Systems Evaluation (SysEval): Applying Systems Thinking to Evaluation

Preprint · January 2021

CITATIONS
0

READS
270

3 authors:

Jennifer Kushner
University of Wisconsin–Madison
11 PUBLICATIONS **61** CITATIONS
SEE PROFILE

Laura Cabrera
Cornell University
51 PUBLICATIONS **515** CITATIONS
SEE PROFILE

Derek Cabrera
Cornell University
73 PUBLICATIONS **911** CITATIONS
SEE PROFILE

This is a draft preprint chapter from a forthcoming manuscript. Citation information as follows: Kushner, J., Cabrera, L. and Cabrera, D. (2021) Systems Evaluation: Applying Systems Thinking to Evaluation. In, Routledge Handbook of Systems Thinking, (Eds) Cabrera, D., Cabrera, L. and Midgley, G. Routledge. London, UK.

Systems Evaluation (SysEval): Applying Systems Thinking to Evaluation

Kushner, J., Cabrera, L. & Cabrera, D.

Evaluation today

The Promise

The challenges we face in today's world are complex and multifaceted, and the interventions and policies to act on these challenges are equally complex. Program evaluation is a means by which we can know if, how, and under what conditions our interventions are making a difference. In her seminal contribution to the field, Carol Weiss (1970) [1] explained, the purpose of program evaluation is "to measure the effects of a program against the goals it set out to accomplish as a means of contributing to subsequent decision making about the program and improving future programming." The Centers for Disease Control (2012) [2] define program evaluation as "the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future program development." Further, "…evaluation should be practical and feasible and conducted within the confines of resources, time, and context. Moreover, it should serve a useful purpose, be conducted in an ethical manner, and produce accurate findings. Evaluation findings should be used both to make decisions about program implementation and to improve program effectiveness." At its core, evaluation is about learning, by using feedback to understand and improve program effectiveness.

Whether intentional or not, program development and evaluation are inherently coupled. Program development involves the selection of specific activities intended to bring about specific outcomes, and in response to a specific situation. A program's design explicitly or implicitly reflects a theory of change, sometimes referred to as program logic. This can be carefully articulated, or emergent and developmental. Either way, a theory of change is the rationale for why certain activities are expected to contribute to certain outcomes. Program implementation tests the theory of change in the real world. Notably, feedback that can be systematically and strategically used to make improvements in the program approach is readily available during the implementation process. Feedback affirms or challenges the assumptions and hypotheses embedded in the theory of change and helps to build an evidence base for what works. This is what enables us to replicate, improve, and evolve our programmatic efforts to have greater impact.

The norm in evaluation today is to focus on localized or isolated program efforts, rather than individual programs within a broader context. Favorite tools, such as log frames, and logic models tend to depict efforts not connected to related programs, issues, or outcomes. Yet for many of our most entrenched societal issues, we would be well served to understand how different programs work in mutually reinforcing ways, and contribute distinctly to an evolving body of knowledge about what works. Evaluations that focus solely on a single, isolated, local program run the risk of duplicating efforts with programs that have come before.

Methods mismatch

The current culture in evaluation focuses on public value, return on investment, scalability, and performance measurement. As such, today's evaluation landscape is populated with frameworks and methods [3]. Evaluators are tasked with determining appropriate methods for measuring the efficacy of a

theory of change. While there are many methods available, there is a tendency to favor certain ones, regardless of their appropriateness. For example, there is a deeply instilled debate in the evaluation community over the use of randomized control trials (RCTs) versus observational studies to evaluate impact. Unfortunately, RCTs have taken on increasing cache within evaluation circles, causing them to be selected even when they are an inappropriate choice. As a result, this method is woefully misapplied at lower levels of validated knowledge about social phenomena and the complexity entailed therein. Evaluators are on both sides of the debate, some believing that RCTs are the gold standard to determine impact and the most rigorous form of evaluation producing the most accurate (i.e., unbiased) results [4].

A counter-movement advocates for qualitative methods that engage participatory evaluations that capture stakeholder input for determining desirable outcomes [5]. These scholars argue that RCTs are not effective at evaluating *all types* of interventions, including those related to public goods [6]. Additionally, they argue that these evaluations are context specific and not always transferrable across populations, geographic areas, etc., which instead may require qualitative data to understand divergent experiences or outcomes [7]. They suggest that qualitative information is useful to provide a broader understanding of the context, and that RCTs provide unique insights into *part* of that context, including to help explain why results occurred [7]. *Thus, in striving for evaluations to appear highly robust and sophisticated, we may sacrifice reliability and validity by not choosing our methods wisely.*

Logic models-a popular convention

Establishing reliability and validity often begins with setting forth program logic via a traditional logic model. Generally, models are abstractions used to depict important elements and relationships within a system [8]. Logic models are the most widely adopted evaluation tool, used across sectors and organization types, to help people map their program activities to their intended outcomes [9]. They are commonly used to depict a change effort usually involving some sort of educational programming or intervention. The parts of a conventional logic model include a situation, inputs, outputs, outcomes, assumptions, and external factors. [10–13], as depicted in Figure 1.

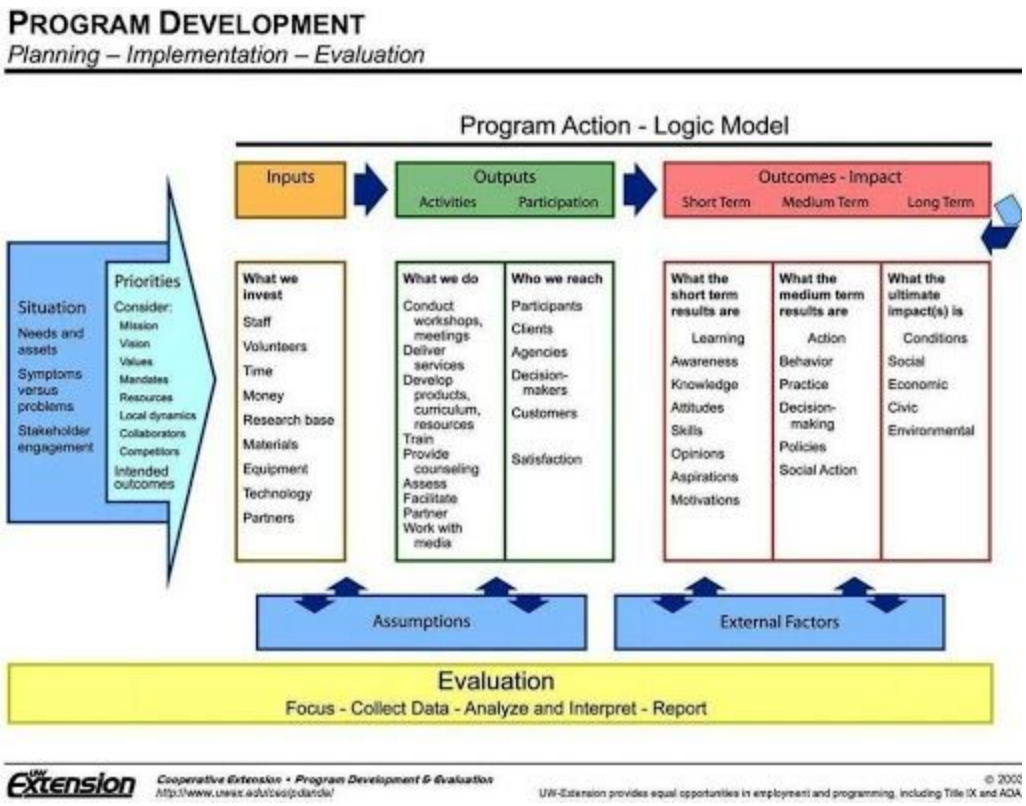

Figure 1. Traditional Logic Model (University of Wisconsin- Extension).

The Elementary and Secondary Education Act of 1965 prompted first official use of logic models as a way for schools to demonstrate accountability for the federal funds they received [14]. Logic models gained a foothold in 1969 when Leon J. Rosenberg of Fry Consultants developed the "logical framework approach" for USAID. The term "logic model" was popularized by Wholey in 1979. They have been called by other names such as "Chains of Reasoning" [15], "Theory of Action," [16], "Performance Framework" [17,18], and the Logical Framework used by Management Systems International in their international development efforts. The logic model was influenced by Schuman (1967) [19] who proposed the processes and mediating factors involved in an evaluation, Weiss (1970) [1] who coined the term "program model" and Fitz-Gibbon and Morris (1975) [20] who introduced theory-based evaluation. Despite work by scholars, the model was largely used in the international development space until 1996 when the United Way of America published a program manual on measuring the outcomes of programs for its local agencies and grantees. Later, large private funders like Kellogg and Hewlett Foundations integrated logic models into their programs. Today, logic models are used at all levels and across types of evaluation (see Figure 2).

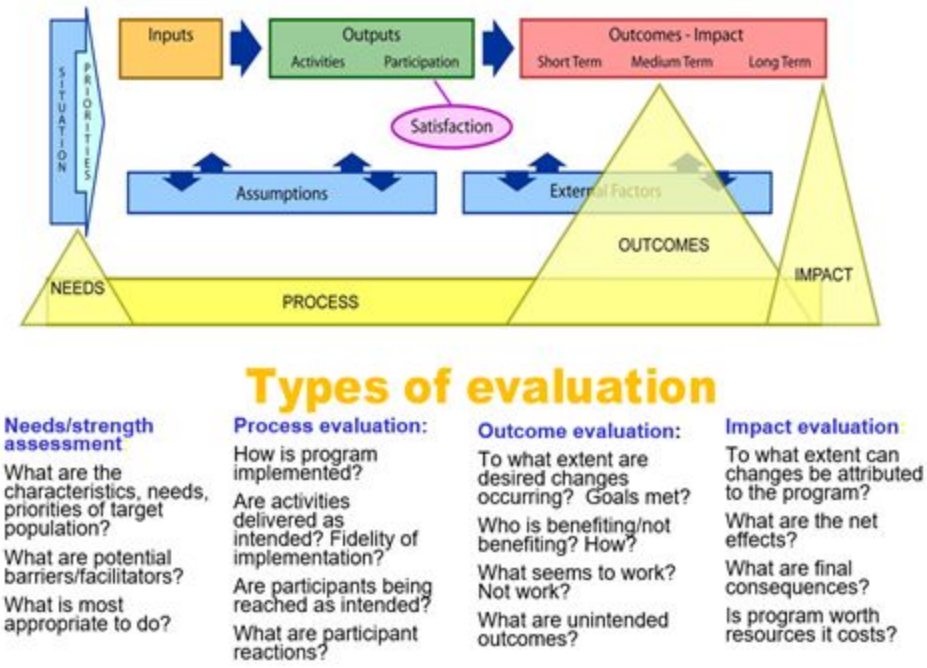

Figure 2. Types of Evaluation (University of Wisconsin-Extension)

The principal intent of a logic model is to depict the logic for the choice of program activities to bring about certain outcomes. When articulated clearly this depiction should be grounded in a clear hypothesis and/or evidence base. Many public and private funding agencies require them because they make the program's theory of change visible in a succinct, albeit limited way. At quick glance, a funder can see the proposed activities and outcomes, and the alignment among them. Logic models force a distilled description of outcomes that are often elusive in grant application narratives but needed by funders to communicate impact and return on investment to congress or their stakeholders.

The popularity of logic models stems from their ease of use and perceived utility in guiding program management. [1,21–23]. For collaborators, logic models facilitate program planning and communication [10] by building a shared mental model and common language representing diverse perspectives. Additionally, sets of logic models can be used to guide multifaceted evaluations by creating coherence across interconnected parts of an intervention [24]. Separate-but-programmatically-related logic models can stand alone or be used to further expand aspects depicted in an overarching programmatic model. Logic models are not scale-dependent; they can be used in large impact studies, or small pilot studies. They also require few resources to develop, making them a cost-effective tool for articulating a theory of change [23,25,26].

Limitations: the paradigm and the tools

Despite evaluation's promise, the multitude of methods available to measure and describe change, and the seemingly simple tools for planning and communicating, evaluation today is challenged. This challenge is rooted in logic itself. At the crux of the matter is that our approaches and tools are grounded in bivalent logic which are being applied to multivalent realities. Thus, we begin our evaluative work in a state of misalignment in our foundational logic. As a result, the common approaches and tools used in evaluation today are letting us down. Although we do our best to accurately portray what is happening, we know at a deep level that our evaluations fall short. Our current tools and approaches fail to capture the complexity

of what is happening within-- and because of --our programs. We also struggle to build a shared evidence base of what works across interventions.

Although popular, logic models have shortcomings that often shift focus to the wrong things in our evaluation efforts. Perhaps most importantly, the structure of logic models is rigid and can lead to thinking errors, and muted results. First, all of the components are presented as equally important to a theory of change, when in fact they are not. Articulating the situation, inputs, outputs, outcomes, assumptions, and external factors undergirding a program might all be important for the purposes of program planning, but not for *explicating the logic of a program*. Users of logics models for evaluation purposes are misled into thinking each of these components is equally important because they are given equal structural weighting. This distracts us, pulling focus to details that are not important for evaluating the *efficacy of a theory of change*, and we lose focus on the components that help us build a deep understanding of what works and why. Figure 3 depicts the most important parts of a theory of change for the purposes of evaluation: *activities* and *outcomes* connected and evolved by *logic* and *feedback*. Note the purposeful *iteration* ('F' in the Figure) of feedback built into this model. In other words, while the relationship between activities and outcomes is of paramount importance, of equal priority is the evolution of one's understanding of effectiveness throughout a program cycle via feedback [27].

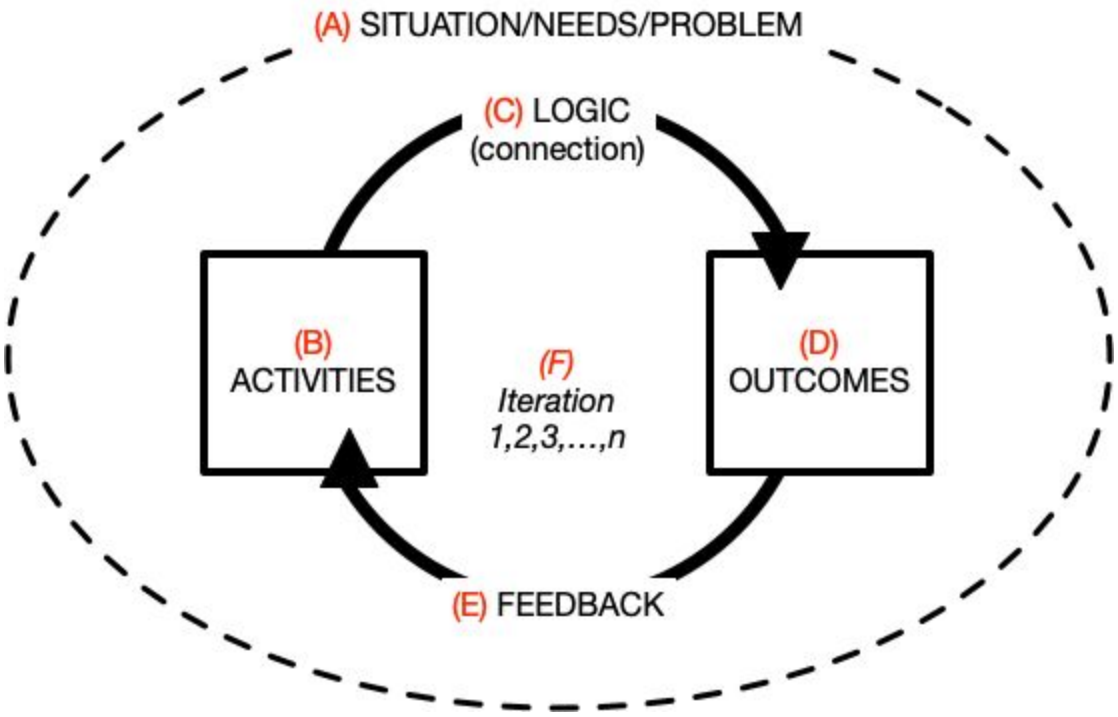

Figure 3: The Steps of SysEval (A through F)

Secondarily, logic models *do not structurally allow for the articulation of relationships between various parts of the program-evaluation system*. Because these relationships are the conceptual glue that binds parts of a theory of change and unifies it into a *theory*, and because they are not made explicit, the logic case we make with a conventional logic model is weak at best. The simplified linear structure of a traditional logic model is made up of *assumed* (not articulated) causal relationships that do not reflect the intricacies and complexities of programs within real-world systems [10]. Usually, logic models are quite binary, relying on simplistic "if…then" relationships that do not make explicit the evidentiary basis for how or why they are related. As evaluators and program planners, we should care deeply about what is known to be effective and should seek to advance that body of knowledge through our own testing and

validation. The format of a traditional logic model also causes users to glean out important contextual variables, such as assumptions, because they occupy the periphery of the model and are not included in the causal relationships that are core. In other words, logic models are popularly understood to depict a complete theory of change but in fact do not accurately reflect the full *logic* of a program.

Equally limiting is that logic models are primarily used to depict isolated, 'local'-level efforts, such as an individual program. Yet we know no programmatic effort is ever isolated, but built on existing knowledge and situated within a larger context of factors, including other programs. It is this larger context that is important in building a knowledge base, grounded in evidence across distinct but related efforts. If we could see those isolated and local programs in relationship to each other, through synergistic activities, and shared outcomes, we would build a deeper understanding of causality. A single logic model alone does not establish causality, but instead must be situated within a system of existing knowledge [9]. Ideally logic models should point us to look beyond our 'local' program to attribute impact, by connecting our evidence base and theory to that of other efforts. Program managers and evaluators seeking to measure collective impact are limited by the same shortcomings that befall singular and isolated logic models. Namely, the relationships and perspectives between programs are not made explicit. When we evaluate programs without recognizing the possible webs of causality for specific outcomes, we sell ourselves short in building meaningful and accurate understandings of what works.

Logic models can help organize an individual or group's thinking about a theory of change. But they are just approximations of reality, not reality. For the model to remain relevant, it must evolve based on the dynamic, complex system that it represents [28]. Another functional challenge is that logic models are often mistaken for a static model or a point-in-time snapshot of the overall effort. Often not updated past their initial creation, they fail to represent changes over time. In the end, this inhibits adaptive management and ultimately true understanding of what works as it is a fundamental failure of the need for evolving mental models for better alignment with reality.

Logic models are ubiquitous and yet much of the time implemented with limited fidelity to meet the needs of funders and decision-makers. In addition to the structural and functional limitations previously discussed, other issues occur when users employ a 'plug and chug' approach to filling in boxes without needing to think deeply about the distinctions, systems, relationships and perspectives embedded in their program's theory of change. In today's accountability landscape, clear and concise articulation of outcomes is essential. Logic models appear to invite that concise articulation, yet confounding the issue is that users get confused by the logic model terminology. Linguistically, the similarity between inputs, outputs and outcomes often leads people to using the wrong terms (and thus meanings) when communicating about outcomes. It is not uncommon for outcomes depicted in logic models to not be outcomes at all, but *outputs*, which are not indicators of change. This can create additional headaches for funders who may look to logic models to quickly glean the intended outcomes of a proposed project. In sum, the logic that should make up the foundation of an evaluation, is only partially represented in a conventional logic model.

These problems can be mitigated by the application of systems thinking to evaluation theory and practice, or systems evaluation. A systems evaluation explicates the full logic of programs by connecting activities to outcomes through an articulated hypothesized or evidentiary basis. This, in turn, increases the internal construct validity within our logic models. Additionally, systems evaluation evolves our understanding of what works by helping us integrate feedback to improve program effectiveness. Thus, systems thinking tools and models can help us be more accurate in both articulating our theory of change, and improve evaluation practice.

This chapter expands original theorizing on systems evaluation which applies four tenets of systems thinking to evaluation [29]. These four tenets are: making distinctions, explicating systems, articulating

relationships, and taking perspectives (or DSRP). These can be applied by program evaluators, planners, funders, facilitators of collective impact, or agents of change to any program, policy, or intervention intended to better a condition or impact a population. Systems evaluation improves our ability to capture *the real-world complexity* within and surrounding programs and has been applied to evaluations across the country--from schools to the U.S. military. Other researchers and practitioners have also recognized the limitations of the current methods for evaluating complex, real-world systems [30]. Yet until now we have not had a simple tool to apply the four tenets of systems thinking to a logic model or help us articulate a theory of change including its complexities.

Systems evaluation: the way forward

DSRP as multivalent logic

If you struggle to articulate your program and its evaluation within the linear confines of a traditional logic model, there is a solution. This solution is based on a different kind of logic, a new and more advanced kind of logic called multivalent logic. Multivalent logic reflects the real world, where causality occurs not in linear fashion but in webs of causality. A new kind of logic model, built upon multivalent logic, enables more accurate and effective modeling of programs and their evaluations... As explained by Reynolds and colleagues (2016) [31]:

> "In accordance with laws of requisite variety good (bounded) models of evaluation need to somehow reflect, accommodate, and adapt to such ongoing change. Using rigid input–output models for evaluation tend not to be adaptive."

This new model uses a complex adaptive system perspective in which programs are dynamic and in which many nonlinear relationships bring about the emergent behavior of the system [31].

Systems evaluation enables evaluators, funders, program planners and others leading complex change initiatives to accurately understand and measure the impact of their work. All programs are systems and should be evaluated using DSRP to yield more robust results. For example, paying attention to distinctions helps us group data and clarify the boundaries of what is inside and outside of the scope. How data is split and grouped by evaluands provides insight into their understanding of the phenomena under investigation. Our brains relate things with and without our awareness. As evaluators, we must be vigilant about the conclusions we draw. This requires knowing which relationships we are paying attention to and which we are not. Reynolds (2016) [31] explained that evaluators sit in very partial positions in attributing reality, which can be very problematic. Applying systems thinking/DSRP to evaluation enables evaluators to step back, consider multiple perspectives, and take an impartial role. The various perspectives taken *within and on* a program or evaluation, shape its analysis, interpretation, findings, and recommendations. In evaluation, we strive to know how something really works, not just how we *thought* it might work. Systems thinking enables us to reveal the biases embedded in our theories and models. It helps unpack what the data is telling us directly and distinguish that from what we are expecting it to say.

A new approach to logic modeling using DSRP includes the same core components of a traditional logic model, but it situates them in the appropriate structural place to reflect their role in the context of articulating and evolving a theory of change. Table 1 shows how traditional logic model structure can be extended through a systems approach.

Table 1. Systems Logic Model Structure

(B) Activities	**(C) Logic**	**(D) Outcomes**	**(E) Feedback**

(F) ITERATION 1 **(A) Situation/Need/Problem:** What is the situation? The assessment of the need or the problem to be solved?			
Participation (sample): What is your participant sample? (demographics, psychographics, etc)	**Assumptions:** What are your eval/research assumptions about (1) Theory of Change and (2) Knowledge-Method Match?	**Short, Mid, Long Range Desired Outcomes:** What are your [hypothesized] short, mid, or long term outcomes for one or more of the activities? (use multiple rows below as needed)	**Information:** What is the data generated?
Inputs (resources): What resources or assets will be used?	**Hypothesis:** What hypothesis are you testing about the relationship between the activities and the outcomes?		
Outputs (activities): What are the program activities? (Use multiple rows below as needed)	**Evidence:** What evidence exists for the above? (e.g., lit review, etc)		**Interpretation:** What is your interpretation of the data? How does it inform your next round? Was any evidence generated that can be added to the logic? **(E) What part of this is useful feedback for model or program evolution?**
Connections: How are your activities connected by logic to your outcomes?			
Activity 1: Add activity 1 here	**relates to:**	**Outcome 1:**	**Relationship 1:**
Activity 2: Add activity 2 here	**relates to:**	**Outcome 2:**	**Relationship 2:**
Activity 3:.Add activity 3 here	**relates to:**	**Outcome 3:**	**Relationship 3:**
Activity 4: Add activity 4 here	**relates to:**	**Outcome 4:**	**Relationship 4:**
(F) ITERATION 2 **(A) Situation/Need/Problem:** What is the situation? The assessment of the need or the problem to be solved?			
Participation (sample): [insert answer here]	**Assumptions:** [insert answer here]	**Short, Mid, Long Range Desired Outcomes:** [insert answer here]	**Information:** [insert answer here]
Inputs (resources): [insert answer here]	**Hypothesis:**[insert answer here]		**Interpretation:** [insert answer here] **(E) What part of this is useful feedback for model or program evolution?**
Outputs (activities): [insert answer here]	**Evidence:** [insert answer here]		
Connections: How are your activities connected by logic to your outcomes?			
Activity 1: [insert answer here]	**relates to:** [insert answer here]	**Outcome 1:** [insert answer here]	**Relationship 1:** [insert answer here]
(F) ITERATION *n* **(A) Situation/Need/Problem** [insert answer here]			
Participation (sample): [insert answer here]	**Assumptions:** [insert answer here]	**Short, Mid, Long Range Desired Outcomes:** [insert answer here]	**Information:** [insert answer here]
Inputs (resources): [insert answer here]	**Hypothesis:** [insert answer here]		**Interpretation:** [insert answer here] **(E) What part**

Outputs (activities): [insert answer here]	**Evidence:** [insert answer here]		**of this is useful feedback for model or program evolution?**
Connections: How are your activities connected by logic to your outcomes?			
Activity 1: [insert answer here]	**relates to:** [insert answer here]	**Outcome 1:** [insert answer here]	**Relationship 1:** [insert answer here]

For example, inputs (participants, resources) are a component of the activities offered by the program. They work in service of the activities, but are not what will create change by themselves. Likewise, the outcomes that a program seeks to achieve should reflect a changed state of a situation. For example, if people are hungry, a home gardening program should aim to reduce hunger.

The model is different in several other important ways. First and perhaps most importantly, it includes *logic*, or the conceptual glue that binds program *activities* to *outcomes*. In use, that logic as it evolves over time, might start with assumptions, move to hypotheses, and develop further based on evidence that becomes increasingly sophisticated. Second, it includes *feedback* (depicted as Item E in the table), or data about what actually happens when the program is implemented. In use, *feedback* is the difference between the original model of the program, and the model based on what actually happens.

In the example in Table 2, we see a program that is designed to address the situation of coffee plantations challenged by low productivity. The program's theory states that if training about plantation management coupled with good farm inputs are provided to coffee farmers, these will contribute to an increase in new trees per year, and greater yields per tree.

Table 2. Systems Logic Model Example: Coffee Production

(B) Activities	(C) Logic	(D) Outcomes	(E) Feedback
(F) ITERATION 1 **(A) Situation/Need/Problem:** Coffee Value Chain is challenged by 1) low productivity and quality, 2) an ineffective market system, and 3) limited capacity of farmers/associations			
Participation (sample): farmers, farmer associations, buyers, roasters, consumers, and investors	**Assumptions:** N/A	**Short, Mid, Long Range Desired Outcomes:** What are your [hypothesized] short, mid, or long term outcomes for one or more of the activities? (use multiple rows below as needed)	**Information:** What is the data generated?
Inputs (resources): Disease-free seedlings, technical experts/trainers, financial credit, post-harvest facilities, technology solutions, farm inputs	**Hypothesis:** Strengthen points and links of value chain and international coffee sales will increase.		
Outputs (activities): Increase production, productivity and quality	**Evidence:** 2011 Catholic Relief Services study		**Interpretation:** What is your interpretation of the data? How does it inform your next round? Was any evidence

Strengthen market system Expand integrated support services	Past performance of alliance partners Lessons learned from Finca et. al. International Coffee Org data		generated that can be added to the logic?**(E) What part of this is useful feedback for model or program evolution?**
Connections: How are your activities connected by logic to your outcomes?			
Activity 1: Renovate plantations and improve management (just one activity as example)	**relates to:** 1	**Outcome 1:** Smallholder farmers increase sales of high quality cherries to their associations.	**Relationship 1:** If farmers adopt BMPs then their productivity and quality improves. Keep refining and focus on weaknesses.
Activity 2: Update processing centers with proven coffee equipment	**relates to:** 1,2	**Outcome 2:** Farmer associations add value to members and deliver consistent products to exporters.	**Relationship 2:** Improved or /increased coffee performance.
Activity 3: Aggressive sales and marketing effort to specialty coffee market.	**relates to:** 2,3	**Outcome 3:** The coffee brand increases its sales on the specialty coffee market.	**Relationship 3:** Increased exports of green coffee.
Activity 4: Develop a digital platform to connect all actors in the value chain.	**relates to:** 4	**Outcome 4:** Integrated support services support the regional coffee sector.	**Relationship 4:** Successful pilot and scale of app and eCommerce platform.
(F) ITERATION 2 **(A) Situation/Need/Problem:** 1) increased productivity and quality, 2) system is strengthened but access is challenged, 3) continued need for capacity development			
Participation (sample): Focus on high performers (supporters/leaders)	**Assumptions:** technology solutions contribute to change insofar as farmers have reliable access to tech and its enabling environment.	**Short, Mid, Long Range Desired Outcomes:** What are your [hypothesized] short, mid, or long term outcomes for one or more of the activities? (use multiple rows below as needed)	**Information:** Continued practice or adaptation of BMPs; yields/cupping scores; #containers; user #s/activity/satisfaction
Inputs (resources): same inputs minus facilities	**Hypothesis:** If each point and link in the value chain is strengthened, the whole system will work more effectively.		**Interpretation:** Areas of non- or slow performance get investigated; activities adjusted. Slow adoption of practices understood better and addressed.
Outputs (activities): same outputs	**Evidence:** Progress toward outcomes 1-4		
Connections: How are your activities connected by logic to your outcomes?			
Activity 1: [insert answer here]	**relates to:** [insert answer here]	**Outcome 1:** [insert answer here]	**Relationship 1:** [insert answer here]

The logic behind this approach is based on past evidence of what does and does not work (gleaned from '*feedback*'). In round two of the program, we see that the situation has shifted slightly. Despite pest management and more trees, productivity is challenged by increasing hurricanes and related weather events. To address this, the program officers shift activities to include training in climate-smart plantation management. The *logic* is that if farmers understand and can implement practices that increase resilience to environmental shocks, their yields will increase. Perhaps in round three of the program, evaluation data (*feedback*) reveals that farmer yields increased—but sales did not. Drawing on past research that suggests productive farmers benefit from farm insurance, round three may include new activities related to farm insurance. *In other words, the program seeks and responds to feedback throughout in an iterative process that improves the program throughout it's life, avoiding the "we missed the mark" conclusion at the program's end.*

In the model the core parts of a theory of change include: *activities, outcomes*, and *logic*. For the theory of change to be validated and evolved the new model explicitly includes *feedback*. These distinct but related parts bring a program into greater ongoing efficacy by tightening the difference between the model of the program and its effect. At the same time, we can build an evidence base of what works by connecting to and building on what is known or learned about the efficacy of interventions. As explained by McLaughlin and Jordan (1999) [23]:

> "It is the measurement of the linkages, the arrows in the logic chart, which allows the manager to determine if the program is working. Monitoring the degree to which elements are in place, even the intended and unintended outcomes, will not explain the measurement or tell the manager if the program is working. What is essential is the testing of the program hypotheses. Even if the manager observes that intended outcomes were achieved, the following question must be asked, "What feature(s), if any, of the program contributed to the achievement of intended and unintended outcomes?"

Method for the state of knowledge

Contributing to a growing body of evidence is what makes evaluation analogous with research methodology. This understanding is elucidated by the Method Matching Matrix (MMM) which matches the method(s) chosen to build increasingly mature evidence of this relationship with the condition of knowledge at the time of the choice [32]. The MMM helps evaluators select an appropriate method given the state of understanding about the evaluand. In particular, the degree of understanding about the program's hypothesized causal relationships is essential for selecting the appropriate method. For example, when knowledge is rudimentary, observational studies are an appropriate choice. Whereas randomized control trials work well when there is already deep knowledge and well-established and validated concepts. It is at this point that mistakes are made. Most often, the mistake evaluators make is to select experimental designs or RCTs when that is not an appropriate method for their context. However, and this is crucially important: an RCT based on dubious construct validity is an invalid RCT.

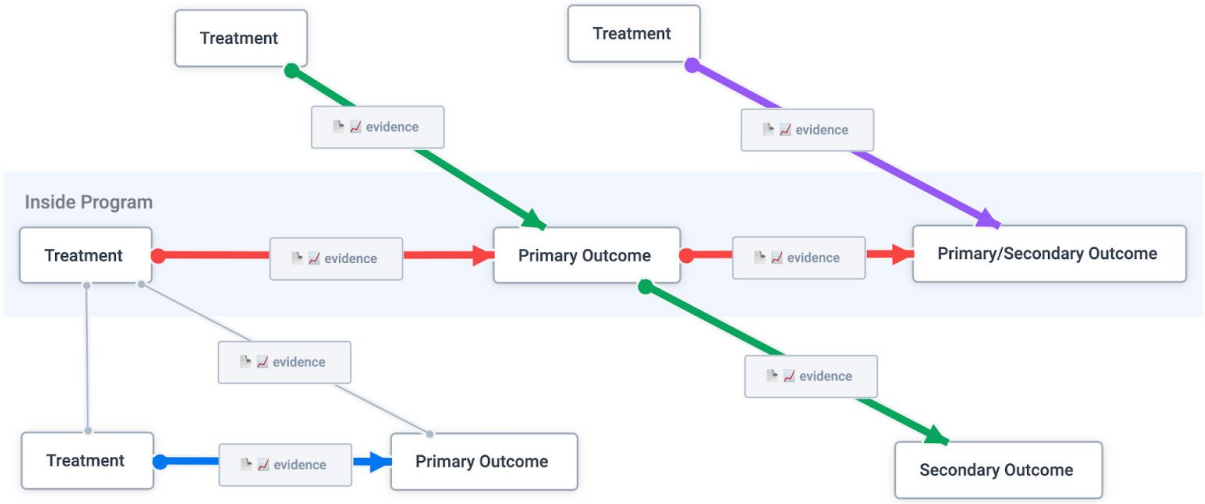

Figure 4. Method Matching Matrix

The model for systems evaluation tells us that a program, and knowledge about the program, will evolve over time if implemented fully. The evolution will be grounded in the "lifespan" (development over time) of the relationship between activities and outcomes (A-R-O) becoming increasingly evidence-based and triangulated. This happens through working up the methodological continuum as the condition of knowledge deepens and becomes more robust.

When the knowledge of the relationship (including the construct validity of the activities and outcomes) are nascent, then the methods one chooses should include observation, case studies, and quasi-experimental design. As the knowledge of the A-R-O relationship develops, one might choose experimental, randomized, or controlled study designs. Based on employing appropriate studies, one can establish greater understanding of the complex constructs at play and eventually develop construct valid concepts for implementation in experimental designs. As knowledge of the A-R-O relationship increases to a mature level, meta-analytical studies can be used. When numerous studies based on numerous methodologies point to the same conclusions, validity increases.

Local-Global and collective impact

We live in a complex and uncertain world. Our programs must be quickly adaptive if they are to stay relevant, and achieve impact efficiently. We need tools and approaches that help us quickly and deeply understand the complexity of what we do, that are simultaneously easy to use. Evaluation exists to build knowledge and the more our models and tools help situate our developing knowledge within a broader system, the 'smarter' and more effective we will collectively be.

The evaluator's perspective is first and foremost, 'the program' s/he evaluates. It can be understood simply as an A-R-O. This equation represents the core structure of a theory of change. The purpose of evaluation is to test that theory, to refine it, and to improve results. The program planner or funder's perspective should be his/her own program but within a larger context of other programs. Those programs could be related in many ways; from the issue being addressed, to the outcomes being sought, or the population the intervention is intended to help. Therefore, those implementing or evaluating programs should seek to understand how their 'local' effort relates to the 'global' context. Or looked at from the

perspective of a complex adaptive system, how the emergent outcomes from many 'local' or individual programs result in collective impact. Taking the 'local-global' perspective prompts the question "how can we more strategically engage to realize collective impact?"

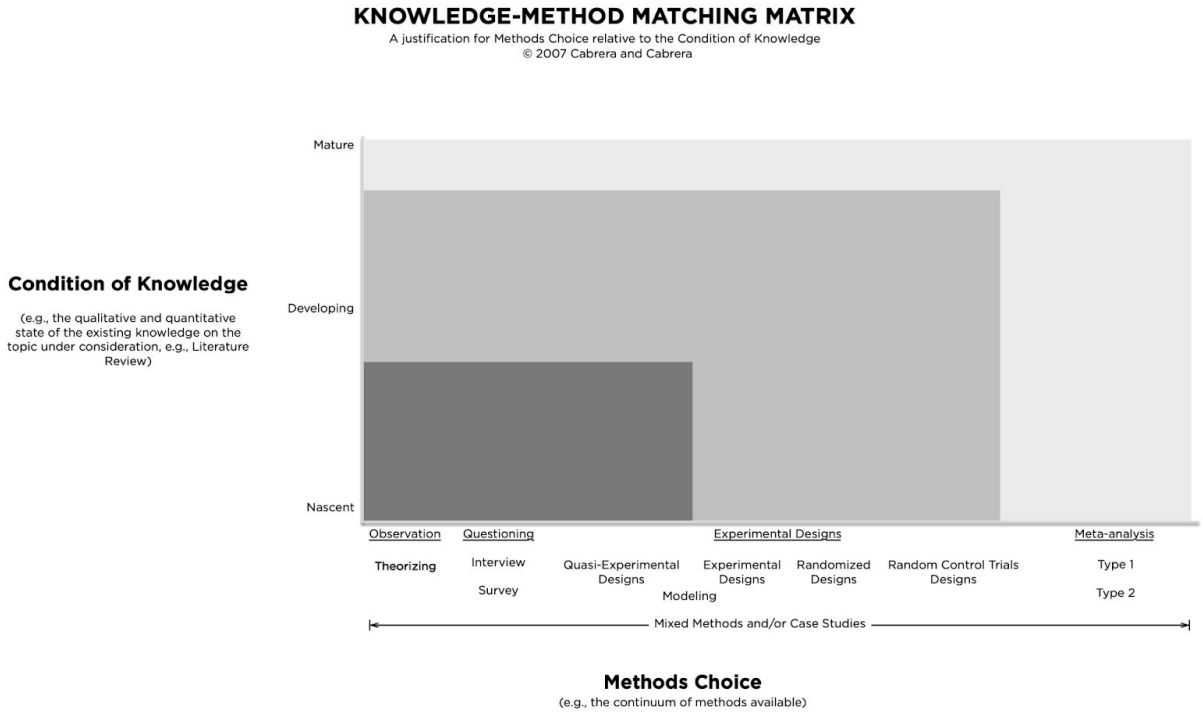

Figure 5. Collective impact

Applying DSRP to evaluation brings focus while enabling leveraging. By mapping a program (here depicted by connected Treatment-Evidence-Outcome), we have a basis to connect our program to others that share elements of the Treatment (Activities), the evidence, and/or the Outcomes (primary or secondary). As seen in this collective impact map, articulating the program boundaries of one clarifies how it is different from, but related to efforts around it. By situating a single, local program inside a larger field of programs, we can see and develop networked pathways, or *netways* [33]. Figure 5 illustrates how any one program can intersect with other programs in a variety of ways. And, thus, taken as a whole, these interconnected programs can show collective impact.

Systems evaluation involves intentionally taking diverse perspectives within a program-evaluation boundary or outside that boundary. As a result, the distinctions of where an evaluator is located become less important. In systems evaluation, an evaluator, regardless of location, would be explicit about taking multiple perspectives as well as examining bias, including confirmation bias when analyzing results. Systems thinking provides the structure to make biases clear, including whose perspectives are centered or excluded.

> "Addressing whose and what interests to privilege should be determined through a process that critically examines alternative options for inclusion while considering implications for who or what may be excluded and/or marginalized by these options. The aim is not for evaluators and evaluations to be impartial or partial towards particular interests, but, instead, to be critically partial—making transparent, questioning, and justifying the selectivity and partiality of an evaluation, and/or whose and what interests are included, excluded, and marginalized." [31]

Evaluation for today's times

Logic models are a popular tool because they help individuals and groups distill and communicate core aspects of a theory of change, which is the backbone of an evaluation. Traditional logic models are problematic because they are insufficient and apply bivalent logic to multivalent realities. Their structure leads us astray in our thinking such that we focus on the wrong things and are inefficient at building or learning from an evidence base of what works. Program planners and evaluators can apply the rules of systems thinking (DSRP) to evaluation to yield more reliable and valid results [33]. At the 'local' individual program level this means evolving your effort based on feedback garnered through methods appropriately matched with the level of knowledge about the evaluand. It also means conceptually connecting the logic of that 'local' individual program to a larger body of knowledge related to the desired outcomes, target population or planned activities.

For many years funding agencies and leaders across sectors and organization types, have expressed the importance of 'evidence-based' programs. They are touted as best practice and considered to be the most reliable and worthy of repute. Yet until now we have been hamstrung in our ability to effectively and efficiently build and *share* evidence about what works and under what conditions. Systems evaluation provides the conceptual basis to start developing tools, platforms and practices to help us as a society do that better. Funders can use systems evaluation to cultivate Mutually Reinforcing Activities [34] amongst programs, and for making strategic decisions based on what they are seeing that depict the larger field of their funding efforts. If evaluation at the program level is about feedback and iteration, that same process is equally relevant at other levels of scale, such as an organization, a consortium, or a nation. As evaluators we should not settle with 'getting by', knowing that so much of the richness and complexity that make our programs important and impactful eludes us. Systems thinking, or DSRP, applied to evaluation helps us strengthen the fidelity of theories of change and their improvement over time.

References

1. Weiss CH. The Politicization of Evaluation Research. J Soc Issues. 1970.
2. Center For Disease Control. Program Evaluation Guide. 16 Apr 2012.
3. Royse D, Staton-Tindall M, Badger K, Matthew Webster J. Needs Assessment. 2009; 1–160.
4. Center for Global Development. When Will We Ever Learn? Improving Lives through Impact Evaluation. 2006.
5. Jane Davidson E. Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation. SAGE; 2005.
6. Hammer J. Randomized control trials for development? Three problems. In: Brookings. 2017.
7. Suter WN. Qualitative data, analysis, and design. Introduction to educational research: A critical thinking approach. 2012;2: 342–386.
8. Morell JA. Evaluation in the Face of Uncertainty: Anticipating Surprise and Responding to the Inevitable. Guilford Press; 2010.
9. United States. Agency for Healthcare Research and Quality, RTI International-University of North

Carolina Evidence-based Practice Center. Community-based Participatory Research: Assessing the Evidence. Summary. 2004.

10. Haynes AS, Derrick GE, Chapman S, Redman S, Hall WD, Gillespie J, et al. From "our world" to the "real world": Exploring the views and behaviour of policy-influential Australian public health researchers. Soc Sci Med. 2011;72: 1047–1055.

11. Knowlton LW, Phillips CC. The Logic Model Guidebook: Better Strategies for Great Results. SAGE; 2012.

12. Wholey JS. Evaluation and Effective Public Management. Little, Brown Boston; 1983.

13. Wholey JS. Evaluability Assessment: Developing Program Theory. New Directions for Program Evaluation. 1987.

14. Provus MM. Discrepancy evaluation for educational program improvement and assessment. McCutchan Berkeley, CA; 1971.

15. Torvatn H. Using Program Theory Models in Evaluation of Industrial Modernization Programs: Three Case Studies. Eval Program Plann. 1999;22: 73–82.

16. Patton MQ. Utilization-Focused Evaluation: The New Century Text. SAGE Publications; 1997.

17. Montague S. The Three Rs of Performance: Core concepts for planning, measurement, and management. Performance Management Network; 1997.

18. McDonald R, Teather G. Science and technology policy evaluation practices in the government of canada. Policy evaluation in innovation and technology: towards best practices. 1997.

19. Schuman SH. Skateboard Injuries in a Campus Community. Clin Pediatr . 1967;6: 252–254.

20. Fitz-Gibbon CT, Morris LL. Theory-based evaluation. Eval Pract. 1996;17: 177–184.

21. Bickman L, Heflinger CA, Pion G, Behar L. Evaluation planning for an innovative children's mental health system. Clin Psychol Rev. 1992;12: 853–865.

22. Julian DA. The Utilization of the Logic Model as a System Level Planning and Evaluation Device. Eval Program Plann. 1997;20: 251–257.

23. McLaughlin JA, Jordan GB. Logic Models: a tool for telling your program's performance story. Eval Program Plann. 1999;22: 65–72.

24. Savaya R, Waysman M. The Logic Model: A tool for incorporating theory in development and evaluation of programs. Adm Soc Work. 2005;29: 85–103.

25. Yampolskaya S, Nesman TM, Hernandez M, Koch D. Using Concept Mapping to Develop a Logic Model and Articulate a Program Theory: A Case Example. American Journal of Evaluation. 2004;25: 191–207.

26. Hernandez M, Hodges S, Cascardi M. The Ecology of Outcomes: System accountability in children's mental health. J Behav Health Serv Res. 1998;25: 136–150.

27. Cabrera D. A Protocol of Systems Evaluation. Cornell University; 2006.

28. Cabrera D, Cabrera L. Systems Thinking Made Simple: New Hope for Solving Wicked Problems in a Complex World. Ithaca: Odyssean Press; 2015.

29. Cabrera D, Colosi L. Distinctions, Systems, Relationships, and Perspectives (DSRP): A Theory of Thinking and of Things. Eval Program Plann. 2008;31: 311–317.

30. Mertens DM, Wilson AT. Program Evaluation Theory and Practice: A Comprehensive Guide. Guilford Press; 2012.

31. Reynolds M, Gates E, Hummelbrunner R, Marra M, Williams B. Towards Systemic Evaluation. Syst Res. 2016;33: 662–673.

32. Cabrera D, Cabrera L. The Knowledge-Method Matching Matrix. Journal of Applied Systems Thinking. 2006;6: 5.

33. Cabrera D, Cabrera L, Lobdell C. Systems Thinking. Journal of Evaluation and Program Planning. 2008;31: 299–310.

34. Kania J, Kramer M. Collective impact. FSG; 2011.