Evaluation and Program Planning 34 (2011) 135–146

Contents lists available at ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

Ten steps to making evaluation matter

Sanjeev Sridharan [a,b,*], April Nakaima [c]

[a] The Evaluation Centre for Complex Health Interventions, The Keenan Research Centre, Li Ka Shing Knowledge Institute, St Michael's Hospital, Canada
[b] Department of Health Policy, Management and Evaluation, University of Toronto, Canada
[c] Independent Consultant

ARTICLE INFO

Article history:
Received 22 December 2009
Received in revised form 19 September 2010
Accepted 23 September 2010
Available online 29 September 2010

Keywords:
Evaluation design
Program theory
Pathways of influence
Learning frameworks
Design
Learning
Spread

ABSTRACT

This paper proposes ten steps to make evaluations matter. The ten steps are a combination of the usual recommended practice such as developing program theory and implementing rigorous evaluation designs with a stronger focus on more unconventional steps including developing learning frameworks, exploring pathways of evaluation influence, and assessing spread and sustainability. Consideration of these steps can lead to a focused dialogue between program planners and evaluators and can result in more rigorously planned programs. The ten steps can also help in developing and implementing evaluation designs that have greater potential for policy and programmatic influence. The paper argues that there is a need to go beyond a formulaic approach to program evaluation design that often does not address the complexity of the programs. The complexity of the program will need to inform the design of the evaluation. The ten steps that are described in this paper are heavily informed by a Realist approach to evaluation. The Realist approach attempts to understand what is it about a program that makes it work.

1. Introduction

In the last few years, there have been policy calls to integrate evaluations with performance measurement systems (Mayne, 2007a, 2007b, 2007c; Mayne & Rist, 2006; Treasury Board of Canada, 2009). Words like "accountability," "results-based culture," and "learning" are becoming common buzzwords in the process of program development and implementation; however, there continues to be limited clarity on how evaluations matter in making policies and programs work. This paper proposes ten steps that program staff and evaluators can implement to help with the integration of evaluations and evaluative thinking into program planning and implementation.

1.1. Motivations for the ten steps approach

The ten steps are motivated with a recognition that there is a need to integrate theory-driven evaluation (Mayne, 2010; Pawson & Sridharan, 2009) with more traditional evaluation designs and methods (Cook, 2000). Additionally, while there is a growing focus on understanding the pathways of influence by which evaluations impact policies and programs (Henry & Mark, 2003; Mark & Henry, 2004), there has been surprisingly limited discussion in the evaluation literature on how the concept of pathways of influence can improve evaluation designs and methods. There is also a need to explore the implications for evaluation of the growing literature on program system dynamics. Emerging from this literature is the notion that programs are dynamic systems that change over time (Pawson, 2006); there is often a lack of clarity of the anticipated timeline of impact of programs (Sridharan, Campbell, & Zinzow, 2006); the need to understand how evaluations can help with the sustainability of programs (Johnson, Hays, Center, & Daley, 2004); and finally, how evaluations can help with the spread of dynamic learning from programs (Massoud, Nielsen, Nolan, Schall, & Sevin, 2006). Fig. 1 summarizes the key questions that inform the ten steps to making evaluations matter.

The ten steps are based on more than fifteen years of evaluation experience in a number of sectors including international development, criminal justice, public health, arts, and community development. In our experience, there exists a somewhat unhealthy divide in evaluation practice between theory and methods. For example, some organizations we have worked with see the role of evaluation as purely problem of design, measurement and methods; whereas theory is considered too eclectic or too much of a 'luxury' for evaluators to be concerned about. Similarly other organizations we have worked with that implement a theory-based evaluation approach often do not integrate other evaluation design or methods to test the theoretical

* Corresponding author at:The Evaluation Centre for Complex Health Interventions, The Keenan Research Centre, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, Canada.
E-mail addresses: sridharans@smh.toronto.on.ca, sanjeevsridharan@msn.com (S. Sridharan).

0149-7189/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.evalprogplan.2010.09.003

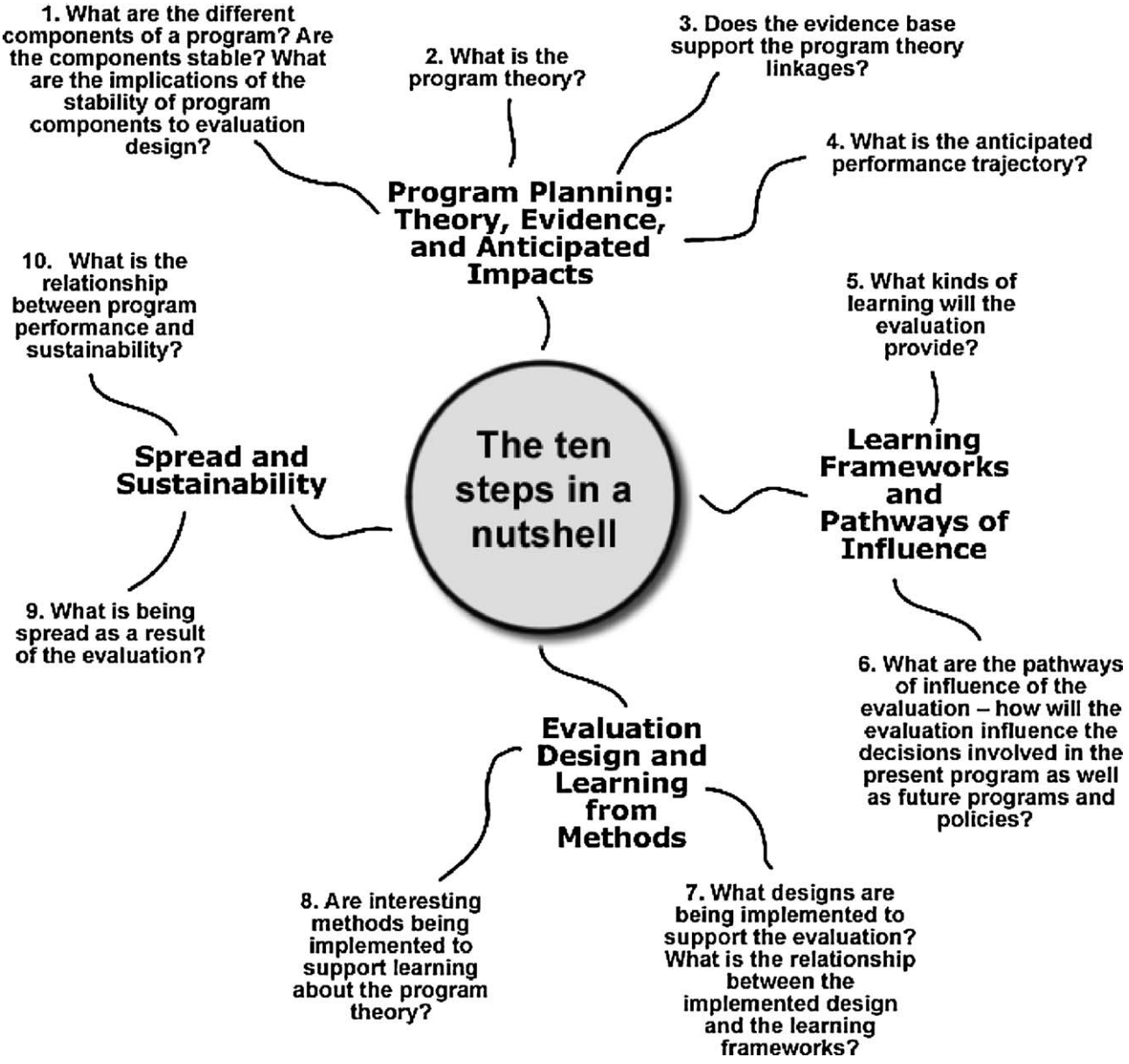

Fig. 1. The ten steps in a nutshell.

propositions embedded in the program theory. We are increasingly convinced that "off the shelf" approaches to evaluation would not suffice in helping bridge this divide between theory and methods.

Similar to the tension between theory and methods, we have also sometimes experienced severe disconnections between program planning and evaluation. Some of our experiences working on a range of program and policy evaluations reinforce an understanding that the purpose of evaluation is often too disconnected from the task of implementing a successful program. We feel that there is a need to undertake more of an explicit and structured process to align the activities of evaluation and program planning. In our view, the ten steps framework provides one approach to building a more interactive relationship between theory and methods, and between evaluation and program planning.

The specific problem that is discussed in this paper is: how can the act of evaluation help with program planning, implementation, and sustainability? In our experience, the conflicts have not been just between program implementers and evaluators but also within the community of evaluators. Different schools of evaluation have had different views of the purpose of evaluations. Sometimes there has been a lack of clarity of what could be learned through evaluations.

We stress that the ten steps offered here are a proposal for a research agenda on "how do we make evaluations matter." While some of the ideas in the ten steps framework are discussed in our prior work (Sridharan, 2008; Sridharan et al., 2006), this paper is our first attempt to systematically synthesize steps needed to "make evaluations matter." The ten steps will inform a broader research agenda. At the most basic level, the research agenda will inform the development of a comprehensive evaluation plan. Further the agenda will need to answer how the evaluation plan needs to be updated in an ongoing manner. What roles do the commissioners of the evaluation, program planners and implementers, and evaluators play in updating the evaluation plan? How can evaluative thinking help with implementing programs? A more detailed research agenda will focus on additional steps that might be required to make evaluations matter.

The proposed ten steps can be applied to a range of programs. The proposed steps will be especially useful for those who view evaluation as a process and activity that happens *with* programs and not just *to* programs. The ideas discussed here will be especially useful in evaluations of complex interventions with multiple components and components that are dynamic in nature.

We recognize that not everyone will agree that the ten steps discussed in this paper fall into the domain of evaluation or should be within the role of evaluators. While there is some merit to this claim, we stress that the ten steps are meant to be a collaborative process between evaluators, program planners and funders, and also the evaluator's role in each of these steps will depend on the particular context of the problem.

The ten steps that are described in this paper are heavily informed by a Realist approach to evaluation (Pawson & Tilley, 1997). The Realist approach attempts to understand why programs work. The Realist approach attempts to understand "what works for whom and under what circumstances." "Realists do not conceive that programs 'work,' rather it is action of stakeholders that makes them work, and the causal potential of an initiative takes the form of providing reasons and resources to enable program participants to change" (Pawson & Tilley, 1997, p. 215).

The focus of this paper is on the conceptual aspects of the ten steps. The operationalization of the ten steps is deferred to a more detailed operational paper on this subject. This paper builds the conceptual ideas using program evaluation rather than policy evaluation as its focus. The ten steps framework can also be extended to policy evaluation settings with suitable modifications.

2. The Realist approach to evaluation

Within a Realist approach, social interventions are "active" (Pawson, Greenhalgh, Harvey, & Walshe, 2004): Active interventions "work through the stakeholders' reasoning, and knowledge of that reasoning is integral to understanding its outcomes" (p. 5). Clearly, this has implications for involving stakeholders, including program recipients, program staff, and funders, in the design of the program and the evaluation.

Why is a Realist approach relevant for evaluations of complex interventions? It explicitly seeks to unpack the complexity of the program. The realist-based approach has many strengths, but most of all, its focus shifts from "does a program work?" to "what is it about a program that makes it work?" Such a focus is consistent with the learning goals of evaluation.

One of the strengths of the Realist approach is its clarity that interventions are dynamic systems. Programs are dynamic (change over time), depend critically on the context in which they are implemented, and change as a result of stakeholder reasoning. Pawson et al. (2004) describe seven characteristics of complex interventions. Table 1 summarizes these seven characteristics and the questions that might emerge in planning a realist evaluation. Not all of these questions can be answered at the start of the intervention; however, it is helpful to reflect on these questions at multiple stages of program planning and implementation because it brings into focus the dynamic complexity of an intervention.

3. The ten steps to making evaluations matter

The proposed ten steps will guide the reader through important conceptual issues related to: Intervention planning and implementation theory; influence of evaluation; design, data and methods, and spread and sustainability.

3.1. Step 1. Demonstrate understanding of the program and the stability of its components

One of the surprising aspects of the evaluation literature is that there is little reflection on the implications of the complexity of the intervention for the evaluation. Will the evaluation/performance system for a simple "aspirin" type intervention follow the same approach as designing a performance system for a complex community initiative? There is often a haste to rush into the evaluation design without a thorough understanding of the program. Programs are "complex systems thrust upon complex settings" (Pawson, 2006; Pawson et al., 2004). Complexity has implications for both the stability and the dynamic nature of the components of a program.

The blueprint for a complex program is rarely very clear at the outset of an intervention. Programs need time to: clarify their program theory, develop a clear blueprint (the program plans), and to respond to changing context and changing understanding of recipient needs.

Even before a program theory is fully explicated, it is important to describe the following:

- What are the different components of the intervention?
- Are the different components stable over time?
- How do the components of the complex intervention relate to each other?

These are not academic questions. The responses to the above questions will help inform how the evaluation design responds to the complexity of a program. There is a need for a greater focus in the evaluation literature on how the evaluation designs of stable

Table 1
Features of complex interventions.

Features of complex interventions (Pawson et al., 2004, p. iv)	Examples of evaluation questions
"The intervention is a theory of theories."	What are the stakeholders' theories of the intervention? Do different stakeholders have different theories of the intervention?
"The intervention involves the actions of people."	How do key stakeholders co-construct the intervention? What are the active ingredients of each of the interventions? Is the actual "journey" of the intervention different from the planned "journey"?
"The intervention consists of a chain of steps."	What are the implications of a complex chain of program activities for impacting long-term outcomes?
"These chains of steps or processes are often not linear, and involve negotiation and feedback at each stage."	How does user involvement change the planned intervention over time?
"Interventions are embedded in social systems and how they work is shaped by this context."	How did the context of the intervention influence the planning and implementation of the intervention? What role did the organizational context play in shaping the eventual intervention?
"Interventions are leaky and prone to be borrowed."	How and why did the intervention change over time? Did the program theory change over time?
"Interventions are open systems and change through learning as stakeholders come to understand them."	How did the experience of implementing a complex intervention change program staff's perceptions of the mechanisms involved in impacting long-term outcomes? What are the implications of such learning for future interventions?

components of a program are different from the evolutionary/dynamic components of programs (Morell, 2010). Consider a health program that consists of the following key steps:

> **Reach** individuals living in the most deprived areas → Invite them to get a **health screening** → Provide a **health coach** to support individuals with personalized care.

A program that has a well defined set of reach activities might need a different evaluation approach from another program in which "reach" activities are still in a developmental stage (Patton, 2010). This insight that evaluation approaches depends on the "developmental" status of the intervention is one of the key principles of developmental evaluation (Patton, 2010).

Who should do this? When should this be done?
The evaluator should help facilitate this process with the program staff leading in taking stock of the intervention.

3.2. Step 2. Develop explicit understanding of the program theory

A program is fundamentally based on a series of "if . . . then" ideas (these ideas are often unfortunately implicit) (Pawson, 2006). Evaluations provide an opportunity to understand such "if . . . then" ideas – specifically the processes (activities, contexts, inputs, outputs) by which programs can make a difference in the lives of individuals and communities. A program theory should describe the hypothesized processes by which a program can bring about change. We stress that these processes are hypothesized. In all likelihood, the actual process of change might be very different. Evaluations provide opportunity to examine whether the actual processes of change are different from the hypothesized processes of change. The conventional understanding of program theory through the use of logic models is useful for explicating the connection between program activities, outputs and outcomes, but there is a need to go well beyond program logic to fully describe a program theory.

How and why is the intervention likely to bring about change? Often a logic model is presented as an answer to this question; however, it is useful to differentiate between the program logic and the program theory (Leeuw, 2003, p. 6):

> "*Sometimes underlying assumptions are referred to as policy theories, but more often the terms program theory or program logic are used. An important difference exists between program theory and program logic. Program logic often specifies the inputs and components of a program, as well as short-term and long-term outcomes, along with the assumed linkages among these. However, program logic rarely outlines the underlying mechanisms that are presumed to be responsible for those linkages.*"

The concepts of context, mechanism, heterogeneity and leverage may be helpful in describing the program theory (Table 2).

Who should do this? When should this be done?
The evaluator should facilitate the development of the program theory, but the program staff should lead in developing the program theory. The initial program theory development should occur at the start of the program, but there needs to be a process to update program theory to reflect changes in stakeholders' understanding of the theory. The changes in stakeholders' understanding of the program theory need to be documented by the evaluator in an ongoing manner. Taking stock of a program theory may be one of the first steps in creating a learning environment that may be needed for developing an evaluative culture.

3.3. Step 3. Integrate a synthesis of evidence within the program theory

The program theory can be further strengthened by synthesizing evidence related to the key linkages of the program theory. This is the approach adopted by a recent method of synthesis called realist synthesis (Pawson, 2006; Pawson et al., 2004). The focus of this method of synthesis is on understanding the mechanisms and contexts under which each of the key linkages in the program theory is likely to fire. Reconsider the health program briefly discussed earlier consisting of the following key steps:

Table 2
Key questions that a program theory needs to answer.

Contexts
What are some of the underlying contexts that might be conducive for the program to work? Did the program plan incorporate such thinking into the design of the intervention?
Contexts are contingent conditions that can alter the relationship between the treatment (the program) and the outcomes. In what contexts are programs most likely to work? How can programs help create conditions for its own success? Context can refer to country policies, community norms, institutional locations, and cultural systems. One of the big gaps of most evaluations is the lack of conceptualization and operationalization of context.
Mechanisms
What mechanisms are needed for the program to work? Did the program incorporate knowledge of such mechanisms in its design? A mechanism is "an account of the makeup, behavior, and interrelationships of those processes that are responsible for the outcome." Understanding program mechanisms is critical in understanding how programs work. How does an evaluation design help explicate and understand the mechanisms by which a program might work?
Heterogeneity
Are the programs likely to impact different groups of individuals differently? Is knowledge of such heterogeneity incorporated in the planning of the intervention? Much of the language of program planning and evaluation assumes that there is a single homogenous group of intended recipients. Program recipients often have complex and heterogeneous needs. Programs consequently need to be designed with a focus on heterogeneous mechanisms. What mechanisms does the program use to impact individuals with such heterogeneous needs? How does the evaluation study the differential mechanisms and effects of a program?
Leverage
Does the program have the "leverage" to impact outcomes by itself? What other inputs are needed for a program to impact the outcomes? Describe the multiple ways in which a program is likely to work. Interventions are often informed by a theory that often implicitly states that by doing a set of activities, changes are possible. However, some problems might be so difficult to address, and the interventions lacking in intensity (due to multiple sets of reasons), that it may be difficult for a program acting solely to impact some "difficult to move" outcomes. While policy initiatives from government and other public agencies are driven by the concerns of specific departments or sectors, their implementation within communities will be such that such boundaries are artificial. The interaction between initiatives, services and programs should ideally be synergistic and mutually reinforcing. Programs are often one of a complex network of interventions that are necessary to impact complex problems. Consider the example of the impact of a "workplace smoking ban policy" on "reductions in health inequalities." It is unlikely that this policy alone without other program inputs can impact health inequalities. Evaluations provide an opportunity to identify other inputs that are needed to impact outcomes.

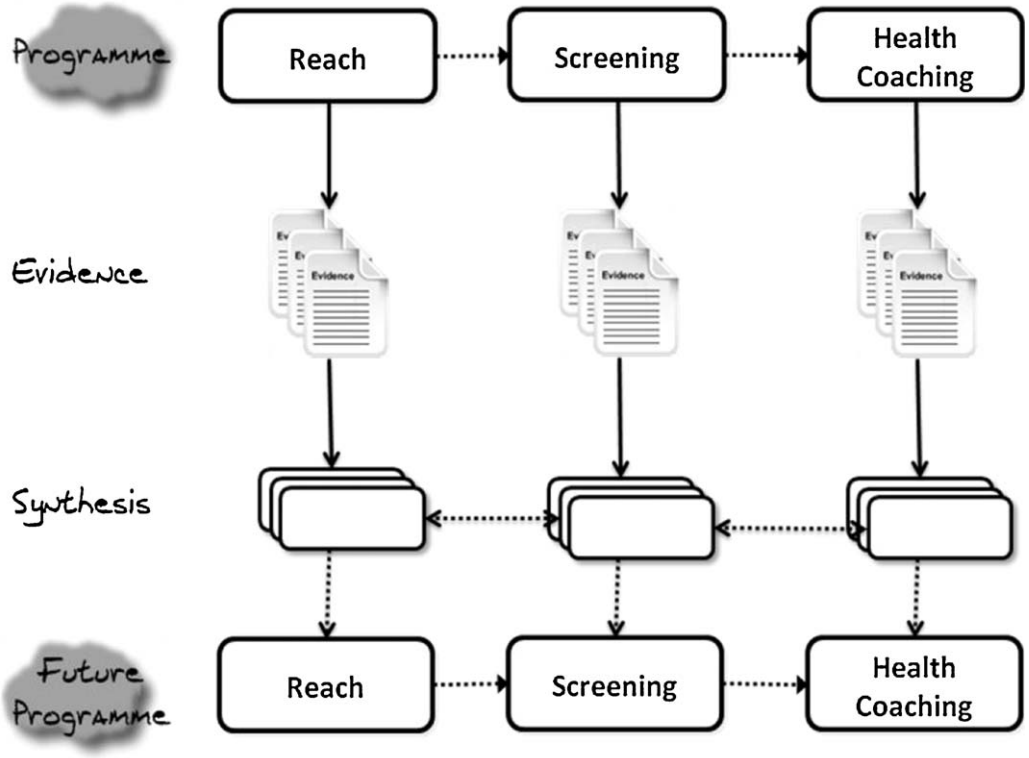

Fig. 2. An example of a realist synthesis.

Reach → health screening → health coaching

Instead of conducting a synthesis of evidence across the entire program, the realist synthesis approach recommends reviewing and synthesizing the literature for each component of the program (in addition to reviewing the entire program; see Fig. 2). A realist synthesis can also help address the following questions: Which are the linkages in the program theory that are supported by a strong evidence base? Which are the linkages that have the greatest areas of uncertainty? As example in Fig. 2, a realist synthesis could help identify insights from the evidence base that could help plan and implement reach activities. The strength of the realist synthesis approach is its focus on integrating evidence within the program theory. Realist synthesis does not subscribe to rigid views of hierarchy of evidence and strongly recommends leveraging the grey literature as needed. We think the strength of this (and as yet underutilized) approach is its utility for program planning and implementation – it helps integrate evidence into thinking about program theories.

Who should do this? When should this be done?
The key program theoretical questions should be driven by the program staff. The questions should identify the program staff's uncertainty about the implementation of the program. The actual synthesis of evidence should be led by the evaluation team. It is critical that there be ongoing meetings to learn between the program staff and the evaluation team. Synthesizing evidence may be an important step towards building a learning culture. In terms of building a learning system, it is critical that distinct bridges are built between the program theory and the existing evidence base.

3.4. Step 4. Develop an understanding of the program's anticipated timeline of impact

Programs aim to bring about change – change usually takes time. The problem of course is that there is often a lack of explicit clarity on the timeframe of change associated with a program. In the absence of a detailed theory, where should such a timeline come from? One approach to developing a timeline of impact is to involve key stakeholders – those most directly involved in the planning, implementation and delivery of a program (Sridharan et al., 2006).

Much of the evaluation literature recommends an active involvement of program stakeholders. As example consider Trochim's view (1998):

> *But my problem is that too many evaluators seem to emphasize the role of academic, social science theory. We seem to think that social scientists or management theorists have an inside track on developing sensible theories. My view is that the people in the best position to tell us about the theory of a program are the ones most familiar with it—the people who initiate it, develop it, implement it and receive it. The problem is that many of those people work in the practical world—they're not used to stating their implicit theories overtly in ways that we can formally evaluate. . . . I think we need to be examining program theory as an integral part of evaluation. But let's include the implicit theories of the people closest to the program and not deify the value that more academic theorizing may have (p. 245–246).*

As described in prior work (Sridharan et al., 2006), the emphasis on a stakeholder-driven timeline is justified (Fig. 3) because stakeholders are the people most familiar with the program. Further, different groups, such as funders, evaluators, and program staff, can have different views of the underlying theories of impact

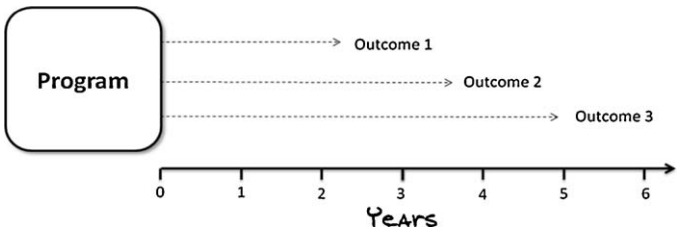

Fig. 3. Anticipated timeline of impact.

(Connell, Kubisch, Schorr, & Weiss, 1995) and the timelines associated with the program. Understanding differences in expectations of timelines may have implications for integrating diverse viewpoints that affect the planning, implementation, and evaluation of programs. Additionally, based on multiple evaluations of community programs, our experience has been that stakeholders in community settings often have a more realistic understanding of the difficulties of community change processes (compared to academic experts or funding agents). Tapping such experience is especially important in evaluations of complex initiatives.

A related idea is that of an anticipated 'performance trajectory' (Fig. 4). Recent work has begun to question standard assumptions by which *expectations* of performance trajectories of outcomes are developed (Woolcock, 2009). As example, Milstein et al. (2007) recommend use of formal system dynamic modeling procedures to understand the anticipated performance trajectories of key outcomes for programs.

> *Popular conceptions about how certain phenomena change over time may often fail to account for real-world sources of inertia and delay and may suggest that things can change more rapidly than is actually possible.*

Milstein et al. (2007) offer a strong critique of the lack of rigor and also a lack of understanding of the system dynamics involved in setting performance targets:

> *Findings from our study indicate that the HP 2010 objective for reducing diagnosed diabetes prevalence by 38% will not be achieved — not because of ineffective or underfunded health protection efforts but because the objective itself is unattainable. Moreover, if current investments in diabetes screening and disease management continue to succeed in diagnosing a greater number of people and in enabling people to live longer with the disease, then diagnosed prevalence will move still farther away from the HP 2010 target.*

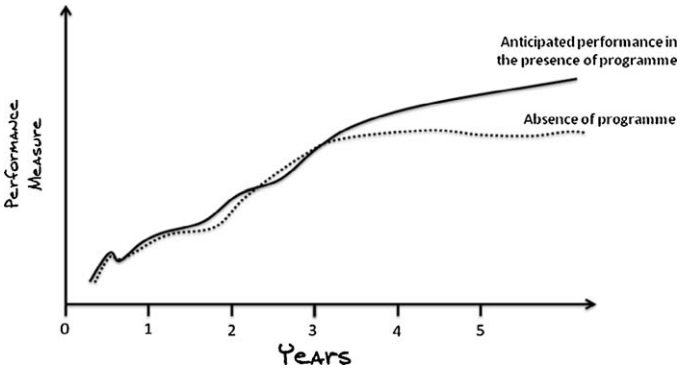

Fig. 4. Trajectory of anticipated performance measures.

Who should do this? When should this be done?
A concern raised by recent critiques of results-based management (Office of Internal Oversight Services, 2008) is the lack of a disciplined process of setting targets. A focus on anticipated trajectory of outcomes and timeline of impact can bring additional rigor and discipline to the process of setting expectations (and targets) for the program. The anticipated performance trajectory/timeline of impact should be developed by the program team, but the evaluation team needs to help facilitate the process through a focus on evidence, program theory and methods (e.g., methods of system dynamics). This should happen at the start of an intervention, and an updating of the anticipated trajectory of outcomes can happen over time.

3.5. Step 5. Develop a learning framework: Be explicit about what can be learned from the evaluation

There is sometimes a tendency in the evaluation literature to think that only a single type of learning is possible from evaluations – as example, the debates around randomization in evaluation suggest that the only learning that matters is individual-level impacts. However multiple kinds of learning are possible from each and every evaluation. Table 3 describes the five examples of learning that may be possible from program evaluations – this table is based on a recent evaluation of a Scottish Demonstration program (Sridharan et al., 2008). We think it is important (to the extent possible) to be explicit about the intended learning of evaluation at the start of the evaluation.

Who should do this? When should this be done?
Once the program theory has been established, the evaluation team and the program team need to work together to develop clarity on a learning framework. What can be learned from a detailed evaluation? How can the evaluation help the program staff achieve their goals? Also, there may be a clash between what program staff wants, and what evaluators are required to focus on (as example, based on the terms of reference in a contract or grant). Sometimes program staff wants information that is consistent with models of continuous improvement; on the other hand evaluators' focus may be on rigorous experimental designs that may not under some contexts help the program navigate towards their long-term goals. Our view is that a greater focus on learning frameworks can help with the complex navigational challenge that programs often face of achieving long-term outcomes – what needs to be learned from evaluations to achieve the program's long-term goals.

3.6. Step 6. Discuss the impact of the evaluation on use and influence

There is a fairly limited literature in evaluation on how evaluations can make a difference – what are the pathways by which evaluations influence programs and policies? The neglect to consider "pathways of influence" is somewhat surprising given that this is the very question evaluators ask of programs. A key insight from the recent evaluation literature is to treat the evaluation process itself as an intervention (Henry & Mark, 2003; Mark & Henry, 2004). Henry and Mark have conceptualized evaluation "as an intervention" and have attempted to understand the "ways in which evaluations, or the evaluation process itself, influences social betterment in the long term."

One of the strongest criticisms of evaluation that is often made by program staff is its lack of clarity of purpose. Routines of

Table 3
Learning frameworks for evaluation.

Policy learning: Every program is an act of translation from an idea of a policymaker to the planning and implementation of the program. Almost always, there is both loss of translation and adaptation in this journey from an idea to program implementation. Evaluations provide a chance to unpack and understand such translation. Key questions for policy learning include: What were the policy aspirations? What was the underlying policy theory? Are the program goals consistent with the policy aspirations? Is the program implementation consistent with the policy aspirations? Surprisingly few evaluations focus on this type of learning even though answers to these questions are important to the policymaker.
Organizational learning: Programs are "complex organizations thrust upon complex settings" (Pawson et al., 2004). Evaluations provide an opportunity to learn more about the organizational structures needed to support the program. What organizational structures (and processes) are needed to support the intervention? Is attention paid to the organizational structures required to sustain the intervention over the long term? Organizational context may be critical to the success of a program in achieving its goals. Yet again, very few evaluations generate knowledge on organizational structures needed to support programs.
Process learning: Evaluations also provide an opportunity to learn about the planning and implementation processes required to successfully deploy the intervention. In addition to the usual focus on planning and implementation, process learning might also focus on planning for sustainability. While a number of evaluations do provide useful learning on process, only a few evaluations have integrated learning about process with understanding of impacts.
Understanding barriers: We often know precious little about the difficulties, constraints, barriers, opportunities that intended program recipients face. Evaluations provide an opportunity to learn more about such "risk landscapes." Knowledge of such a "risk landscape" might not only help plan future interventions but might be important in explaining the success or failure of an intervention.
Individual-level impacts: Finally, evaluations provide a chance to examine what differences an intervention makes in the lives of intended program recipients. Did the program improve the lives of its intended recipients? Most of the standard views of evaluation design fall in this category of learning. Much of our debates on experimental designs fall into this category of learning.

drudgery and filling of forms to comply with evaluation policies is unlikely to lead to any serious learning. In our view, having an influence plan can help bring greater clarity and buy-in among program staff for the evaluation systems.

Who should do this? When should this be done?
Early in the life of the program, the program team and the evaluation team need to develop an influence plan. The influence plan needs to address: How will the evaluation make a difference to the planning and implementation of the program? Such an influence plan needs to go beyond the boilerplate of dissemination and communication strategies and discuss well thought out mechanisms by which the evaluation system can make a difference to navigating the program.

3.7. Step 7. Develop a comprehensive evaluation design

A surprising number of evaluations do not explicitly discuss the type of *evaluation design* implemented in understanding if and how programs work. In this section, we briefly discuss some general principles of design – space constraints prevent us from discussing the data and analytical methods that are needed to support the design (Bamberger, Rugh, & Mabry, 2006; Davidson, 2005; Hatry, Wholey, & Newcomer, 2004; Mark, Henry, & Julnes, 2000; Patton, 2010; Rossi, Lipsey, & Freeman, 2004; Shadish, Cook, & Leviton, 1991). The following are some important points to keep in mind when thinking about evaluation design:

- A lot of the discussion on evaluation design often focuses on a single facet of the learning framework – i.e., learning on individual-level impacts. There are many occasions where the focus of the evaluation should be on other types of learning. This needs to be kept in mind before evaluations are dismissed because of poor designs.
- Design is not just about measurement or methods or even theory. What is surprising is to find the very large number of evaluations that do a good job of describing the measurement system but do a poor job of describing the design that will be implemented to study the causal impacts of programs.
- Designs often make implicit assumptions about program stability. As far as possible, considerations about program stability need to be made as explicit as possible.
- A good evaluation design also needs to shed light about the actual program's mechanism of change (or alternatively test the hypothesized mechanism of change). Such knowledge might be extremely critical to assess generalizability of the program in order to make decisions about replicating or adapting a program to a new setting. A good design should shed light on the contexts needed and the mechanisms by which programs work.
- Studying the impacts of programs often requires a process of ruling out alternative explanations for why change could have happened. Threats to internal validity provide a checklist of alternative explanations that need to be seriously considered before impacts can be attributed to programs. Evaluation designs can improve significantly as a result of the consideration of such threats to internal validity.
- Understanding program impacts requires knowledge (and information) of what could happen in the absence of the intervention. Evaluation designs help provide such information. This view of program impacts is connected to a counterfactual view of causality.
- Threats to external validity deal with the problem of generalizing findings from a single evaluation study. Addressing threats to external validity requires consideration of settings, units of intervention, treatments, and multiple outcomes (Shadish, Cook, & Campbell, 2002). Given the important role of contexts in shaping the success of an intervention, it is unlikely that a finding from an intervention will be generalizable across all settings. For the most part, considerations of external validity have not informed design choices as much as they perhaps should.
- Good design does not negate the need for program theory. The mantra of "no causation without theory" should always be kept in mind as we develop and implement evaluation designs.

One of the challenges of integrating evaluation with program needs is that most discussions of evaluation design has focused on studies of effectiveness – *for the most part, evaluators have not focused on models of evaluation that combine a focus on effectiveness with models of continuous improvement*. Program staff, on the other hand, might need evaluation designs that follow a continuous improvement model. One exception to the literature is Morell (2000) who argues for integrating continuous improvement from industrial engineering with traditional program evaluation models.

Table 4
Examples of "innovations" in evaluation of simple and complex programs.

	Analytical techniques	Examples of evaluation questions
Innovations in simple programs	Observational studies (propensity scoring techniques)	In the absence of randomized designs, what alternative analytical techniques can help assess causal effects of programs?
	Methods of developmental trajectories	Does the same intervention have very different impacts over time for different groups of individuals (or communities)?
	Respondent-driven sampling	How does the program (and evaluation) reach individuals who are hard to reach and not part of a sample frame?
Innovations in complex interventions	Network analysis	How does the program operationalize networks of interventions? What role does a network of partners play in the delivery of services in a program?
	Event structure analysis	How does the "event structure" (for example, the structure of interactions between doctors and clients) impact on outcomes over time?
	Concept mapping	How do different groups of stakeholders conceptualize the intervention? Do these conceptualizations change over time?
	System dynamics	What are the anticipated trajectories of the impacts of the overall program? How would changing key components of the intervention alter the expected trajectory of outcomes?
Approaches to complexity	Theories of change	What are the pathways by which the program achieves its outcomes? What are the linkages between program inputs, program contexts, outputs and outcomes?
	Adaptive/sequential designs	How can a design help an intervention adapt/evolve from large complexity with many components to an intervention of a few "effective" components?
	Realist synthesis	How does evidence synthesis help identify the mechanisms by which an intervention works?

Who should do this? When should this be done?
The evaluation design needs to be developed in collaboration between the program and the evaluation teams at the start of the program – ideally at the program-planning phase. However, the evaluation design needs to be updated in an ongoing, sometimes phased manner, especially given that programs themselves will change over time. A well conceived (and implemented) evaluation design can help bring relevance to the evaluation by connecting measurement more closely to theory (and program needs), and will ensure that attention is paid to the outcomes that are most closely connected to the program theory.

3.8. Step 8. Demonstrate emergent, dynamic learning about the program learning as the program is implemented

A well-implemented evaluation design will help understand if a program is working. A range of other methods are available that can shed light on additional learning on program theory that may be possible through an evaluation of a complex intervention. Table 4 summarizes some examples of innovative methods – a few of them are relatively new and might help augment learning from evaluations. This table was generated through an email survey of evaluation methodologists.[1]

There is a tendency in the evaluation literature and a number of evaluation reports to speak of programs as static, stable entities. Even with programs that have a strong evidence base, programs are likely to change over time (both in the planning and implementation). One of the unfortunate implications of assuming that programs are stable, static entities is that evaluators rarely document how interventions might change over time. How much of this change was driven as a response to the particular context in which the program was operating? What part of the change suggests a more general process of adaptation that is likely to hold in other settings? Were such changes driven by contextual factors or a means of aligning the complex intervention with the evidence base? For the most part, most evaluations have treated programs as static entities. Evaluators need to be open to raising new questions as the program unfolds and implement a range of methods as the situation demands.

One of the "signatures" of complex interventions might also be the lack of well-developed conceptualization or operationalization of the intervention theory. Examples of methods that can help conceptualize and operationalize the interventions include concept mapping, network analysis, and event structure analysis. Conceptualizing and operationalizing the complex intervention has very practical consequences: a network analysis might throw light on the type of collaboration that is needed to make an intervention work; and a concept mapping might show that a Federal agency and members of a local community view the same complex intervention quite differently. Some of the methods described in Table 4 (e.g., network analysis, concept mapping, theories of change, system dynamics) can be helpful in understanding changes in programs over time.

The standard view of evaluation that describes a linear path between program theory and methods may not be sufficient. Within such a view the relationship between program theory and impact runs as follows (scenario 1 in Fig. 5):

> Program theory → Program planning and implementation → Performance measurement and evaluation → Program impacts

What is needed is a more dynamic view of the relationship between theory and methods (scenario 2 in Fig. 5) – in such a view there is greater clarity and honesty about the areas of the program theory where there is greater uncertainty in the evidence base (both in planning and implementation) at the start of the program.

Who should do this? When should this be done?
Innovative methods that help understand how programs work need to be implemented both at the start of the evaluation and also in an ongoing way. As these methods often require strong analytical training, the evaluation team needs to lead on implementing the methods. There, however, needs to be close collaboration between the evaluation team and the program

[1] 50 evaluation methodologists were sent an email survey requesting feedback on recent innovations in evaluations of complex interventions. 35 of the 50 methodologists responded to the email.

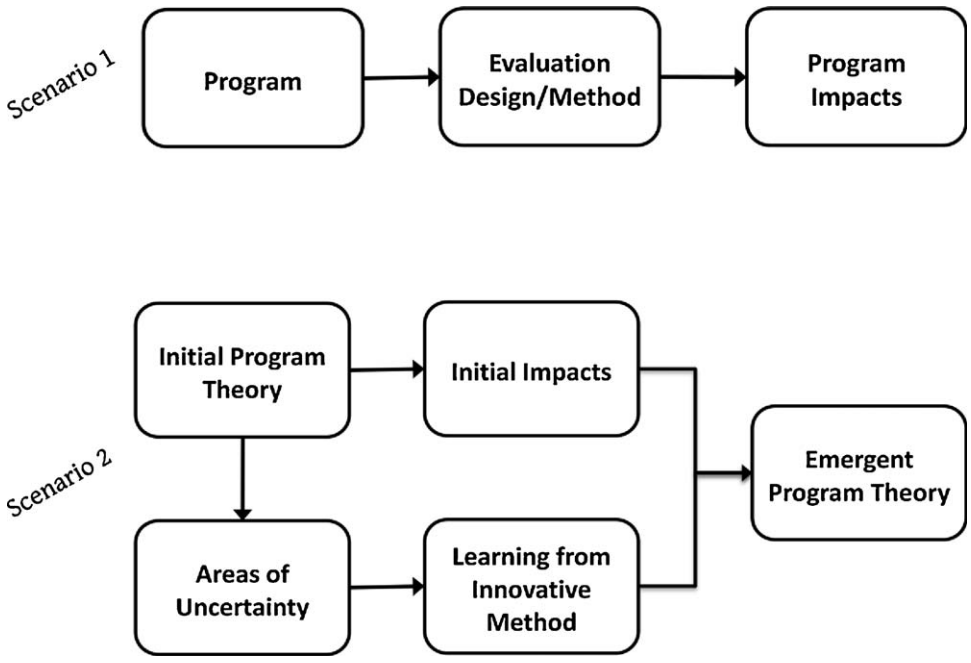

Fig. 5. Learning from evaluation methods.

team to ensure that the questions addressed are relevant to the program. These methods can assist with learning from the evaluation. An explicit focus on 'learning from methods' can help make the evaluation less rigid and mechanical.

3.9. Step 9. Discuss the relationship between evaluation and a 'framework of spread'

One of the more surprising things about evaluation is the lack of clarity in what is being "spread" at the end of an evaluation. For example, does an evaluation aim to provide recommendations regarding scaling up, scaling out or replicating the project. Evaluators often are advised to design an evaluation with a focus on generalizability (e.g., reduce the threats to external validity), but there is rarely clarity on what results are generalizable from an evaluation of specific interventions. What will be spread at the end of the evaluation?

The Institute of Health Care Improvement has begun to use a framework of spread to help "develop, test, and implement a system for accelerating improvement by spreading change ideas within and between organizations" (Massoud et al., 2006). Some of the relevant questions this framework raises for evaluators include: How will the knowledge learned from an evaluation of a specific initiative in one organization be spread to other organizations? What role is there for spread of innovation in a 'results-based' culture? Who is in charge of such spread? What role does the funder of the program and the evaluation play in the spread of innovative practices?

Are we doing the evaluation purely as routine, or is there a plan to spread learning from the evaluation? There needs to be clarity (both at the outset and over time) on the learning from the evaluation that is likely to be spread. A spread framework can focus on many possibilities including lessons learned about the entire program, components of the program, innovative practices that are part of the program, or perhaps about contextualized learning (see Fig. 6).

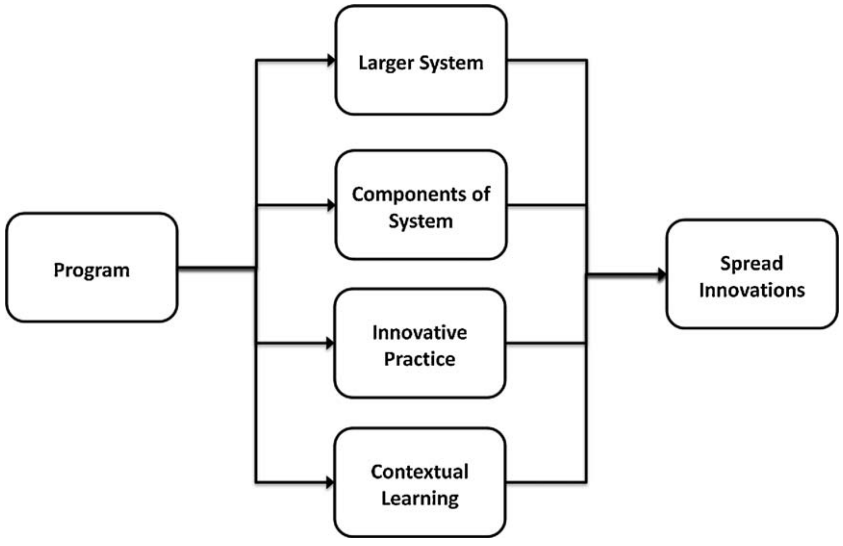

Fig. 6. A framework of spread.

Who should do this? When should this be done?
A framework of spread should be developed both by the evaluation team and the program team soon after the implementation of a program. This framework should be updated on an ongoing manner. One concern with evaluations and others results based management system has been the lack of consequences of poor or good performance (Office of Internal Oversight Services, 2008). Our view is that this is not just a failure of not having an organizational structure or process in place to take relevant decisions but also a lack of focus on the type of learning and innovation that is being spread as a result of the evaluation. A focus on spread can also address a concern about the lack of strategic direction and cross-organizational performance incentives that often occur with a results based management framework (Office of Internal Oversight Services, 2008).

3.10. Step 10. Demonstrate consideration of the relationship between performance and sustainability

The decision to sustain programs is often decided at the highest levels of the corridors of power. However, the question of the relationship between performance and sustainability remain quite fundamental to the purpose of the evaluation.

What role should program effectiveness play in the decision to continue and sustain a program? Perhaps this question is poorly conceptualized. It implies a model of program activities that occur along a linear sequence of planning, implementation, and sustainability. As Johnson et al. (2004, p. 146) argue: "The sustainability process needs to begin early after decisions have been made to adopt or experiment with an innovation." Further, the proposed approach also resonates with a recent result by Pluye, Potvin, and Denis (2004, p. 127): "We propose that program implementation and sustainability are not distinct and successive phases but are concomitant processes." Further, arguing against a distinct and successive view of implementation, Pluye et al. (2004, p. 126) state: "This model does not take account of the recursive or reflexive character of sustainability and learning or of the continuous adjustments that shape the sustainability process." The implication of such a view is a need for discussion early in the life of an evaluation on the role of evaluation in sustainability decisions. The decision to sustain a program should not only be driven by the effectiveness of the intervention – other types of learning might also have a bearing on the decision to sustain an intervention.

Second, clearly both the results of the evaluation as well as the match between the actual performance and anticipated trajectories should be taken into consideration. As this paper is a reflection on evaluation, performance measurement and sustainability, we raised the following question to five thought and practice leaders in evaluation[2]: Should programs that do not have a "successful" trajectory of 'performance measures' be terminated?

This is quite a difficult question to answer because:

- The trajectory of even a successful program might be quite non-linear (things might get worse before they get better), so one needs to be a little careful in answering this question;
- "Success" in the trajectory of some outcomes might come at the expense of others (see Milstein et al., 2007; outcomes might form a system – and interdependence in outcomes suggests that some outcomes might move in opposite directions);
- There is no reason for the trajectory of performance outcomes to be linear or monotonic over time – this has important implications for an evaluation system.

A key response was that the decision to sustain a program goes well beyond the performance results. Key themes in their responses included the need to:

(a) Understand the drivers of the performance using the lens of theory. This view is consistent with the realist framework that has guided this paper.
(b) Pay attention to the process by which the goals and targets of the programs are being set. This is consistent with Milstein et al.'s (2007) critique of the lack of clarity and rigor with which program targets are set. Who is setting the performance targets? Are the targets being set with a process that involves a range of diverse stakeholders?
(c) Recognize that without clarity on the "drivers" of performance, it is hard to conclude if programs need more investment or need to be downsized.
(d) Pay attention to the 'system dynamics' involved in the process of program implementation. The nature of program impacts might be such that the benefits of innovation will take time to accrue.

Who should do this? When should this be done?
One of the problems in implementing programs and deploying evaluations is a lack of structure and process to address questions linking performance and sustainability. One of the concerns with many conducted evaluations is the lack of consequences of poor or good performance. Having a plan that links sustainability to performance can help address this problem. However, the bigger problem is: Who is responsible for developing a sustainability plan? This question needs to be addressed by the funders, the program staff and the evaluators at an early stage of the evaluation.

4. Conclusion

A focus on the ten steps can result in better planned evaluations and can also help in developing more rigorous evaluation practice and evaluation training.

4.1. Implications for developing evaluation plans

The ten steps described in this paper can help with developing comprehensive evaluation plans. A few points are worth noting in developing an evaluation plan:

- Most of the ideas for the evaluation plan need to be developed in collaboration between the evaluator, the program staff and other stakeholders.
- While some of the ideas for a plan can be developed at the start of a program, more realistically much of the plan development needs to take place on an ongoing basis. It is important to have an organizational structure (and a system of commissioning evaluations) that can support the collaboration between program and evaluation teams.

[2] Of the five leaders of evaluation surveyed, two of the respondents were editors of leading evaluation journals; one respondent was a leader in application of systems models in evaluation; a fourth was a leader in application of evaluation frameworks in translating knowledge to policy; fifth was an evaluator in an international development agency and a leader in evaluation capacity building.

- While the ten steps are presented sequentially, a comprehensive evaluation plan needs to develop in a non-linear, iterative way.
- Much of the discussion of the steps is quite conceptual – each of the steps will need to be further unpacked. Space constraints prevent us from developing each step in detail in this paper.

4.2. Implications for evaluation practice and training

A focus on the ten steps can result in a closer integration of evaluation within program planning and implementation. We are struck with the absence of explicit and upfront thinking of important concepts of pathways of influence or sustainability of programs in much program implementation or evaluation practice. An integration of broader evaluative thinking early in program planning will not only result in improved evaluation but can only result in more rigorous programs. In a number of complex interventions there is a lack of theoretical understanding of how the programs are supposed to work. Evaluators are increasingly called upon to explicate the pathways by which programs work. We believe that a stronger focus on program theory and planning for sustainability should be a part of the training of evaluators. Explication of program theory is not merely a matter of learning the substantive aspects of the program but also requires considerable facilitation and communication skills.

In our experience, both the practice of evaluation and evaluation training tends to be methods heavy or theory heavy. Evaluations informed by strong program theory as well as a sound set of methods tend to be rare. Our view is that paying attention to the ten steps will result in more balanced evaluations that are both theoretically informed and methodologically rigorous.

Program theory provides insights on how a program should work; methodological approaches can help explicate if programs actually work the way they are supposed to work. The ideas presented in this framework are intended to move the discussion forward on "what is good enough theory." The evaluation literature is rich on discussions of best practices in methods. In our view there needs to be a similar dialogue on what constitutes good enough program theory that can be implemented (Miller, 2010).

Additionally, we think the field can benefit from more elaborate discussion on how evaluation approaches respond to the complexity of the real world. In our experience, we find a disconnect between the clean sterile view of programs that are presented in a number of text books and the muddled complexity of programs of the "real word". Part of the strength of the ten step approach is that it takes the complexity seriously and thinks explicitly about methods that can respond to such complexity.

4.3. Implications for knowledge translation

The ten steps framework also stresses the need for a literature on the relationships between complexity, evaluation and knowledge translation. As example, recent literature on knowledge translation addressed problems of real world complexity by making persuasive arguments to incorporate ideas of "knowledge use" more directly into the knowledge development process (Carlile, 2004; Kitson, 2009; Nutley, Walter, & Davies, 2007).

The implementation of evaluation designs needs to go beyond evaluation methods. Even though methods do matter in matters of guidance, improving evaluation practice should not rely solely on the development of "new and improved" methods. We concur with Carden (2007, 2010) that there is need to think more broadly about what is needed to address the "evaluation gap" (Carden, 2007; Carden, 2010): "The evaluation gap in my view is not found in the methodological debates of which the evaluation community is so fond, but in the lack of focus on supporting the development of evaluation as a profession ... (Carden, 2010, p. 220)."

As example, Sridharan and De Silva (2010) argue that the evaluation field needs to go beyond the narrow pre-occupation with methods and pay greater attention to issues of understanding the pathways of influence by which evaluations impact policies and programs:

> "It is also important that we build a better knowledge base of how our evaluations are leading to better decision making in our respective communities and countries. Although we often ask the question on the pathways by which programs work, we often do not reflect on the pathways by which our evaluations have influenced policies and practice. It is vital that we spend time reflecting on features and characteristics of evaluations that have 'made a difference.' The reality is that the evidence base for how evaluations have made a difference to improving policy and programmatic decision making both in the West and the South is very scant" (Sridharan & De Silva, 2010, p. 247).

4.4. Limitations of the proposed approach

We have presented our framework in multiple settings and some of the critiques of the proposed framework include:

(i) Even though the framework is very comprehensive, it also leaves out many key evaluation concepts and approaches. For example, we make no mention of economic evaluations as part of our ten steps even though there is a growing interest in the cost effectiveness of programs. This in our judgment is a fair criticism, though we think this can be handled at the design stage (step 7).

(ii) A second critique that is more common is that this framework is too broad. Evaluators already are asked to do a lot with very little and it is both unwise and infeasible to expect evaluators to "do it all." For example, one common feedback is that it is not for the evaluator to come up with a sustainability plan. We agree that there is a need for greater discussion on roles and boundaries of evaluation, but thinking about sustainability should be part of thinking evaluatively about programs. Our goal is not to increase the workload of evaluators – rather, we are increasingly struck by the growing need for evaluations to focus on questions of theory, learning, influence, design, methods, spread and sustainability. We concur that it will be unwise and unfair to expect evaluators to do it all. Instead it is our view that the ten steps need to be part of the dialogue of making the most of evaluations. The specific issues of 'who should do what' should be decided by funders, program planners and evaluators based on the specifics of each problem.

(iii) Another insightful feedback that we have received is that the ten steps are not consistent with standard practice of commissioning evaluation. Evaluators get hired to do specific pieces of work and not raise such a comprehensive set of questions. Once again there is merit in this critique. Changes in commissioning practice might hold one of the keys to changing evaluation culture. We think one of the implications of the ten steps is the need for a broader dialogue on how best to commission evaluations especially in light of incomplete and very general program theories at the outset of a program.

The ten steps presented are conceptual and need to be unpacked further. The goal of these ten steps is to provoke discussion within the evaluation and program planning communities on the steps required to make evaluations matter. While each of the ideas needs additional development, a focus on the concepts of program theory, linking evidence to program theory,

anticipated performance trajectory, learning frameworks, anticipated timeline of impact, design, innovative methods, spread, and sustainability can help with the difficult task of making evaluations matter.

References

Bamberger, M., Rugh, J., & Mabry, L. (2006). *RealWorld evaluation: Working under budget, time, data, and political constraints*. Thousand Oaks, CA: Sage.

Carden, F. (2007). The real evaluation gap. *Alliance Magazine, 12*, 4.

Carden, F. (2010). Introduction to the forum on evaluation field building in South Asia. *American Journal of Evaluation, 31*(2), 219–221.

Carlile, P. (2004). Transferring translating and transforming: An integrative framework for managing knowledge across boundaries. *Organizational Science, 15*, 555–568.

Connell, J., Kubisch, A., Schorr, L., & Weiss, C. (Eds.). (1995). *New approaches to evaluating community initiatives: Concepts, methods and contexts*. Washington, DC: Aspen Institute.

Cook, T. D. (2000). The false choice between theory-based evaluation and experimentation. *New Directions in Evaluation, 87*, 27–34.

Davidson, J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.

Hatry, H. P., Wholey, J. S., & Newcomer, K. E. (2004). *Handbook of practical program evaluation*. San Francisco, CA: Jossey-Bass.

Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation, 24*(3), 293–314.

Johnson, K., Hays, C., Center, H., & Daley, C. (2004). Building capacity and sustainable prevention innovations: A sustainability planning model. *Evaluation and Program Planning, 27*, 135–149.

Kitson, A. (2009). Knowledge translation and guidelines: A transfer, translation or transformation process? *International Journal Evidence Based Healthcare, 7*, 124–139.

Leeuw, F. L. (2003). Reconstructing program theories: Methods available and problems to be solved. *American Journal of Evaluation, 24*, 5–20.

Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation, 10*(1), 35–57.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA: Jossey-Bass.

Massoud, M. R., Nielsen, G. A., Nolan, K., Schall, M. W., & Sevin, C. (2006). *A framework for spread: From local improvements to system-wide change*. IHI Innovation Series white paper. Cambridge, MA: Institute for Healthcare Improvement.

Mayne, J. (2007a). Evaluation for accountability: Reality or myth? In M.-L. Bemelmans-Videc, J. Lonsdale, & B. Perrin (Eds.), *Making accountability work: Dilemmas for evaluation and for audit*. New Brunswick, NJ: Transaction Publishers.

Mayne, J. (2007b). Challenges and lessons in implementing results-based management. *Evaluation, 13*(1), 89–107.

Mayne, J. (2007c). *Best practices in results-based management: A review of experience*. A Report for the United Nations Secretariat. New York: UN Secretariat.

Mayne, J. (2010). Contribution analysis: Addressing cause and effect. In R. Schwartz, K. Forss, & M. Marra (Eds.), *Evaluating the complex*. New Brunswick, NJ: Transaction Publishers.

Mayne, J., & Rist, R. (2006). Studies are not enough: The necessary transformation of evaluation. *Canadian Journal of Program Evaluation, 21*(3), 93–120.

Miller, R. L. (2010). Developing standards for empirical examinations of evaluation theory. *American Journal of Evaluation, 31*(3), 390–399.

Milstein, B., Jones, A., Homer, J. B., Murphy, D., Essien, J., & Seville, D. (2007). Charting plausible futures for diabetes prevalence in the United States: A role for system dynamics simulation modeling. *Preventing Chronic Disease 4(3) Available from: http://www.cdc.gov/pcd/issues/2007/jul/06_0070.htm*.

Morell, J. A. (2000). Internal evaluation: A synthesis of traditional methods and industrial engineering. *American Journal of Evaluation, 1*, 41–52.

Morell, J. A. (2010). *Evaluation in the face of uncertainty: Anticipating surprise and responding to the inevitable*. New York, NY: Guilford Publications.

Nutley, S. M., Walter, I., & Davies, H. T. O. (2007). *Using evidence. How research can inform public services*. Bristol: The Policy Press.

Office of Internal Oversight Services. (2008). *Review of results-based management at the United Nations*. http://www.un.org/ga/search/view_doc.asp?symbol=A/63/268.

Patton, M. Q. (2010). *Developmental evaluation*. New York, NY: Guilford Press.

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. Sage Publications.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2004). Realist synthesis: An introduction. ESRC Research Methods Programme, University of Manchester, Paper 2.

Pawson, R., & Sridharan, S. (2009). Theory-driven evaluation of public health programmes. In A. Killoran & M. Kelly (Eds.), *Evidence-based public health: Effectiveness and efficiency* (pp. 43–61). Oxford, England: Oxford University Press.

Pawson, R., & Tilley, N. (1997). An introduction to scientific realist evaluations. In E. Chelimsky & W. R. Shadish (Eds.), *Evaluation for the 21st Century: A handbook* (pp. 405–418). Thousand Oaks, CA: Sage Publications.

Pluye, P., Potvin, L., & Denis, J. L. (2004). Making public health programs last: Conceptualizing sustainability. *Evaluation and Program Planning, 27*, 121–133.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach seventh edition*. Thousand Oaks, CA: Sage.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin.

Sridharan, S., Campbell, B., & Zinzow, H. (2006). Developing a stakeholder-driven timeline of change for evaluations of social programs. *American Journal of Evaluation, 27*(2), 148–162.

Sridharan, S. (2008). Making Evaluations Work. *American Journal of Evaluation, 29*(4), 516–519.

Sridharan, S., & De Silva, S. (2010). On Ambition, learning, and co-evolution: Building evaluation as a field. *American Journal of Evaluation, 31*(2), 246–325.

Sridharan, S., Gnich, W., Moffat, V., Bolton, J., Harkins, C., Hume, M., et al. (2008). *Evaluation of primary prevention intervention: Have a heart Paisley Phase 2*. Glasgow: NHS Health Scotland.http://www.chs.med.ed.ac.uk/ruhbc/evaluation/hahpfinal/Reports/detailed_reports/Primary_Prevention_HAHP2.pdf.

Treasury Board of Canada Secretariat. (2009). Policy on Evaluation. http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=12309.

Trochim, W. (1998). An evaluation of Michael Scriven's Minimalist Theory: The least theory that practice requires. *American Journal of Evaluation, 19*(2), 243–249.

Woolcock, M. (2009). Toward a plurality of methods in project evaluation: A contextualised approach to understanding impact trajectories and efficacy. *Journal of Development Effectiveness, 1*, 1–14.

Sanjeev Sridharan is Director of the Evaluation Centre for Complex Health Interventions at St. Michaels Hospital and Associate Professor of Health Policy, Management and Evaluation at the University of Toronto. Prior to his position at Toronto, he was the Head of the Evaluation Program and Senior Research Fellow at the Research Unit in Health, Behaviour and Change at the University of Edinburgh. Sanjeev has a strongly interdisciplinary background with a Doctorate in Social Ecology from the University of California at Irvine, Master's in Public Policy from Purdue University, and a Bachelor of Technology degree in Civil Engineering from the Indian Institute of Technology. His work over the last decade has been funded from a variety of sources including the Scottish Executive, NHS Health Scotland, U.S. Department of Health and Human Services, UNICEF South Asia and U.S. Department of Justice. His primary substantive focus has been on evaluation influence, evaluation methodology, health inequities, cardiovascular disease and other co-morbid conditions. He is presently working closely with the China Health Economics Institute to build evaluation capacity in the health sector in China. He is also working on an initiative to develop a post-graduate program in evaluation in five S. Asian countries. He is also advising the Ministry of Health in Chile on utilizing evaluation approaches to redesign health policies in Chile. Additionally, he has also focused on developing evaluation strategies for organizations including National Health Service Health Scotland and the Forestry Commission in Scotland. He is on the Board of the American Journal of Evaluation and Evaluation and Program Planning.

April Nakaima is an independent evaluation consultant based in Toronto, Canada. She holds a Master of Fine Arts degree from the University of California, Irvine, and a Bachelor of Arts degree in Theatre Arts from the University of California, Santa Cruz, and is a graduate of the Kamehameha Schools in Honolulu, Hawaii. Ms. Nakaima has developed and delivered health interventions in the form of dance activity in Scotland as part of Scotland's National Health and National Sports Strategies, through Artlink Edinburgh, and through the Outreach Programme at Dance Base, Scotland's National Dance Centre. April has been a practicing evaluator for the past five years. Her interest is in examining the pathways by which arts and dance interventions impacts health outcomes.