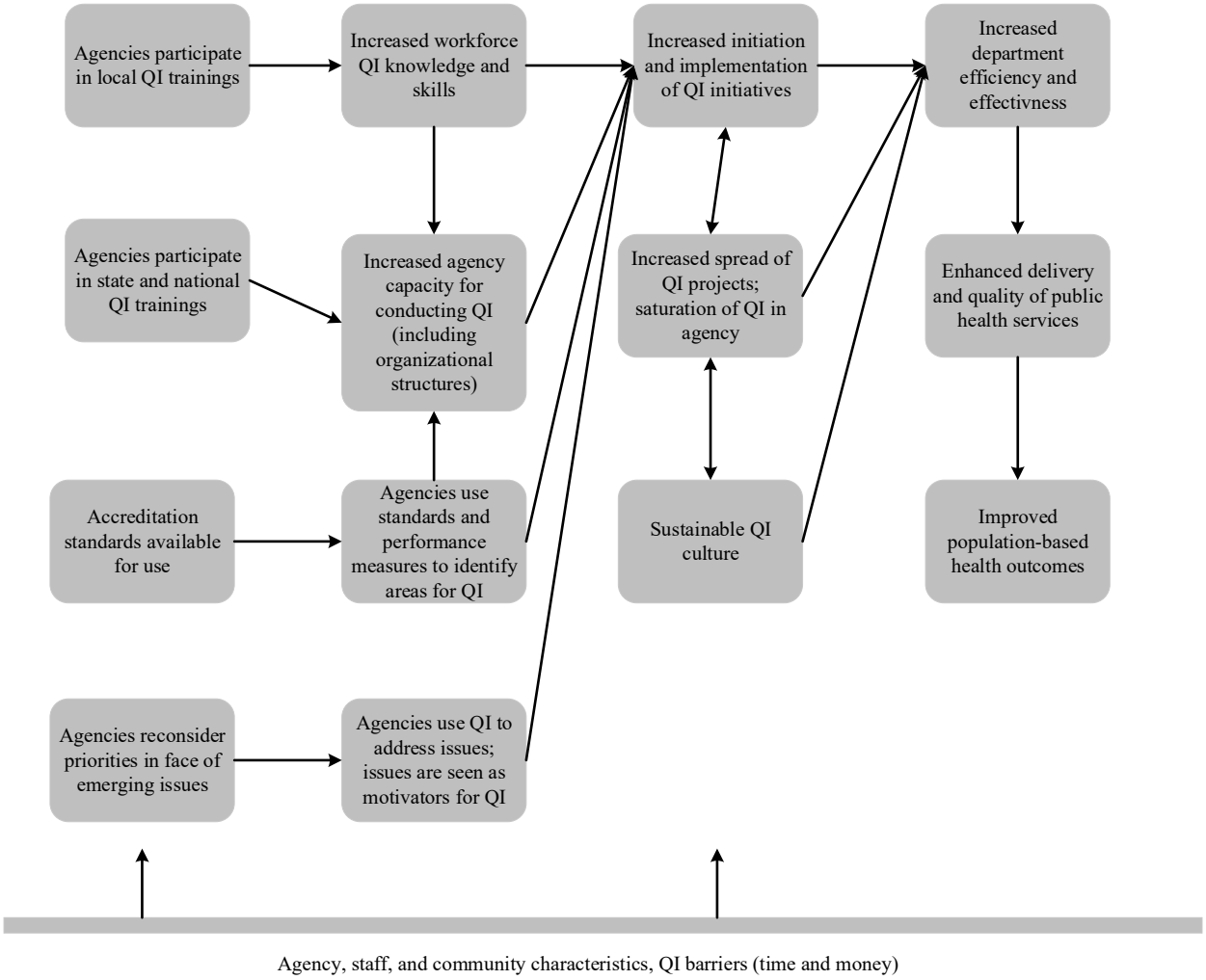
Figures 1 through 5 in the "Results from Content Analysis" section offered illustrations of models encountered and how they were coded to describe mediation. Three variables were used to understand mediation: the presence of mediators, the length of the mediator chain, and whether or not the pathway of causation was depicted specifically or generally. A key tenet in

theory-driven evaluation is breaking open the "black box" of improvement, presenting the presumed process by which changes takes place. The presence and length of the mediator chain are two ways to look at the contents of the box. In the sample, the majority of the models did indeed make an effort to open the "black box." Only 16.4% of the models omitted a mediator; 83.6% outlined the mediators expected to lead to change. Among the models with mediators, most included one mediator (23.3%) or two mediators (24.1%); there were others with three mediators (14.7%) or more (15.5%). The number of mediators may depend on the nature of the intervention, e.g., if it is more simple or complicated, so the interpretation of these findings is not that a longer chain is better. Of primary interest was to see if an effort was made in the models to chart the path of change. The third variable also ties to this interest; it focused on whether the path was outlined specifically or generally. Among the models in the study, the path was outlined specifically in 49 cases (50.5%) using either arrows, positioning of text, or rows in a table. The findings connected to these three variables indicate that more evaluators are identifying the mechanisms that may lead to desired change and presenting how an intervention is expected to affect outcomes visually in a graphic conceptual model.

Regarding moderators, the coders looked to see if moderators were included in the model, and if they were positioned in such a way as to indicate the relationship being moderated. Figures 1 through 5 also provided examples of moderators in a model. Moderators were included in 43.1% of graphic conceptual models. In 48.3% instances, moderators were included in the text of the article but not in the model. The fact that moderators were identified in 91.4% of the interventions evaluated—whether in a model or in the text—is encouraging. It indicates that evaluators and their stakeholders frequently are reflecting on the circumstances under which a change may occur.

There is a balance to consider in making models: conveying an appropriate amount of information so that the model is easily understandable while still conveying accurately how an intervention works and under what circumstances. Including multiple moderators in a model may be a challenge on occasion, so the option of calling them out separately in a narrative may be desirable. Figure 10 below shows a common way that moderators were mentioned in models in the sample: above or below the model as a whole, with arrows pointing in the direction of the mediation paths in the model. The specific mediation path influenced by the moderator was not always called out; one possible implication being that all paths were influenced. Also, the language describing the moderator was often general. For example, in Figure 10, "agency, staff, and community characteristics" are indicated as influencing change. The specific characteristics are not called out in the model. (In some cases, they were described in greater detail in the narrative.)

Figure 10. Model of an Effort to Create Quality Improvement Culture in a Department (Davis, et al., 2014)



The content analysis as undertaken in this study revealed the frequent presence of mediators and moderators in graphic conceptual models. Moderators, however, were included less frequently and less specifically in many models but then often discussed further in a narrative. The content analysis did not look at the strength of the relationships identified. This would be important to undertake in a future investigation because the strength of relationships is a critical factor in the amount of change in a desired outcome. This study looked at the surface of the models, but there is room to examine deeper. Such a deeper investigation could involve

looking at the results of testing mediation and moderation in those articles that included the results. Another way would be to review more intensely the social science theory, research and evidence that informed the development of the model.

Types of graphic conceptual models.

The design of the models was most often linear in both the published articles on evaluation studies and survey responses of evaluators. The finding was expected; however, this study also sought to look at some of the points raised by constructive critics of linear models of program theory. For example, Patton (2008) has questioned the effectiveness of linear models in complex programs, which he describes as programs with a large number of interacting and interdependent elements in which there is no central control and where cause and effect are unpredictable and difficult to understand. While complex programs may pose the greatest challenge to evaluators, complicated programs with many components may also present difficulties (2008). In the content analysis, several variables were coded to try to identify programs that might be characterized as complex (Table 23). The characteristics were based on descriptions in the literature of complicated and complex programs (Rogers, 2008; Patton, 2008).

Table 23. *Variables and Coding Schema to Identify Complicated and Complex Interventions*

Variable	Coding Schema
Number of sites	single site; multi-site (same model implemented); multi-site (variation of model implemented)
Single or multiple organizations	single or multiple organizations
Interdisciplinary or intersectoral collaboration	yes or no
Level of change	individual, interpersonal, organization, community, public policy
Number of causal strands	(Enter number)
Proportion of impact	linear causality with proportional impact, recursive with feedback loop(s), tipping point(s)
Author describes project as "complex"	yes or no

Rogers (2008) offers these suggestions for labeling an intervention as "complicated," "complex," or "simple." A "complicated" intervention may display the following characteristics: involve multiple, often interdisciplinary agencies; have multiple simultaneous causal strands; or have different causal mechanisms operating in different contexts. A "complex" intervention may have non-linear and disproportionate outcomes through a recursive feedback loop or a critical tipping point or outcomes may be emerging instead of being pre-defined. With the exception of emerging outcomes, these characteristics were included in the coding protocol. An additional code was added to help identify complex interventions: whether or not the authors described the intervention as complex. In contrast, a "simple" intervention may involve a single organization with a single or very few causal strands and linear impact (Rogers).

The majority of the articles in the sample included characteristics of complicated interventions. Only two met a strict definition of a simple intervention: carried out by a single

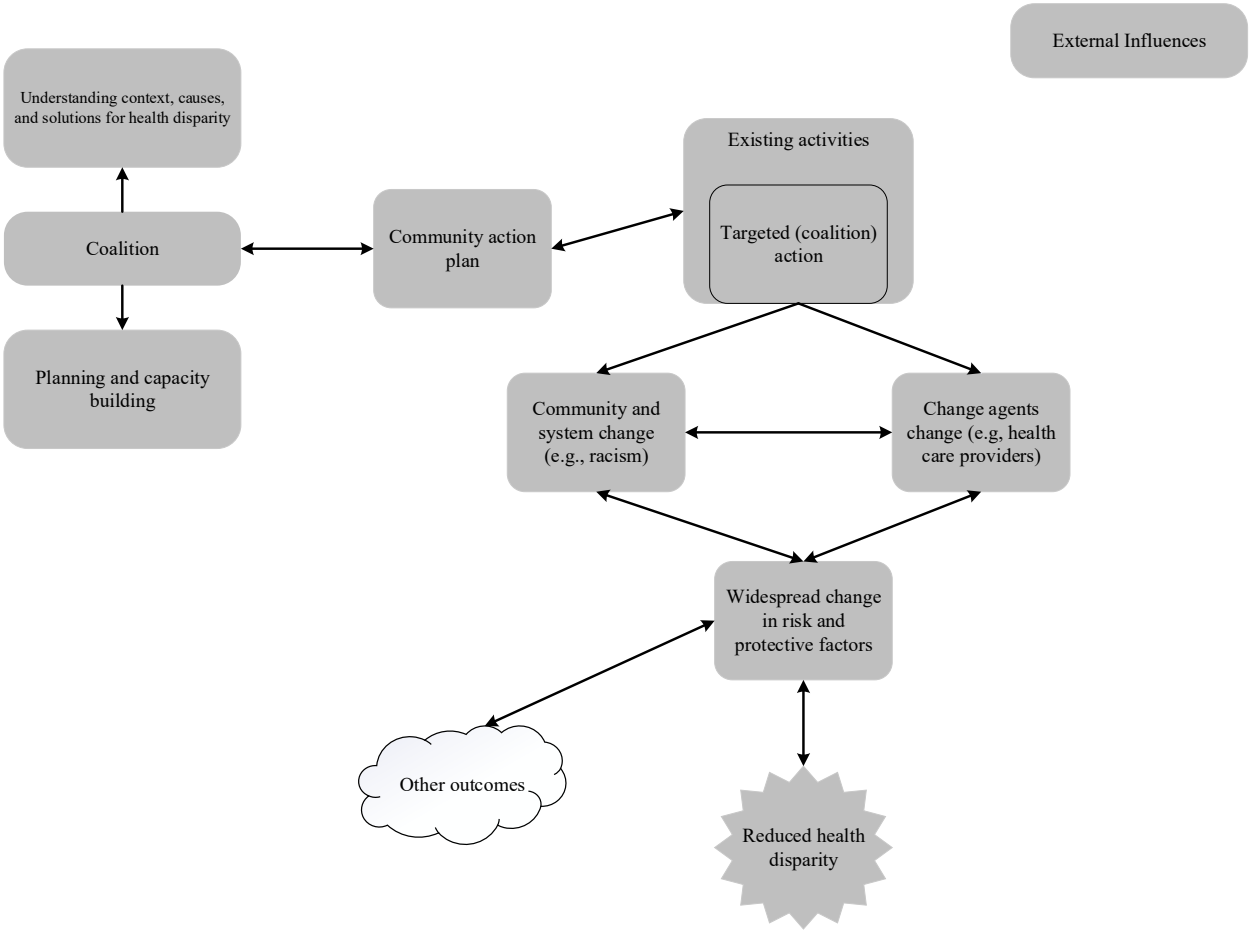
organization in one location, with few causal strands (two and three), and linear causality. Forty-three interventions (37.0%) met all the coded characteristics to be considered a "complicated" intervention. There were other interventions that exhibited at least one characteristic of a complicated intervention but not all. Thirty-six interventions (31.0%) included recursive relationships and three experienced tipping points (2.6%), making them "complex" interventions by the coding protocol. (The majority of interventions [100 cases, 94.8%] expressed linear causality.)

A closer look at the data uncovered some nuances in describing complicated and complex interventions. For example, there was some overlap among those interventions that met all criteria for being "complicated" and for being "complex." Sixteen "complicated" interventions also included "recursive relationships" and two included a tipping point. Then, the additional code for complexity revealed that the term "complex" is one around which a widespread shared definition does not yet seem to exist. There were twenty-eight articles where the authors indicated that the intervention was complex. However, in fourteen of the interventions, the causality depicted was linear only. In twelve there was non-linearity and disproportionate impact and two did not describe impact. The majority of interventions described as complex did not include a distinguishing feature of being complex: non-linearity and disproportionate impact. (The authors may have used other definitions of complex, but these were not collected or analyzed in this study.)

With the data showing that the majority of interventions were "complicated interventions" expressing "linear causality," it may not be surprising that of the 116 published articles reviewed, a linear model was used in 101 instances (87.1%). Even among those interventions expressing recursive relationships and tipping points, all graphic conceptual models

were linear. Figure 11 offers an example of an intervention described as complex by the authors and that also met characteristics of complexity in the coding protocol but that used a graphic conceptual model that was linear. The intervention was designed to apply and expand promising strategies to improve health disparities in cardiovascular disease and diabetes in 14 urban communities. It drew on a multi-sectoral coalition that sought to impact change at various levels: personal, organizational, community, and policy. While there were some expected outcomes leading to a larger effect through reinforcing feedback loops, there were others that could not be pre-determined as they emerged in part from the developing interactions of coalition partners throughout the life of the initiative.

Figure 11. Community Health Project (Plescia, M., Herrick, H, and Chavis, 2008)



Survey findings revealed to a slightly greater degree than did the content analysis that evaluators are beginning to explore alternative ways of depicting programs whose path of change is less predictable. While linear models still dominated the 141 responses (69.5%), there were also evaluators who explored alternative models (30.5%) when the path of change was less certain. Also, working with adaptive or emergent programs was one theme among open-ended responses cited as a challenge to implementing theory-driven evaluation. Given that few interventions in the content analysis met the characteristics of a complex intervention, and given the uncertainty around definitions used to describe an intervention as complex by the authors, there is need for additional study of the use of graphic conceptual models to describe the program theory of complex interventions.

Another line of constructive criticism of graphic conceptual models used in theory-driven evaluation focuses on their cultural responsiveness. Some have argued that linear models may not be the most appropriate to use in certain cultural contexts (Johnston, 2002; LaFrance, 2004; Frazier-Anderson et al., 2011). The survey once again proved more revealing in this regard than the content analysis. Several survey respondents indicated that cultural context of the program and its participants often influences how they approach designing graphic conceptual models. Of the 141 evaluators, 19.1% said that cultural context "always" influences their model design, 22.7% said that cultural context "often" influences design, and 27.0% said that cultural context "sometimes" does so.

A little over half of evaluators surveyed (51.7%, 73 of 141) reported modifying graphic conceptual models in response to different cultural contexts. An open-ended question allowed for modifications to be described by respondents. Seven themes around modification approaches emerged from the responses: simplify the complexity of the model; reduce the linearity of the

model; modify language used in the model; use culturally-relevant metaphors; use culturally-relevant images and colors; make culture explicit in the model; and provide an alternative to the model.

The most common modification, used by 26.7% of evaluators, involved changing the language used in the model; such adjustments include simplifying the language, making culturally-sensitive word choices, using more positive and empowering language, and direct translation. *"We're found traditional logic models sometimes are confusing for community members,"* shared one respondent. *"Wording that tells more of a story, particularly with examples, seem to be better received."* Using culturally-relevant images and colors was the next most frequently used approach (22.9% of respondents). A different evaluator recounted matching the design of the model to the organization's brand (e.g., colors, logo, and themes). Simplifying the model and reducing linearity were two approaches used by 14.3%, respectively. Culturally-relevant metaphors were used by 12.4%. *"We have used a growing plant with farm-workers, a ripple-effect type graphic to indicate rippling effects in the community, and 'seeds' as parts of the theory of change,"* offered another person.

In a different example, the writer explained that *"we deliberately avoided tables, arrows, and diagrams and went with a tree image. Cyclical representations (versus linear) can be appropriate for working with tribes (although I do not want to make categorical statements here; tribal communities and tribal leadership can be very different)."* The illustration and reflection in this quote raise an important point when contemplating how best to develop graphic conceptual models in different communities. There is no one type of model that works better with one community over another. Each community is distinct and it is necessary to deeply listen to its members and come to know them—their values, interests, and preferred ways of learning

and expressing ideas—to help determine the best model that will recognize the culture and be a representation to which the client can relate. Another evaluator reinforced this idea, making note of the following:

> *Sometimes it is more clear to folks to ensure that their language and their graphics are infused. Also, sometimes due to cultural contexts, graphics are not used traditionally and sometimes, infographics, pictures, or just text are used. It is very dependent. For example, in some contexts, it is important that nothing look too "researchy" or academic. Anything that looks academic is immediately discounted. Whatever you do has to be something that they can "see" themselves in or the buy-in and ownership will not be there (which will result in it being a wasted exercise).*

If evaluators develop models that key stakeholders do not find relevant or accessible, it is less likely that they will be interested in using the program theory to guide their thinking about the program and its evaluation. It is important for evaluators to know their audience and adapt graphic conceptual models in different circumstances. Knowing more about how evaluators modify graphic conceptual models to make them more culturally-sensitive and user-friendly is helpful for both training and future research. As part of this, additional skills may be needed in areas like data visualization and graphic design. Indeed, one theme among open-ended responses in the survey to a question that asked about barriers to theory-driven evaluation were skills and tools for developing visually appealing models.

While graphic conceptual models may take different forms, there has been growing interest in identifying program archetypes (Funnell and Rogers, 2011; Lemire, Whynot, and Montague, 2018). Programs may focus on different issues and serve different people, but there may be similarities across design and implementation characteristics and impact pathways. If such archetypes could be identified, they could serve as a heuristic device for practioners and help build knowledge about programs in a systematic manner. This study took a step towards understanding possible archetypes. To date, only Funnell and Rogers have proposed a series of

archetypes that could be used to classify programs. They have developed in greater detail five archetypes: advisory, public information, and education; carrots and sticks, case management, direct service delivery, and community capacity building. They also suggest a sixth, network theory, that must be refined further.

Initially, the research team coded inductively program theories from the published evaluation articles in this study. The resulting codes, however, closely resembled the archetypes identified by Funnell and Rogers so the coding team re-coded the program theories using those archetypes. The archetypes of Funnell and Rogers could be seen as the foundation for almost all of the program theories in the articles in this study. This suggests that the archetypes of Funnell and Rogers are good starting points for more research. In the sample, 61.5% of models followed the advisory, public information, and education archetype. These models described programs that took different approaches to providing information and educating others, including training, community outreach and education, and social marketing. This finding is not surprising given the dominance of public health interventions described in the articles and the frequency of individual or interpersonal theories of behavior change mentioned as the source of program theory, including social cognitive theory, the theory of planned behavior, and the health belief model.

Identifying archetypes for social interventions could be seen as having roots in structuralism, a school of thought that sees individual and collective behaviors as emerging from some underlying structure. Archetypes isolate elements of an intervention and their interrelationships. A critique of structuralism is that it runs the risk of being too vague to be useful. On the other hand, there could be educational value in categories based on unique patterns. For archetypes of social interventions to be more helpful, different classification schemes may need to be considered. The one used by Funnell and Rogers is based on form,

much like different genres of literature (e.g., poetry, fiction, nonfiction, and drama). Topic classification of social interventions could complement genre classification, e.g., youth leadership development, health education, organizational capacity building, and other substantive focus areas. Studying and understanding archetypes is a potentially useful starting point both when designing and evaluating social interventions.

Development Process of Models to Describe Program Theory

Through a survey of evaluators, this study also looked at the process for creating graphic conceptual models. Among the empirical studies on theory-driven evaluation, very few look at the development process for models of program theory (Torres, Hopson, and Casey, 2013). In this inquiry, all 141 survey respondents shared a description of their process for developing graphic conceptual models. The majority of respondents engaged in a participatory process to co-design models with key stakeholders such as organizational leadership, program staff, or clients. The strategies that they used, which were categorized thematically in Table 16, underscored the importance of facilitation skills among evaluators. Just as facilitation skills are necessary in other forms of participatory evaluation, so they are when stakeholders are involved in program theory development. (While many survey respondents engaged in participatory theory-based evaluation, it must be noted that the approach can also be non-participatory if desired or if required by the situation.)

Some evaluation theorists propose that co-designing program theory depicted in a graphic conceptual model can result in shared understanding of how an intervention is expected to solve social problems (Donaldson, 2003). Among survey respondents, 56.1% felt that stakeholders engaged in a process to develop program theory achieved such shared understanding to a "great extent" or a "very great extent." Discussions facilitated by evaluators encouraged deeper

thinking about the program among stakeholders. For example, 72.3% of participants reported that they discussed with stakeholders the strength of connection between identified strategies and outcomes; 60.3% reported that they discussed whether or not the duration and sequence of chosen strategies were sufficient to accomplish desired outcomes; and 58.9% reported that they discussed beliefs or assumptions that underlie the program. These are topics of conversation essential to theory-based evaluation that encourage stakeholders to develop better understanding between process, outcomes, and mechanisms that can lead to program success (or lack thereof). Other topics, however, were discussed to a lesser degree. For instance, only 25.5% of evaluators noted that they discuss how social science theory or previous research and evaluation may influence choice of strategies and desired outcomes. Bringing social science theory into the conversation about program theory is strongly within the purview of evaluators. Some surveyed evaluators recognized this, but cited access to social science theories as a challenge in the open-ended question asking about barriers to theory-driven evaluation.

Use of Program Theory in Evaluation

In this study, the content analysis of published evaluation studies provided the most insight into use of program theory in evaluation. Overall, findings indicate that graphic conceptual models often were used to develop evaluation questions and that constructs and relationships outlined in the model were frequently investigated.

Several of the results affirm findings of Coryn et al. (2011), while others differ slightly. For example, in 93.2% of the cases where evaluation questions were stated explicitly in articles, the questions were tied to the program's underlying logic or theoretical foundations. This was higher than that observed by Coryn et al., who found that program theory guided question formulation in 76.0% of instances reviewed.

Another difference involved construct measurement. In this study, the evaluations in 56.0% of the articles reviewed assessed constructs articulated in the program theory. Among the evaluations measuring constructs from the program theory, process constructs were measured in 50.8% of cases, outcome constructs in 80.0%, and contextual constructs in 53.8%. The percentages for measuring outcome constructs and contextual constructs are higher in this study than in that of Coryn et al. That investigation found that outcome constructs were measured in 49.0% cases, and contextual constructs in 36.0% (process constructs were measured in 45.0% of instances; a percentage similar to that found in the present study.)

Measuring outcomes alone does not get at a core consideration for theory-driven evaluation: looking at how an intervention works and under what conditions. In the sample, 24 evaluations (20.7%) measured only the extent to which outcomes were attained. On the contrary, 73 evaluations (62.9%) collected data to describe the relationships in the program theory and 49 evaluations (42.2%) provided greater explanation of the relationships by analyzing mediation and moderation.

This study did not assess the specific method or approach used to test a relationship between mediators or moderators. The sample included studies that used quantitative and qualitative methods, statistical tests, and techniques like contribution analysis (Mayne, 2012) and process tracing (Beach and Pedersen, 2013) to explain how an outcome or set of outcomes occurred. Overall, it is encouraging to learn that over half of the evaluations undertaken made an effort to understand the relationships in the model. It is also promising to see the use of approaches like contribution analysis and process tracing, emerging methods for causal inference that may be more feasible and practical in certain situations. Contribution analysis and process tracing both involve searching for evidence that increases confidence in the existence or non-

existence of the causal mechanism or theory of change by increasing confidence in the existence of its component parts (Mayne and Befani, 2014).

The findings of this study regarding analysis of mediation and moderation partially match those discovered by Coryn et al. Both studies found that in approximately 65% of published articles on evaluation, mediator relationships were investigated (in 65.3% of the articles in the current study and in 67.0% of the articles in that by Coryn et al.). A point of difference had to do with moderators. In the present study, methods were used to test the extent to which one construct moderates the relationship between other constructs in 81.6% of evaluations that tried to explain cause-effect associations. Coryn et al. found that to be true in a smaller percentage of the articles that they reviewed (53.0%). The increased attention to moderators in the articles may reflect a growing recognition that context matters for interventions to yield desired outcomes. Pawson and Tilley (1997) emphasized the importance of context when they argued that for an evaluation to be useful for decision makers, they need to identify "what works in which circumstances and for whom" rather than merely "does it work?"

Strengths and Limitations

This study contributes to the field of evaluation in several ways. First, it adds to the empirical research on theory-driven evaluation, a popular approach to evaluation but that has not been empirically studied often. Second, the content analysis of a larger sample of published evaluation articles than previously investigated (Coryn et al., 2011) allows for confirmation and contradiction of earlier findings that can add to the knowledge base and point to directions for future research. Third, this study investigates issues that have been raised in relation to theory-driven evaluation, including the need to adapt models for more complex programs and in different cultural settings and the possible benefits of identifying program archetypes. While

these topics have been addressed theoretically, they have not been studied empirically on a large scale. Lastly, this study gathered information on the process of developing graphic conceptual models, including facilitation strategies and topics of conversation among stakeholders. Again, while the process of developing models has been described in the literature, published empirical studies on the model development process are not available. This investigation took a comprehensive approach to understand theory-driven evaluation in practice, as evidenced in published evaluation articles and in evaluator responses to a survey.

While this project has several strengths, it is not without limitations. One limitation of this study is that the graphic conceptual models analyzed came only from peer-reviewed journals. It is possible that graphic conceptual models were not presented in full detail because of constraints around length of articles by different publications. The same could be true of descriptions of the evaluation. Looking at complete evaluation reports could be another way to study the use of graphic conceptual models in evaluation. However, even evaluation reports may underrepresent graphic conceptual models. Models may be created as part of the evaluation process, but not included in the final evaluation report.

Then, looking at published articles provides information about the behavior of evaluators with a desire to publish. Reviewing models developed by evaluators who do not publish is also needed. The survey in this study was meant as a complement to the content analysis, building on topics that came up in the content analysis or addressing new ones that the content analysis could not cover, such as the process to develop models. However, the survey did not make a distinction between evaluators who publish or not, nor did it ask for sample models to be uploaded.

Another limitation is the exclusion of articles using realist evaluation from the sample. Many of the articles on realist evaluation that were identified as the study sample was being

assembled were theoretical in nature, explaining what realist evaluation is and why it should be considered. In some cases, examples were included of the CMO (Context Mechanism Outcome) configuration used in realist evaluation, but these configurations were sometimes incomplete or key components were not identified explicitly. In other instances, the CMO configuration was not presented in the context of a full evaluation, which was the criteria for inclusion in this study. Ten articles using realist evaluation were identified that could have been included in the sample, but during training with coders it became clear that the coding guide as constructed would not capture what very well may be a great strength of realist evaluation—explaining the mediator and moderator relationships that underlie a program. While this is a limitation of this present study, it could be seen as a starting point for future research.

Finally, this study is also limited by a research design more favorable to breadth than depth, particularly in the content analysis. This investigation looked at several key points in the use of graphic conceptual models in theory-driven evaluation and revealed some useful findings that add to the field. However, the largely deductive, quantitative approach limited obtaining a deep understanding of the design and quality of the models, and how and why certain actions were taken in developing and using models. Utilizing pre-determined codes based on theory and prior research in a content analysis can mask important contextual aspects in the object of study. The method is also inherently reductive and susceptible to researcher bias. It would be beneficial to use an inductive approach and qualitative methods in future studies that look more deeply at aspects of theory-driven evaluation.

Future Research

One area where a more inductive approach could be beneficial would be in better understanding the use of graphic conceptual models in complex interventions and in interventions carried out in different cultural contexts. The insight offered by the open-ended question in the survey on the modifications by evaluators to models used with different cultural groups shows the depth of information that could be obtained using a more qualitative approach. Qualitative research on the use of graphic conceptual models with specific cultural groups may be a fruitful next step. The same may be true of research on complex interventions. Case studies of specific complex interventions and how graphic conceptual models are used to guide planning and evaluation could uncover details that would advance the conversation on the value of models in such instances.

Then, as noted in the discussion on archetypes, alternative classification schemes could be explored. For instance, instead of categories focused on form there could be a focus on substance. Such classifications also could build on the social science theories that frequently appear in certain classes of interventions. A benefit of this approach is that the exemplars created would offer guidance on the incorporation of solid social science theory in intervention design. The idea of developing and using archetypes in evaluation is relatively new so there is room for additional research in this area.

A third possible area for deeper investigation is the use of social science theory. Explanatory theories from the social and behavioral science literature can offer valuable insight when developing program theory. This study revealed growing adoption of social science theory to develop program theory. A next step would be to look at *how* social science theories are used to inform program theory. Leeuw and Donaldson's (2015) "theory-knitting" and "theory-

layering" approaches are two new ways to examine the relationship among theories. In theory knitting, previous theories are integrated into a single higher order theory. In theory layering, mechanisms of change are analyzed by different theories but viewed as part of a nested system with "upward causation". These approaches could be useful in improving the use of theory in evaluation, including in the evaluation of more complex interventions.

A final area of future study involves realist evaluation, a promising direction for theory-driven evaluation. Realist evaluation asks "what works for whom in what circumstances and in what respects, and how?" (Pawson and Tilley, 1997, p. 2). Realist evaluation emphasizes the behaviors of individuals, which may influence the outcome of a program. Programs or policies provide resources to individuals; how those actors interpret and act upon the ""mechanism of change" then determines success or failure of the intervention. To understand the relationship between two events, it is necessary to understand the mechanism connecting the two and the context in which the relationship occurs. Realist evaluation also attempts to address issues of complexity, recognizing that interventions may be complex because of volition; implementation chains; intended and unintended outcomes; and emergence, among other reasons. While realist evaluation may seem diametrically opposed to the idea of archetypes of program theory, Lemire el al. (2019) found "mechanism archetypes" to be possible in a review of realist evaluations. This study of 195 published realist evaluations is the most comprehensive review of realist evaluations to date. However, given the interest in realist evaluation, there is need for more investigation to better understand how realist evaluation is implemented and if it more effectively could be used to explain what works, for whom, and under what conditions. Leeuw and Donaldson (2015) found that realist evaluation had the highest number of "hits" in a search

of theory-related terms in the journal *Evaluation* between 2000 and 2015, highlighting the increasing interest in this approach to evaluation, and need for research on it.

Conclusion

Graphic conceptual models help to clarify thinking about an intervention. They are tools that help to articulate and make explicit assumptions about a program's context and what stakeholders expect to achieve. They are effective tools for planning and evaluation. Graphic conceptual models can be used to intentionally and strategically plan programs. They also can be used to improve programs and document their effects through focused evaluation.

While the models are popular, there has been limited effort thus far to study them. This research found that graphic conceptual models are developed using multiple sources, including social science theory and research and stakeholder theory; they frequently are used to develop evaluation questions; and the evaluations often measure constructs and relationships outlined in the model. In addition, evaluators are using models to have conversations with stakeholders about how programs are expected to function and lead to change. While more research is needed to understand how to better use models with complex interventions and with different cultural groups, this study underscores that models have a useful role in these situations also.

As a tool to depict program theory, graphic conceptual models are important to the field of evaluation. Among the methodological competencies for evaluators outlined by the American Evaluation Association (2018) are the identification of assumptions that underlie program logic (2.5) and the use of program logic and program theory as appropriate (2.9). Overall, this study contributes to a better understanding of how evaluators develop and use graphic conceptual

models in evaluation practice and suggests possible future directions for research on theory-driven evaluation practice.

REFERENCES

Alkin, M.C. (2004). *Evaluation roots: Tracing theorists' views and influences.* Thousand Oaks, CA: Sage Publications.

Astbury, B., and Leeuw, F.L. (2010). Unpacking black boxes: Mechanisms and theory building in evaluation. *American Journal of Evaluation, 31*(3), 363-381.

Bamanyaki, P.A., and Holvoet, N. (2016). Integrating theory-based evaluation and process tracing in the evaluation of civil society gender budget initiatives. *Evaluation, 22*(1), 72-90.

Beach D and Pedersen R (2013) *Process-Tracing Methods: Foundations and Guidelines*. Michigan, MI: The University of Michigan Press.

Berg, B.L.(2001). *Qualitative research methods for the social sciences*. Boston: Allyn and Bacon.

Bronfenbrenner, U. (1994). Ecological Models of Human Development. *International Encyclopaedia of Education, 3*, Oxford, Elsevier.

Carvalho, S., and White, H. (2004). Theory-based evaluation: The case of social funds. *American Journal of Evaluation 25*, 141-160.

Centers for Disease Control and Prevention. (1999.) *A Framework for Program Evaluation*. Retrieved on April 27, 2014, http://www.cdc.gov/evaL/framework/index.htm

Centerss for Disease Control and Prevention, (2007). *Interim pre-pandemic planning guidance: community strategy for pandemic influenza mitigation in the United States: early, targeted, layered use of nonpharmaceutical interventions*. https://stacks.cdc.gov/view/cdc/11425. Retrieved on October 2, 2020.

Chen, H.T. 1990. *Theory-Driven Evaluations*. Newbury Park, CA: Sage.

Chen, H. (2005). *Practical program evaluation*. Thousand Oaks, CA: Sage Publications.

Chen, H.T., and Rossi, P.H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7,* 283-302.

Chen, H.T., and Rossi, P.H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning, 10,* 95-103.

Christie, C.A., and Fleischer, D.N. (2010). Insight into evaluation practice: A content analysis of designs and methods used in evaluation studies published in North American evaluation-focused journals. *American Journal of Evaluation, 31*(3), 326-346.

Coryn, C.L.S., Noakes, L.A., Westine, C.D., and Schroter, D.C. (2011). Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009. *American Journal of Evaluation 32*(2). 199-226.

Crano, W.D. (2003). Theory-driven evaluation and construct validity. In S.I. Donaldson and M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 111-142). Mahwah, NJ: Lawrence Erlbaum Associates.

Davis, R. (2003). Network perspectives in the evaluation of development interventions: More than a metaphor. Paper presented at the EDAIS Conference New Directions in Impact Assessment for Development, November 24-25.

Davis, M.V., Mahanna, E., Joly, B., Zelek, M., Riley, W., Verma, P., and Solomon Fisher, J. (2014). Creating quality improvement culture in public health agencies. *American Journal of Public Health, 104*(1), 98-104.

Donaldson, S.I., (2001). Mediator and moderator analysis in program development. In S. Sussman (Ed.) Handbook of program development for health behavior research (pp. 470-490). Newbury Park, CA: Sage.

Donaldson, S.I. (2003). Theory-driven program evaluation in the new millennium. In S.I. Donaldson and M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 111-142). Mahwah, NJ: Lawrence Erlbaum Associates.

Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. Mahwah, NJ: Erlbaum.

Donaldson, S. I. (2020). Exemplary evaluations in a multicultural world. In L. Bickman, and Rog, D. (Eds.), *The evaluation handbook: An evaluator's companion*. Newbury Park, CA: Sage.

Etzkowitz, H. (2002.) *The triple helix of university-industry-government implications for policy and evaluation* (Working Paper 1650-3821). Stockholm, Sweden: Science Policy Institute.

Frazier-Anderson, P., Hood, and Hopson, R. (2011). Preliminary considerations of an African American Culturally Responsive Evaluation System. In Lapan, S.D., Quartaroli, M.T., and Riemer, F.J. (Eds.). Qualitative Research: An Introduction to Methods and Designs (pp. 347-372). San Francisco, CA: Jossey-Bass.

Funnell, S.C., and Rogers, P.J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco, CA: John Wiley and Sons.

Gargani, J. (2003). *The history of theory-based evaluation: 1909 to 2003*. Paper presented at the American Evaluation Association annual conference, Reno, NV.

Government Accountability Office. (2009). *A variety of rigorous methods can help identify effective interventions* (GAO Publication No. GAO-10–30). Washington, DC: U. S. Government Printing Office.

Harris, D [@drewaharris]. (2020, February 27). *Important to remember that #Covid-19 epidemic control measures may only delay cases, not prevent. However, this helps limit surge and* [Tweet, https://twitter.com/drewaharris/status/1233267475036372992]. Twitter.

Henry, G. T., and Mark, M. M. (2003). Toward an agenda for research on evaluation. In C. A. Christie (Ed.), The practice-theory relationship in evaluation (pp. 69-80). New Directions for Evaluation, No. 97. San Francisco, CA: Jossey-Bass.

Hurtado, M., Spinner, JR, Yang, M., Evensen C., Windham A., Ortiz G., et al. (2014). Knowledge and behavioral effects in cariodvascular health: Community health worker health disparities initiative, 2007-2010. *Prevention of Chronic Disease*, *11*.

Jacobson, M.R., Azzam, T., and Baez, J.G. (2013.) "The Nature and Frequency of Inclusion of People with Disabilities in Program Evaluation." *American Journal of Evaluation 34 (1)*, 23-44.

James Bell Associates (2009, October). *Evaluation brief: Measuring implementation fidelity.* Arlington, VA: Author.

Johnston, A. (2002). *Building capacity: Ensuring evaluation findings contribute to program growth*. Paper presented at the annual meeting of the American Evaluation Association, Washington, D.C.

Keene, M. (2011, April 17). EPE Week: Matt Keene on Fuzzy Logic Models – Embracing and Navigating Complexity. Message posted to AEA365, archived at http://aea365.org/blog/?p=3254#sthash.vr3JoXmx.dpuf

Krippendorff, K. (2004). *Content analysis: An Introduction to Its Methodology*. Thousand Oaks, California: Sage Publications Inc.

LaFrance, J. (2004). Culturally competent evaluation in Indian country. *NewDirections for Evaluation, 102,* 39-50.

Leeuw, F. L., and Donaldson, S. I. (2015). Theories in evaluation: Reducing confusion and encouraging debate. *Evaluation: The International Journal of Theory, Research, and Practice*, 21(4), 467–480.

Lemire, S., Whynot, J., and Montague, S. (2018). Why we model matters: A manifesto for the next generation of program theorizing. *Canadian Journal of Program Evaluation, 33.* Retrieved September 20, 2019, from https://evaluationcanada.ca/canadian-journal -program-evaluation/volume-33-2018/special-issue

Lemire, S., Kwako, A., Bohni Nielsen, S., Christie, C., Donaldson, S., and Leeuw, F. (2019). What is this thing called a mechanism? Findings from a review of realist evaluations. Retrieved September 20, 2020 from https://www.researchgate.net/publication/337657654_What_is_this_thing_called_a_mechanism_Findings_from_a_review_of_realist_evaluations

Lipsey, M.W. (1988). Practice and malpractice in evaluation research. *Evaluation Practice*. 9(4), 5-24.

Lipsey, M. W. Wilson, D.B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*(12), 1181-1209.

Mayne J (2012) Contribution analysis: Coming of age. *Evaluation* 18(3): 270–280.

Mayne, J., and Befani, B. (November 2014). Process tracing and contribution analysis: A combined approach to generative causal inference for impact evaluation. *IDS Bulletin, 45*(6).

McLaughlin, J. A., and Jordan, G. B. (1999). Logic models: A tool for telling your program's performance story. *Evaluation and Program Planning, 22*, 65-72.

Miller, R. L., and Campbell, R. (2006). "Taking stock of empowerment evaluation: An empirical review." *American Journal of Evaluation, 27*, 296-319.

Milstein, B., and Wetterhall, S. CDC Working Group. (2000). A framework featuring steps and standards for program evaluation. *Health Promotion Practice*, 1, 221-228.

Munter, C., Cobb, P., and Shekell, C. (2016). The role of program theory in evaluation research: A consideration of the What Works Clearinghouse standards in the case of mathematics education. *American Journal of Evlauation, 37*(1), 7-26.

Neuendorf, K. A. (2002). The Content Analysis Guidebook. Thousand Oaks, CA: Sage.

Patton, M.Q. (2008). *Utilization-focused evaluation*, 4th edition. Thousand Oaks, CA: Sage.

Patton, M.Q. (2015) *Qualitative research and evaluation methods: Integrating theory and practice.* 4th edition. Thousand Oaks, CA: Sage.

Pawson, R andTilley, N. (1997) *Realistic evaluation.* London: Sage.

Pawson, R. (2006). Evidence-based policy: A realist perspective. London, England: SAGE.

Peck, L.R., Kim, Y., and Lucio, J. (2012.) An Empirical Examination of Validity in Evaluation. *American Journal of Evaluation, 33*(3) 350-365.

Plescia, M., Herrick, H., and Chavis, L. (2008). Improving health behaviors in an African American community: The Charlotte Racial and Ethnic Approaches to Community Health Project. *American Journal of Public Health, (98)*9, 1678-1684.

Practical Concepts, Inc. (1979.) *The logical framework: A manager's guide to a scientific approach to design and evaluation.* New York: Practical Concepts.

Rockwell, K. and Bennett, C. (2004). Targeting outcomes of programs: A hierarchy for targeting outcomes and evaluating their achievement. *Faculty Publications: Agricultural Leadership, Education and Communication Department.* Paper 48. Retrieved July 20, 2013, from http://digitalcommons.unl.edu/aglecfacpub/48

Rogers, P.J., Petrosino, A., Huebner, T.A., and Hacsi, T.A. Program theory evaluation: Practice, promise, and problems. *New Directions for Evaluation, 87,* 5-13.

Rogers, P.J. (2007). Theory-based evaluation: Reflections ten years on. *New Directions for Evaluation, 114,* 63-81.

Rogers, P.J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation* (14)1, 29-48.

Rollison, J., Hill, G., Yu, P., Murray, S., Mannix, D., Matthews-Younes, A., and Wells, M.E. (2012). Evaluation of a complex, multisite, multilevel grants initiative. *Evaluation and Program Planning, 35*, 273-279.

Rosenberg, L.J. and Fry Consultants. (1970.) Project evaluation and the project appraisal reporting system. Final Report to U.S. Agency for International Development, Contract No. csd-2510.

Rossi, P.H., Freeman, H.E., and Lipsey, M.W. (1999). *Evaluation: A systematic approach*. 6th edition. Thousand Oaks, CA: Sage Publications.

Scriven, M. (1997). Minimalist theory: The least theory that practice requires. *American Journal of Evaluation 19*(1), 575-604.

Shadish, W., Cook, T.D., and Leviton, L.C. (1991). *Foundations of program evaluation*. Newbury Park, CA: Sage Publications.

Sobelson, R.K., and Young, A.C. (2013). Evaluation of a federally funded workforce development program: The Centers for Public Health Preparedness. *Evaluation and Program Planning, 37*, 50-57.

Stame, N. (2004). Theory-based evaluation and types of complexity. *Evaluation, 10*(1), 58.76

Stufflebeam, D. L. (2001). Evaluation Models. *New Directions for Evaluation, 89* (complete issue).

Torres, R.T., Hopson, R., and Casey, J. (October 2013). Stakeholder engagement with graphic conceptual models: Results from the field. Paper presented at the annual conference of the American Evaluation Association, Washington, D.C.

Stokols, D 1992, 'Establishing and Maintaining Healthy Environments: Toward a Social Ecology of Health Promotion', *American Psychologist,* vol. 4, no. 1, pp. 6–22.

Stokols, D 1996, 'Translating Social Ecological Theory into Guidelines for Community Health Promotion', *American Journal of Health Promotion,* vol. 10, no. 4, pp. 282–298.

Taylor-Powell, E., Jones, L., and Henert E. (2003.) *Enhancing Program Performance with Logic Models*. Madison, Wisconsin: University of Wisconsin Extension, Program Development and Evaluation Unit.

United Way of America. (1996.) *Measuring program outcomes: A practical approach.* Alexandria, VA: United Way of America.

Van Mierlo, B., Arkesteijn, M., and Leeuwis, C. (2010). Enhancing the reflexivity of system innovation projects with system analyses. *American Journal of Evaluation, 2*(31), 143-161.

Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. Connell, A. Kubisch, L. B. Schorr, and C. H. Weiss (Eds.), *New approaches to evaluating community initiatives: Volume 1, concepts, methods, and contexts* (pp. 65-92). New York, NY: Aspen Institute.

Weiss, C. H. (1997a). How can theory-based evaluation make greater headway? *Evaluation Review*. 21(4): 501-524.

Weiss, C. (1997b) Theory-based Evaluation: Past, Present and Future. In D. J. Rog (ed.) *Progress and Future Directions in Evaluation*, *New Directions for Evaluation*, 76. San Francisco, CA: Jossey-Bass.

W.K. Kellogg Foundation (2001.) *Logic model development guide.* Battle Creek, MI: Author.

W. K. Kellogg Foundation. (1998). W. K. Kellogg Foundation evaluation handbook. Battle Creek, MI: Author.

Wray, R., J., Jupka, K., and Ludwig-Bell, C. (2005). A community-wide media campaign to promote walking in a Missouri town. *Preventing Chronic Disease, 2*(4).

APPENDIX A

Linear Models Describing Impact Theory

Figure A1. Outcomes hierarchy for fictional "Apple a Day" program (Funnell and Rogers, 2011)

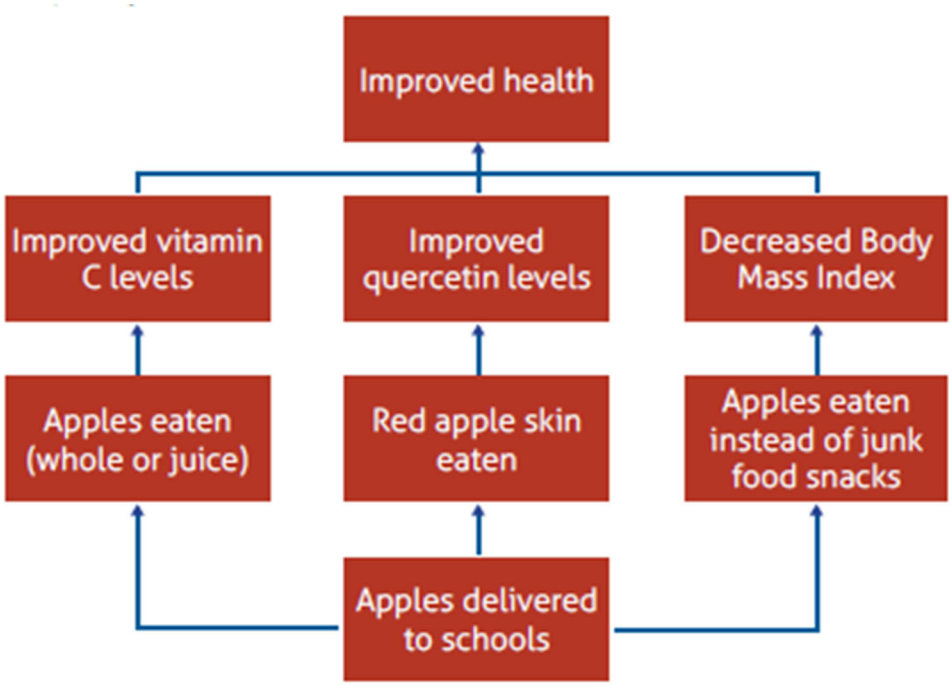
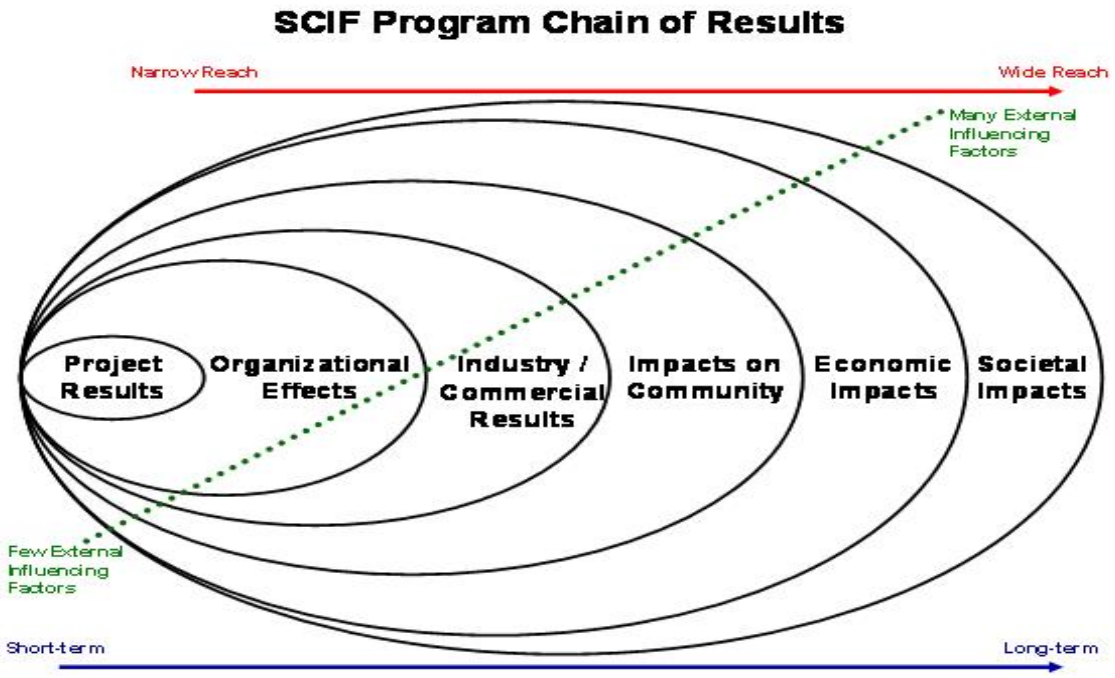


Figure A2. Results chain for the Strategic Communications Investment Fund (http://www.acoa-apeca.gc.ca/eng/accountability/auditsandevaluations/pages/scif_final_report.aspx)



APPENDIX B

Linear Models Describing Process Theory and Impact Theory

Figure B1. University of Wisconsin Extension – Program Development and Evaluation (http://www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html)

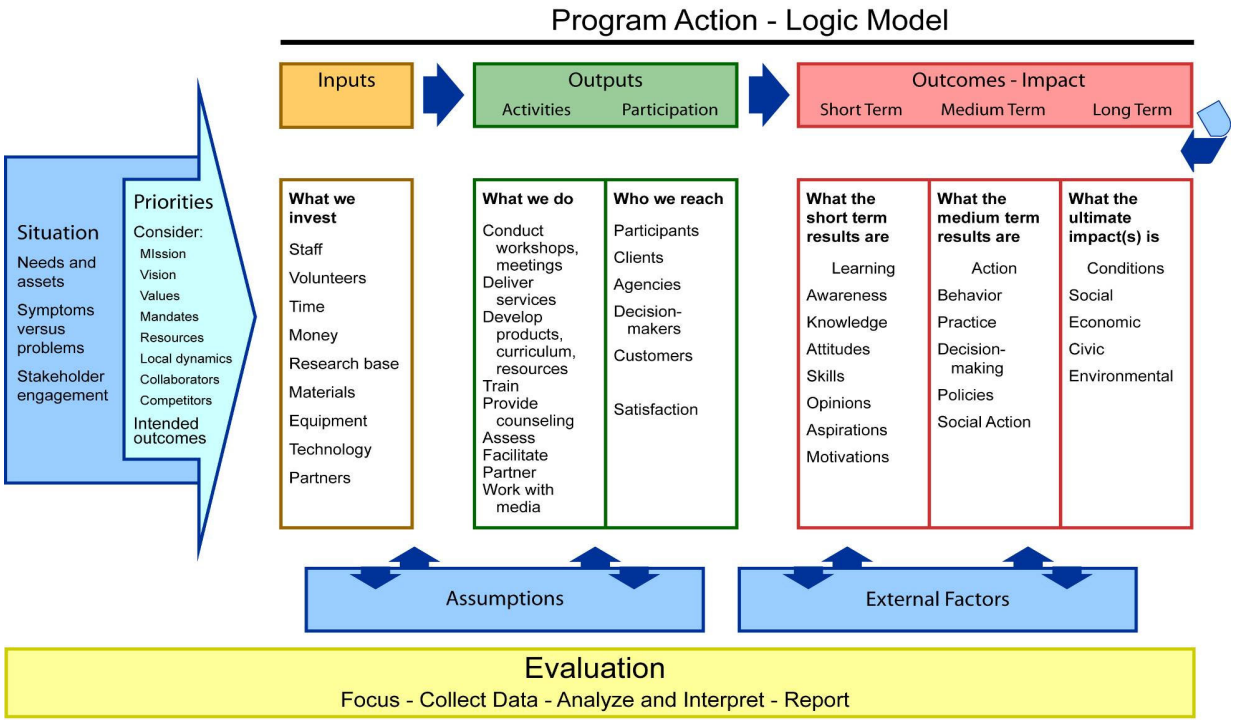


Figure B2. Bennett's Hierarchy (https://alfinfanther.wordpress.com/category/tipstrick/page/2/)

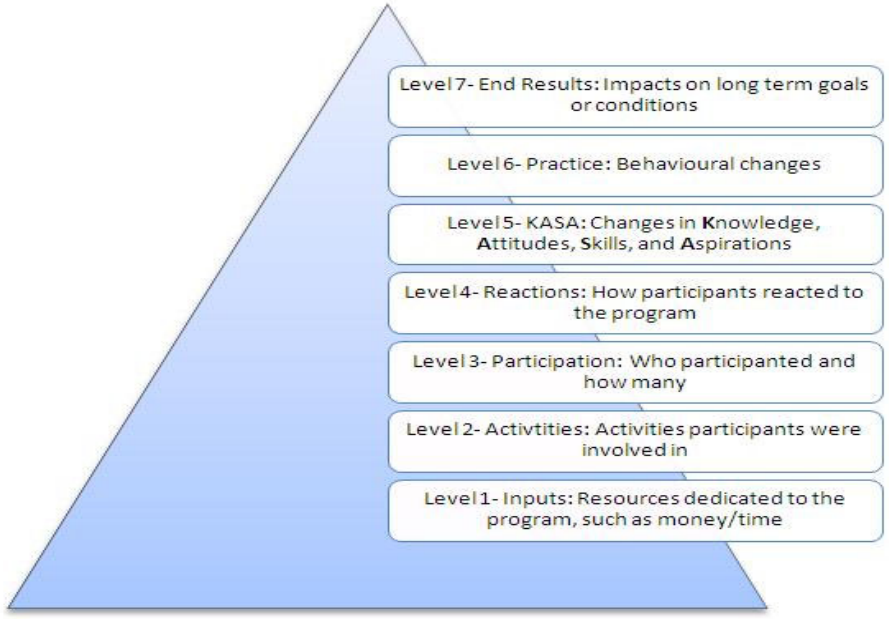


Figure B3. Log Frame Matrix (http://www.sswm.info/category/planning-process-tools/implementation/implementation-support-tools/project-design/logical-f)

Project Description		Objectively verifiable indicators of achievement	Sources and means of verification	Assumptions
Goal	What is the overall broader impact to which the action will contribute?	What are the key indicators related to the overall goal?	What are the sources of information for these indicators?	What are the external factors necessary to sustain objectives in the long term?
Purpose	What is the immediate development outcome at the end of the project?	Which indicators clearly show that the objective of the action has been achieved?	What are the sources of information that exist or can be collected? What are the methods required to get this information?	Which factors and conditions are necessary to achieve that objective? (external conditions)
Outputs	What are the specifically deliverable results envisaged to achieve the specific objectives?	What are the indicators to measure whether and to what extent the action achieves the expected results?	What are the sources of information for these indicators?	What external conditions must be met to obtain the expected results on schedule?
Activities	What are the key activities to be carried out and in what sequence in order to produce the expected results?	**Means:** What are the means required to implement these activities, e. g. personnel, equipment, supplies, etc.	What are the sources of information about action progress? **Costs** What are the action costs?	What pre-conditions are required before the action starts?

APPENDIX C

Models Based on Realist Evaluation

Figure C1. CMO Configuration for Health Literacy
(http://openi.nlm.nih.gov/detailedresult.php?img=2706798_1472-6882-9-19-1andreq=4)

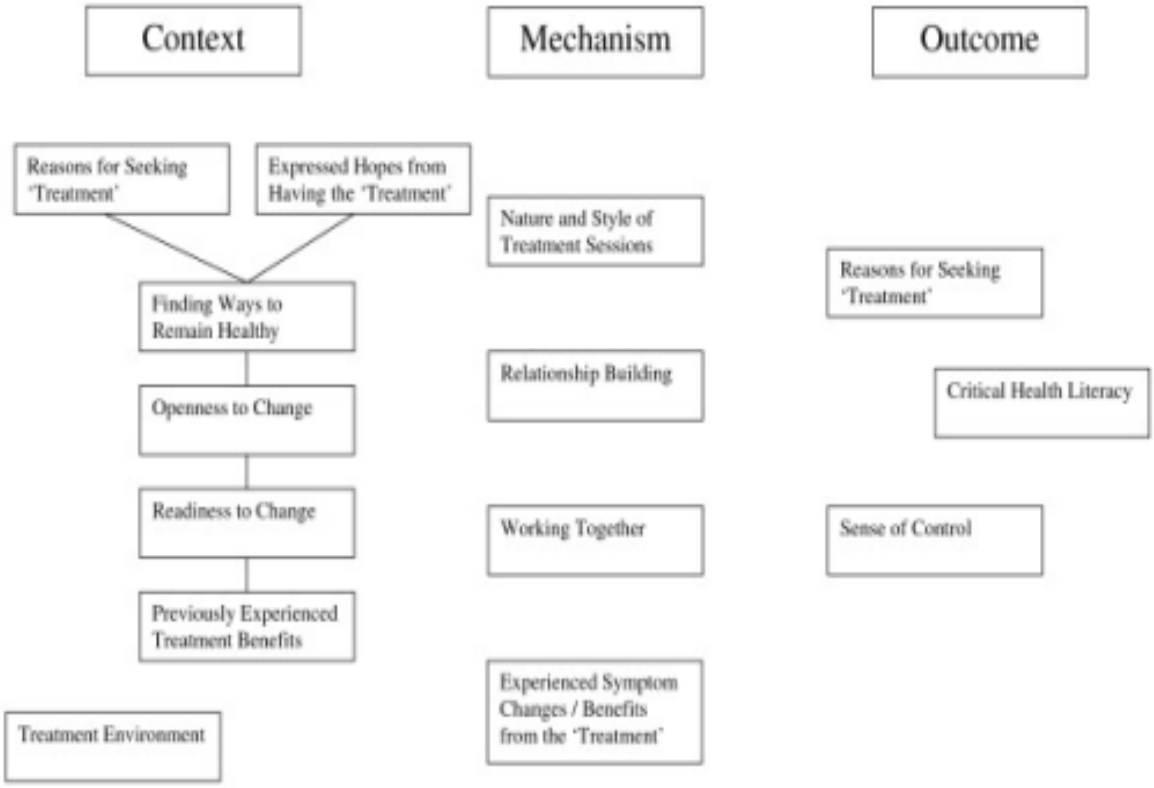


Figure C2. Fictional Realist Matrix (Funnell and Rogers, 2011)

Context	Mechanism	Outcome
Conscientious students with strong literacy skills but no prior computer experience	Skill development	Achieve a threshold level of computer skills that makes it possible to start an internship
Students with literacy and numeracy problems and poor behaviour	Skill development does not occur	Do not achieve a threshold levels of computer skills
Students with existing computer experience, skills and confidence	Skill development does not occur.	No change to level of skills

APPENDIX D

Ecological Models

Figure D1. "Eco-logic" model by Center for Community Based Research (http://www.communitybasedresearch.ca/Page/View/Logic_Modeling_Innovation.html)

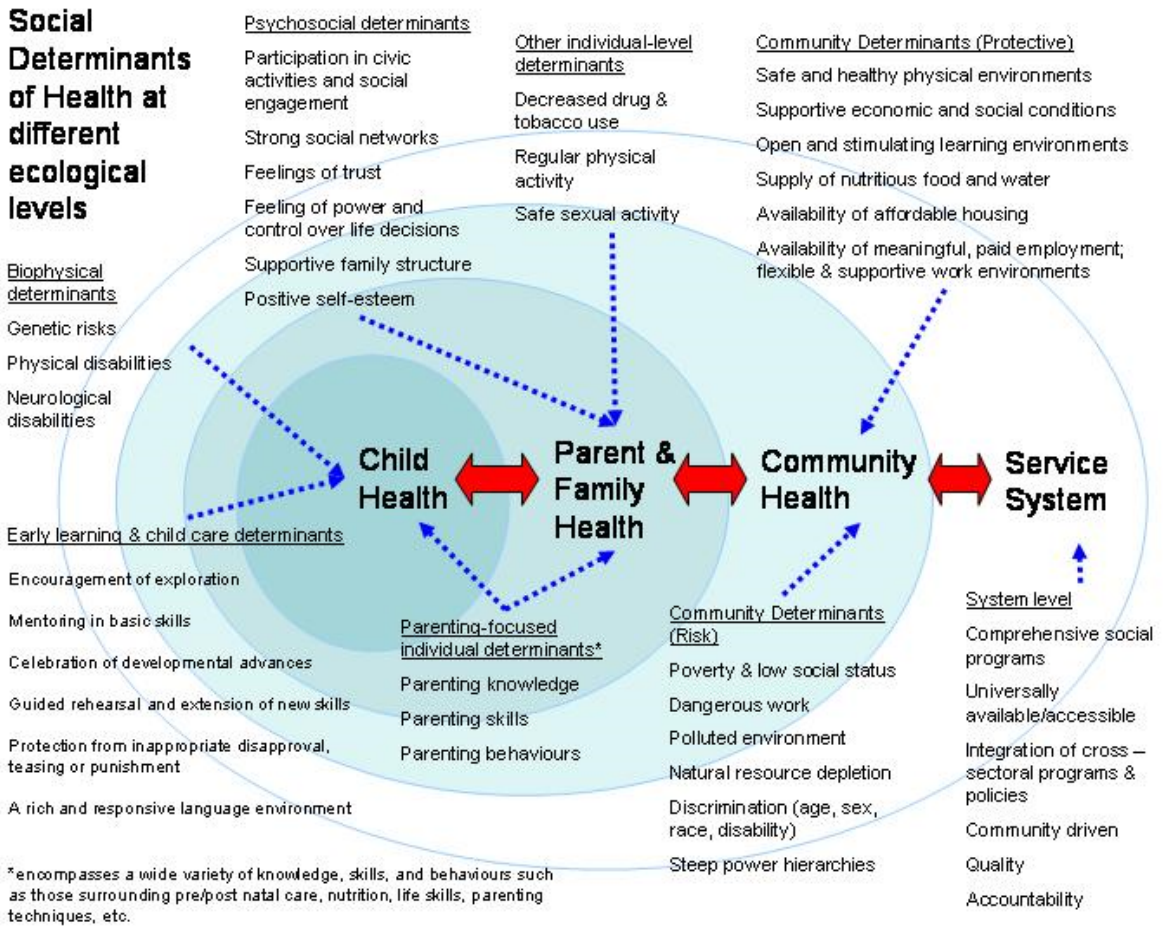


Figure D2. Composite Logic Model of a Healthy Chesapeake Bay
(http://www.scag.ca.gov/wptf/pdfs/wptf112907_NAPA_EPA.pdf)

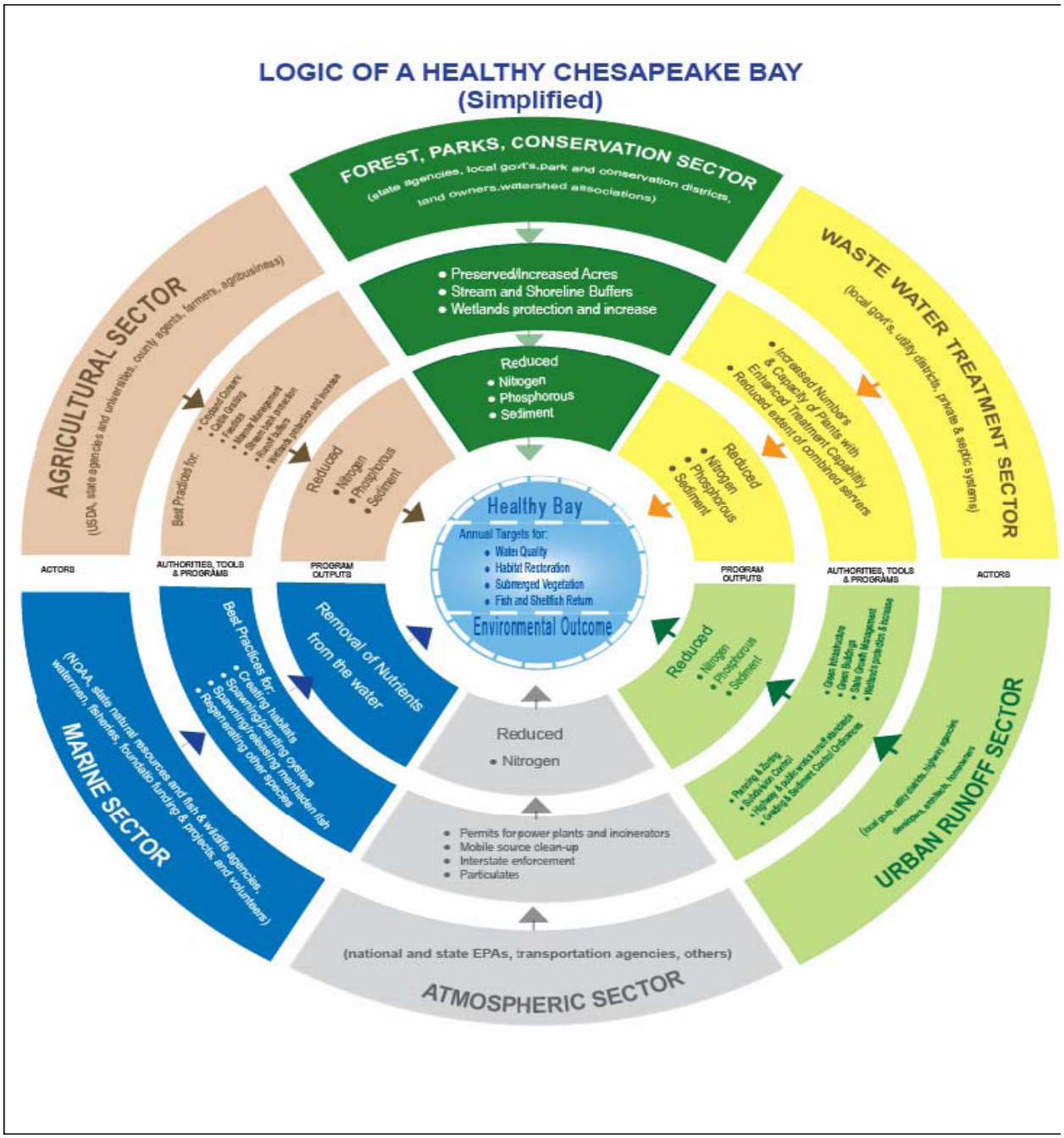
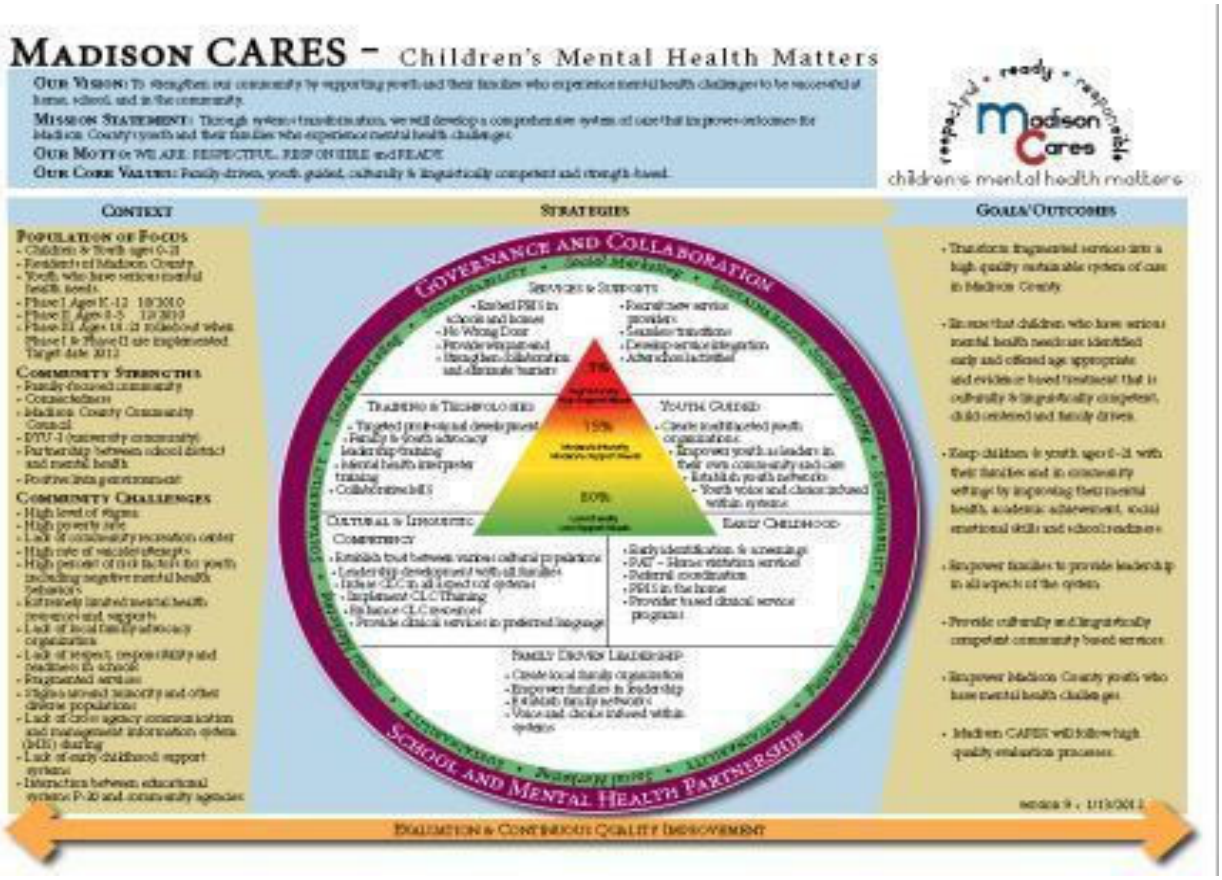


Figure D3. Madison School District 321 Logic Model
(https://ax.d321.k12.id.us/apex/f?p=WEB:CONTENT:::::P2_PAGE_ID:656)



APPENDIX E

Multi-dimensional Models

Figure E1. Translating Research and Innovation Lab (TRAIL) Activities (http://trail.ulster.ac.uk/activities/)

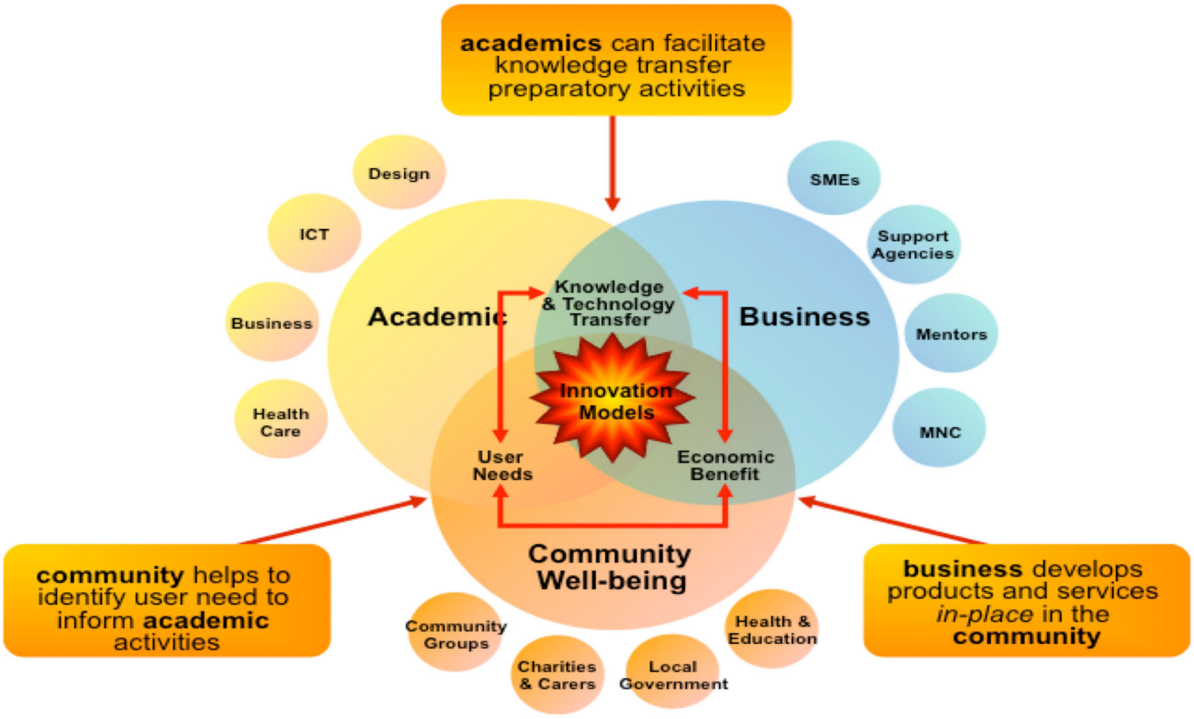


Figure E2. The Knowledge-Based Economy and the Triple Helix Model (http://www.leydesdorff.net/arist09/)

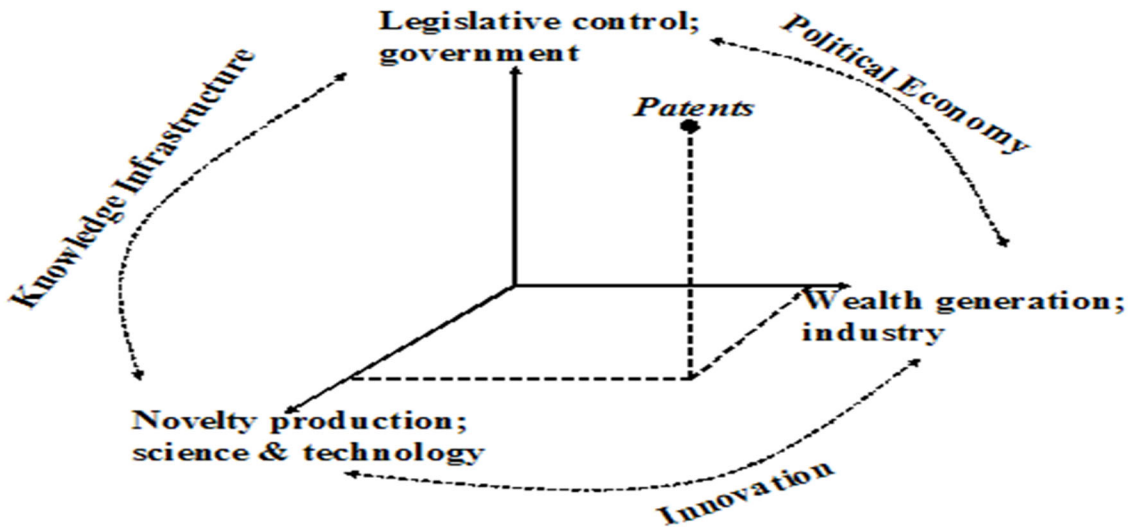
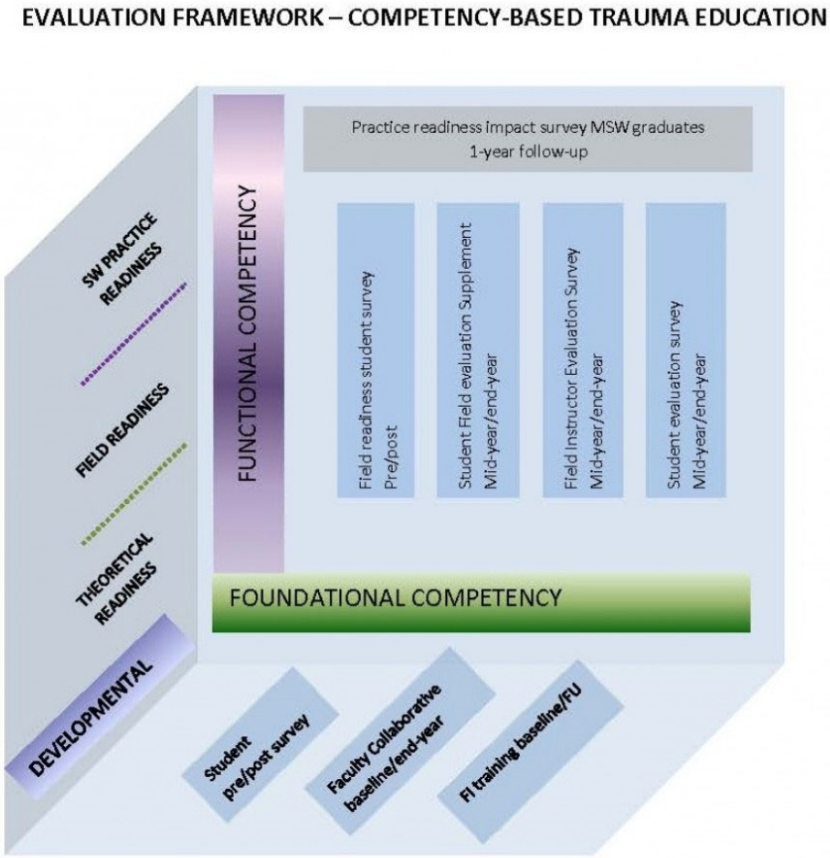


Figure E3. National Center for Social Work Trauma Education and Workforce Development Evaluation Framework (http://www.ncswtraumaed.org/evaluation)



APPENDIX F

Network Models

Figure F1. Sample network of participants and events (http://www.mande.co.uk/networkmodels.htm#Example%20networks)

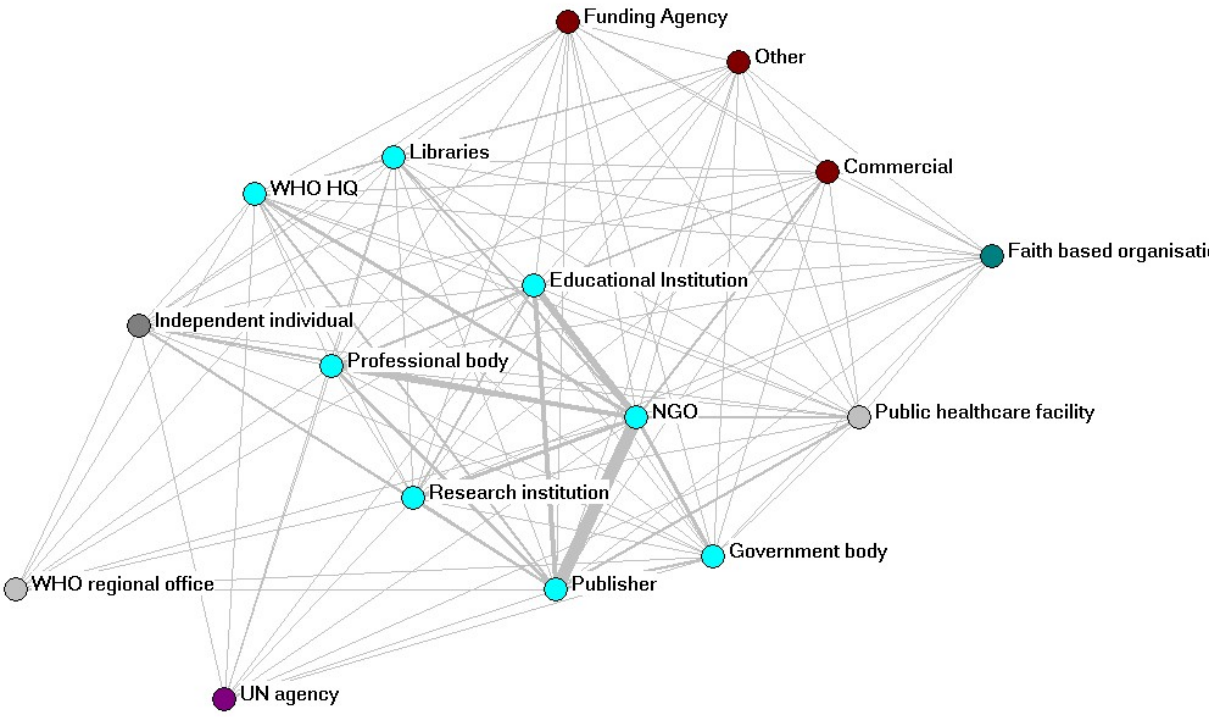


Figure F2. Innovation System (IS) Framework – system analysis showing systems of imperfection (grey) and windows of opportunity (white) (Van Mierlo, Arkesteijn, and Leeuwis, 2010)

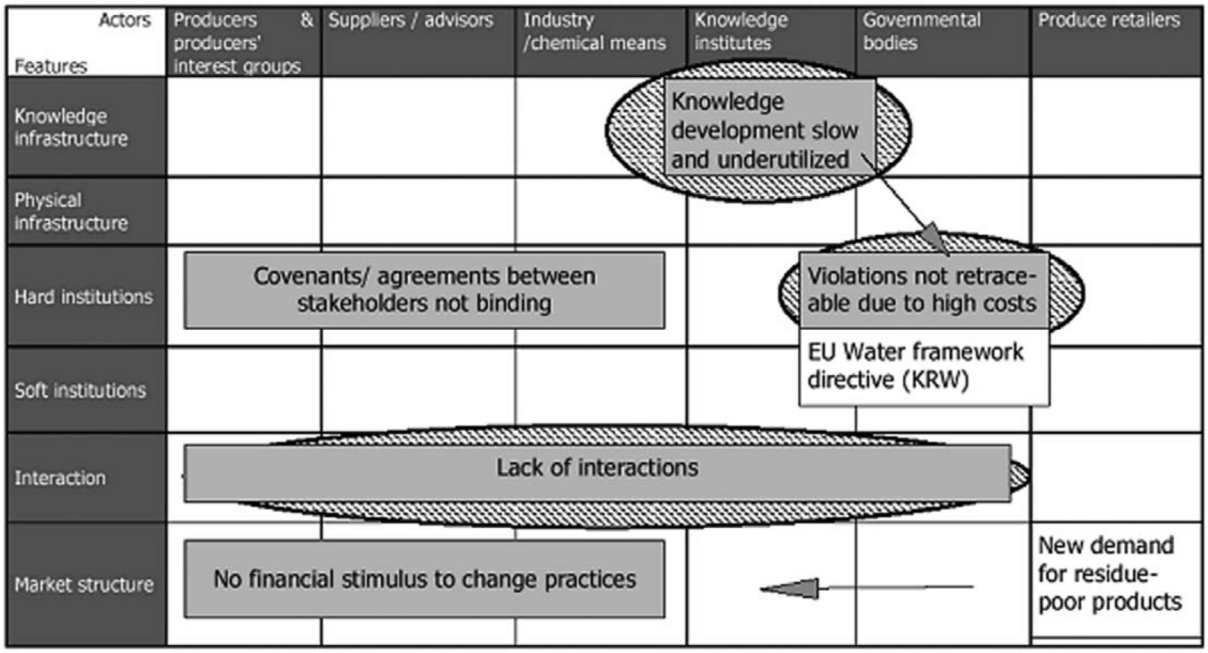
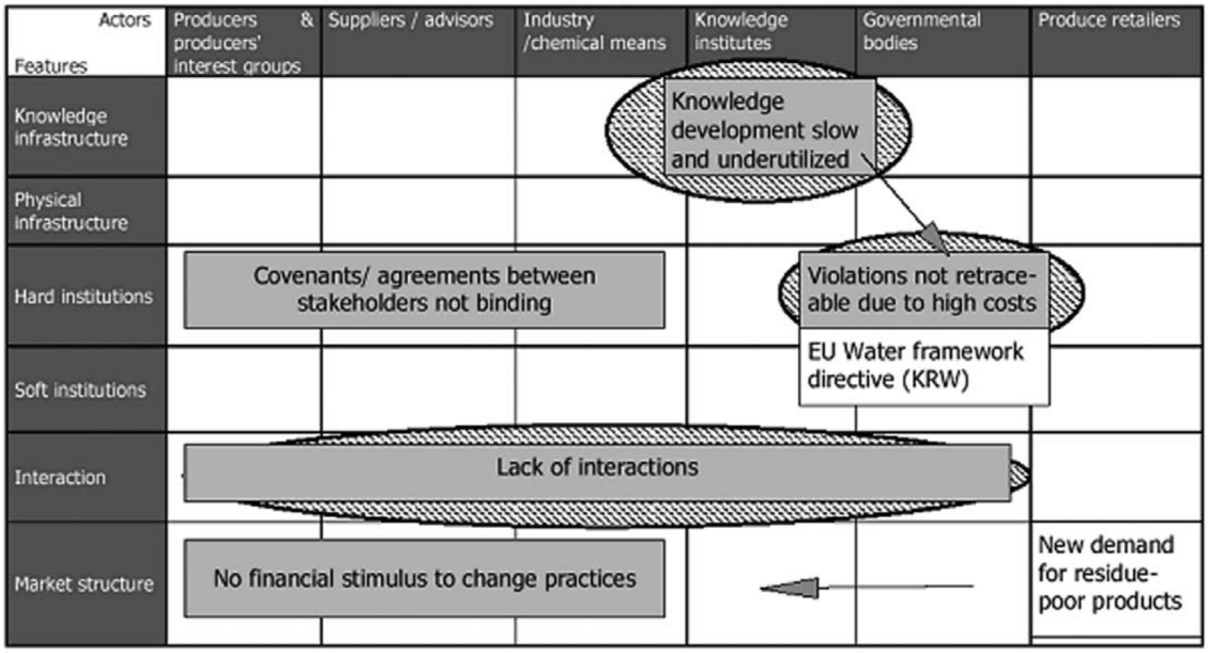


Figure F3. Innovation System (IS) Framework - match between envisioned project actions (in circles) and the barriers in the glass sector (Van Mierlo, Arkesteijn, and Leeuwis, 2010)



APPENDIX G

Models that Build on Visual Metaphors

Figure G1. Oregon Paint Stewardship Pilot Program
(http://www.paintstewardshipprogram.com/)

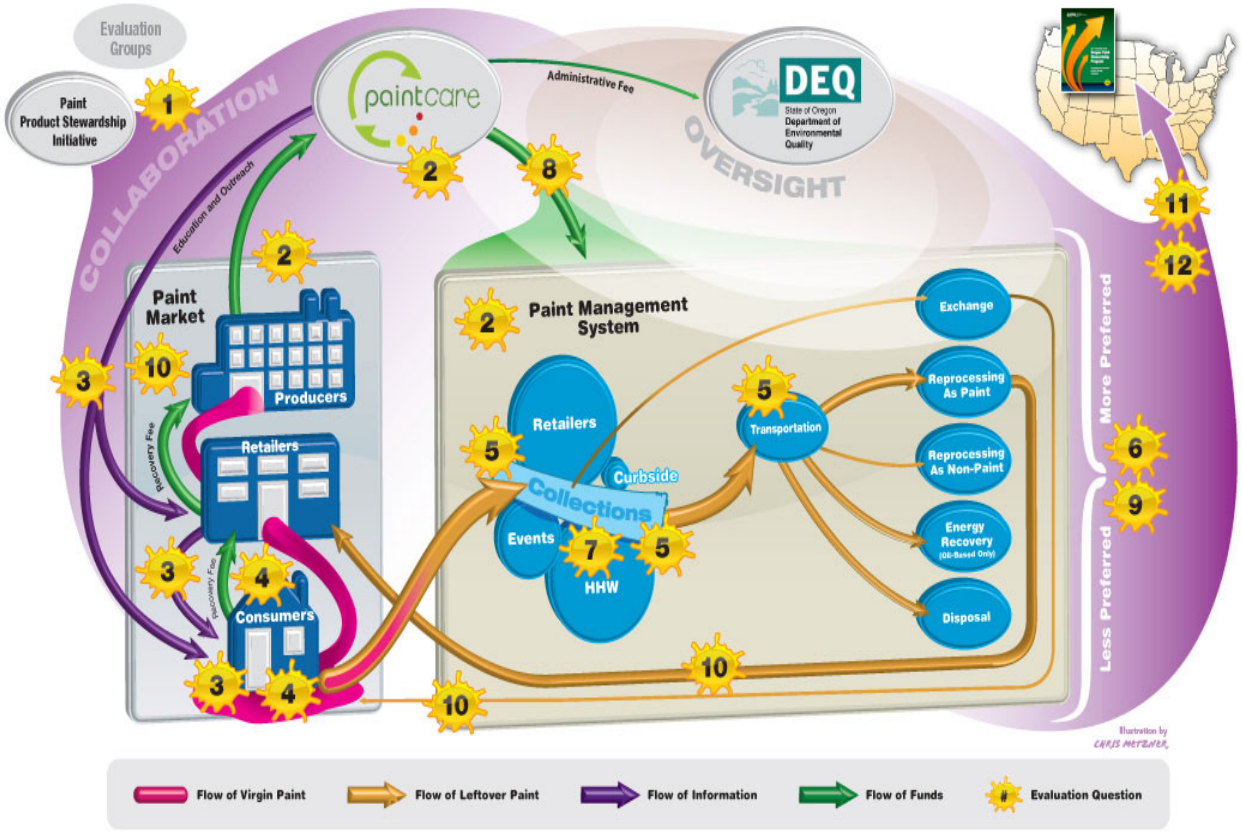


Figure G2. Wayne Food Initiative Logic Model
(http://waynefoods.wordpress.com/home/program-logic-model/)

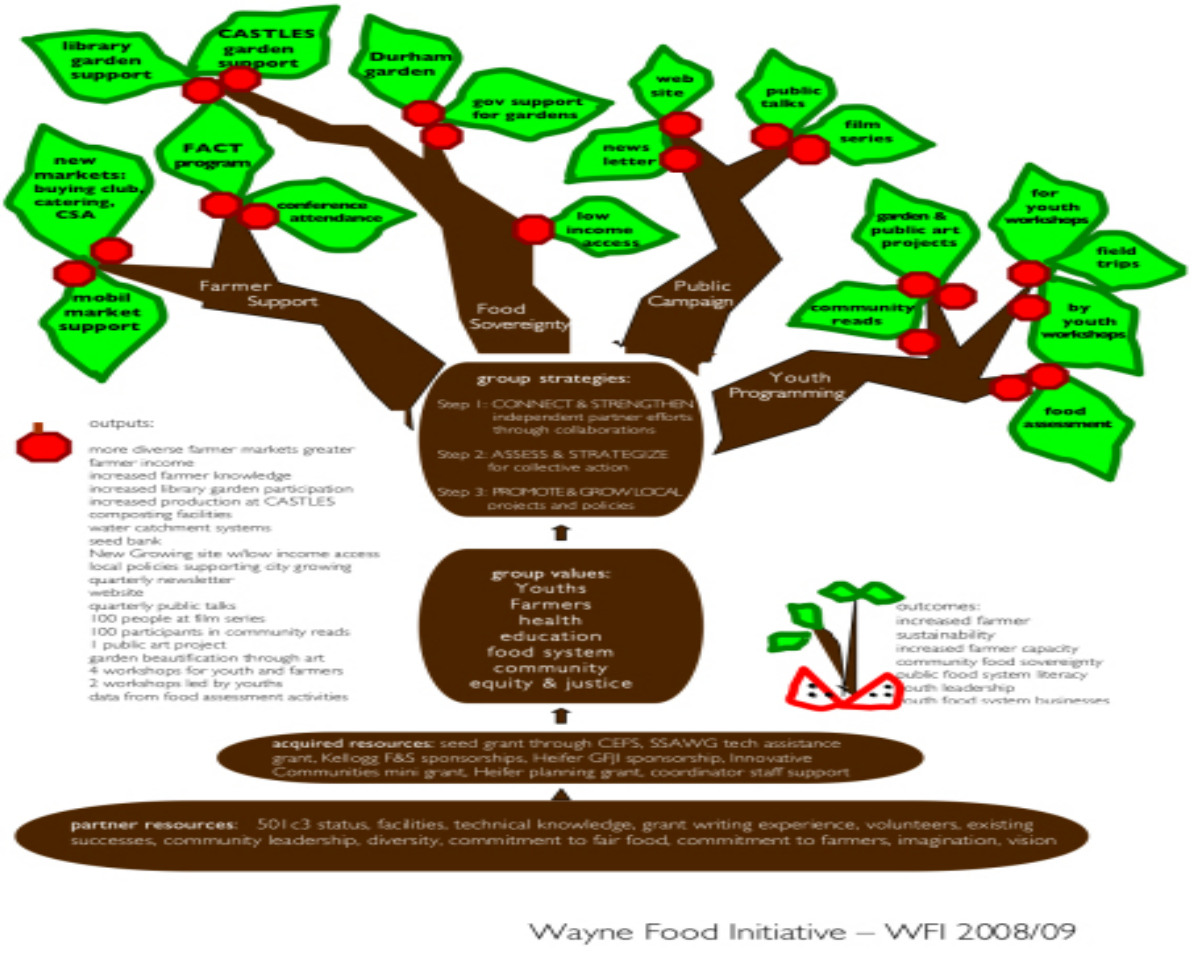
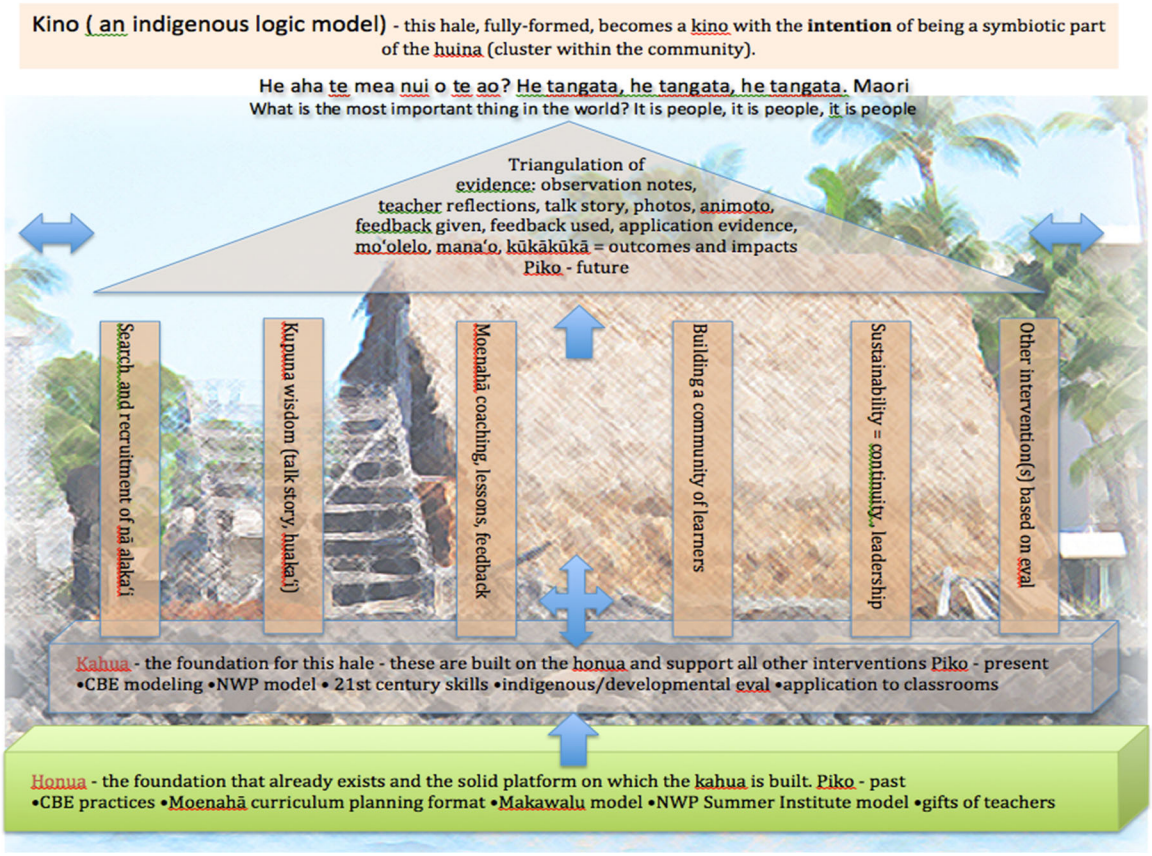


Figure G3. Kino Logic Model (http://puremanao.blogspot.com/2012/07/kino-indigenous-logic-model-post-1-of-4.html)



APPENDIX H

Coding Guide –

Investigation of Graphic Conceptual Models in Published Evaluation Studies

The study has a descriptive component (questions 1a-1e) and an exploratory component (question 2):

3. How do evaluators use graphic conceptual models in studies that they publish?
    a. How frequently are graphic conceptual models included in evaluation studies? (answered during development of sample)
    b. To what extent are core principles of theory-driven evaluation applied in graphic conceptual models?
    (Associated variables: 21 – 24, 26 – 37)
    c. What types of graphic conceptual models are most frequently used?
    (Associated variables: 35-37)
    d. Do the model types vary by substantive field; by cultural context; and/or by program ecology (e.g., simple, complicated, or complex)? (Associated variables: 2-20, 26-27)
    e. How are graphic conceptual models used to design evaluation? Do studies with graphic conceptual models that are strongly rooted in theory and that depict clear relationships result in more comprehensive evaluation studies that explain "how" and "why" an intervention works? (Associated variables: 38-50)

4. What are common impact pathways that could inform the development of archetypes? (Associated variable: 25)

#	Variable	Code	Definition / Guidance for Coding
1	ID		
Responses to questions 2 – 17 most likely can be found in the abstract and/or opening sections of the article that describe the intervention.			
2	Publication year	*Open-ended*	
3	Evaluation team	All of the evaluators are from universities. (1)	

#	Variable	Code	Definition / Guidance for Coding
		Some of the evaluators are from universities and some are from private research firms, a government agency, or are independent consultants. (2) cannot determine (0)	
4	Evaluator role	internal (1) external (2) internal and external mixed (3) cannot determine (0)	• An evaluator who is internal to the organization they are evaluating. • An evaluator who is external to the organization they are evaluating. • An evaluation team that is a mix of internal and external evaluators.
5	Evaluator geographic area	U.S. (1) Canada (2) Europe (3) Latin America (4) Caribbean (5) Africa (6) Australia and New Zealand (7) Asia (8) Middle East (9) cannot determine (0)	• One way to determine this can be by looking at the location of the university with which the evaluator is affiliated.
6	Evaluand	program or project (1) policy (2) initiative (3) other (4) *[please specify]*	• Program or project: organized work intended to advance the social, health, or economic conditions of an individual or community, delivered by one or more entities • Policy: laws, regulations, rules, mandates, or budgets/funding (within government or an organization) • Initiative: A program of a funding body that serves as the frame for funding. The

#	Variable	Code	Definition / Guidance for Coding
			initiative may have goals and objectives, but funded programs have the flexibility to design their programs to align to these goals and objectives. Funded programs may have their own goals and objectives. Mark "initiative" when the evaluation is being done on behalf of the funding body and looks across programs or projects within an initiative. <u>This will most likely come up if the author uses the word "initiative" to describe the evaluand.</u> • Example of initiative vs program/project: The California Endowment has a statewide initiative known as Building Healthy Communities. Individual organizations have received funding to carry out projects aligned to the goals of BHC but that have goals and objectives of their own. • If the author identifies the evaluand using a specific term, mark that term.
7	Geographic location of evaluand	*Please mark all that apply.* U.S. (1) Canada (2) Europe (3) Latin America (4) Caribbean (5) Africa (6) Australia and New Zealand (7) Asia (8) Middle East (9) cannot determine (0)	• It is the program, policy, or initiative that must be located in one of these areas, not the evaluator.

#	Variable	Code	Definition / Guidance for Coding
8	Target population of the evaluand	*(Open ended)*	• Describe the target population as described in the article, e.g. Latino girls in middle school.
9	General substantive field	education (1) public health (2) other (3) *[please specify]*	• Education – Programs or policies that fall under the education category are designed to improve learning of academic or vocational content or skills by individuals of all ages. They are usually carried out in schools, colleges, and universities, as well as in community-based organizations. The programs can influence education either directly, e.g., by working with students, or indirectly, e.g., by training and organizing residents to advocate for educational policy or systems change. • Public health – Public health refers to all organized measures to prevent disease, promote health, and prolong life among the population as a whole. This may be done through promotion or education of healthy lifestyles and recommendations for policy or systems change. Because social, environmental, and biological factors interact to determine health, public health comprises a broad range of interventions.
10	Specific substantive focus area	*(Open ended)*	• Education – Examples: elementary math education program, STEM after-school

#	Variable	Code	Definition / Guidance for Coding
		Please specify the focus area of the intervention (e.g., elementary math education, post-secondary education, substance abuse treatment, housing):	program, adult English language program, continuing education/professional development • Public health - Examples: diabetes education, housing inspection policy, substance abuse treatment, vaccination campaign, affordable housing development
11	Number of sites	single site (1) multi-site – same model implemented (2) multi-site – variation of model implemented (3) cannot determine (0)	• The focus is on the intervention not the evaluation; in how many sites is the intervention carried out? • Mark 2 if the program is implemented in the same manner in different sites. • Mark 3 in cases where variations of a program or different programs are implemented in different sites.
12	Area	neighborhood(s) (1) city/cities (2) county/counties (3) state(s) (4) region(s) (5) nation(s) (6) continent (7) cannot determine (0)	• Neighborhood must be specifically stated. • Region must be specifically stated. • Region – an area or division having definable characteristics but not always fixed boundaries. • In the case where the program is implemented in various locations, look at the highest known level where sites fall. For example, if a training is offered in different neighborhoods or locations in a city, mark city. If it takes

#	Variable	Code	Definition / Guidance for Coding
			place in several cities of a state, mark state, and so forth.
13	Single or multiple organizations	single organization (1) multiple organizations (2) cannot determine (0)	• The focus is on the organization or organizations involved in implementing the program (not evaluating it). • The distinguishing feature is collaboration. A collaborative would be considered multiple organizations. • Needs be explicit
14	Number of organizations involved	enter number: _____ cannot determine (0)	• Enter 1 for single organization.
15	Interdisciplinary / intersectoral collaboration	yes (1) no (2) not applicable (such as with single organization (3) cannot determine (0)	• Sector examples – school district, higher education system, health care system, nonprofit organization, business sector • Discipline examples – education, psychology, medicine • May refer to either individuals or organizations involved in designing / carrying out the intervention. • The focus is not on the composition of the evaluation team.
16	Disciplines / sectors involved	*Open-ended, if 15 is 1*	• List substantive field and sector, e.g. education – nonprofit afterschool program

#	Variable	Code	Definition / Guidance for Coding
			provide or education – elementary school
17	Author describes project as "complex"?	yes (1) no (2)	• The word "complex" must be used in the article; synonyms should not be accepted.
18	Stage of program	innovation (1) fine-tuning (2) established (3) cannot determine (0)	(Rossi, Freeman, and Lipsey) • Innovation: This includes pilot programs and other activities to help guide the development of programs. • Fine – tuning: The program is at an early phase of implementation (after any pilot testing or development activities have been completed). Modifications may still be common but they are a result of lessons learned during initial implementation. The program may have been implemented only once or in effect for a short period of time. The evaluation at this stage will most often be formative. Established: The program is at a more mature phase of implementation where the design is more stable. The program may have been repeated several times or in place for an extended period of time. The evaluation at this stage will most often be summative.

#	Variable	Code	Definition / Guidance for Coding
19	Primary commissioner of the evaluation	*Please mark all that apply.* government agency (1) private foundation (2) other (3) *(specification not needed)* cannot determine (0)	• In this code we are looking for the primary funder of the program and/or the evaluation. • For example, an article might directly say that a government agency or a private foundation requested that an evaluation be completed. • Or, an article might say that a program falls under a particular initiative of either a government agency or a private foundation. • Often when this information is provided, it appears in the opening section or as a footnote at either the start or end of the article. • Foundation must be a specific philanthropic foundation. **INCLUDE NOTE:** • If there is a different funder for program and evaluation. • If the funder is a university.
The response to question 20 may be found in a description of the program at the start of an article. It may also be seen in the model. In the model, a helpful place to look is at the targets of outputs and outcomes.			
20	Level of change	*Please mark all that apply.* individual (1) interpersonal (2) organization (3)	• The response may be found either in the model or in the text of the article.

#	Variable	Code	Definition / Guidance for Coding
		community (4) public policy (5)	• The aim is to determine at what level(s) an intervention is trying to promote change. This may be apparent in either the description of the program or in the model. • Individual – individual knowledge, attitude, behaviors • Interpersonal – family, friends • Organizational – churches, stores, community organizations, local health departments, school districts, etc.; the rules, regulations, policies, and structures within these organizations • Community – relationships among organizations, e.g., collaborations and coalitions; broad social networks of individual community members and/or organizations; community norms (community regulations) • Public Policy – local, state, and federal policies and laws that regulate or support practices/actions
Responses to questions 21 – 25 most likely will be found early in the article in a section describing the program. Sometimes there may be a section entitled "program theory." The needed text is often not far from the graphic conceptual model and is meant to be a companion to the model.			
21	Program theory narrative	yes, the program theory of change is described in the narrative (1)	• In describing the program theory of change, the author should discuss how inputs are expected to lead to outcomes—*to discuss the anticipated mechanisms of change and/or relationships*

#	Variable	Code	Definition / Guidance for Coding
		no, the program theory of change is not described in the narrative (2) cannot determine (0) **Page number: _____**	*between mediators and moderators*. It is not enough to just describe program components and expected outcomes. The narrative complements the graphic conceptual model, expanding on elements and relationships in the model and providing evidence of each principle. In the published articles in this study, the program theory narrative may be short. • Please note page number where program theory narrative can be found.
22	Details of theory formulation	*Please mark all that apply.* existing social science theory (1) evidence base (2) stakeholder theory (3) program observation (4) program document review (5) other (6) *[please specify]*: cannot determine (0)	• When 1 (social science theory) is present, 2 (evidence base) is also likely to be present. However, it may be possible for 2 (evidence base) to be present without 1 (social science theory to be present). • Existing social science theory: It is sufficient if the authors say that they used social science theory to develop the graphic conceptual model. However, it is preferable if they reference the specific theories that serve as the basis for the model. Many articles will include a literature review as a background to the program. What is important to determine is that social science theory influenced the design of the model. (Please note that

#	Variable	Code	Definition / Guidance for Coding
			number 23 asks for the social science theories mentioned, if applicable.) • Evidence base: The evidence base may include results from research and evaluation and programs or practices considered models or exemplary. • Where citations are placed is important to consider. They should be included in, or close to, the discussion of the program theory. • Stakeholder theory: Individuals involved in the delivery of the program or affected by the program share their views of program theory, and these views inform the final graphic conceptual model. • Program observation: Observation by the evaluator • Program document review: Review by the evaluator
23	Overview of theory formulation	one approach is used to develop the program theory (1) a combination of approaches is used to develop program theory (2) cannot determine (0)	• This variable is a "summary" of number 22. Mark 1 if only one approach of theory formulation is used in number 22; mark 2 if more than one approach is used.
24	Social science theory	*Open-ended*	• Please list the social science theories and/or evidence base

#	Variable	Code	Definition / Guidance for Coding
			mentioned as underpinning the program theory described in the model.
25	Archetypes (archetypical outcome pathways)	*Open-ended*	• Include a brief description of the program and the outcome pathway.
Responses to question 26 – 37 are to be found in the model itself.			
26	Number of causal strands	number of arrows: _____ cannot determine (0)	• The causal strand in program theory shows the hypothesized relationship between a program component/activity and immediate outcomes, between immediate and intermediate outcomes, and between intermediate outcomes and ultimate outcomes or impacts. • When the causal strands are not separated by boxes or arrows (e.g. all inputs are in one box, all outputs in another, all short-term outcomes, in another, and so on), <u>mark 0</u>
27	Proportion of impact	*Please mark all that apply.* linear causality with proportional impact (1) recursive with feedback loop(s) (2) tipping point(s) (3) cannot determine (0)	• Recursive with feedback loops – The implementation and attainment of higher-level objectives interact with the implementation of lower-level objectives through feedback mechanisms. The cause-effect relationship may be mutual, multidirectional, or multilateral. There may be disproportionate impact at critical levels (Patton, 2007; Rogers 2008).

#	Variable	Code	Definition / Guidance for Coding
			• Tipping points - A small initial effect leads to a large ultimate effect through a reinforcing loop or attainment of critical levels/threshold. This may be rare to see. The authors may talk about a tipping point in the article as well. Tipping points are often seen in "complex" projects. • Look for information in the model. **INCLUDE NOTE:** If information appears in text.
28	Presence of mediators	yes, the model includes mediators [e.g., program-mediator-outcome] (1) no, the model does not include mediators [direct effect: program-outcome] (2)	• Mediator: a variable that is affected by the program, which in turn affects an outcome of interest. • To be considered yes, the items in the model must be true mediators. (To distinguish from "chains" that may include outcomes that do not have a mediating relationship).
29	Length of mediator/outcome chain	the model includes one mediator (1) the model includes two mediators (2) the model includes three mediators (3) the model includes more than three mediators (4) not applicable; no mediators (5) cannot determine (0)	• One mediator – e.g. program-mediator-outcome • Two mediators – e.g. program; short-term outcome; intermediate outcome; long-term outcome • Three mediators – e.g., program; short-term outcome; intermediate outcome; long-term outcome; impact

#	Variable	Code	Definition / Guidance for Coding
			• In the event of more than one causal chain, focus on the longest causal chain. • See "Supplement to Coding Guide 1/2/17." • For models using realist evaluation, mark 6 (cannot determine). By design, mediator chains are not the focus of realist evaluation.
30	Pathway of causation	yes, a specific pathway of causation is explained using arrows or other means (1) yes, a general pathway of causation is explained using arrows or other means (2) no, the pathway of causation is not explained using arrows or other means (3)	• Mark "1" for models that have more refined paths, e.g. specific boxes with one or a small number of outcomes connected by arrows. In this model, the path explains how discrete components in the model are connected. • It may be common to see models that connect entire categories by arrows, e.g. short-term outcomes listed in one column connected by an arrow to another column listing long-term outcomes. This would be considered a general pathway of causation ("2").
31	Moderator included	yes, the model includes moderators (1) no, the model does not include moderators (2)	• Moderator: a variable that affects the direction or strength of the relationships between the program and a mediator, or a mediator and an outcome • **INCLUDE NOTE:** If moderators are mentioned in

#	Variable	Code	Definition / Guidance for Coding
			the narrative but not in the model. • In realist evaluation, moderators may be described in the context section of CMO configurations.
32	Type of moderator	*Please mark all that apply.* assumptions (1) external factors (2) participant characteristics (3) provider characteristics (4) characteristics of the setting the intervention (5) strength (dosage) of the intervention (6) intervention attendance (7) other (8) *[specify other]* not applicable; no moderators (9)	• Assumptions: beliefs about the intervention, the people involved, and the context, and the way providers think the program will work • External factors: factors in the environment in which the intervention exists that interact with and influence action, e.g. conflict among agencies to which intervention participants are referred; the political climate around a particular policy targeted by the intervention • Participant characteristics: e.g., gender, ethnicity, socioeconomic status • Provider characteristics: e.g., education / credentials of person delivering the intervention • Characteristics of the setting of the intervention: e.g., accessibility via public transportation, equipment for an activity • Strength (dosage) of intervention: frequency and duration of program activities

#	Variable	Code	Definition / Guidance for Coding
			• Intervention attendance: how often program clients attend intervention activities
33	Explanation of path moderators	yes, arrows or other means are used to identify the particular mediation relationship affected by the moderator (1) no, arrows or other means are not used to identify the particular mediation relationship affected by the moderator (2) not applicable; no moderators (3)	• Sometimes moderators might be included in a model in a box off to the side, or elsewhere. Arrows or other means are not used to explain how the moderator influences certain paths in the model. In such a case, mark 2. • In realist evaluation, the influence of context may be explained using text. The explanation could take place in the context and mechanism parts of the CMO configuration.
34	Description of the intervention process	*Please mark all that apply.* needs (1) inputs (2) activities (3) outputs (4) other (5) *[specify other]* the intervention process is not described in the model analyzed (6)	• Needs: Problems or conditions being addressed by the program • Inputs: The human, financial, organizational, and community resources a program has available to direct toward doing the work • Activities: What the program does with the resources: processes, tools, events, technology, and action that are an intentional part of program implementation. • Outputs: The direct products, goods, and services provided to a program's direct

#	Variable	Code	Definition / Guidance for Coding
			customers as a result of program activities. Outputs may include types, levels, and targets of services to be delivered by the program. • **INCLUDE NOTE:** If the intervention process is described in the narrative but not in the model, as was the case of the article by Chen on the tobacco prevention program. In cases of realist evaluation: • In realist evaluation, activities might be mentioned when describing the mechanism. • Also, realist evaluation does not distinguish between outputs and outcomes, so outputs may be described in the outcome section of the CMO configuration. • For realist evaluation, the language needs to be very specific with regard to these components of process in order for it to be counted that they are present in the description.
35	Design of model	linear model describing process theory (1) linear model describing impact theory (2) linear model describing process theory and impact theory (3)	• In linear models, there needs to be more than one box describing either process or outcome for it to count as process or impact theory.

#	Variable	Code	Definition / Guidance for Coding
		socio-ecological model (4) multi-dimensional model (5) model built on a visual metaphor (6) combination of designs (7) other (8) [*please specify*]	Example 1: Only impact theory described Activity – Outcome 1 – Outcome 2 – Final Outcome Example 2: Process and impact theory described Resources – Activity – Outputs - Outcome 1 – Outcome 2 – Final Outcome Example 3: Only process theory described Resources – Activity – Outputs – Outcome • See "Supplement to coding guide 1/2/17." • See document "GCM Study_Definitions of Variables_Model Type_Images" • See document "GCM Study_Definitions of Variables_Model Type_Narrative" • We may see a few logframes and systems dynamics models. In such an event, mark "other" and specify as such (9).
36	Logic model	yes, the model is specifically referenced as a "logic model" (1)	

#	Variable	Code	Definition / Guidance for Coding
		no, the model is not specifically referenced as a "logic model" (2)	
37	Term used to describe model (if not logic model)	*Open-ended*	
Responses to question 26 – 37 are to be found in the model itself.			
38	Evaluation purpose	*Please mark all that apply.* assess merit and worth of program and its value to society (1) improve the program (2) ensure program compliance with mandates (3) build knowledge and expertise for future programs (4) other (5) *[specify other]* not included (6) cannot determine (0) **Page number: _____**	• The purpose needs to be stated by the author using terms similar to the ones in the codes; we should not infer purpose. If we cannot determine purpose, we should indicate it as not specified. • **INCLUDE NOTE:** If building knowledge and expertise (4) is mentioned only in the discussion section. If it is mentioned as a purpose sooner, mark as code. • **INCLUDE NOTE:** If implementation/process checks or fidelity measures are done as either part of improving the program (2) or to build knowledge and expertise (4). • Specify a needs assessment as other. • Mark the page number with the text so we can refer back to it if needed. • "Not included" is a code because there are a few articles that focus on model development that do not

#	Variable	Code	Definition / Guidance for Coding
			discuss program evaluation. The sample is more inclusive since the models themselves are a significant focus of the study. • For articles that do not include a discussion of the evaluation, the coding ends here.
39	Evaluation purpose text	Please copy the statement in the article that best describes the evaluation.	
40	Evaluation approach	theory-driven evaluation (1) other (2) *[specify other]* not specified (3)	• If we mark "1", **the authors must specifically state that the approach they are using is theory-driven evaluation.** There are different terms used for theory-driven evaluation, but most often the word "theory" appears in the description of the evaluation, e.g., "theory-driven", "theory-based", "program theory-driven evaluation science," etc. • For an approach other than theory-driven evaluation, please note the approach used as an open-ended response after marking 2.
41	Evaluation questions 1	yes; stated explicitly in question form (1) no; can surmise (2) no the authors do not state the evaluation questions in question form and it is too difficult to surmise what type of questions the evaluation would be answering. (3)	
42	Evaluation questions 2 (tied to program theory)	If the evaluation questions are explicitly stated (1 in 41), then:	

#	Variable	Code	Definition / Guidance for Coding
		yes, the evaluation questions are tied to the program's underlying logic or theoretical foundations (1) no, the evaluation questions are not tied to the program's underlying logic or theoretical foundations (2) not applicable; answer to 41 was not 1 (3)	
43	Focus of the evaluation	effects of intervention only (outcomes, impact) (1) effects of implementation factors only (process, context) (2) effects of intervention and implementation factors but not the causal chain (3) effects of intervention and implementation factors, including mediation/moderating relationships between components of program theory (4)	
44	Evaluation design	non-experimental (1) quasi-experimental (2) experimental (3)	• Non-experimental / descriptive – no comparison or control group; designs include cross-sectional ("snapshot in time", like a political poll); time-series (like the cross-sectional, but the "snapshot" is taken multiple times in order to describe trends and look for changes over time), and case study (focused on a selection case or cases, desire for an in-depth understanding of an

#	Variable	Code	Definition / Guidance for Coding
			issue, collecting data in multiple ways, but a focus on qualitative methods). • Quasi-experimental – comparison group consists of individuals or cases considered similar to those who received the intervention and not randomly assigned; designs may include post-test only, pre-test / post-test, interrupted time-series, and regression discontinuity • Experimental – random assignment into two different groups (treatment and control); may involve post-test only or pre-test and post-test
45	Measure of implementation fidelity	*Please check all that apply.* adherence (1) exposure – frequency (2) exposure – duration (3) quality of delivery (4) participant responsiveness (5) none of the above (0)	Fidelity is the extent to which the delivery of an intervention adheres to the program model as intended by the developers of the intervention. Several dimensions are important to review in relation to fidelity—and these may often be done as part of a process evaluation. Adherence refers to the extent to which program components are delivered as prescribed by the model. Adherence indicators can include program content, methods, and activities. Data are typically reported as the proportion of program components that were delivered compared to the number prescribed. Exposure (dosage) is the amount of program delivered in relation to the amount prescribed by the program model. Exposure can include the

#	Variable	Code	Definition / Guidance for Coding
			number of sessions or contacts, attendance, and the frequency and duration of sessions. Frequency – how often programs meet or how often an activity is done. Duration – how long a program runs Quality of delivery refers to the quality of how a program is delivered. It may look at provider preparedness and delivery, perhaps from the perspective of clients (e.g., through a satisfaction survey). Participant responsiveness refers to the manner in which participants react to or engage in a program. Aspects of participant responsiveness can include participants' level of interest in the program; perceptions about the relevance and usefulness of a program; and their level of engagement, enthusiasm, and willingness to engage in discussion or activities. Information on attrition may also be a part of looking at participant responsiveness.
46	Theory-guided construct measurement 1	Yes, constructs articulated in the program theory are assessed (1) Yes, constructs articulated in the program theory are assessed, and additional constructs are also assessed (2) No, constructs articulated in the program theory are not assessed (3)	• "Yes, extra" (2) means that there is something looked at in the evaluation that was not part of the model. • Methods checks or manipulation checks don't count as "extra". • In cases where a construct in the model is expanded upon in the evaluation (to operationalize the construct in more ways), this would not count as "extra," but would

#	Variable	Code	Definition / Guidance for Coding
			rather be 1. Example: the article on Fathers' Clubs in Haiti
47	Theory-guided construct measurement 2	*If answer to 46 is 1 or 2, please mark all that apply.* process constructs articulated in the program theory are measured (1) outcome constructs articulated in the program theory are measured (2) contextual constructs articulated in program theory are measured (3)	
48	Focus of the analysis	whether effects were found without discussion of cause (1) description of cause-effect associations between theoretical constructs (2) explanation of cause-effect associations between theoretical constructs (3) not applicable (e.g., only process evaluation) (4)	• In the case of 1, the evaluation only measures the extent to which outcomes are attained. • In the case of 2, there is an acknowledgment and discussion of relationships but the relationships are not tested. • For 3, the relationships among mediators or moderators are tested.
49	Mediator analysis	yes, the evaluation tests the extent to which one construct accounts for/mediates the relationship between other constructs (1) no, the evaluation does not test the extent to which one construct accounts for/mediates the relationship between other constructs (2)	

#	Variable	Code	Definition / Guidance for Coding
50	Moderator analysis	yes, the evaluation tests the extent to which one construct moderates the relationship between other constructs (1) no, the evaluation does not test the extent to which one construct moderates the relationship between other constructs (2)	

APPENDIX I

Sample Recruitment Email for Survey

Subject: Survey Request – Your Experience with Program Theories of Change

Dear X,

Greetings from Claremont Graduate University!

I am writing to you because (state relationship, e.g. member of X evaluation association, author of X article, recommended by). My name is Susana Bonis, and I am a Ph.D. student at Claremont Graduate University. As part of my dissertation, I am conducting a survey that focuses on the design and use of program theories of change and/or logic models in evaluation.

I would like to invite you to participate in a 15 to 20-minute survey on your own experience developing and using program theories of change and/or logic models. The survey is anonymous. You may stop and return to the survey at a later time, if needed. For completing the survey, you will be given the chance to enter a lottery to win one of three $100 gift cards for Amazon.com. The lottery will be drawn upon the close of the survey.

The link to the survey is as follows: (enter link). Please complete the survey by (enter date).

A summary of survey results may be shared with you, if you desire.

Thank you for taking the time to complete the survey.

Respectfully,

Susana Bonis

APPENDIX J

Survey –

Graphic Conceptual Models in Evaluation Practice

Start of Block: Default Question Block

Q1 Introduction (Study Leadership and Purpose) Hello! My name is Susana Bonis and I am a Ph.D. student at Claremont Graduate University. I am carrying out a study on theory-driven evaluation in practice. My research advisor is Dr. Stewart Donaldson, Professor of Psychology, School of Social Science, Policy and Evaluation, and Executive Director of the Claremont Evaluation Center. This survey looks at factors that may influence design and use of graphic conceptual models in evaluation (visual representations of program theory). Eligibility This survey is designed for individuals over the age of 18 who have carried out evaluations of programs or policies in various fields. Participation It should take you about 15-20 minutes to complete the survey. Questions focus on your background in evaluation and your experience in designing and using graphic conceptual models. It may be helpful for you to have accessible sample graphic conceptual models that you have developed, but this is not necessary. Your participation in the survey is voluntary, and you are free to stop the survey at any time. Risks and Benefits There are minimal risks to taking this survey, although you may feel inconvenienced by the amount of time that it takes to complete the survey. While the survey doesn't offer you direct benefits, your responses may help us learn more about how evaluators implement theory-driven evaluation. Also, your participation benefits me personally by helping me to finish my doctoral program. Compensation At the end of the survey, you will be given the chance to enter a lottery to win one of three $100 gift cards for Amazon.com. The lottery will be held around May 16. Confidentiality Your responses will be stored in a password protected file to which only I have access. At the end of the survey you will be asked if you are interested in participating in a lottery for a gift card, and if you are interested in receiving results of the survey. If you choose to provide your email address, this will only be used for the lottery or to share results with you. Your name will not be associated with your responses. Your responses to this survey will remain confidential and no names or identifying information would be included in any publications or presentations based on these data.

Contact If you have questions at any time about the study or procedures, please feel free to contact Susana Bonis at susana.bonis@cgu.edu. Consent
If you have read the information above, if you are 18 years of age or older, and if you voluntarily agree to participate in this survey, please mark "agree."

○ Agree (1)

End of Block: Default Question Block

Start of Block: Education and Training

Q2 **Education and Training**

Please note that the survey does not have a back button.

Q3 What is the highest degree or level of education that you have completed? (If you're currently enrolled in school, please indicate the highest degree you have received.)

- ○ Less than a high school diploma (1)
- ○ High school degree or equivalent (e.g., GED) (2)
- ○ Some college, no degree (3)
- ○ Associate's degree (e.g., AA, AS) (4)
- ○ Bachelor's degree (e.g., BA, BS) (5)
- ○ Master's degree (e.g., MA, MS, MEd) (6)
- ○ Doctorate (e.g., PhD, EdD) (7)
- ○ Other (Please specify) (8) _____

Q4 In what substantive field is your highest degree?

- ○ Business (1)
- ○ Economics (2)
- ○ Education (3)
- ○ Evaluation (4)
- ○ Psychology (5)
- ○ Public Health (6)
- ○ Public Administration or Public Policy (7)
- ○ Sociology (8)
- ○ No degree (9)
- ○ Other (Please specify) (10) _____

Q5 How did you receive training in program evaluation? (Please select all that apply.)

- ☐ On-the-job training; no formal training in evaluation (1)
- ☐ Through professional development opportunities (e.g., workshops, one-time courses) (2)
- ☐ Through a certificate program (3)
- ☐ As part of a degree program (4)
- ☐ Other (Please specify) (5) _____

End of Block: Education and Training

Start of Block: Work Context

Q6 **Work Context**

WHEN RESPONDING, PLEASE THINK OF YOUR WORK CONTEXT IN THE LAST FIVE YEARS.

Q7 In what parts of the world have the interventions (e.g., programs, policies) that you have evaluated been located? (Please select all that apply.)

- [ ] Africa (1)
- [ ] Asia (2)
- [ ] Australia/New Zealand (3)
- [ ] Canada (4)
- [ ] Caribbean (5)
- [ ] Europe (6)
- [ ] Latin America (7)
- [ ] Middle East (8)
- [ ] Pacific Islands (9)
- [ ] United States (10)

Q9 How would you describe the individuals served by the interventions you most frequently evaluate, in terms of culture and/or language (e.g., low-income immigrant Latino)?

_____

_____

_____

_____

_____

Q10 How frequently have you performed program evaluations in the following fields?

	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)
Agriculture (1)	○	○	○	○	○
Arts (2)	○	○	○	○	○
Community Development (3)	○	○	○	○	○
Criminal Justice (4)	○	○	○	○	○
Economic Development (5)	○	○	○	○	○
Education (PreK-12) (6)	○	○	○	○	○
Health (7)	○	○	○	○	○
Higher Education (8)	○	○	○	○	○
Housing (9)	○	○	○	○	○
Workforce Training and Development (10)	○	○	○	○	○

Carry Forward Unselected Choices from "How frequently have you performed program evaluations in the following fields?"

Q40 In each field where you conduct an evaluation, to what extent are you knowledgeable of the field, especially as related to problems and treatments?

	Not at all (1)	To a small extent (2)	To a moderate extent (3)	To a large extent (4)	To a very large extent (5)
Agriculture (x1)	○	○	○	○	○
Arts (x2)	○	○	○	○	○
Community Development (x3)	○	○	○	○	○
Criminal Justice (x4)	○	○	○	○	○
Economic Development (x5)	○	○	○	○	○
Education (PreK-12) (x6)	○	○	○	○	○
Health (x7)	○	○	○	○	○
Higher Education (x8)	○	○	○	○	○
Housing (x9)	○	○	○	○	○
Workforce Training and Development (x10)	○	○	○	○	○

Q45 Are there other fields in which you have carried out evaluations that were not mentioned in the previous question?

- ○ Yes (1)
- ○ No (2)

Skip To: Q12 If Are there other fields in which you have carried out evaluations that were not mentioned in the p... = No

Q48 How frequently have you performed evaluations in this/these other field(s)?

	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)
Other 1 (Please specify) (1)	○	○	○	○	○
Other 2 (Please specify) (2)	○	○	○	○	○
Other 3 (Please specify) (3)	○	○	○	○	○

Q47 In this/these others field(s) where you conduct evaluations, to what extent are you knowledgeable of the field(s), especially as related to problems and treatments?

	Not at all (1)	To a small extent (2)	To a moderate extent (3)	To a large extent (4)	To a very large extent (5)
Other 1 (Please specify) (1)	○	○	○	○	○
Other 2 (Please specify) (2)	○	○	○	○	○
Other 3 (Please specify) (3)	○	○	○	○	○

Q12 Do you consider yourself primarily a specialist of generalist evaluator? Specialist evaluators mainly choose projects within their own field to evaluate. Generalist evaluators often choose projects from varying fields.

- ○ Specialist (1)
- ○ Generalist (2)

End of Block: Work Context

Start of Block: Design of Program Theory 1

Q13 **Design of Program Theory 1**

Q41 WHEN RESPONDING TO THE FOLLOWING QUESTIONS, PLEASE CHOOSE THE ITEMS THAT ARE TRUE FOR THE MAJORITY OF PROGRAM THEORIES THAT YOU HAVE DEVELOPED IN THE LAST FIVE YEARS. "A program theory is an explicit theory or model of how an intervention, such as a project, a program, a strategy, an initiative, or a policy, contributes to a chain of intermediate results and finally to the intended or observed outcomes (Funnell and Rogers, 2011)."

Q14 How frequently have you developed a program theory as part of your evaluation, when an intervention does not yet have one?

- ○ Never (1)
- ○ Rarely (2)
- ○ Sometimes (3)
- ○ Often (4)
- ○ Always (5)

End of Block: Design of Program Theory 1

Start of Block: View of Program Theory

Q61 **View of Program Theory**

Q62 Please share why you do not use, or have not yet used, program theory in your evaluations.

_____

_____

_____

_____

_____

End of Block: View of Program Theory

Start of Block: Closing Questions

Q59 **Closing Questions**

Q43 If you would like to participate in the lottery for a $100 gift card to Amazon, please enter your email address. (Three individuals will receive gift cards.)

_____

Q44 If you would like to receive results of this survey, please enter your email address.

_____

Q60 If you have any additional comments, please share them here.

_____

_____

_____

_____

_____

End of Block: Closing Questions

Start of Block: Design of Program Theory 2

Q54 **Design of Program Theory 2**

PLEASE REMEMBER TO CHOOSE THE ITEMS THAT ARE TRUE FOR THE MAJORITY OF PROGRAM THEORIES THAT YOU HAVE DEVELOPED IN THE **LAST FIVE YEARS**.

Q15 In most program theories that you have developed, how frequently did you use the following sources of information?

Please note that social science theory is defined as "a logically interrelated set of propositions about

empirical reality" (Schutt, 2011). Certain programs may be intentionally designed with certain social science theories in mind, e.g. the theory of planned behavior or social cognitive theory.

Stakeholder theory is based stakeholders' observations and experiences in working with clients.

	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)
Social science theory (1)	○	○	○	○	○
Research on or evaluation of similar programs (2)	○	○	○	○	○
Stakeholder theory (3)	○	○	○	○	○
Program observation by evaluator (4)	○	○	○	○	○
Program document review by evaluator (5)	○	○	○	○	○

Q17 Please describe how you have typically facilitated conversations around program theories among stakeholders (for example, for what length of time did you convene stakeholders, and what were common questions and activities that you included in the session).

_____

_____

_____

_____

_____

Q18 When you have facilitated conversations around program theories, to what extent did stakeholders typically have a dialogue on the following topics?

	Not at all (1)	To a small extent (2)	To a moderate extent (3)	To a great extent (4)	To a very great extent (5)
Beliefs or assumptions that underlie the program (1)	○	○	○	○	○
How social science theory or previous research and evaluation influences choice of strategies and desired outcomes (2)	○	○	○	○	○
If identified resources are sufficient to implement strategies to desired level (3)	○	○	○	○	○
If sufficient numbers of people are being served to expect influence on the desired outcomes (4)	○	○	○	○	○
Strength of connection between identified strategies and outcomes (5)	○	○	○	○	○
If duration and sequence of chosen strategies are sufficient to accomplish desired outcomes (6)	○	○	○	○	○

External factors that may influence desired outcomes (7)	○	○	○	○	○

Q19 In the majority of your experiences facilitating conversations around developing program theories, to what extent did stakeholders involved in the process seem to develop a shared understanding of how the program is expected to work?

- ○ Not at all (1)
- ○ To a small extent (2)
- ○ To a moderate extent (3)
- ○ To a great extent (4)
- ○ To a very great extent (5)

Q42 When designing program theories, what stakeholder groups have you usually engaged (e.g., program staff, parents, students, etc.)?

_____

_____

_____

_____

_____

Q20 At what point in a program have you most often been brought in to develop a program theory?

- ○ At the start of a program—when the program is being designed (1)
- ○ At a time when the program is being tested and can still be modified (2)
- ○ At a time when the program is established, and few modifications are possible (3)
- ○ Not applicable; the program theory has already been developed by the time I am brought in to evaluate an intervention (4)

End of Block: Design of Program Theory 2

Start of Block: Representation of Program Theory

Q21 **Representation of Program Theory**

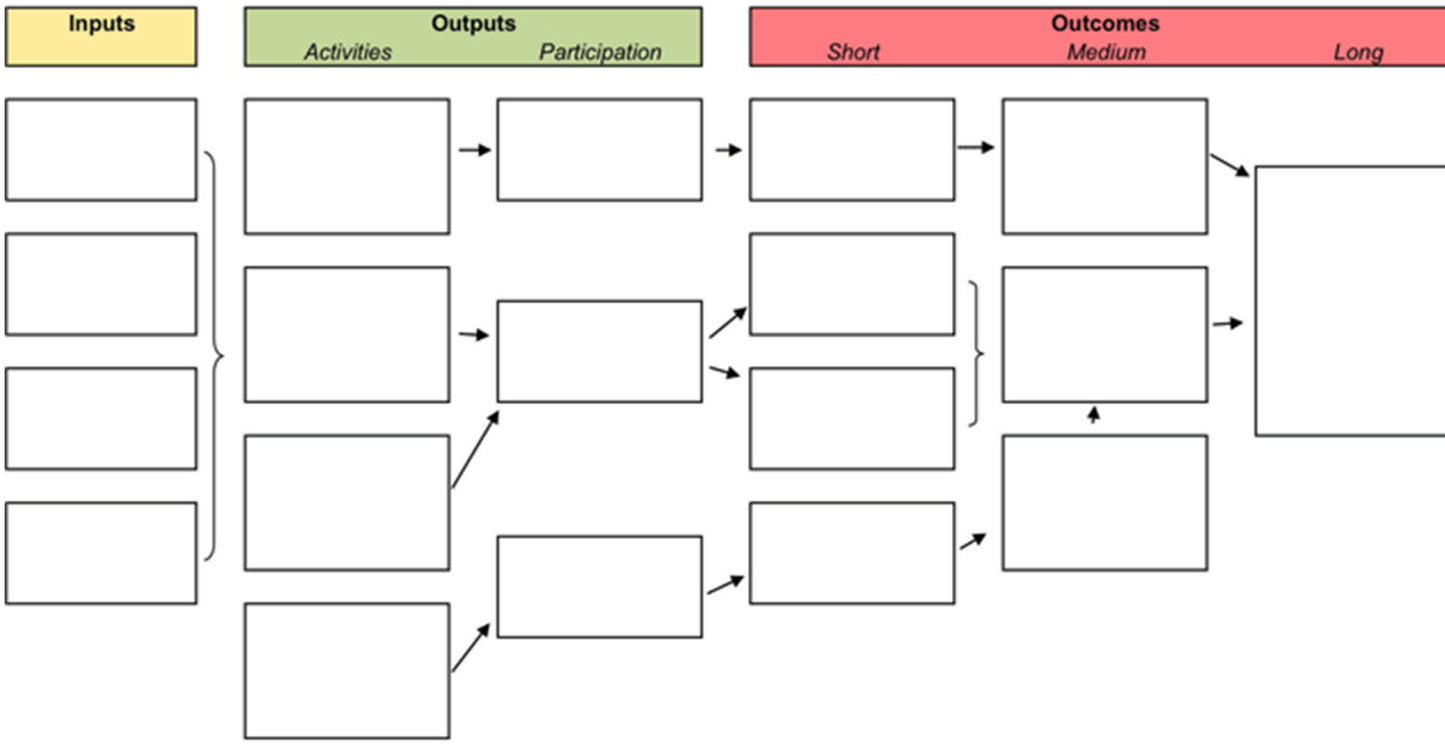
You're halfway done! Thank you so much for helping with my research!

WHEN RESPONDING TO THE FOLLOWING QUESTIONS, PLEASE CHOOSE THE ITEMS THAT ARE TRUE FOR THE MAJORITY OF PROGRAM THEORIES THAT YOU HAVE DEVELOPED IN THE LAST FIVE YEARS.

Q50 In theory-driven evaluation, the program theory is often depicted in a graphic conceptual model. For the next few questions, it is suggested that you have accessible sample graphic conceptual models that you have developed in the last five years (e.g., logic models, tables, flow charts, or other diagrams describing program theory). A sample model is shared below.

Q55

Program: (name) **Logic Model**
Situation:



Q49 How frequently have you developed the following types of models in your evaluation work?

	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)
Table (1)	○	○	○	○	○
Diagram with shapes and arrows (2)	○	○	○	○	○
Model built on a visual metaphor (e.g., a program as a tree or building) (3)	○	○	○	○	○
Multi-dimensional model (e.g., cube) (4)	○	○	○	○	○
Other (Please specify) (5)	○	○	○	○	○

Q24 How often has cultural context of the program/participants/location influenced some of the graphic conceptual models that you have designed?

- ○ Never (1)
- ○ Rarely (2)
- ○ Sometimes (3)
- ○ Often (4)
- ○ Always (5)

Skip To: Q26 If How often has cultural context of the program/participants/location influenced some of the graphi... = Never

Q25 Please describe the most common modifications that you have made to graphic conceptual models in response to different cultural contexts in which you have worked.

_____

_____

_____

_____

_____

Q26 What type of model have you most frequently designed for interventions where the path to reach outcomes is known, e.g., knowledge change resulting from a well-tested and implemented training program?

- ○ Table (1)
- ○ Diagram with shapes and arrows (2)
- ○ Model built on a visual metaphor (e.g., a program as a tree or a building) (3)
- ○ Multi-dimensional model (e.g., cube) (4)
- ○ Other (Please specify) (5) _____

Q27 What type of graphic conceptual model have you most often constructed for interventions where the causal pathway may be adaptive or emergent—where it is not easy to identify in advance details of what will be done or accomplished?

- ○ Table (1)
- ○ Diagram with shapes and arrows (2)
- ○ Model built on a visual metaphor (e.g., a program as a tree or a building) (3)
- ○ Multi-dimensional model (e.g., cube) (4)
- ○ Other (Please specify) (5) _____

Q28 In the models that you have developed, what aspects of the program have you most frequently described?

- ○ Process theory only (the program's inputs, activities, and outputs) (1)
- ○ Impact theory only (the outcome chain) (2)
- ○ Both process theory and impact theory (3)

Q29 How frequently have you included moderators in your model (variables that affect relationships in the model, such as the relationship between program activities and outcomes, or the relationship between two outcomes)?

- ○ Never (1)
- ○ Rarely (2)
- ○ Sometimes (3)
- ○ Often (4)
- ○ Always (5)

Skip To: Q58 If How frequently have you included moderators in your model (variables that affect relationships in... = Never

Q30 Which of these variables that could affect relationships in your model have you most often included in the model itself? (Please select all that apply.)

- ☐ Participant characteristics (e.g., age, gender, race, ethnicity) (1)
- ☐ Characteristics of program providers/staff (e.g., level of education, experience) (2)
- ☐ Characteristics of the setting of program implementation (3)
- ☐ Program attendance (4)
- ☐ Strength of program activities (frequency and duration of activities) (5)
- ☐ External factors (e.g., government policies, funding) (6)
- ☐ Assumptions (7)
- ☐ Other (Please specify) (8) _____

Q58 Among evaluations that you have carried out in the last five years, how frequently have you revised a model of program theory during the evaluation period in response to changing circumstances?

- ○ Never (1)
- ○ Rarely (2)
- ○ Sometimes (3)
- ○ Often (4)
- ○ Always (5)

Q51 Among evaluations that you have carried out in the last five years, how frequently have you developed different versions of a model of program theory to better meet the needs of stakeholders?

- Never (1)
- Rarely (2)
- Sometimes (3)
- Often (4)
- Always (5)

End of Block: Representation of Program Theory

Start of Block: Use of Program Theory in Evaluation

Q32 **Use of Program Theory in Evaluation** WHEN RESPONDING TO THE FOLLOWING QUESTIONS, PLEASE CHOOSE THE ITEMS THAT ARE TRUE FOR THE MAJORITY OF EVALUATIONS THAT YOU HAVE CONDUCTED IN THE LAST FIVE YEARS.

Q33 How frequently have you used program theory to develop your evaluation questions?

- Never (1)
- Rarely (2)
- Sometimes (3)
- Often (4)
- Always (5)

Q34 How frequently have the following been a focus of evaluations that you have conducted?

	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)
Process and implementation (1)	○	○	○	○	○
Outcomes and Impact (2)	○	○	○	○	○
Context / Moderators (3)	○	○	○	○	○
Relationships in the outcome chain (mediation) (4)	○	○	○	○	○

Q38 When reporting on evaluation findings to stakeholders, how frequently do you discuss the findings in relation to the program theory?

- ○ Never (1)
- ○ Rarely (2)
- ○ Sometimes (3)
- ○ Often (4)
- ○ Always (5)

End of Block: Use of Program Theory in Evaluation

Start of Block: Barriers to Using Theory-Driven Evaluation

Q35 **Barriers to Using Theory-Driven Evaluation**
ALMOST DONE!

Q36 How much of a barrier have each of the following been in your attempts to implement theory-driven evaluation?

	Not at all (1)	To a small extent (2)	To a moderate extent (3)	To a large extent (4)	To a very large extent (5)
Program/Organization is not interested in critically examining the program theory—they believe the program is fine as it is. (1)	○	○	○	○	○
Program/Organization only wants to know about final outcomes--little interest in relationships among outcomes along the causal chain or in variables that could affect outcomes. (2)	○	○	○	○	○
Funder (e.g., foundation, government agency) is only interested in knowing about final outcomes--little interest in relationships among outcomes along the causal chain or in variables that could affect outcomes. (3)	○	○	○	○	○
My own content knowledge of the field in which the program is based. (4)	○	○	○	○	○
My own technical knowledge about theory-based evaluation and/or research methods. (5)	○	○	○	○	○
Time allocated to conduct a theory-driven evaluation. (6)	○	○	○	○	○
Financial resources allocated to conduct a theory-driven evaluation. (7)	○	○	○	○	○

Q37 Do you face any other barriers to conducting theory-driven evaluation?

_____

_____

_____

_____

_____

Q63 Are you are interested in participating in the lottery, receiving survey results, or sharing additional comments, please mark "yes."

- Yes (1)
- No (2)

End of Block: Barriers to Using Theory-Driven Evaluation