DOI: 10.1002/ev.20443

ORIGINAL ARTICLE

WILEY

The Systems Evaluation Protocol for evaluation planning

Jennifer Brown Urban[1] | **Monica Hargraves**[2] | **Jane Buckley**[3] | **Thomas Archibald**[4] | **Claire Hebbard**[2] | **William M. Trochim**[2]

[1] Montclair State University, Montclair, NJ, USA

[2] Cornell University, Ithaca, NY, USA

[3] JCB Consulting, Rochester, NY, USA

[4] Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

Correspondence
Jennifer Brown Urban, Montclair State University, Montclair, NJ, USA.
Email: urbanj@montclair.edu

Funding information
John Templeton Foundation, Grant/Award Number: 60483; National Science Foundation, Grant/Award Numbers: 0535492, 0814364, 1811214

Abstract

In this paper, we present a comprehensive integrative approach to evaluation planning based on Relational Systems Evaluation (RSE). We describe the major implementation approaches and the tools and resources that support evaluation planning in RSE embodied in the Systems Evaluation Protocol (SEP), a step-by-step guide to the tasks that every evaluation should address. The SEP provides both a conceptual framework for thinking about evaluation and a set of specific methods and tools that expand awareness of the influences of multiple parts of complex systems. The SEP is implemented through a partnership approach, joining evaluators and program professionals in an iterative, phased approach to evaluation planning, which also often emphasizes evaluation capacity building (ECB) as an intentional goal. This paper provides a practice-oriented overview of the specific steps comprising the three stages outlined in the SEP: preparation, model development, and evaluation plan development.

Evaluation is a complex endeavor attempting to address a complex problem. It takes place in "real world" settings where life is subject to sudden changes and unpredictable influences. It deals with "programs" that are multi-faceted, sometimes inherently fluid, diffuse in intention, and difficult to administer with consistency. Evaluation exists in contexts where controls or comparisons are often not feasible or sustainable. It is subject to pressures that can lead to ignoring results in some cases or undermining results in others. Evaluation often occurs in situations where both participants and staff are highly mobile and shift locations over time. It requires perseverance and time commitments that compete with other needs often perceived as higher priority. At the heart of the problem is a theme that unifies all of the factors that impinge on or influence evaluation: the idea of complex systems.

This paper presents a comprehensive integrative approach to evaluation based on Relational Systems Evaluation (RSE). The approach is general and designed to be useful in any evaluation context. We describe the major implementation approaches and the tools and resources that support evaluation planning in RSE. The Systems Evaluation Protocol (SEP; Trochim et al., 2016) is, in the simplest terms, a step-by-step guide to the tasks that every evaluation should address. Yet, to say that this approach is only or even primarily just a step-by-step guide to evaluation would be to mischaracterize both its intent and its manifestation in practice. To be sure, the approach has a step-by-step protocol and describes a detailed series of discrete tasks for accomplishing each step. This sequence of steps provides an orderly and powerful process that helps simplify the many challenges in planning and conducting evaluations, as well as in utilizing results. The steps in the Protocol may mistakenly convey the sense that it is inflexible and static—that each step must be completed in the given order in a specific time frame. On the contrary, the SEP is intended to be a flexible and adaptive protocol that is contextually applied. While all evaluations should address the issues raised in the SEP, different contexts may look very different in terms of implementation. At its core, the SEP recognizes no program exists in isolation; rather, all programs are part of many integrated systems. The SEP provides both a conceptual framework for thinking about evaluation and a set of specific methods and tools that expand awareness of the influences of multiple parts of complex systems, thereby enhancing one's ability to accomplish high-quality evaluation with integration across organizational levels and structures.

The SEP incorporates the RSE components of Evolutionary Evaluation principles, a systems perspective, and a partnership model to consider the complex factors that are inherent in the larger systems within which a program is embedded (Trochim et al., 2016; Urban & Trochim, 2009; Urban, Hargraves, & Trochim, 2014; Urban, Hargraves, Hebbard, Burgermaster, & Trochim, 2011). The SEP emphasizes: the importance of creating a causal diagram that illustrates the programmatic theory of change; incorporating the perspectives of both internal and external stakeholders of the program; recognizing how the program is related to other programs either in the same system or other systems, in part by identifying research on similar or related outcomes that can help link the program to more universal long-term goals; developing and using a comprehensive evaluation plan; and continually assessing and revising the theory of change and evaluation plan based on knowledge gained through evaluation efforts (Urban & Trochim, 2009; Urban et al., 2014). The SEP addresses the three primary phases of evaluation: Planning, Implementation, and Utilization. Each phase is comprised of several stages, each of which includes a detailed list of components or steps. This paper emphasizes the SEP's first phase, Evaluation Planning, which has three stages: (1) Preparation, (2) Modeling, and (3) Evaluation Plan Development. The complete summary of the steps in the three planning stages is presented in Figure 1. Although the foundations of the approach are complex and rooted in theories of developmental systems (e.g., Lerner, 2006; Overton, 2006, 2010), evolution (Darwin, 1859; Mayr, 2001), evolutionary epistemology (Bradie & Harms, 2006; Campbell, 1974, 1988; Cziko & Campbell, 1990; Popper, 1973, 1984), and systems thinking (Bertalanffy, 1995; Laszlo, 1996; Midgley, 2003; Ragsdell, West, & Wilby, 2002), the process of planning a high-quality evaluation is manageable, replicable, and teachable.

PREPARATION STAGE

The primary objectives of the Preparation stage are to establish a partnership between the evaluators (whether internal or external) and key program staff, acquaint the partnership

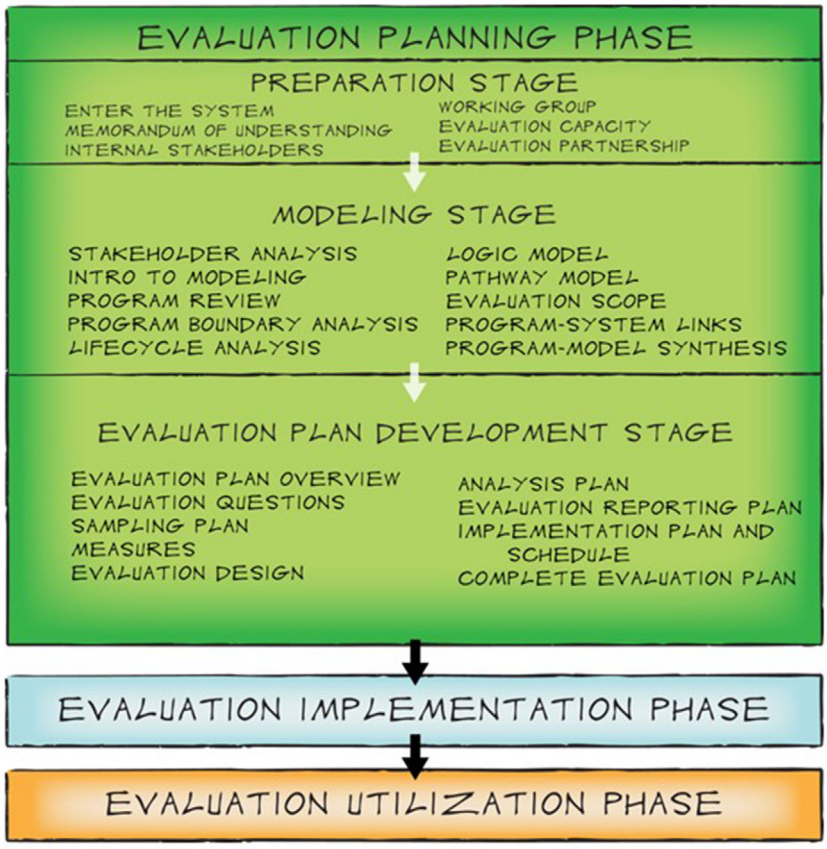

FIGURE 1 The phases and steps in the Systems Evaluation Protocol

team with the SEP process, build a shared understanding of the nature and vision of this approach to evaluation and program development, identify people's roles in the evaluation process, and collect basic information about the program. The relationships and expectations established during this stage are critical to the success of the partnership as well as the outputs and outcomes of the SEP process. Both evaluation and program professionals frequently approach evaluation work with deeply held assumptions and values related to evaluation and program development as well as the context and systems surrounding a program (Mertens & Wilson, 2019). Attending to these factors early in the process allows all members of the partnership to establish common ground as well as individual ownership in the collaborative process. Lack of attention to this stage could lead to confusion and conflicting priorities later on. Therefore, careful thought should be given to selecting who will be included in the partnership team. In addition to the evaluator, the partnership team should have between two and five representatives from the focal program/organization, representing a diversity of perspectives on the program.

In RSE, partnerships are integral and Evaluation Capacity Building (ECB) is often also emphasized. Evaluators facilitate the implementation of the SEP and provide knowledge of research methods and the research literature. In contrast, practitioners provide critical programmatic, organizational, and contextual expertise and can serve as Evaluation Champions within their organizations. During the Preparation stage, the partnership team members begin to build shared understanding and expectations. They gain common perspec-

tive on the program and organization, as well as a grounding in a systems approach by discussing (and perhaps documenting in a Memorandum of Understanding, or MOU) roles, timelines, and expectations; in addition, the members engage in system mapping activities. System Mapping often occurs prior to or at the first in-person meeting, based on an organization diagram—an illustration of the structure and relationships within the organizational system. With the System Map as a base, SEP facilitators may work with program partners to expand or layer additional system components onto the diagram including: external stakeholders, ecosystem features, internal communication patterns, and organizational knowledge flow.

As mentioned above, because RSE emphasizes a partnership model, ECB is also often an intentional goal. Therefore, to establish the organization's current level of evaluation capacity and to be able to track subsequent increases in capacity, baseline data and a needs assessment may also be conducted during the Preparation stage. While each of these exercises has value for any program, responsive and strategic use of one or more of these activities can help the partnership team establish system awareness, including the strengths and challenges they may face when planning, implementing, and utilizing evaluation.

MODEL DEVELOPMENT STAGE

The Model Development stage is a central and distinguishing component of the SEP, focused on iteratively surfacing and articulating a deeper understanding of the program through: Stakeholder Analysis (creating and making meaning of a visual depiction of the stakeholders and their relationship to each other); Program Review and Boundary Analysis including group discussion and development of a written program description; identification of program and evaluation lifecycle phases; and structured Program Modeling in two forms: the familiar columnar logic model (with Program Assumptions, Context, Inputs, Activities, Outputs, and Outcomes), and a corresponding visual pathway model (depicting the theory of change, or how the program is thought to work). Each of these steps is described in greater detail below.

STAKEHOLDER ANALYSIS

The goal of Stakeholder Analysis is to identify all potential people and/or groups that have a stake in the program and its evaluation and to begin to understand their perspectives on the program. This activity may be done on a sheet of paper, a whiteboard, a sticky wall, or poster paper, as long as it can either be saved or photographed for later reference. First, an initial list of program stakeholders is generated. It is a good idea to use a moveable object—such as a sticky note or repositionable card—for each stakeholder because the stakeholder names will likely be shifted around the map as the relationships start to take form. The program name is placed in the center of a set of concentric circles that form the basis for the "map," and the stakeholders are placed in positions around the program reflective of their relationships to each other and to the program. Stakeholders with a comparatively more vested interest or direct voice in the program and participant outcomes (such as the participants and their families, program staff, the local community and its organizations) will be settled closer to the program at the center. Similarly, the most distal stakeholders, in the outer ring of the concentric circles, are those who may be indirectly impacted by or have an indirect impact on the program. These stakeholders may not necessarily be aware of the specific program (likely to be national and global organizations and interest groups,

government agencies, and broader funding organizations). The remaining stakeholders (perhaps including similar or related programs and organizations, state and regional audiences, future educators, and employers of the participants, researchers in related fields, and immediate funders) will be placed somewhere in-between according to their relationship to the program. The relationships among stakeholders may be reflected in the appearance of "spokes" or "slices" emanating from the program in the center. Stakeholders within each slice are reflective of stakeholders' relationships with each other. For example, each slice may represent a specific sector, such as the community, topical interest groups, education, government, or funders.

The ultimate goal is for the map to be meaningful and informative for both modeling the theory of change and identifying evaluation questions. The brainstorming and discussion of the stakeholder map layout help bring otherwise overlooked perspectives into awareness and consideration. The partnership team should consider (and record for later use) each stakeholder's (or each sector's) perspective on the program, how they would describe it, and why it is important. Often, members of the partnership team realize the program's purpose may be seen very differently, even among program staff. By generating this simple social network analysis of the program, the stakeholder map activity encourages evaluation partnership members to view the program as part of larger programmatic and political systems, sharpens their understanding of the different perspectives stakeholders have on the program and its evaluation, and helps them appreciate the program from these different perspectives.

PROGRAM REVIEW AND BOUNDARY ANALYSIS

The goal of the Program Review step is to gain a firmer understanding of the components and characteristics of the program and its parent organization, including how the program operates and whom it serves (Trochim et al., 2016). In one sense, this step mirrors what any evaluation consultant would do in the early stages of an evaluation as they seek to understand the program contours. Yet, influenced by systems thinking, the approach outlined in the SEP emphasizes the importance of relationships, multiple perspectives, and boundaries when describing a program.

To facilitate this step, the evaluator can lead a guided brainstorm session or distribute a worksheet with guiding questions so partnership team members have time to reflect on their program before hearing from the rest of the team. Guiding questions include (but are not limited to): Why does the program exist? What is the program's mission or vision? What are the program's activities? Who participates in these activities? When does the program take place? Where does the program take place? Ideally, a good program description provides insights on program implementation (e.g., information about the target audience, program scale, activities), program context (e.g., information about the social, cultural, physical context in which the program takes place), intended outcomes or goals, and even program background (i.e., information on the history of the program's development and references to relevant research). Usually, the program review activity yields a written description, but could alternatively be represented in artistic or graphical forms.

Program Review is closely related to the next activity, Boundary Analysis. Boundary Analysis is a clear operationalization of systems thinking in evaluation, drawing on the notion of part-whole and nested relationships between programmatic and organizational units (as introduced in Trochim and Urban, this volume). The goal of this step is to determine what is "in" and what is "outside" the program. While this may seem simple or unnecessary at

first glance, the boundaries of a given program are seldom made explicit. They frequently vary with context as well as various stakeholder perspectives. When facilitating this activity, an evaluator could use the deceptively simple approach of drawing a vertical line down the middle of a piece of flip chart paper, and asking team members to brainstorm which program elements are in (on the left) and which are out (on the right). Some guiding questions that help people complete this step include: What is the essence of the program (in terms of activities and outcomes)? What is not? Who are the participants? Who are not? How is the program being delivered, and is that an essential part of the program? How many and which of the possible activities must be included to have a full view of the program? How many and which of the outcomes must be included to get a full view of this program? What scale and scope are accurate?

The metaphor of a microscope is apt here, as the evaluator helps program implementers zoom in and out on their program until they arrive at just the right level of magnification. This is akin to Schwandt's (2018) use of concepts from critical systems heuristics to frame evaluative thinking as a social, participatory process of boundary judgement making. Taken together, these two SEP activities—Program Review and Boundary Analysis—are integral to helping program professionals move toward creating a theory of change Pathway Model diagram.

LIFECYCLE ANALYSIS

In addition to thinking about program boundaries, it is important to think about the lifecycle phase of the program and its evaluation. Evolutionary Evaluation recognizes program development and evaluation as evolutionary processes with inherent lifecycle qualities (Trochim et al., 2016; Urban et al., 2014). Program designs may originate in research or in practice, but in either case, a program must be initiated and piloted in some context for the first time. From those beginnings, a program typically goes through a period of rapid change and growth as it is refined; it may then stabilize and be expanded; it may become formalized and disseminated widely, and at any point along the way it may be retired or revised substantially in light of changing circumstances or needs. This process can usefully be delineated in terms of four program lifecycle phases: Phase 1 Initiation; Phase 2 Development; Phase 3 Stability; and Phase 4 Dissemination. At each of these phases, the kind of information most useful to the program's development is different—in the initiation phase, when the program is changing rapidly, it is most useful to have quick feedback on program implementation, participant and facilitator satisfaction, and so on. Once any early roughness in program design is smoothed out and the program becomes more stable from one iteration to the next, it is most useful to explore and test the relationship between the program and its intended outcomes so that the program's effectiveness can be improved. At any phase, appropriate evaluation can contribute to the state of knowledge about a program. Evaluation itself can also be usefully characterized as proceeding through a set of four phases: Phase 1 Process and Response; Phase 2 Change; Phase 3 Comparison and Control; and Phase 4 Generalizability (see Figure 2 for detailed definitions of the program and evaluation lifecycle phases).

As we introduced in Trochim and Urban (this volume), Evolutionary Evaluation highlights the importance of alignment between program and evaluation lifecycles. Alignment ensures that the program evaluation yields the information the program needs most and that both program and evaluation resources are being put to their best use. Misalignment can be costly in several ways. For example, if an early-phase, rapidly evolving program is subjected to an inappropriately advanced phase of evaluation (i.e., 'premature experimen-

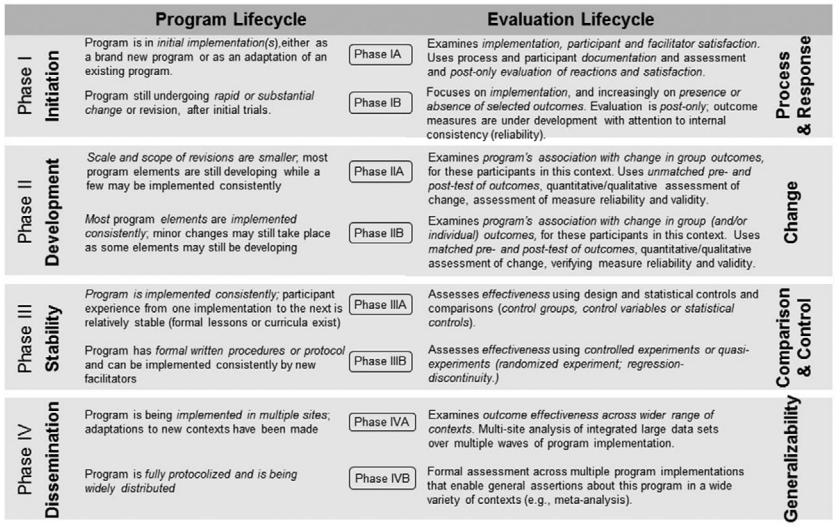

FIGURE 2 Program and evaluation lifecycle definitions

tation'), there is a greater risk that decisions to affirm or reject the program may be based on short-lived false positive or negative results. Similarly, if a widely implemented, stable program is still only being evaluated using simple satisfaction surveys, there is a risk that an ineffective (but popular) program may be extended when in fact it should be retired or significantly revised, or that an effective program may be undervalued and underutilized (see Urban et al., 2014 for a more extensive discussion of the costs of misalignment).

Given the importance of program and evaluation lifecycles for guiding the choice of evaluation methodology, a critical step in the Systems Evaluation Protocol is a Lifecycle Analysis. Assessing the program's lifecycle phase, the evaluation's lifecycle phase, and their alignment or misalignment, provides critical guidance for what kind of evaluation should be undertaken in the next evaluation cycle. Aligned evaluation should be the gold standard as it will provide the most useful data in the most efficient manner. If the program and evaluation lifecycles are not aligned, effort should be made to align them over successive evaluation cycles, and evaluation results should be interpreted with caution so as not to put excessive weight on findings that may be subject to correction as the program evolves. The evaluation partnership team should determine both the current program and evaluation lifecycle phases (based on any prior evaluation work) using the definitions provided in Figure 2.

LOGIC AND PATHWAY MODELS

A pathway model is a visual program model that utilizes components of the columnar logic model (activities and outcomes) but also specifies the theory of change underlying the program. Figure 3 provides an example of a pathway model. The notion of hypothesized causal relationships is critically important in RSE. Note that here we are using 'causal' in the more vernacular sense, suggesting that one element *contributes* to the other, not that it necessarily *causes* it in the positivist empirical sense of the term. In a program logic model, this general idea of causality is in the background—activities are expected to lead to short-, medium-, and long-term outcomes and ultimately to impacts—but standard logic mod-

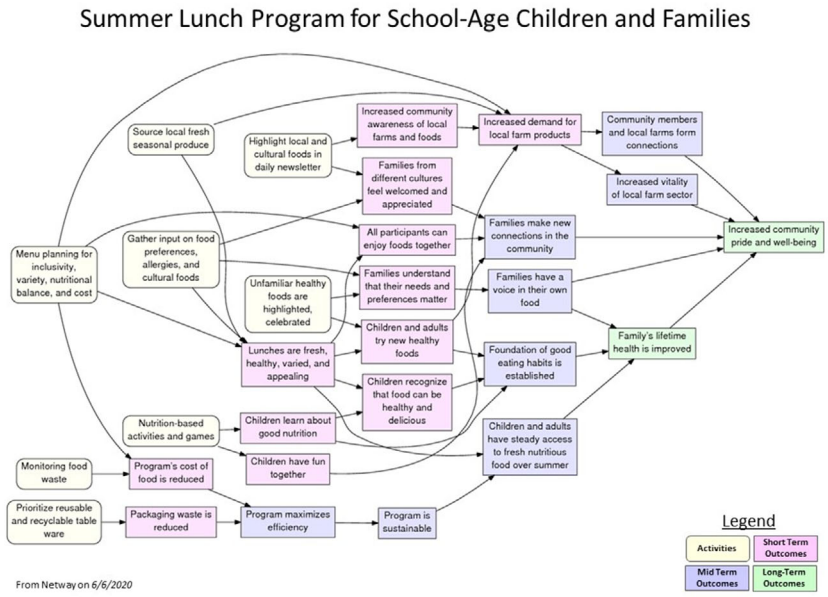

FIGURE 3 Example of a pathway model for a hypothetical program

els do not spell out the specific cause-effect relationships. A pathway model illustrates this assumed causation, incorporating the logical connections believed to link an activity to one or more short-term outcomes, and from there to other short- or medium-term outcomes, and on to long-term outcomes. It tells the story of how a program works.

Illustrating thinking is a powerful tool (Brookfield, 2012; Buckley, Archibald, Hargraves, & Trochim, 2015). The knowledge or insight expressed via the pathway model is often held subconsciously or tacitly by people involved with the program; the opportunity to articulate these insights and combine the insights of a number of key people is quite rare and powerful. The pathway modeling process is an effective method for establishing a common way of understanding and communicating about how and why a program "works." Moreover, the process of working through the Model Development stage of the SEP (particularly when done within a partnership) builds evaluation capacity, deepens staff understanding of the program and ability to communicate with stakeholders, and cultivates evaluative thinking (see Buckley et al., 2015; Chauveron et al., this volume).

The pathway model building process is described in more detail elsewhere (Archibald, Sharrock, Buckley, & Cook, 2016). In brief, it consists of having the partnership team brainstorm program activities, outputs, and short-, medium-, and long-term outcomes and writing them on sticky notes or note cards, with just one element per note. Many people begin on what will become the left side of the model, with program activities, while others prefer starting with the end in mind, focusing on what will become the right side of the model where the long-term outcomes of the program will be posted. Then the partnership team positions and configures the elements either up on a wall (i.e., with flip chart paper taped together to make a large canvas, or a sticky wall, whiteboard, etc.) or on a paper-covered tabletop. After some trial and error, and much reflective dialogue, the team can then pencil in the arrows representing hypothetical causal linkages. An interesting phenomenon that almost always occurs, which must be addressed with care by the facilitator, is the tendency for some people to split activities or outcomes into their smallest units. In contrast, others tend to lump elements together into larger, more abstract units. Like with the Goldilocks

metaphor, there is a sweet spot somewhere in the middle, and the facilitator must help members of the partnership team to find it; otherwise, the model will become less legible, less useful, or both. Another common challenge for people building models for the first time comes from our ingrained tendency to think chronologically. It is important to keep in mind that the arrows in the pathway model represent causal and not simply sequential connections. After a draft model is created on the wall or tabletop, the model components can be entered into the Netway software[1]; a clean version can then be printed for further review and use.

Pathway Modeling is frequently cited as the most valuable aspect of the SEP, and the pathway model as the most valuable output (see Hargraves et al., this volume). Program professionals frequently use the pathway model for multiple purposes, including for communication with stakeholders and in funding proposals. They also frequently use Pathway Modeling again (often without a facilitator) with another program or project (see our discussion of the notions of 'resonance' and 'stickiness' in Hargraves et al., this volume). The value of Pathway Modeling, when done in conjunction with the other SEP activities, is hard to overstate. All of the SEP steps leading up to pathway model development contribute to the accuracy and quality of the model in some way, and all of the steps following pathway model development benefit from an accurate and high-quality model. Evaluation partnership teams often experience "aha moments" as part of the model building process that are sometimes immediately actionable, making them both an evaluation planning tool as well as a kind of informal evaluation result (Urban, Linver, Thompson, Davidson, & Lorimer, 2018).

EVALUATION SCOPE

Evaluation seldom occurs as a one-off endeavor; typically, evaluation takes place in multiple stages or waves over the life of a program. In any given wave or evaluation cycle (e.g., annual evaluation), a program usually only has the resources to focus on a subset of the activities and outcomes on a logic or pathway model. Therefore, the partnership team will need to select the activities or outcomes that are most important, relevant, or feasible for the upcoming evaluation cycle. That is, the partnership team must determine what the "scope" of the evaluation will be at any given point in time and on which elements of the program to focus. Typically, the scope of earlier lifecycle phase evaluations focuses on assessing program activities (e.g., implementation and process), outputs, and perhaps a few key short-term outcomes. Over time the scope tends to move to the right on a pathway model, encompassing later short-, medium-, and sometimes long-term outcomes.

This step in the Protocol represents a change in focus for the planning of an evaluation. Prior to this, the evaluation partnership team has considered stakeholders, program components and boundaries, lifecycle stages, and the components of the logic and pathway models. Thinking about scope combines and builds on these prior steps to make strategic decisions about what aspects of the program to focus on during the current evaluation cycle. Feasibility requires consideration of program activity, time constraints, funding constraints, and/or reporting requirements. In addition, evaluation time and resources are typically limited, so it is essential to direct the evaluation efforts toward generating information about the program that is useful and credible to both internal and external stakeholders.

[1] The Netway is an online tool that includes tools for developing pathway models. It can currently be accessed free of charge at www.evaluationnetway.com.

The process of determining scope highlights one of the particular strengths of the pathway model, namely that it provides a visual representation onto which the partnership team can "layer" the various considerations of stakeholder priorities, lifecycle considerations, and insights from the theory of change itself. This process is referred to as "Mining the Model." To complete this step, the evaluation partnership team should begin by revisiting the stakeholder map, and identifying the key stakeholders for the current evaluation cycle—asking, who needs to know what, at this time, to inform important decisions? These key stakeholders could be funders, collaborators, organizational decision-makers, or others. With a shortlist of 3–5 key stakeholders in hand, the partnership team should mark the outcomes or activities on the pathway model that each of these stakeholders is most interested in, using stickers or highlighters.

Similarly, the team should then mark outcomes or activities that the program team identifies as internal priorities for evaluation. The next consideration is the program logic itself, made visible in the model. The team marks outcomes that have the most numbers of arrows going into them ("key destinations"), outcomes or activities that have a large number of arrows going out ("gateways"), and those that have both – a large number of arrows in and out ("key hubs"). Although the number of arrows alone does not necessarily indicate importance, the visual cues from the model direct our attention to outcomes that may be particularly useful to measure either because a lot of effort is going into achieving them, or because they are important gateways for subsequent outcomes. The partnership team should also mark "key links"—arrows representing particularly important causal connections. Once these key outcomes and key links are marked, the partnership team can move to identify the "key pathways"—a small number of meaningful throughlines from activities through these key nodes and links all the way out to the ultimate long-term outcomes. These few key pathways are then given descriptive labels. They reflect the core strategies inherent in the program's theory of change.

With all these considerations mapped onto the visual model, the partnership team can then step back—literally or metaphorically—and use this layered representation to review and balance strategic considerations for focusing the current evaluation. Often there are clear visual cues about what a program should focus on. However, the areas of the model that "jump out" visually should be considered further before making a final determination about scope. Are there outcomes or key pathways that multiple stakeholders are interested in? Are there any outcomes or pathways that none of the stakeholders have an interest in? Do certain stakeholders' perspectives get precedence over others? Is the program's lifecycle phase aligned with the emerging evaluation interests? Finally, there are important practical considerations that must be taken into account before finalizing the scope: are there particular activities, outcomes, or pathways on which the program is required to report? Those need to be within the chosen scope. Preexisting or low-cost data collection activities should also be acknowledged when choosing the scope if resources are limited, and the information would be valuable. Once these individual components and questions have been considered, the partnership team can determine the scope for the evaluation. The evaluation should be constructed to include as many of the identified key elements (activities, outputs, or outcomes) as can reasonably be accommodated within the next evaluation cycle. It is often useful to show the scope of an evaluation graphically by drawing a line that encloses the portions of the pathway model that will be included in the scope of the evaluation.

PROGRAM-SYSTEM LINKS

Once the pathway model is complete, it can ground additional analyses that facilitate links to the research literature, already existing measurement tools, and program resources (e.g.,

curricula, handbooks). This part of the SEP allows users to take a step back and explore the broader system within which the program is situated. It can also provide insights that guide evaluation design and implementation. There are four primary steps in Program-System Links: 1) Model Validation, 2) Evidence Mapping, 3) Measurement Mapping, and 4) Practice Mapping. Any one of these steps can be used in isolation. Depending on resources, SEP users may choose to do all, some, or none of the steps in this section of the SEP.

Model Validation

As described above, the evaluation partnership team typically consists of an evaluator who serves as the facilitator and a small team of program professionals. Although stakeholders can be brought in to provide feedback on the model during its development, it can be helpful to have a more systematic review of the pathway model by a larger number of key stakeholders who can offer alternative perspectives. One way to do this is to conduct focus groups with key stakeholders. When facilitating focus groups, the evaluator first orients the participants to the pathway model by explaining that they are designed to show a process of change and clarify that they are not flow charts demonstrating a sequence based on chronological time. The evaluator should also identify the nodes in the model (activities, short-, medium-, and long-term outcomes) and explain that the arrows show the logical connections that lead from an activity to one or more short-term outcomes that activity contributes to, and from there to other short-, medium-, and ultimately long-term outcomes. It is important to emphasize that the goal of the focus group session is to understand whether the model makes sense to participants, whether anything is missing or should be removed, and what they view as the most important nodes and links. The results of the focus group(s) should be used to make edits to the pathway model.

Evidence Mapping

In addition to validating the pathway model with program stakeholders, Evidence Mapping validates the model against the existing literature. The primary goal of Evidence Mapping is to search the research literature to determine if there is support for the links in the pathway model. Some evaluators may choose to focus only on the links along the key pathways identified in "Mining the Model." Evaluators should keep a careful log of the evidence they find and what it suggests about the corresponding link in the model (i.e., there is support, there is no support, or the link should be edited). The results of Evidence Mapping should be used to make edits to the model. The revised pathway model can then be annotated with symbols or icons to mark links that are supported in the existing research literature. Evidence Mapping can also reveal where there are gaps in the research literature that the evaluation could address, thus informing decisions about evaluation scope. One of the benefits of Evidence Mapping is that it provides additional support for the logic of the connections laid out in the pathway model. Although the evaluation partnership team may not have the time or funds to collect data on medium- or long-term outcomes, the research literature can pick up where the evaluation efforts end and make a strong case for why one can expect to ultimately achieve longer-term outcomes (Urban & Trochim, 2009).

Measurement Mapping

In Measurement Mapping, the evaluation partnership team turns again to the literature to determine whether there are any existing tools that can be used or adapted for the evalua-

tion. Once measures have been selected, icons can again be added to the pathway model to indicate which constructs will be measured in the evaluation. It is helpful to use different icons that correspond to different measurement strategies. For example, an image of a telephone can indicate constructs measured with an interview, while an image of a checkmark can indicate constructs measured with a survey. Mapping the measures onto the pathway model illustrates where there is measurement triangulation and where there are potential gaps.

Practice Mapping

The goal of Practice Mapping is to determine whether there is consistency between the pathway model and program artifacts such as curricula, handbooks, or training guides. For example, if a program outcome is to increase youth employment, is there evidence in program materials that youth employment is addressed? To conduct Practice Mapping, evaluators (or the partnership team) conduct a review or audit of programmatic activities as well as the pathway model nodes that are addressed. A priori thematic coding can be used to determine which nodes are (or are not) addressed and to what extent. Gaps in the model should be noted, and the model and/or the program should be revised accordingly.

EVALUATION PLAN DEVELOPMENT

Once the program pathway model has been refined to the extent possible given project needs and resources, and the evaluation scope has been identified, it is time to develop an evaluation plan. The steps in this stage will be familiar to most evaluators. The products of the Modeling stage—the stakeholder map, lifecycle determinations, and the logic and pathway models—form the foundation for strategic decision-making about evaluation scope, evaluation purpose, and ultimately the specific evaluation questions and other components of the evaluation plan. An evaluation purpose statement explains the reason for conducting the evaluation, including goals and boundaries. The evaluation scope that emerged from the "Mining the Model" step described above will guide the purpose statement by clarifying the stakeholder, lifecycle, and theory of change considerations that inform the evaluation priorities. Within this scope, and after reviewing resources and feasibility, evaluation questions should be articulated carefully as they will direct measurement, sampling, design, and analysis decisions. When finalizing evaluation questions, particularly those relating to program outcomes, it can be helpful to refer to the program pathway model. The full evaluation plan will include sections covering: measures, sampling, design, data management and analysis, a reporting plan (informed by the stakeholder analysis conducted earlier), and an implementation plan and schedule.

DISCUSSION

The evaluation plan developed in the third and final stage of the SEP planning phase integrates Evolutionary Evaluation (through attention to program and evaluation lifecycle alignment) with related systems considerations and insights, to yield a methodologically rigorous evaluation that will serve the program's practical, strategic, and developmental needs.

An inherent benefit of working through the SEP is that the process builds evaluation capacity both in terms of specific skills and knowledge (e.g., modeling, lifecycle analysis, evaluation methodology) and in terms of instilling skills, patterns of thought, and commitment to evaluation that constitute evaluative thinking. By design, RSE incorporates many of the strategies that specifically promote Evaluative Thinking (Buckley et al., 2015).

The integration of systems evaluation approaches within a systematic and repeatable evaluation protocol offers the promise of improving practical evaluation work while incorporating the best that evaluation theory has to offer. The Systems Evaluation Protocol is a practical tool that helps evaluators and organizations move through evaluation in an organized way while allowing for flexibility and adaptability to local circumstances. It can be used to guide the development of a single one-off evaluation study. However, because the SEP is general and applicable to virtually any setting, it is especially useful when a consistent approach to evaluation is required for a system of multiple organizations or programs.

ACKNOWLEDGMENTS

This work supported in part by the John Templeton Foundation, grant # 60483, and the National Science Foundation, grant #s 0535492, 0814364, and 1811214.

ORCID

Jennifer Brown Urban https://orcid.org/0000-0001-8405-3078
Monica Hargraves https://orcid.org/0000-0002-6074-4445
Jane Buckley https://orcid.org/0000-0003-4206-2341
Thomas Archibald https://orcid.org/0000-0002-3567-9143
Claire Hebbard https://orcid.org/0000-0001-6429-7788
William M. Trochim https://orcid.org/0000-0003-0369-2922

REFERENCES

Archibald, T., Sharrock, G., Buckley, J., & Cook, N. (2016). Assumptions, conjectures, and other miracles: The application of evaluative thinking to theory of change models in community development. *Evaluation and Program Planning, 59*, 119–127. https://doi.org/10.1016/j.evalprogplan.2016.05.015

Bertalanffy, L. V. (1995). *General system theory: Foundations, development, applications* (Revised ed.). New York: Braziller.

Bradie, M., & Harms, W. (2006). Evolutionary epistemology. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2016 ed). Stanford, CA: The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University.

Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation, 36*(3), 375–388. https://doi.org/10.1177/1098214015581706

Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (pp. 413–463). LaSalle, IL: Open Court.

Campbell, D. T. (1988). Evolutionary epistemology. In E. S. Overman (Ed.), *Methodology and epistemology for social science: Selected papers of Donald T. Campbell*. Chicago, IL: University of Chicago Press.

Cziko, G. A., & Campbell, D. T. (1990). Comprehensive evolutionary epistemology and bibliography. *Journal of Social and Biological Structures, 13*(1), 41–82. https://doi.org/10.1016/0140-1750(90)90033-3

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.

Laszlo, E. (1996). *The systems view of the world: A holistic vision for our time*. Creskill, NJ: Hampton Press.

Lerner, R. M. (2006). Developmental science, developmental systems, and contemporary theories of human development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (6th ed., Vol. *1*, pp. 1–17). Hoboken, NJ: John Wiley & Sons.

Mayr, E. (2001). *What evolution is*. New York: Basic Books.

Mertens, D. M., & Wilson, A. T. (2019). *Program evaluation theory and practice: A comprehensive guide* (2nd ed.). New York: Guilford Press.

Midgley, G. (2003). *Systems thinking*. Thousand Oaks, CA: Sage.

Overton, W. F. (2006). Developmental psychology: Philosophy, concepts, methodology. In R. M. Lerner (Ed.), *Theoretical models of human development* (6th ed., Vol. *1*, pp. 18–88). Hoboken, NJ: Wiley.

Overton, W. F. (2010). Life-span development: Concepts and issues. In W. F. Overton & R. M. Lerner (Eds.), *Handbook of life-span development: Cognition, biology, and methods* (Vol. *1*, pp. 1–29). Hoboken, NJ: Wiley.

Popper, K. (1973). Evolutionary epistemology. Paper presented at the Sections I-VI of "The rationality of scientific revolutions" given at the Herbert Spencer lecture, University of Oxford, Oxford.

Popper, K. (1984). Evolutionary epistemology. In D. M. Miller (Ed.), *Popper selections* (pp. 78–86). Princeton, NJ: Princeton University Press.

Ragsdell, G., West, D., & Wilby, J. (2002). *Systems theory and practice in the knowledge age*. New York: Kluwer Academic/Plenum.

Schwandt, T. A. (2018). Evaluative thinking as a collaborative social practice: The case of boundary judgment making. In A. T. Vo & T. Archibald (Eds.), *New Directions for Evaluation: No. 158. Evaluative thinking* (pp. 125–137). San Francisco, CA: Jossey-Bass.

Trochim, W.M., Urban, J. B., Hargraves, M., Hebbard, C., Buckley, J., Archibald, T., … Burgermaster, M. (2016). *The guide to the Systems Evaluation Protocol (V 3.1)*. Ithaca, NY: Cornell Digital Print Services. Retrieved from https://core.human.cornell.edu/research/systems/protocol/index.cfm

Urban, J. B., Hargraves, M., Hebbard, C., Burgermaster, M., & Trochim, W. M. (2011). Evaluation in the context of lifecycles: 'A place for everything, everything in its place'. Paper presented at the annual meeting of the American Evaluation Association, Anaheim, CA.

Urban, J. B., Hargraves, M., & Trochim, W. M. (2014). Evolutionary Evaluation: Implications for evaluators, researchers, practitioners, funders, and the evidence-based program mandate. *Evaluation and Program Planning*, *45*, 127–139. https://doi.org/10.1016/j.evalprogplan.2014.03.011

Urban, J. B., Linver, M. R., Thompson, J., Davidson, R., & Lorimer, D. (2018). Evaluating youth character education using evolutionary evaluation and the Systems Evaluation Protocol. *Applied Developmental Science*, *22*(4), 245–257. https://doi.org/10.1080/10888691.2017.1285236

Urban, J. B., & Trochim, W. M. (2009). The role of evaluation in research-practice integration: Working toward the "golden spike". *American Journal of Evaluation*, *30*(4), 538–553. https://doi.org/10.1177/1098214009348327

AUTHOR BIOGRAPHIES

JENNIFER BROWN URBAN is professor of Family Science & Human Development, and co-director of the Institute for Research on Youth Thriving and Evaluation, both at Montclair State University. Her research focuses on positive youth development and Evaluation Capacity Building in youth-serving organizations.

MONICA HARGRAVES is the Associate Director for Evaluation Partnerships in the Cornell Office for Research on Evaluation, at Cornell University, where her work focuses on a collaborative approach to evaluation and Evaluation Capacity Building using the tools of Relational Systems Evaluation.

JANE BUCKLEY of JCB Consulting is an evaluation planning facilitator and Evaluation Capacity Building consultant whose focus is on fostering a culture of Evaluative Thinking in partner organizations.

THOMAS ARCHIBALD is an associate professor and extension specialist in Virginia Tech's Department of Agricultural, Leadership, and Community Education, where his research and practice focus primarily on Evaluation Capacity Building and Evaluative Thinking in community development contexts.

CLAIRE HEBBARD is the Project Coordinator at the Cornell Office for Research on Evaluation (CORE), where she collaborates on research on evaluation methods, assists in facili-

tating Evaluation Capacity Building, and helps implement evaluations in small outreach programs.

William Trochim is professor of Policy Analysis and Management at Cornell University who specializes in research methodology and program evaluation.

How to cite this article: Urban, J. B., Hargraves, M., Buckley, J., Archibald, T., Hebbard, C., & Trochim, W. M. (2021). The Systems Evaluation Protocol for evaluation planning. *New Directions for Evaluation*, 1–15. https://doi.org/10.1002/ev.20443