THE GUIDE TO

THE SYSTEMS EVALUATION PROTOCOL

PHASE I: PLANNING










V 3.1

The Guide to

The SYSTEMS EVALUATION PROTOCOL

Phase I: Evaluation Planning

Cornell Office for Research on Evaluation (CORE)

Cornell University

Research on Evaluation and Developmental Systems Science Lab (REDSS)

Montclair State University

This work was funded in part by The National Science Foundation Award #0535492: Building Evaluation System Capacity for STEM Programs: Enhancing Capability and Advancing Practice, and Award #0814364: A Phase II Trial of the Systems Evaluation Protocol for Assessing and Improving STEM Education Evaluation, and Award 1322861: Collaborative Research: Impact of the Summer Institutes on Faculty Teaching and Student Achievemnt. It was also funded in part by Cornell Cooperative Extension, Cornell University.





V3.1 Copyright ©2016 by Cornell Office for Research on Evaluation (CORE) at Cornell University, and the Research on Evaluation and Developmental Systems Science Lab, (REDSS Lab) at Montclair State University. All rights reserved.

Citation Information:

Trochim, W., Urban., J. B., Hargraves, M., Hebbard, C., Buckley, J., Archibald, T., Johnson, M., and Burgermaster, M. (2016). The Guide to the Systems Evaluation Protocol (V3.1). Ithaca, NY.

No part of this publication may be reproduced in any form or by any means without the prior written permission of one of the authors. For copyright inquiries please contact Cornell Office for Research on Evaluation, MVR Hall, Ithaca, NY 14853. email: evaluationresearch@cornell.edu. website: www.core.human.cornell.edu

Illustrations - Warren Greenwood - warrengreenwood.com

Digital color & Post-processing – V1 Steve Rokitka - spinetree.org; V2.1 , V2.2, and V3.1 updated or created by Cornell Office for Research on Evaluation (CORE).

TABLE OF CONTENTS

Preface v
History of the Systems Evaluation Protocol vii
Acknowledgements ix
I. Introduction to Systems Evaluation 1
A Systems Perspective on Programs and Program Evaluation 2
II. The Systems Evaluation Protocol 5
Phase 1: Evaluation Planning 5
Stage 1: Preparation 6
1.01 Enter the System 7
1.02 Memorandum of Understanding 7
1.03 Internal Stakeholders 10
1.04 Working Group(s) 10
1.05 Evaluation Capacity 11
Stage 2: Program Model Development 12
2.01 Stakeholder Analysis 12
2.02 Program Review 14
2.03 Program Boundary Analysis 14
2.04 Lifecycle Analysis 17
2.05 Logic Model 27
2.06 Pathway Model 30
2.07 Evaluation Scope 33
2.08 Program-System Links 36
2.09 Reflection and Synthesis 38
Stage 3: Evaluation Plan Creation 39
3.01 Evaluation Plan Overview 39
3.02 Evaluation Purpose 40
3.03 Evaluation Questions 40
3.04 Measurement and Measures 42
3.05 Sampling Plan 48
3.06 Evaluation Design 51
3.07 Data Management and Analysis 53
3.08 Evaluation Reporting Plan 54
3.09 Implementation Plan and Schedule 55
Evaluation Planning Summary 56
Glossary 57
Appendices 61

Sidebars: Systems Concepts

Greater than the Sum - Part-Whole Relationships 2
The Local and the Global - Scale 4
Simple Rules - Complexity and the Idea of a Protocol 7
The Rock and The Bird - Static and Dynamic Processes 8
The Eye of the Beholder - Multiple Perspectives 11
Inside-Outside--Boundaries 16
In the Course of a Lifetime - Ontogeny 22
The Flower and the Bee - Symbiosis and Co-Evolution 26
The Survival of Programs with “Fitness” - Evolution and Evaluation 28
And so on, and so on... - Causal Pathways 34
Driving With Your Eyes Open - Feedback 44

Activities

Evaluation Café 9
Launch Meeting 12
Stakeholder Affinity Diagram 14
Program History 18
Evaluation History 20
Lifecycle Alignment Review 25
Logic and Pathway Model Peer Review 32
Mining the Model - Part 1 35
Mining the Model - Part 2 37
Stakeholder Interview 38
Question-Claim Match-up, Stakeholder Review 42
Peer Evaluation Plan Brainstorm 43
Finalize Evaluation Plan 55

Figures

Figure 1 Phases of Program Evaluation 1
Figure 2 Protocol Terminology for System Hierarchy 3
Figure 3 Phases of Evaluation Planning 6
Figure 4 Hypothetical Map of Stakeholders 13
Figure 5 Phases in Program Lifecycle 18
Figure 6 Phases in Evaluation Lifecycle 19
Figure 7 Lifecycle Alignment 21
Figure 8 Program and Evaluation Lifecycle Definitions and Alignment 23
Figure 9 Sample Logic Model 29
Figure 10 Sample Path from Activity to Outcome 31
Figure 11 Pathway from Activity to Outcome 31
Figure 12 Program Pathway Model 32
Figure 13 Stakeholders Located on Program Pathway Model 35
Figure 14 Evaluation Scope 36
Figure 15 Literature Mapping (Golden Spike) 37

PREFACE

This *Guide to the Systems Evaluation Protocol* is intended to serve several purposes: it is first and foremost designed to be a step-by-step guide for both program and evaluation professionals who wish to integrate a systems evaluation perspective into their evaluation work in order to enhance the quality and appropriateness of program evaluations. The information provided in this Guide is designed to be comprehensive enough to allow the non-professional evaluator to use the materials, and in-depth enough to serve as a useful reference for the professional who is new to the Systems Evaluation Protocol (frequently referred to simply as the "Protocol").

This Protocol was created in the context of education and outreach programs generally and specifically for programs in Science, Technology, Engineering and Mathematics (STEM) education sponsored by the National Science Foundation (NSF) and in programs sponsored by Cornell Cooperative Extension. While many of the examples will be related to STEM and Extension education and outreach contexts, we have designed the Protocol to be generally applicable for any type of program evaluation context and we hope that a broader audience will find it useful.

We start from the assumption that the basic unit of interest in the use of this Protocol is a "program." The term "program" might be defined generally as "a series of activities conducted with the intention of producing some effect (outcomes) on participants." But even though this is the focal unit, it is important to recognize that from a systems perspective a program is always a part of a larger whole and is a whole to its subparts. That is, programs are often parts of collections of similar programs (or program areas) that are parts of their larger organizations which are in turn parts of networks and systems. In addition, programs have parts that consist of activities, people (both deliverers and participants), and so on. This Protocol continually incorporates these multiple system levels into the focus on a specific program.

The Introduction below goes beyond simply laying the groundwork for the steps of this Protocol. It is here that we address a second goal – that of providing an overview of our view of a "Systems Perspective" that shapes our approach to evaluation. The Systems Evaluation Protocol (SEP) has its foundations in the literatures of evaluation theory, systems theory, and evolutionary epistemology..

The Systems Perspective that shapes the Protocol highlights the value of having multiple voices and perspectives included in the evaluation process. Accordingly, this Guide is written with the assumption that the steps will (usually) be completed by a working group made up of internal program staff and possibly some external stakeholders who are close to the program. The working group may be large or small, balancing the risk of being unwieldy against the benefits of multiple perspectives.

It is important that a lead person be designated to guide the process. This could be someone hired by the organization as an external evaluator, an internal staff member assigned to the task, or someone selected by the working group. We will refer to this person as the "Evaluation Champion." This term refers to the person in this leadership role, rather than to any specific professional title or qualification. Because the Protocol consciously adopts a systems perspective, the Evaluation Champion should be thought of not only as a facilitator of the Protocol, but also as a driving force for addressing contextual factors of both the organization and the larger systems within which the program is embedded. Ideally, by engaging in the process of this Protocol, the organization will increase their evaluation capacity and build an evaluative culture, and program evaluation will become a routine aspect of program management.

For readers who are interested in learning more about our systems approach to evaluation, the Introduction should provide some insight and knowledge about the theoretical underpinnings of our Protocol. Throughout the Protocol there are green sidebars that will be of interest to

systems evaluation theorists. These are added for their supplementary value and will enhance understanding of the foundations of our approach. However, the reader should be able to use the Protocol even without this material.

At the same time, we hope that practitioners who simply want to start in and walk through utilizing the Protocol are able to do just that by beginning in Section II: The Systems Evaluation Protocol

New in this version: throughout v3.1, you will be directed to the accompanying *Workbook for The Systems Evaluation Protocol - Phase I Planning (v1.1)* for resource materials and worksheets for many of the tasks in the Protocol. We simply refer to it as the "Workbook". We have also identified areas throughout the Guide where accessing the freely available online software - The Netway - can make evaluation planning easier. As in the previous versions, throughout the Guide there are red inset boxes describing activities that could be used to guide a working group through a particular step. These are optional and are meant to provide suggestions or ideas on the process and can be adapted as needed. Additional information on these resources can be found at the web links included at the bottom of this page.

For more information about evaluation methodologies, terms and other background information the reader may wish to refer to the Research Methods Knowledge Base, a comprehensive web-based textbook that addresses all of the topics in a typical introductory undergraduate or graduate course in social research methods. The Research Methods Knowledge Base can be found online at: www.socialresearchmethods.net.

Get more information on this Guide, and its companion resources, online at:

https://core.human.cornell.edu/research/systems/protocol/index.cfm

and

https://core.human.cornell.edu/research/systems/netway.cfm

or

http://evaluationnetway.com

History of the Systems Evaluation Protocol

In 2004 Dr. Trochim was part of a collaborative effort at Cornell University to explore the role of Evidence-Based Practice (EBP) in Extension and Outreach. He began to articulate ideas on how program planning and evaluation methods could be better integrated into a variety of group and organizational contexts, and how they could be utilized to help research systems better translate science into practice. In 2005 he was appointed the responsibility of improving the evaluation capacity of Cornell Cooperative Extension. Throughout 2005 and 2006 he worked with twenty-one programs across two program areas at Cornell University Cooperative Extension in New York City (CUCE-NYC) (Nutrition & Health, and Family & Youth Development) to build evaluation capacity and to link evaluation more integrally with strategic planning within CUCE-NYC. He also began collaborating with an international team of researchers to investigate systems thinking approaches (e.g., systems organizing, systems dynamics modeling, network analysis, knowledge management) in public health that could enable better translation of science to practice. In his work with CUCE-NYC he began to outline the steps of the protocol for programs to develop evaluation plans.

After initial work with evaluation planning in CUCE-NYC, Dr. Trochim received NSF Award #0535492 to support inclusion of outreach programs in Science, Technology, Engineering, and Mathematics (STEM). In 2007 and 2008 thirty more programs across 8 organizations participated in evaluation capacity building trainings and developing evaluation plans. It became clear that having an evaluation plan in hand was not sufficient for many organizations to conduct evaluation, so workshops were also conducted to provide support on implementation of the evaluations. At the end of the project, *The Evaluation Facilitator's Guide to: Systems Evaluation Protocol* (which was actually only for evaluation planning) was printed and distributed, and was written for experienced evaluators interested in a systems approach to evaluation.

In August of 2008 Dr. Trochim entered into collaboration with Dr. Jennifer Brown Urban, at Montclair State University in New Jersey, to work on a 5-year follow-up project on the Systems Evaluation Protocol (NSF Award #0814364). (Dr. Urban had actually worked on the project while still a graduate student at Cornell, so she was completely familiar with the work.) Monica Hargraves and Jane Earle Buckley, two individuals who were program staff at organizations doing evaluation planning in the previous project, joined the Trochim team as facilitators of the Protocol. Their perspectives as recipients of the training, and their experiences with facilitating the Protocol with an additional 44 programs across 29 organizations, resulted in much new content being created. In addition to training materials, the Protocol steps had evolved significantly from the original version. In January 2012, the *Guide to the Systems Evaluation Protocol* was printed. Version 2.1 was targetted towards program practitioners. September 2012 saw some final edits to the evaluation planning Protocol, and was released as *Guide to the Systems Evaluation Protocol (Version 2.2*). It was a major output of the second NSF award.

In the summer of 2015, when the *Netway* - supporting cyberinfrastructure for evaluation planning - was publicly released (http://evaluationnetway.com) we updated the appendices from the Guide 2.2 and integrated them, along with with Netway resources (worksheets, FAQs), into a publication we called *The Workbook for the Systems Evaluation Protocol, Phase 1: Evaluation Planning* (also available at the website, opposite). The purpose of this current version (3.1) of the Guide is to remove the outdated appendices and to connect to the resources in the *Workbook* and the *Netway*.

We anticipate that the Guide to the Protocol will continue to undergo changes as our understanding of system interactions evolves and becomes more refined and that materials in the *Workbook* will be adapted to these changes. Our expectations are to make our materials available on our website (http://core.human.cornell.edu/research/systems/protocol/index.cfm) between publications of both the Guide and the Workbook, and we encourage feedback and discussion of our approach to systems evaluation.

For comments or questions please contact:

Dr. William M. Trochim
Professor, Policy Analysis & Management
Director, Cornell Office for Research on Evaluation (CORE)
wmt1@cornell.edu
CORE phone: 607-255-0397

Mailing Address:
Policy Analysis and Management
Martha Van Rensselaer Hall
Cornell University
Ithaca, NY 14853

Companion resources:

Hargraves, M. J. & Buckley, J. C. (Eds) (2015) Workbook for the Systems Evaluation Protocol. Cornell Office for Research on Evaluation, Ithaca, NY.
Available at: https://core.human.cornell.edu/research/systems/protocol/index.cfm

The Netway - for program and evaluation planning. Online at http://evaluationnetway.com.

Related Publications:

Urban, J. B., & Trochim, W. (2009). The Role of Evaluation in Research-Practice Integration Working Toward the ''Golden Spike''. American Journal of Evaluation, 30(4), 538-553.

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(0), 127-139.

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.

Urban, J. B., Burgermaster, M., Archibald, T., & Byrne, A. (2015). Relationships Between Quantitative Measures of Evaluation Plan and Program Model Quality and a Qualitative Measure of Participant Perceptions of an Evaluation Capacity Building Approach. Journal of Mixed Methods Research, 9(2), 154-177.

ACKNOWLEDGEMENTS

Many people worked on this project in its various stages. Although it is next to impossible to mention everyone who has influenced this work, we would like to be sure to mention the key collaborators who were supported by the sponsoring agencies.

CORE's Evaluation Facilitator roles have been filled at different times by Monica Hargraves (Extension Associate), who had also served as Evaluation Project Manager for CCE Tompkins County in the first project; Cathleen Kane, who had also served as Evaluation Project Manager for CCE New York City; and Jane Earle Buckley, who had also served as Evaluation Project Manager for the Cornell Center for Materials Research (CCMR).

Staff in project management and administration roles at Cornell included: Claire Hebbard (CORE Project Manager), Laura Colosi (Extension Associate, Policy Analysis and Management), Claire Lobdell, and Debra Perlmutter.

Graduate Research Assistants working on these projects over the years have included: Thomas Archibald (Cornell University, Education), Marissa Burgermaster (Montclair State University, Nutrition and Food Science), Derek Cabrera (Cornell University, Policy Analysis and Management, who also worked as a post-doc), David Carucci (Montclair State University, Education), Wanda Casillas (Cornell University, Human Development), Alison Cathles (Cornell University, CIPA), Larry Daffin (Montclair State University/ New York University, Administration, Leadership, and Technology), Sarah Hertzog (Cornell University, Human Development), Margaret Johnson (Cornell University, Policy Analysis and Management), Linda McMahan (Montclair State University, Education), and Jamila Walida Simon (Cornell University, Natural Resources). Most of these students are now professionals working in the field of evaluation.

Undergraduate Research Assistants on this NSF project at Cornell University and Montclair State University were vital for literature searches, finding measures and other organizational tasks, and included: Timothy Au, Amanda Bates, Lindsey Bober, Kelsey Brykman, Natalie Cook, Haixin Dang, Stacey Eilbaum, Miranda Fang, Joshua Hatcher Mullins, Anna Hays, En Chi (Angela) Hsu, Darlene Kim, E. Blythe McCoy, Nicole Monahan, Matthew Negrin, Oluwarantimi Damilola Oluwasegun, Stephanie Phelan, Sabrina Rahman, Drew Stadlen, Marissa J. Steinmetz, Helen Zaitsev and Anthony Zhu.

Special recognition goes to participants in the Evaluation Partnerships: Cornell Cooperative Extension (at different times under the leadership of CCE Director Dr. Helene Dillard, or Dr. Christopher B. Watkins); NSF Materials Research Science Engineering Centers (with organizational collaborations led through Cornell Center for Materials Research (CCMR), the Cornell MRSEC under the leadership of Dr. Melissa Hines); and the Museum of the Earth at the Paleontological Research Institution.

Special thanks go to Gorges Websites (Ithaca, NY) for their development on the Netway – our cyberinfrastructure: Matthew Clark, Jim Tse, John Sammis and Chris Grant.

Additionally, recognition for collaboration on the first phase of this project go to: Stephen Hamilton (Professor of Human Development, Cornell University), Nancy Trautmann (Senior Extension Associate, Cornell University), and Anna Waldron (Waldron Educational Consulting).

I. Introduction to Systems Evaluation

We use a systems perspective as the framework for developing evaluation capacity, enhancing evaluation quality and ultimately improving programs. Several theoretical systems constructs have guided our work including complexity theory, evolutionary theory and natural selection, general systems theory, ecology, system dynamics, developmental systems theories, and ideas on research-practice integration (including evidence-based practice and translational research).

The field of evaluation can itself be viewed as an evaluation system, and refers to the comprehensive and integrated set of capabilities, resources, activities and support mechanisms for conducting evaluation work. This should not be confused with systems evaluation, which refers to the assessment of the functions, products, outcomes and impacts of a system (set of programs, activities or interventions). Systems evaluation is an approach to conducting program evaluation that considers the complex factors that are inherent within the larger "structure" or "system" within which the program is embedded. Systems evaluation provides both a conceptual framework for thinking about evaluation systems and a set of specific methods and tools that enhances our ability to accomplish high-quality evaluation with integration across organizational levels and structures.

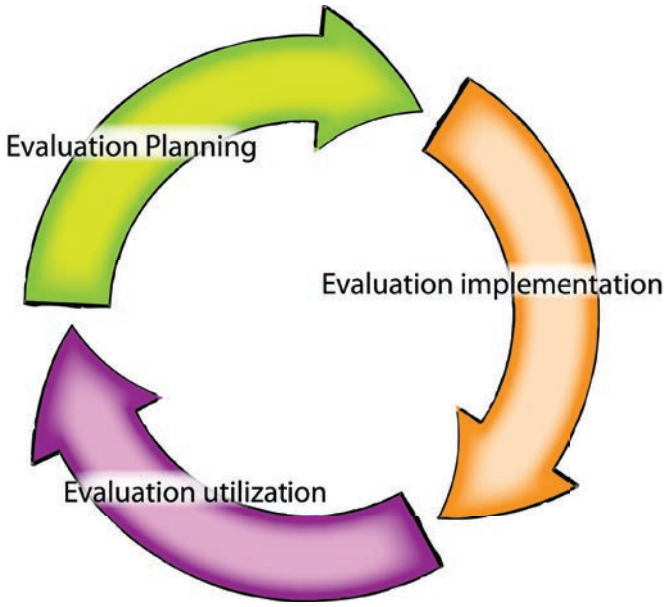
In the Systems Evaluation Protocol (SEP) we have tried to integrate principles associated with systems theories into program evaluations in order to assure that programs that use the SEP will incorporate such principles when developing program pathway models , identifying key pathways and nodes (outputs and outcomes), determining the boundary conditions for program models, assessing program lifecycles, and selecting evaluation designs that are appropriate to program evolution.

It was the examination of some of our beliefs about evaluation that led to our work in systems evaluation. Here are some of our assumptions about evaluation, which were precursors to developing a systems perspective of evaluation:

- Evaluation is a dynamic on-going process that is applied to programs that are (themselves) dynamic.
- Evaluation is a form of feedback that can be used for program or organizational improvement (see sidebar "*Driving with your Eyes Open*," page 44).
- A formalized and standardized evaluation framework utilizing a systems perspective is needed to develop consistent and high-quality evaluations.
- Program Evaluation is best viewed as a three-phase process, beginning with Evaluation Planning, followed by Evaluation Implementation, and completing the cycle with Evaluation Utilization (which leads back to planning for the next iteration of the program). **Figure 1** provides a pictorial representation of the Phases of Evaluation.
- The Evaluation Planning phase is a critical step for systems evaluation. This is where introducing the systems perspective will shape how evaluators and program staff view the program, program boundaries, stakeholders, and its evaluation.
- Because of the need to evaluate multiple programs within an organization, there is value in developing systems for evaluation that encompass multiple programs, rather than conducting isolated evaluations of individual programs.

Throughout the Guide there are sidebars that will be of interest to systems evaluation theorists. These are added for their supplementary value and will enhance understanding of the foundations of our approach. However, the reader should be able to use the Protocol even without this material.

FIGURE 1. PHASES OF PROGRAM EVALUATION



Greater Than the Sum - Part-Whole Relationships

Systems are by their very nature collections of multiple things, so it's not surprising that one of the most fundamental distinctions in systems theory is that of "part" and "whole." What do we mean by "part" and "whole" in a systems context? A part can be almost anything. For instance, it might be a piece in a machine – a wheel is a part of the larger whole that we call an automobile. Or it might be an organ in an organism – a heart is a part of a body. In most systems part-whole relationships exist in nested hierarchies. For instance, a hubcap is a part of the whole wheel which is in turn a part of the automobile which might in turn be considered part of a fleet of vehicles. Or, a cell is a part of the heart which is part of the body which is in turn part of a class or group of organisms. But the part-whole distinction is more than just a physical one. We can also talk about part-whole hierarchies in concepts. For instance, our idea of the concept of humanity consists of parts like nationalities and sub-parts like people from different states or towns. Or, we might divide the whole concept of humanity into the parts of those who were born in different years, subgroups who are male and female, and subgroups of those who have brown hair or are right-handed. As these examples show, the part-whole concept is a universal one that can be applied to virtually anything.

When thinking of parts and wholes it is also important to keep in mind that in addition to the whole and the parts that it is made up of, we also can think about the relationships between these as something that is distinguishable and meaningful (almost as if the relationships are separate "parts" of the part-whole distinction!) For instance, in a car there is the whole of the car, its various parts (e.g., wheels, engine) and the relationships between these. This idea of the importance of relationships is central to systems thinking. It gives rise to the famous saying that "the whole is greater than the sum of its parts" that you have probably heard of. We have to be careful about this saying, however. In mechanical systems, the whole is very often precisely the sum of its parts. You can take a car apart and reassemble it and it will work. However, in dynamic or living systems this is not the case – you cannot take a human body apart and then reassemble it and expect to have a working system. So, the phrase "the whole is greater than the sum" really is meant to refer to dynamic systems, not mechanical ones.



Part-whole relationships are everywhere in evaluation and it is critically important that we use these concepts in our evaluation work. For instance, an organization (whole) will often operate in multiple program areas (e.g., an educational or outreach organization might have programs for children, teens, adults and the elderly; or programs in health, education, environment, science, etc.); each program area (whole) might have multiple programs (parts); each program (whole) will usually consist of multiple activities (parts); each activity (whole) can typically be broken down into different tasks (parts); and so on. Or we can view different levels of part-whole hierarchies in terms of stakeholder groups. In an educational or outreach program we might think of stakeholders at the program level (participants, their families and program deliverers), the organizational level (program managers and organizational administrators), the local context (local officials or the local public), the funders level (e.g., state or national) and even the societal level (Congress or society as a whole). The idea of part-whole relationships is essential in the development and implementation of our programs. It is central to our description of the program, the development of logic or pathway models, and the analysis of stakeholders and their interests.

A Systems Perspective on Programs and Program Evaluation

Below are some key points that we believe should anchor a systems evaluator's perspective.

- An organization is a system, and is composed of a collection of parts (see sidebar *Greater Than the Sum*, page 2). Systems involve parts, wholes, and their interrelationships.
- Any program necessarily occurs within a complex environment composed of "nested systems". "Nested systems" refers to the structure where a system is embedded within another system, which is embedded within yet another system. For example, Ms. Smith's third grade class is a system within the entire third grade, which is part of the elementary school, which is part of the school district, which is part of the state school system, and so on.
- Human systems are dynamic (see sidebar *The Rock and The Bird*, page 8). A dynamic system is

necessarily composed of evolving relationships and programs.

- Consequently, evaluation needs to be dynamic and should change in order to successfully link with the needs and maturity of the program being evaluated (see sidebar *The Flower and the Bee,* page 26).
- Programs have lifecycles, and move through various phases. Different evaluation approaches are appropriate for different program phases. In other words, like programs, evaluations should evolve (see sidebar *In the Course of a Lifetime*, page 22).
- Many organizations have multiple programs and many programs are implemented in multiple organizations – perhaps the third grade consists of Ms. Smith's class, Mrs. Jones' class, and Mr. Perez's class, and at the same time there are many schools with numerous other third grade classes. But each school also has multiple grade levels, and you quickly see that systems can rapidly become complicated.
- Nested and dynamic systems create an environment where there are multiple perspectives. Each stakeholder has their own perspective (see sidebar *Eye of the Beholder*, page 11).
- Each stakeholder of a program has specific expertise, and brings a distinct perspective and motivation for evaluation. The comprehensive set of stakeholders should be identified and included in the evaluation design and/or evaluation planning process. Placing the program being evaluated at the center of a circle, and surrounding it with stakeholders placed closer to or further away from the program of focus – depending upon their relationship to that program – can create a representative map of stakeholders (see *Protocol step 2.01 – Stakeholder Analysis*). The perspectives of these stakeholders are described in relationship to their association with the program. A stakeholder within the organization would have a local perspective, and one that is placed further away would have a more global perspective (see sidebar *The Local and the Global*, page 4).

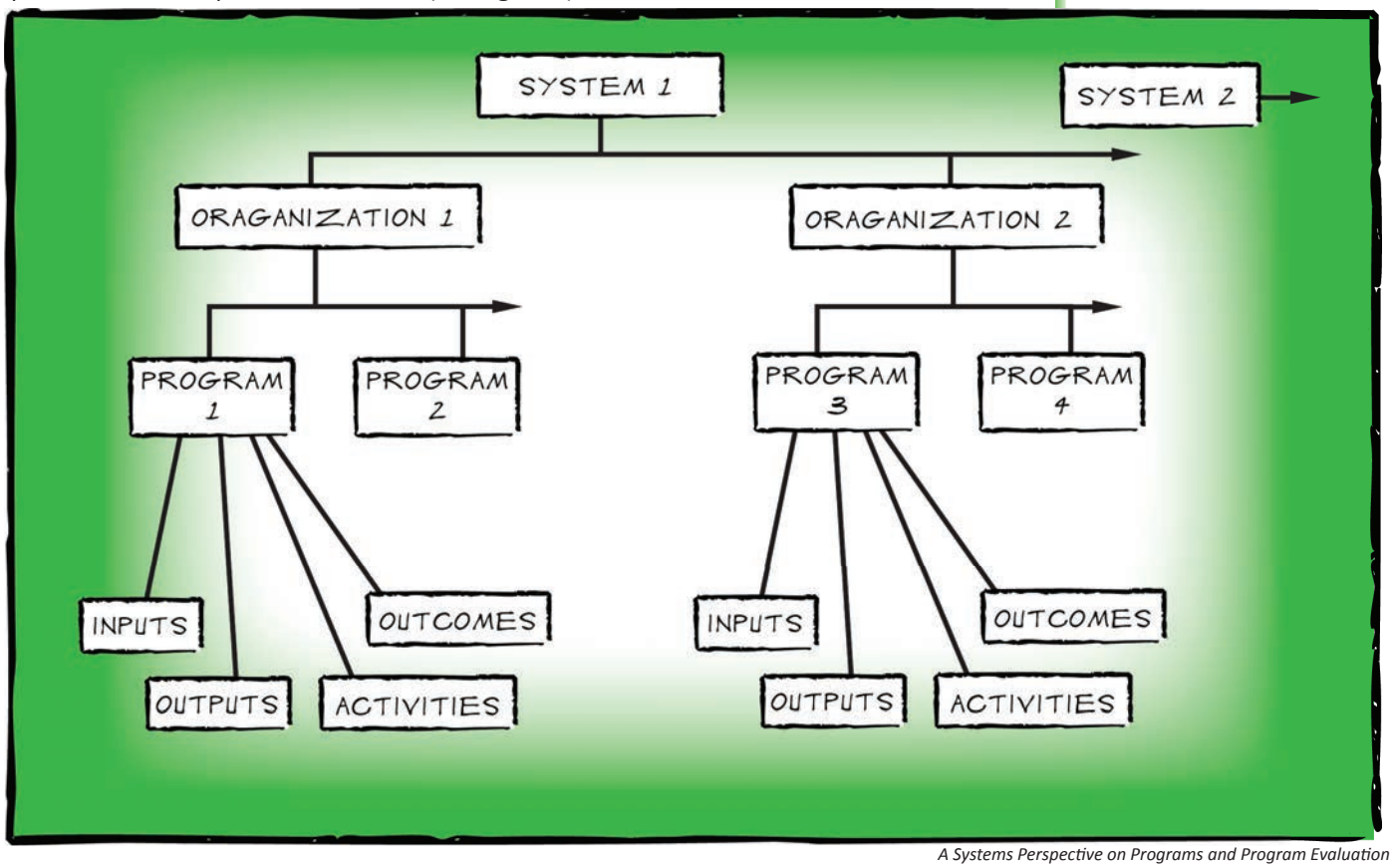
Red Activity boxes

Throughout the Guide the red activity boxes will suggest some exercises the Evaluation Champion may want to use as a group exercise in order to work through specific steps. These are optional, and are based on the authors' experience with facilitating the Protocol.

Terminology

Before proceeding further we would like to clarify/define some of our hierarchical (nested systems) terminology. This is not meant as a standard hierarchy applicable to all systems, but is simply a convenient hierarchy for describing how the steps of this Protocol might be applied when working with an organization. The broadest level is that of the system (rather like a network of organizations). Within a system there are assumed to be multiple organizations. Each organization may have one or more programs. Programs in turn are made of up multiple components, including inputs, activities, outputs and outcomes (see **Figure 2**).

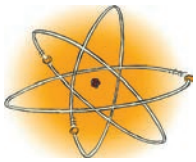
FIGURE 2. PROTOCOL TERMINOLOGY FOR SYSTEM HIERARCHY



The 'Local' and the 'Global' - Scale

In systems thinking we are always dealing with hierarchies of part-whole relationships. For instance, activities are parts of programs which are parts of collections of programs in an organization which may exist in a system of many similar organizations. When we think or talk about different levels of this kind of hierarchy we are operating at different levels of scale in the system. Physical part-whole hierarchies can exist from the subatomic level to the scale of the universe as a whole. Conceptual hierarchies can exist from the most general level (programs in general) to the most specific subcategory (the summer science youth camp in Ithaca, New York).

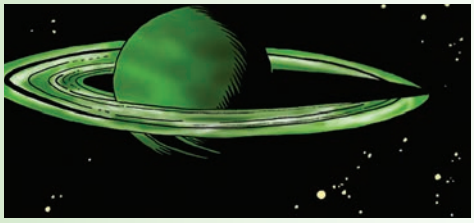
ATOM



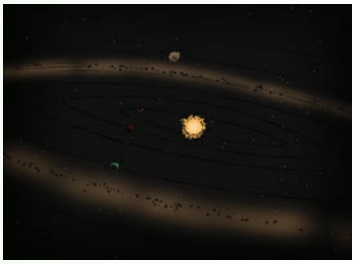
We can look at any system from many different viewpoints. For instance, if we are looking at an organization with multiple programs, each program can be viewed as a "part" in the system that constitutes the organization. When we talk about the relationship between a program and its organization, we can think of the program as "local" and the organization as "global" in relation to each other because they are at different but related levels in the hierarchy. On the other hand, when we compare or contrast two programs within an organization we can think of that as a "local–local" relationship because both are at the same level of scale in the hierarchy. If we shift our perspective to a higher level of scale, we are also shifting what we consider "local" and "global". For instance if we think about an organization as one part in a larger system of similar organizations (e.g., a county office in a state-wide system of such offices), then the organization becomes "local" to the system's "global". When we compare two county level offices, we are looking at a "local-local" relationship. When we look at the county level office in relation to the state office we have a "local-global" relationship because we are looking across different levels of scale.

Why are the ideas of scale and of local and global relationships important in evaluation? Different parts of a system don't exist in isolation. If we don't take them into consideration throughout our evaluation efforts we can run into significant problems that can jeopardize the whole endeavor. For instance, very often something in one part of a system may be in conflict with something in another part of a system. A program activity may conflict or compete with the activity of another program (a local – local relationship in a system) or with an organizational policy or effort (a local – global relationship). Or the expectations that stakeholders at one level of scale have for an evaluation may be very different than those of stakeholders at a different level of scale. Funders may expect that the evaluation will focus on accountability and impact while program implementers may be more interested in how evaluation can contribute more immediate feedback that can be used to improve programs. Evaluation should address both perspectives, and the process of evaluation can assist stakeholders to appreciate the complexity of the system.

PLANET



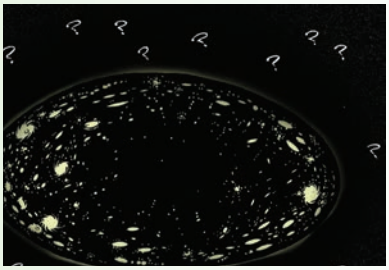
SOLAR SYSTEM



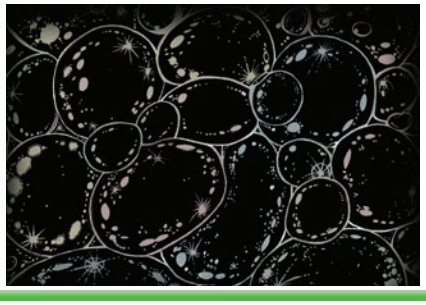
GALAXY



UNIVERSE



MULTIVERSE



II. The Systems Evaluation Protocol

The Guide to the Systems Evaluation Protocol (SEP) is more than just the sequence of steps and a list of factors to be considered when designing an evaluation – it describes the process of developing an evaluation plan. Throughout this Guide, we will refer to the reader as the "Evaluation Champion." The Guide specifically articulates the unique facilitation techniques and strategies that the Evaluation Champion may use, as well as the role that he or she plays when conducting systems evaluation. This term is intended to be inclusive, and applicable to any professional who may be using this Guide to plan or help plan an evaluation. An Evaluation Champion should be thought of not only as a facilitator of the Systems Evaluation Protocol but as a driving force encouraging everyone to think about evaluation, and to build evaluation activities into all program management and practice within the organization. In addition to the Evaluation Champion, we will also refer to the "Working Group." This is also intended to be an inclusive term, describing any members of the organization who are working together through the steps of the Protocol. In some cases this may include collaborating program staff exclusively, while in other cases this term may refer to members of the organization from various levels in the organizational hierarchy (program staff, administrators, funders) as well as participants and related stakeholders.

The process of working through the Protocol will consist of collaborative meetings that will seemingly spiral through several focal points over time, as well as ongoing work around building a culture of evaluation in the participating organization. This process is essential to the nature of the SEP. It is through these discussions that members if the organization and its program practitioners will develop a new outlook on their work that will change both their understanding of how the program stakeholders perceive the program, as well as their sense of purpose in what they are doing and why.

The SEP is a standardized protocol that nevertheless enables any program to develop an evaluation uniquely tailored to that program. In this sense it addresses the administrative need in an evaluation environment to standardize evaluation approaches while respecting the variety of contexts within which programming is conducted.

Putting evaluation concepts into a simple set of steps which we call the Systems Evaluation Protocol requires that we present the Guide in a linear format. In fact, an important objective for us in this work has been to instill the idea that effective modern evaluation requires evaluators to move beyond a linear mindset. Good evaluation requires feedback, and is embedded within a dynamic changing system. Although any written document is by definition linear, systems evaluation is a non-linear and iterative process (see sidebar *Simple Rules*, page 7). We expect that in various contexts it will be appropriate to perform steps out of the presented sequence or in tandem, as well as to revisit steps repeatedly throughout the process.

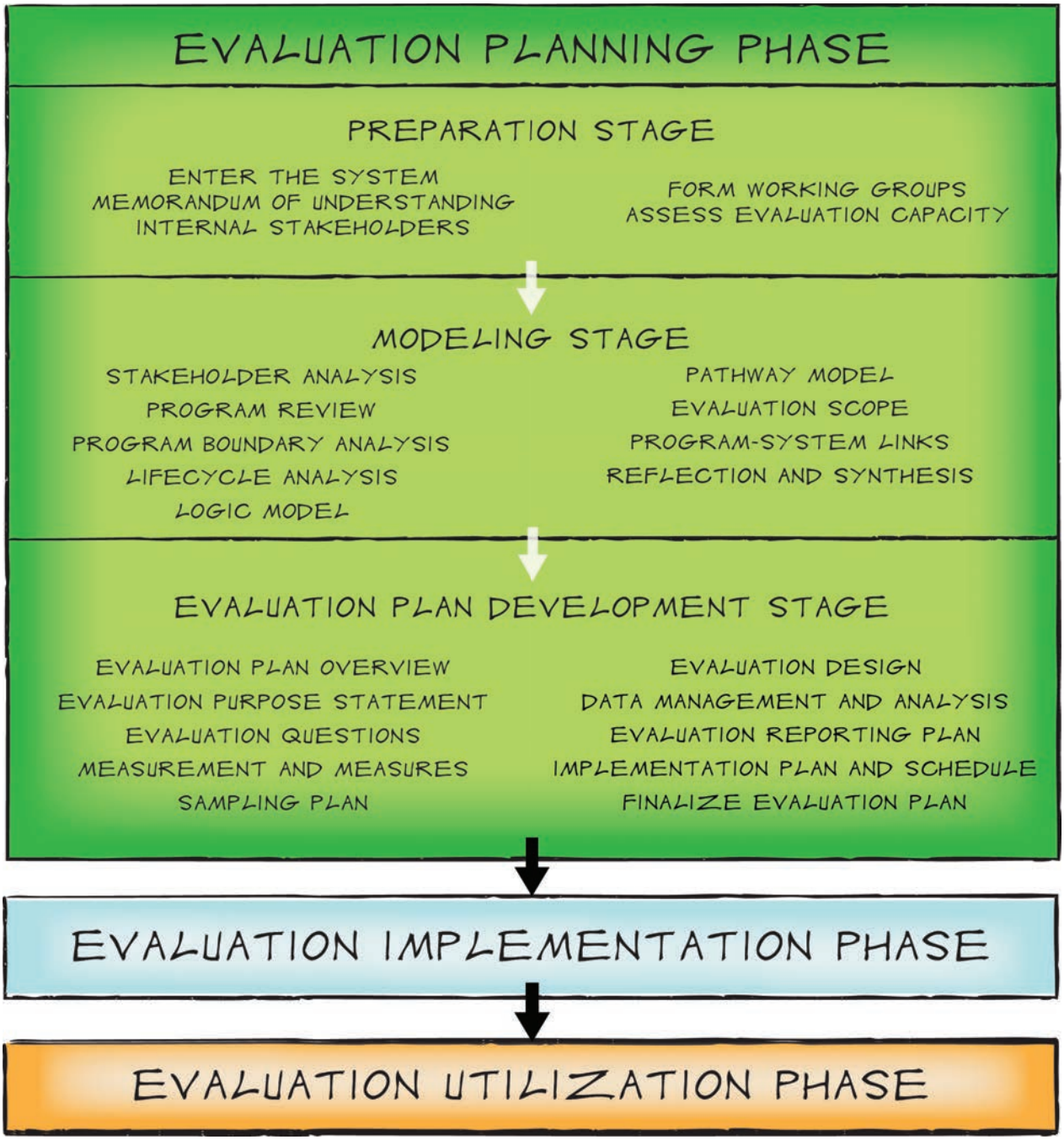
As a reminder, there are three phases to evaluation: Planning, Implementation, and Utilization. This Guide presents the Protocol for the first phase only.

Phase 1: Evaluation Planning

The purpose of the Evaluation Planning Phase is to create a logic and pathway model as well as an evaluation plan for the program.

We identify three separate stages within the Evaluation Planning Phase: 1: Preparation; 2: Modeling; and 3: Evaluation Plan Development (see **Figure 3**). Each of these stages, in turn, contains a number of individual steps. Once these stages are completed, the organization will have a comprehensive evaluation strategy that will guide the complete evaluation effort.

FIGURE 3.
PHASES OF EVALUATION PLANNING



Stage 1: Preparation

Together, the Evaluation Champion, and the participating program/organization form what is referred to as the Evaluation Partnership (EP). The Preparation Stage is intended to help you, the Evaluation Champion, enter into an organization (if you are not already an internal member of the organization) and establish the Evaluation Partnership (EP); or, if you are already a member of the organization, to help you establish your role as Evaluation Champion and make expectations about this process clear to your colleagues. During this stage you will identify people's key roles, acquaint the participants with the SEP process and establish baseline information regarding the target program(s) and organization.

Simple Rules - Complexity and the Idea of a Protocol

How do we get complex phenomena in nature? How do birds fly in formation or ants build complicated anthills? How is the ecosystem regulated? How do our bodies adapt to changes in the environment? All of these are questions related to complex adaptive systems. The recently developed field of complexity theory attempts to address these types of questions scientifically. Complexity theory holds that when we have multiple independent agents that follow simple rules and are provided with feedback, complex phenomena will emerge. For instance, when ants build an anthill or bees build a beehive, they don't first get together and do strategic planning! Instead, each ant or bee does what it is genetically and biologically programmed to do and the environment provides ongoing feedback. The complex structures that are anthills or beehives result from the millions of behaviors that are undertaken. Similarly, in human behavior, no one group planned a city like New York or Paris (although certainly city planners try to plan at least some aspects of them). These cities have evolved based on the choices made by millions of individuals over centuries. Each individual made countless decisions based on local circumstances and feedback in their own lives, and the complex cities we know today emerged as a result. And, the individuals who live and work in a city change over time and yet the city continues to exist despite a constant churning of specific inhabitants. Complexity theory seeks to model and identify the simple rules that can lead to the emergence of complex adaptive systems like cities.

The notion of "simple rules" can be extremely valuable in evaluation. If we recognize that an evaluation is undertaken in an ecosystem that involves a wide variety of different stakeholders (autonomous agents) each making their own decisions based on their understanding of local circumstances and feedback, we can readily see why evaluations can be such challenging endeavors. If we can provide a set of simple rules that multiple independent stakeholders can follow and that incorporate feedback as the process unfolds, we can help to shape an evaluation without trying to force it into a "one-size-fits-all" framework.

The idea of "simple rules" is a lot like the idea of a protocol in medicine. A medical protocol is simply a set of "rules" that one or more medical staff apply consistently whenever the circumstances call for it. They don't have to recreate the rules every time the triggering situation arises. Wonderfully complex and adaptive results can emerge from simple protocols. For instance, in basic first aid, there is a standard set of "rules" for first responders who use the acronym ABC: check for a clear AIRWAY; make sure the person is BREATHING; check CIRCULATION. The protocol helps individuals concentrate, makes sure they don't skip a step, and means that they don't have to reinvent processes each time they respond to a new emergency. From these simple rules complex phenomena can emerge, including saving someone's life!

In systems evaluation we are using the idea of simple rules to develop a standard protocol that anyone can follow when doing an evaluation. The protocol does not predetermine the result – each evaluation, like each life threatening emergency, is a unique event. The protocol can be useful for ensuring that we don't miss key steps in an evaluation and can lead to the emergence of unique and adaptive evaluation systems.

1.01 Enter the System

The SEP process begins with contact between the Evaluation Champion and the organization's and programs' key decision makers. Whether the Evaluation Champion is an external evaluator or an internal staff member, it is important in this step to make contact with all the key decision maker(s) in order to lay the foundation of a working relationship with the person(s) responsible for approving the Evaluation Partnership (EP). Part of this foundation should be a clear summary of what this process includes and its primary objectives.

1.02 Memorandum of Understanding

The Evaluation Champion and relevant decision maker(s) will need to identify, negotiate, and outline the responsibilities and expectations of the Evaluation Partnership (EP). All EP members should put the finalized commitments into a written Memorandum of Understanding (MOU). Why is a written MOU important? First, in order to develop such a product, the group needs to reach consensus about the roles of the key participants and the expectations of each. While this could be done implicitly without setting it out in written form, we assume that an evaluation is a dynamic

The Rock and the Bird - Static and Dynamic Processes



There is a parable in systems thinking that illustrates well the difference between static and dynamic processes. If you throw a rock into the air, you can predict with some accuracy where it will go. The harder you throw it the farther it will generally go. The higher you aim it, the higher its trajectory. And, if we eliminate the variability of the human thrower and use mechanical devices like a catapult we can predict even more accurately where the stone will go. A rock is a static object, one that cannot direct itself. On the other hand, if you throw a bird (gently, please!), there is virtually no way to predict which way the bird will go and where it will land. The bird can sense its surroundings and may head off in any direction. A bird is a living dynamic system that gathers and processes input and interacts with its environment.

This distinction between static and dynamic processes is important in systems theory and in evaluations that are done from a systems perspective. Since programs involve people and organizations they are inherently dynamic. It is difficult to predict where they will go and what will happen. As programs unfold the directions they take are influenced by the surroundings and by the interactions of the participants. In this sense, programs are more like birds than like rocks. On the other hand the idea of a "program" suggests that we are trying to do something systematic, that we are attempting to follow a pre-determined set of steps in order to achieve some predictable result. In this sense, programs are more static, they are more like the stone in the parable.



So, which is it? Are programs static or dynamic? Should our evaluations be constructed for one or for the other? The short answer is: both are important. Both the rock and the bird can be understood from a systems perspective. Both are parts in a larger whole. Both have relationships to the other parts. Over time programs are likely to evolve through different phases, some more static and others more dynamic. For instance, when a program is first being developed and piloted it is likely to be very dynamic and unpredictable. In fact, that dynamism is essential for learning and adaptation, for enhancing the focus and quality of the endeavor. Over time many programs will tend to become more static. They become routinized and develop standard operating procedures that can be implemented consistently. They can be transferred to new contexts and settings with some degree of predictability. This standardization is also essential. Over even a longer period of time the program may become too static or rigidified, or it may lead to insights that suggest even better variations that might be tried. In either case, we might be motivated to begin other cycles of dynamic-static program development and evolution.

Understanding the interplay of static and dynamic systems is essential for systems evaluation. We need to recognize that both have their place in evaluation and identify how our evaluation approaches need to evolve over time to encourage program evolution as well as provide feedback and learning about it.

and evolving entity and that such a document can help the participants to remain focused on the nature of what they have undertaken. While the understanding of roles and expectations may evolve over time – and the MOU can be revised accordingly – without such a document it is likely that there will be opportunities for misunderstanding and confusion. Second, a written MOU is itself a form of feedback. It signals to the entire organization what is happening. It encourages various internal stakeholders (**see step 1.03**) to react to it, and to state any interests or concerns they might have up front. Additionally, it helps these stakeholders to adapt to the changing circumstances that any evaluation necessarily introduces. Finally, an MOU is a way to be transparent. Without it, different organizational stakeholders might tend to misinterpret or misunderstand the nature of the work. With it, the organization has an objective written statement that describes who will be responsible for what in the evaluation.

The development of the MOU may take as long as a couple of weeks or more, depending upon the number of people involved and the scope of the work envisioned. The discussion (and eventual document) should identify the working group members by name and specific role. In addition, the MOU should detail the goals for the project, expectations for the Evaluation Champion, working group members, organization administrators, and organization staff.

There are several key points to negotiate at this time:

Project Goals. The goals of the project should be agreed upon as part of the MOU. If the primary focus is to plan and implement an evaluation for a single program, then the scope of work is smaller. If this is viewed as part of a larger process of building evaluation capacity and evaluation culture within the organization, then more time would be required for trainings, communication, broader staff meetings, and so on.

Who will be in the evaluation working group? (**See step 1.04**) Deciding who will be part of the working group that works directly with the Evaluation Champion is an important consideration. The Evaluation Champion can work with one or more working group members to train them on aspects of evaluation so that they become an evaluation resource to the organization.

Responsibilities and roles of the EP members. How will the key decision maker(s) support the evaluation working group and Evaluation Champion? What expectations do the key decision makers have of the evaluation working group and Evaluation Champion reporting back to them? What expectations do the Evaluation Champion and the working group have of each other?

Time commitments.

- General timeline of the expected process for a quality evaluation. Depending on priorities and time availability, it may take several months -- even up to 6 months or more -- just to create the evaluation plan, then more time to implement the plan and analyze the results. The time put into planning will pay off in the long run.
- Time commitments of staff and working group members. The members of the working group are typically comprised of staff members who are directly engaged with the program and are committed to evaluation. The working group members are the core team that works with the Evaluation Champion. Additional staff members may also assist at various points in the process. There will be at least one meeting with internal stakeholders, and several meetings with the working group members. There may be a need to educate the working group and other staff about evaluation, and eventually about the specifics of program modeling, measure development, and the like. These meetings will have to occur within the busy and demanding schedules of program staff. The time commitment of working group members will vary depending on how much support and collaboration the project goals require.
- Time commitments of the Evaluation Champion. The Evaluation Champion's role and responsibilities will be shaped by the project's goals and scope of work, and by the availability of and roles assigned to members of the working group. In cases where the Evaluation Champion is a staff member internal to the participating organization, this individual will be spending additional hours planning meetings and addressing questions. The additional responsibilities of the Evaluation Champion should be considered and planned for in light of their existing duties and this should be documented in the MOU.
- Policy regarding evaluation. Although many organizations have some sort of policy stating that evaluation is a job requirement, most managers do not think about the amount of time that this will take. Ten to fifteen percent of an employee's working time is not unreasonable. We encourage organizations to make their commitment more explicit by writing it down either as part of the MOU or elsewhere.

Costs and Budget. Costs could include fees for an external professional evaluator, if used, as well as the planning costs – travel, time for meetings, phone calls, room rental, fees for licenses, printing, etc. Also consider equipment – laptops, projectors, copiers, etc. Getting agreement in writing about the expected costs and how they will be covered is critical to determining the scope of and commitment to evaluation.

What is the timeline for this MOU? We have traditionally used a single MOU specifically for the evaluation planning phase, and then created a new one for the implementation phase – but this is not the only option. You may wish to address the entire evaluation process in a single MOU. But it is very important to be clear about expectations on the length of time each phase takes. Some programs have moved relatively quickly through the evaluation planning phase (4 months), while

Activity: Evaluation "Café"

This discussion format allows participants the opportunity to express their expectations and concerns out loud, and may be utilized with all internal stakeholders, or only within your working group (see page 10). During your first in-person meeting, team members should be broken up into small groups (2-5 people in each group). A discussion prompt is presented and at each "café" table individuals are given 2-3 minutes to think about and take personal notes on the prompt, then 5-7 minutes to share and summarize the group's responses. Participants then re-shuffle themselves, creating new small groups for the next prompt. This process can be repeated several times, though 3-4 prompts generally allow for ample discussion. At the end of the café sessions, small groups share the notes created at their tables with the entire group. Possible prompts include: "What comes to mind when you think about evaluation?" "What do you expect to get out of the evaluation partnership?" and "What are your concerns about evaluation and/or the evaluation partnership?"

others may take a year or more. We often begin by asking staff about their annual cycles – when are reports due? When is the next program being held? The answers to these questions will help determine a timeline that will work for the project.

At a minimum, the MOU is meant to be a vehicle to plan evaluation activities, and to help assure that all parties are clear about what is intended or expected, understand their individual roles as they pertain to evaluation, and agree on what can feasibly be accomplished in the timeframe allotted. This written agreement does not necessarily have to be a formal legal document, but it should be a consensus between all the involved parties, be put in writing, and be made available for easy reference. A sample MOU (titled *Memorandum of Understanding Template)* can be found in **Appendix I**. Although this will ideally be accomplished through a series of face-to-face discussions, we have often worked through this step via teleconferencing and email.

1.03 Internal Stakeholders

An important first step in the launch of the EP is to bring together the internal stakeholders of the organization and its programs in an introductory meeting. This typically goes beyond the working group to include directors, administrative staff, and staff from the target and other programs in the organization. The introductory meeting will help the organization understand the evaluation process, and what to expect as a result of participation. Not all of the people attending this meeting will participate in each and every step of the process, but it is particularly important that this first meeting be inclusive. One of the most common challenges of implementing the SEP is working with organizations and staff who hold inaccurate expectations about the process. Though the MOU does address many of these on paper, we have found that it is essential to talk directly with stakeholders and participants, either in person or over the phone. Thus, the Evaluation Champion has the opportunity to address any questions that may arise early in the process. One way of addressing these issues and allowing expectations, questions and concerns to be expressed is to use an activity such as the **Evaluation Café** (in the sidebar on page 9). In the Modeling stage of evaluation planning, you will consider the consider the perpectives of these stakeholders more indepth, along with other stakeholders' priorities for this program, the task for now is to focus on addressing questions or issues that evaluation planning may contribute to within the organization. Sometimes this group simply wants to be kept informed of what's going on, sometimes they want some additional training on evaluation so they can transfer these skills throughout the organization.

We offer a companion Workbook, which will be referenced frequently. The Workbook holds materials applicable to many of the SEP steps, including worksheets, FAQs, and other resources. Materials previously contained in the Appendices to this Guide are now in the Workbook (as well as on the Netway - more on that later).

For more information on locating these resources, see the Preface, page vi.

1.04 Working Group(s)

The working group will consist of the people who should be responsible for and directly involved in evaluation planning, implementation and utilization. This group should represent a range of perspectives from within the program, as well as the organization. If you are working with an organization to conduct evaluations on multiple programs you will need to determine if each program needs its own working group, or if they can start off together then break apart later. Things to consider when deciding which staff members to include in the working group are:

- Who needs to be present in order to obtain a complete picture of everything and everyone involved in and affected by the organization and its relevant program areas and programs, and the roles of key players?
- Who cares about the program areas and program, and why?
- Is there anyone who might be upset to later find out that they were not included in a conversation about evaluation?
- Who from the organization or program is able to participate in this process? The significant time required to participate should be carefully considered.
- If evaluation capacity will be measured, who is the best person or group of people to fill out the Organizational Evaluation Capacity Survey? This may be the organization's director or key decision maker. Who should complete the Program Evaluation Capacity Survey? This is usually the Evaluation Champion, program leader and/or staff.

As the evaluation effort continues, the working group(s) will likely become smaller – and more focused on the program staff. Maintain flexibility about who to involve in each step throughout the planning process.

The Eye of the Beholder - Multiple Perspectives

A system can be viewed from many different perspectives. Almost everyone is familiar with the famous drawing from the psychology of perception that shows either faces or a vase depending on how you look at it. When you stare at this picture you can actually experience the shift in perspective that psychologists have described as the "figure-ground" effect. The same system can seem very different when looked at from different viewpoints. We cannot really understand the system and its interdependencies unless we look at it from multiple points of view.

The issue of multiple perspectives is essential in evaluation for a number of reasons (depending on your perspective!). For instance, all program evaluations involve a multiplicity of stakeholders including the participants, program developers, administrators, support staff, families and community, funders, policymakers, politicians, and the general public. One of the most important things an evaluator can do is to help the different stakeholders see the system of a program from the perspectives of other stakeholders. For instance, program deliverers may not perceive why they are being pressured to evaluate their programs "from an outside perspective" or why they need to demonstrate outcomes and impacts. If they understand the system pressures on different stakeholders, in this case the funders, they may gain a greater appreciation of how their view fits into the larger system. Conversely, funders may not understand why the organizations they fund are resisting their calls for evaluation. If they can begin to view the program through the eyes of those who deliver it or participate in it they are likely to understand the system better. In this example it's easy to see that the issue of perspective is intimately related to the motivations and incentives of different stakeholders. The field of evaluation has long emphasized the values of participatory evaluation approaches, in part because of this critical importance of multiple perspectives.



But multiple perspectives are also critical for understanding the content and meaning of programs. Throughout an evaluation it is valuable to have key stakeholders look at different parts of the program, to share their views, and to consider how others might perceive them. For instance, it is surprising how many times even in simple programs different people will have remarkably different views of what they are trying to do or what the program is affecting. We find that when people share their perspectives they can uncover such differences and that this learning is critical for informing the evaluation.

When working with multiple programs simultaneously, it can be helpful to create and share a contact list of all participants so that they can contact each other easily and begin building an evaluation network.

1.05 Evaluation Capacity

It is generally a good idea to assess the organization's and program's capacity surrounding evaluation prior to developing and implementing an evaluation plan. This process can help to elucidate resources available within the organization and program as well as identify areas that may need additional attention. In addition, it helps to establish a baseline assessment of evaluation capacity prior to implementing the SEP. By working through the SEP, the Evaluation Champion and the working group will hopefully help to build the evaluation capacity of both the program and organization.

Assess Evaluation Capacity

Organizational capacity can be assessed formally by completing an evaluation capacity survey that focuses on organizational resources. It should be completed by individuals who have the information available to accurately assess evaluation capacity at the level of the organization. Typically, in order to assess organizational capacity, an organization leader (who has a broad perspective on the organization and its programs) should be asked to reflect on: a) the evaluation resources available within the organization (personnel, budget, and technology), b) any organizational evaluation policies, and c) the current evaluation capacity of the staff.

Program capacity can be assessed formally by completing a capacity survey focused on the program. The survey should be completed by staff who are familiar with how the program currently handles evaluation. The Evaluation Champion and the working group should be involved in assessing program evaluation capacity. The survey should be completed for each of the programs that will be evaluated. This includes gathering basic program information such as a description of the program, its participants (i.e. numbers, and demographic characteristics) and other key data that will be used for the logic model - the inputs (i.e. staff, curricula), activities, outcomes, context, and assumptions. It also includes gathering information on any evaluations of the program that have occurred to date.

The purpose of this step is to establish a baseline and also to get a sense of what is actually involved in running each of the programs. The benefit of using a survey is that it can be revisited after the organization has implemented the evaluation and used to assess whether any changes have occurred after participating in the SEP. **Appendix II**, *Assessing Evaluation Capacity*, outlines many of the issues that you may wish to integrate into your own survey(s).

Activity: Launch Meeting

This working meeting or "workshop" is an event that may include several of the activities described in Stage 2. The purpose of this meeting is to introduce the Systems Evaluation Protocol and begin working through some of the modeling steps. The initial contact (described in Stage 1) and the "evaluation café" activity presented earlier help to establish expectations for the evaluation partnership and a working definition for "evaluation." Though these concepts will be revisited, the launch meeting is the time when these expectations and ideas are first put into practice. The Launch may or may not include all internal stakeholders, but definately will include all members of the working group who will be involved in the evaluation planning process. Other internal stakeholders, including organization administration, need not attend the entire workshop. However, inviting these stakeholders to learn about the organization's commitment to evaluation and how it will affect them, as well as to participate in conversations around 2.01 -2.03 (Stakeholder Analysis, Program Review, and Program Boundary Analysis) may be useful. This workshop could take on a variety of different formats depending on the needs of the working group. In some cases, the "launch" meeting might last several days and include the completion of a majority of the steps in the Protocol. In other instances, the working group may decide to work through only the logic modelling (Step 2.05), then pause to take time to more fully consider, revise and share their program model with stakeholders. (Additional meetings will not be listed specifically in this Guide because there is no specific organized timetable for the completion of the steps in this Protocol. An example of how we scheduled meetings when we worked with multiple programs as a single cohort going through evaluation planning together is shown on page 6 of the sample MOU in Appendix I.)

Stage 2: Program Model Development

This stage is intended to enhance the working group members' knowledge of evaluation concepts and how to use systems approaches for analyzing programs, and to develop and fine tune their program's Logic Model and Pathway Model upon which the evaluation will be based. In the previous stage you (the Evaluation Champion) were working with the organization as a whole. During this stage you will be working primarily at the level of the program (but there may be some overlap between programs if you are working with more than one program in an organization). Of course, because the SEP takes a systems view, you will be considering the broader organizational stakeholder context. However, in this part of the Protocol you will primarily work with staff of each individual program.

Program documents - including stakeholder maps, program descriptions, logic models, and pathway models - as well as evaluation plans can be built online in the Netway, allowing you to develop valuable program development and evaluation tools in collaboration with colleagues and make the information available to your team members. The Netway is available free at http://evaluationnetway.com

This stage involves several steps that address: Stakeholder Analysis, Program Review, Program Boundary Analysis, Lifecycle Analysis, Logic Model, Pathway Model, Evaluation Scope, Program-System Links, and Reflection and Synthesis. Many of these steps *could* be completed by the Evaluation Champion alone. However, the process of developing a program model depends on the contribution of stakeholders and is strengthened by collaboration and discussion among program staff. Collaboration also builds support for the evaluation. Many of the following steps may be performed in a different sequence or in tandem and will build upon each other in an iterative manner.

2.01 Stakeholder Analysis

The goal of this step is to identify all of the potential people and/or organizations that have a stake in the program and its evaluation, and to begin to understand their perspectives on the program and its evaluation. This should be a broad and inclusive brainstorming exercise. The working group should be encouraged to name every possible entity at all levels of the system, from program participants to state and national funding sources. Use of a whiteboard or sticky notes are suggested so that the stakeholder names can be physically placed and moved on a diagram. The following questions may help guide the conversation:

- Who are the people/types of people who have a stake in the program? Who benefits? Who is responsible for the program? Who takes part in it? Who encounters those who take part in it? Who experiences it indirectly? Whose lives are affected by it? Who has to pay for it? Who has to make decisions about it? Note: Local stakeholders may include the funding agencies, participants, program leaders, parents, administrators, staff, board, local and county government, the press, etc.

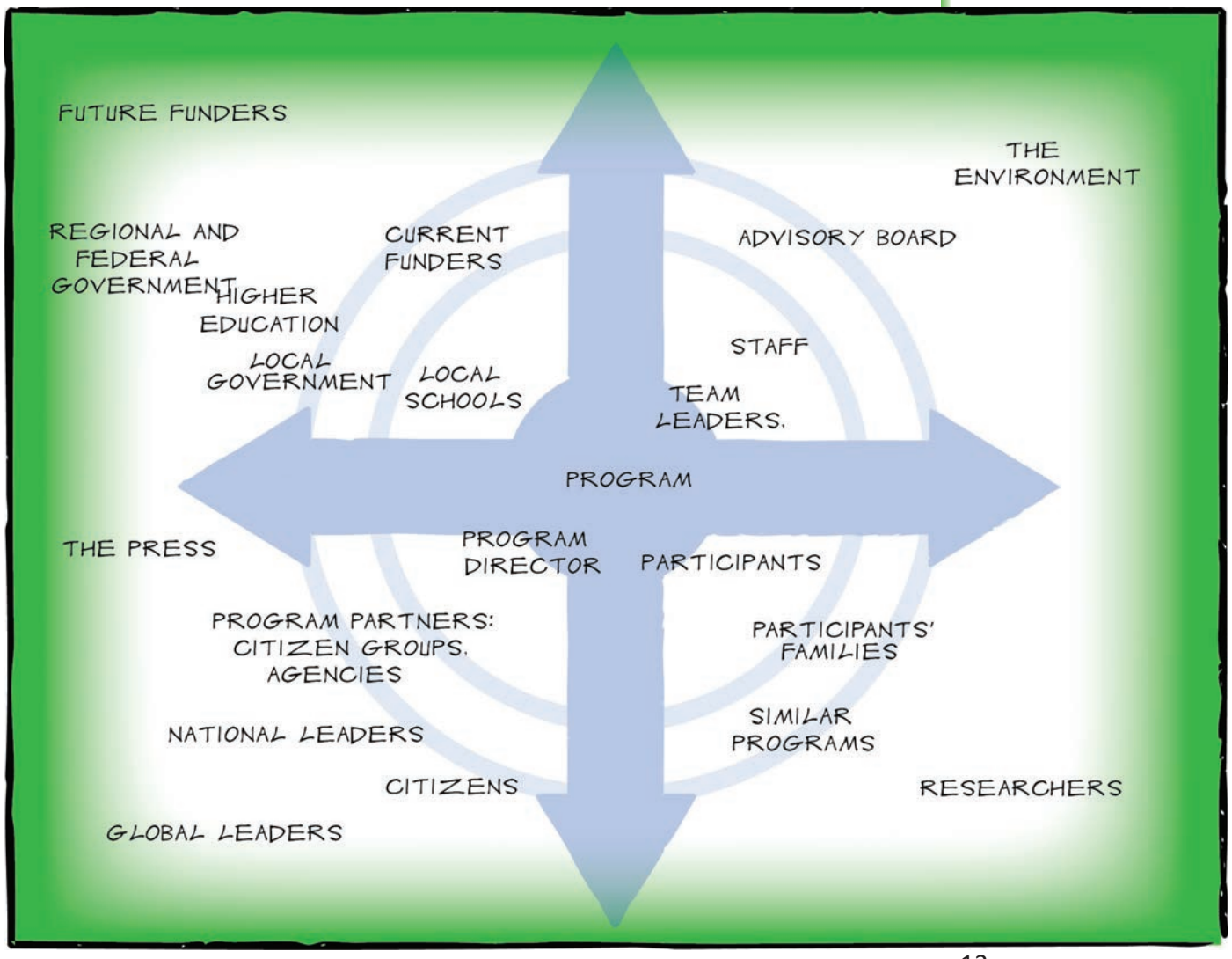
- Who else cares about the program, or at least the program's general scope? This refers to people beyond the immediate scope of the program, and could include the community, schools, policy makers, researchers, a potential future funding agency, other organizations with similar or related programs, global issue leaders, etc.

Building your Stakeholder Map in the Netway makes editing easy - you can use colors to indicate sub-groups or systems colors and move stakeholders about the map to arrange them.

Create a "Map of Stakeholders" - a visual depiction of the stakeholders and their relationship to each other (see **Figure 4: Hypothetical Map of Stakeholders**). This is an informal map that is designed to show all the key stakeholders or stakeholder groups at a glance. In general the stakeholders most centrally involved with the program should be nearer the center of the map, and others who are more remotely related should be at the outer circles. You might want to arrange the stakeholders so that similar groups are near each other, but the most important thing is to identify all of the relevant stakeholders and ensure that everyone in the working group is comfortable with the map. You can find a *Stakeholder Worksheet: Blank Map Template* in the **Workbook (page 5)**. It is likely you will have to go through several iterations to produce a map that everyone is comfortable with.

FIGURE 4. HYPOTHETICAL MAP OF STAKEHOLDERS

Guide the group to consider which aspects of the program each stakeholder is most interested



Activity: Stakeholder Affinity Diagram

For larger groups, and especially colleagues working on different but related programs, an additional level of analysis of stakeholders can be useful. This is usually done at the launch meeting. After brainstorming a list of stakeholders, have one program member list the stakeholders individually onto separate pieces of paper. Ask them to place the stakeholders on the wall, grouping their stakeholders near their peers' similar/common stakeholders. (We use index cards and a sticky wall, but post-it notes should also work for any wall.) Rather than taking turns in a formal sense, participants should just add their items and possibly move others as they find a place to do so. It is important to allow participants to use their own criteria for similarity so that affinity clusters develop organically. When the diagram is complete (when everyone is satisfied that none need to be moved), participants should reflect on and discuss the process and what they noticed. They may assign identifiers or titles to the clusters that have developed - this may help them to identify key stakeholders on their own diagrams and additional stakeholders they had not previously recognized.

in learning about. It can be helpful for later steps if you record the identified interests of each stakeholder. See the *Stakeholder Perspectives Worksheet* in the **Workbook on page 7**. Another activity may be to conduct stakeholder interviews - consider using the *Stakeholder Interview Guide* on **pages 9-10 in the Workbook**.

2.02 Program Review

The goal of this 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. The output will be a fairly comprehensive program description, which will probably be more helpful than any benchmark initial description previously used for a program. This is an iterative step - any tools that have been used or developed in previous steps in order to describe the program should be discussed and reviewed – such as the capacity surveys and stakeholder map. It is also appropriate to examine any curricula, guidelines, manuals or prior evaluations that might exist in relation to the program. They are likely to come back to this step, later.
The following questions may help guide the discussions:

- 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?

Even though these questions are listed from the broadest to the more specific, that is not necessarily the best order to follow when working with a group. Sometimes the best place to start is with what program staff actually do – brainstorm the activities of the program – and then build from there to the more general issues of program mission/vision. This may be best if the program is underway or well-established and the working group includes active program managers. For programs that are being developed, or that are especially mission-driven, it may work best to start with broad themes and work toward discussing specifics. Regardless of the sequence, the discussion should be inclusive and ideally should help build relationships within the working group and between the working group and the Evaluation Champion.

This step should be approached knowing that some of these questions can probably never be completely answered to everyone's satisfaction. There may be some areas of disagreement or responses for which answers are uncertain, but it is not necessary to have complete agreement at this point. The descriptions outlined should be fairly comprehensive so as to "fill out" a relatively complete picture of the elements of the program. This effort to define the program can be surprisingly difficult. This is an important discussion to have before moving on to the next step, which will be to set some boundaries for the program.

In our experience this step has been an exciting one for many programs. It offers staff the opportunity to step back and reflect on ALL that goes into their program. Sometimes some unexpected connections arise as they recognize the contributions of program components they may previously have undervalued. When you're doing this step it is probably a good idea to use a white board or black board and draw things out as the group is saying them.

The working group shouldn't get ahead of themselves —don't start drawing the logic or pathway models yet. Stay focused on encouraging brainstorming. The next few steps will give shape to what is generated in this step. The result of this step will typically be a tangled mess that shows the level of detail in the program. The complexity of the program may even surprise many in the working group, who previously may have had a rather narrow, organized view of the program.

2.03 Program Boundary Analysis

Program names are simply labels given to a set of related activities and goals. Understanding the meaning behind these labels and constructs is the purpose of this exercise. This task builds off the

previous one, and is frequently one of the most difficult steps because there are many ways to define, structure, and parse the elements of a program. The goal of this step is to determine what is "in the program" and what is "outside of the program". While this may seem simple or unnecessary at first glance, the boundaries of a given program are seldom made explicit and frequently vary with the context. Different people within the same organization often draw a program's boundaries quite differently, and other stakeholders define it still more diversely.



The Program Boundary Analysis exercise is focused on language and terminology, and asks participants to clarify and make precise the statements they make about their program and may take for granted. Boundaries are artificial constructs created by humans, and now is the time to question exactly what those boundaries are. (For resources that may help guide this discussion see **Workbook pages 11-15***, Guidance for Boundary Analysis, and Program Boundary Worksheet.)* The output should be an improved program description, improved understanding of the program, and possibly new common ground for those who participated in this conversation.
Keep in mind that staff cannot realistically evaluate every aspect of their program during this evaluation cycle, but they may evaluate different parts at different times. Participants should be encouraged to be broad and to include all the things that matter and go into or come out of the program even though they may not be evaluated in this evaluation cycle. Below are several guidelines for thinking about what the program is, and developing a comprehensive program description.

The level of detail may lead to frustration or confusion when trying to make sense of the complexity of even the simplest program.

When it's more articulated, the Netway provides an ideal place to record your program's mission and description.

- Ask the program staff exactly what it is that they do, and ask other stakeholders to describe what they 'think of' as the program effort. The chances are that you will get different answers, but all of those things are part of clarifying the boundaries of the program.
- What are the "elevator stories" of the program – if someone were to describe the program in two sentences during a short elevator ride, what would they say?
- Is there a formal definition of the program? If there is such a definition, this can be the working program definition and boundary (or at least a starting point for a new one). If not, are there any informal descriptions of the program? Did you come up with a better definition during the previous step (2.02 Program Review?) Places to look include program descriptions on a website, or various types of promotional literature. Do these say the same thing, or are they different? Synthesize a working definition from these descriptions and move on.
- Another thought to consider is what statements would they like to make about their program? If they want to say that their program increases community health, then the community should be within the boundaries of their program.
- If they were to package up the program and hand it over to someone else, what elements would it consist of? Usually the program staff training is not considered part of the program itself, assuming that the staff come to the program adequately trained to carry out the program (note, in that case, this would be one of your assumptions for later on in the Logic Model).
- Look at the Map of Stakeholders. How would these various stakeholders define the program? Would strangers, reporters, board members, funders, etc. be able to "get" what the program is doing if they read your description? Would they include or exclude elements that have been included in the description? Have the working group assume the perspectives of different stakeholders or stakeholder groups and play the roles of those people. How are they similar to or different from how another stakeholder would describe the program? Perspective taking is absolutely critical to understanding the program. For example, how does a teacher define "school", as opposed to a student, the parents, or even the government? Why do different stakeholders, including program staff, describe the program differently (i.e. draw the program boundaries in different places)? What different information or values inform these differing descriptions? Can the group reconcile these

IT IS IMPORTANT TO NOTE THAT THERE IS NO ONE CORRECT WAY TO DRAW PROGRAM BOUNDARIES, BUT THERE ARE WAYS OF DRAWING BOUNDARIES THAT WILL BE MORE OR LESS USEFUL FOR YOUR PURPOSES.

differing views in a meaningful way?

- Do you describe your program differently in internal communications (such as memos or program plans) than in external communications (such as websites or mailings)? Does the program description include the information from both types of sources?
- How would the program evaluation be affected if elements were included or excluded from the program description? For example, what would happen if you included or excluded activities that are aimed at different audiences but share similar resources and goals?
- Look back at any previous program logic models. Is the information in there (including inputs, participants/audience, activities, outcomes, and assumptions) still in the picture with this new program description? If you have excluded or included elements, are you able to justify that?
- Write a new program description with the boundaries clearly established, and in language that would be understandable to someone with no knowledge of the program.

One way to think about the drawing of common boundaries is to consider it from the perspective of a biological scientist. Imagine that the program is a specimen and you are examining it under the microscope. You can zoom all the way in and see all of the intricate details of a portion of the specimen. Alternatively, you can zoom all the way out so that you can see the broader structure

Inside-Outside - Boundaries

All systems have boundaries that distinguish the system from what's outside it. That sounds simple enough, and for many systems it is a relatively simple thing to define what's in the system and what's not. But for other systems the boundaries are continuous (not abrupt) and are not easily defined. For instance, in nature how do we define the exact borders of the system that constitutes an organism, or a river, or a cloud? What is the "boundary" between two different breeds of dogs or two different species of animals? In living systems, the boundaries may be different depending on how you look at the system and the level of scale or precision at which you look.

In systems evaluation, defining boundaries is a very challenging endeavor. For instance, where does one draw the boundaries on who is a stakeholder to a program? In an educational program, for instance, do you limit the stakeholders to the program participants and deliverers? Do you include family members? Administrators? Funders? The public? Or, how do you determine what the boundaries of your program are? In many situations, we think of the program as a set of activities that we can list. However, when we actually try listing program activities we can often find that even co-workers involved in delivering the same program may have different items. For instance, one person might say the planning of the program or training of program staff is an essential "part" of the program, while others would say the program just consists of what is done once the program is planned and the staff is trained. Is one right and the other wrong? Even if we take the narrower version, we can run into difficulties. Two trained staff members who try to do exactly the same set of activities will inevitably do things slightly differently. A teacher will adapt the way they are presenting material depending on the reactions of students. A doctor will adapt the way they are treating someone depending on their pain level or initial response to treatment. Is that adaptation part of what we call the "program"? What exactly is the boundary of the program? The same kind of boundary problem occurs in relation to outcomes. If we have a science outreach program that is trying to influence children's attitudes towards science, where do we draw the boundaries on what that means? Does that mean that children become more interested in science? And what does that mean? What do we include in "science?" What do we mean by "more interested?" All of these questions involve determining boundaries, often in circumstances where there simply are no fixed and easily determined borders between what is or is not in the system.

Developing an understanding of boundary issues is an important part of systems evaluation. There are no simple answers and often reasonable people involved in the same program will disagree. In some sense, boundary discussions require that stakeholders negotiate a consensus about what they mean by their "program." For instance, in a teacher-training program, is the program just the set of activities used in training teachers or does it also include the activities that the teachers subsequently do in training their students? Discussions about program boundaries often become important learning events for stakeholders because they lead to discussions about the meaning of what they are doing with their programs and the evaluations of them.

of the whole specimen. Or, you might want your level of magnification to be somewhere in the middle so that you maintain a sense of the broader structure while still gaining an understanding of some of the more nuanced details. It is important to note that there is no one "correct" way to draw program boundaries, but there are ways of drawing boundaries that will be more or less useful for your purposes. To continue the biology example, the boundaries of the system may be the incubator full of petri dishes, a single dish, a single colony, a single cell, or a cell organ. These are all systems that are nested within other systems, and the job of the working group is to define their boundaries for their purposes. They should be guided by the kinds of statements they wish to be able to make at the end of the analysis (i.e. "X% of sample dishes share evidence of…" vs. "cell walls in the treated colonies showed the effect of…").

This conversation can help guide the efforts toward a tighter boundary with finer detail, or a wider boundary with broader components. The objective is to get the definition of the program to a point where it is not too detailed and not too vague. Here another helpful metaphor from the story of Goldilocks and the Three Bears comes to mind - Goldilocks' quest to find the perfect bowl of porridge which was neither too hot nor too cold, but just right. Also demonstrated in this example is that we did not focus on the whole Goldilocks' story – instead we chose to focus on the porridge (rather than on the chairs or the beds, or on the Bears' return).

Another example of different possible boundaries - consider how someone might define "school"? Does school begin when the first class begins, or does it begin when a student walks out of their home or gets on the bus? Is lunch part of school? Does school end after the last class, or when a student gets home? Are after-school activities – such as drama or sports – part of school? The selection among these options is ultimately a practical one driven by what participants want to do with their evaluation results. Under the circumstances, the decision about boundaries tends to get revisited several times as the working group moves through the modeling and evaluation planning steps.

For a program that is designed to "train the trainer" a key boundary question is whether the program includes the activities done with the trainers or only the subsequent activities with their students. Another classic boundary question is whether program planning and preparatory training is considered part of the program, or do program activities only involve what happens after program staff is trained?

The output of this whole step should be a newly revised and even more precise program description upon which the working group has reached consensus. At this point everyone should agree on what activities are considered to be part of the program, and which ones are not. In addition, the working group should develop a mission statement for the program. Often, the organization will have an existing mission statement for each of its programs. The working group should keep a copy of this statement with their evaluation files. If no mission statement exists, or the existing statement needs revision, this should be done as part of this step. (See *Guidance for Program Descriptions*, **Workbook page 17**.)

Some groups get very ambitious about defining the program boundaries, and they may pay a price later when trying to evaluate it. If their boundary includes an extended view of the program, their whole program model becomes much bigger and more complicated. Some groups will try to scale back and define more precisely what they are trying to do and they run the danger of leaving key program components out. This is all an iterative process that will be revisited throughout the other steps. Redefining the boundaries will likely continue as you work through other steps, such as when completing the logic model, when determining the evaluation scope, or prior to the next cycle of evaluation.

2.04 Lifecycle Analysis

Programs change over time. In fact, like organisms, programs can be viewed as progressing through lifecycle stages: they are initiated (born); they typically go through phases of rapid change and growth; they may stabilize and become more "settled"; they may be disseminated widely; and at any point along the way they may be retired or replaced. Integrating principles from both systems

Activity: Program History

The purpose of this activity is for participants to reflect on the origins of their program, and to discuss how and why the program has changed over time. Participant working groups should begin by discussing their program's evolution. The following prompt is designed to help the working group bring the program's history to light:

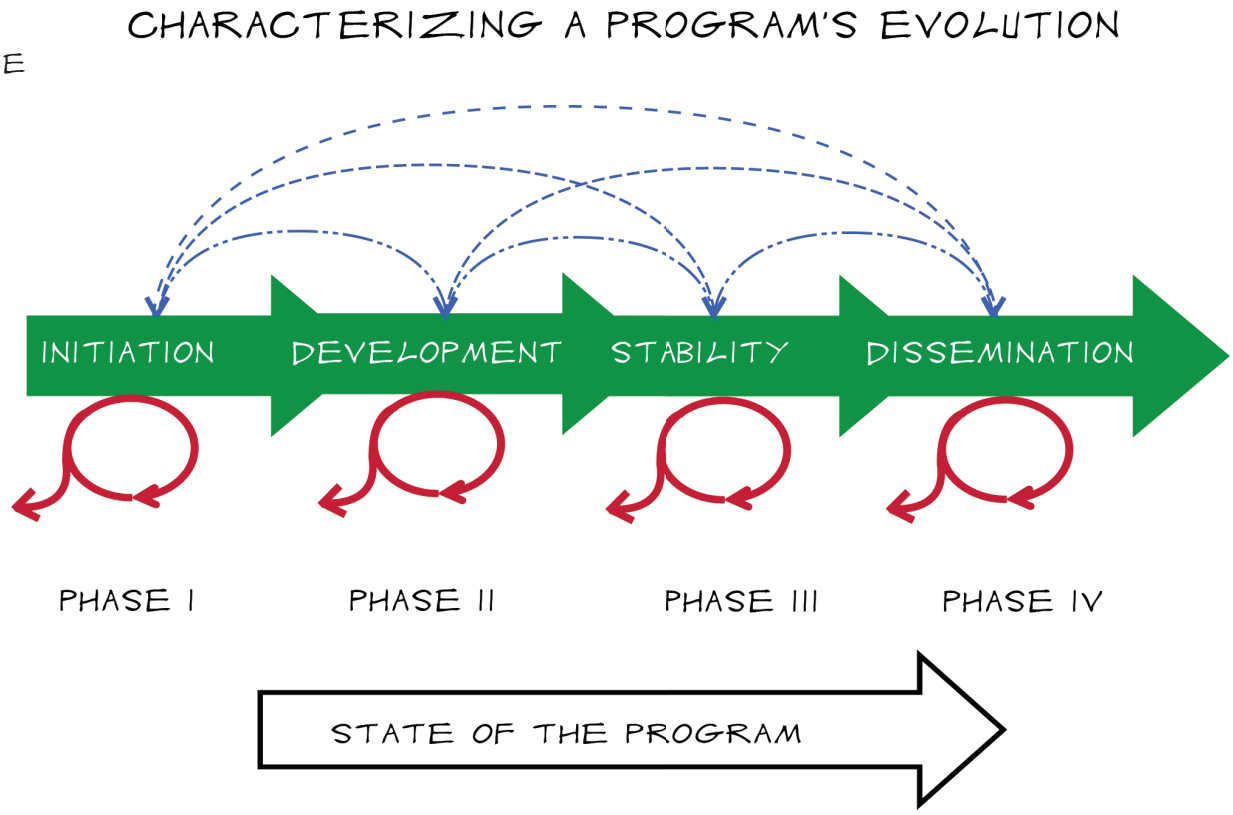
• Think of yourself as biographers working to tell the life story of your program. As you recall your program's history, specific considerations might include: how and why the program began, how it has changed over time and why, and the degree to which it continues to change.

Following the discussion, the group should try to capture the program's history in a schematic way - consider using either Workbook page 27, or any other graphic representation that enables the group to best communicate the story of the program's history. If there are several working groups engaged at the same time, use the opportunity to have working groups take turns sharing and describing their illustration and program history with the larger group. These brief presentations should include a description of how the program has evolved so far and a brief description of the factors that have influenced its evolution - funding changes, community needs, and so on. Answers to FAQs and/or Program History and Lifecycle Worksheet can be found in the Workbook (pages 20-28).

theories and evolutionary theories, the SEP was explicitly designed to identify where a program is in its lifecycle, and to encourage a progression through evaluation phases appropriate to its lifecycle phase.

Figure 5 offers a way of characterizing a program's evolution. The "State of the Program" arrow emphasizes that it is not just the passage of time that marks a program's evolution. Decisions are made throughout a program's lifecycle by program staff, organization leadership, key funders and others. Over time, these decisions contribute to a substantive progression that includes refinement and stabilization of program content and approach (reducing the variability of the program from one round to the next as a program settles into its essential components). In other words, as the state of the program moves from left to right, the internal stability of the program increases. This progression in the state of the program also reflects decisions that are made along the way about a program's expansion, continuation, or contraction, and also generally reflects a move from smaller-

FIGURE 5. PHASES IN PROGRAM LIFECYCLE



scale pilot trials of a program to more widespread use.

For practical purposes we define four broad lifecycle "phases": initiation; development; stability; and dissemination. In actuality, program evolution is a continuous and dynamic process. In fact, a program is never "done evolving." Moreover, an individual program's evolution is not always linear, and different activities may be at different developemental levels. The iterative, "regroup-and-try-again" possibilities symbolized in the blue dashed lines are realistic (and important) paths. There can be "backward" reversions to an earlier phase at any point in a program's lifecycle, even for mature programs. Programs may also stay in one phase or move incrementally within it for some time (symbolized by the red circles); and they may be retired at any point (the exiting red arrows).

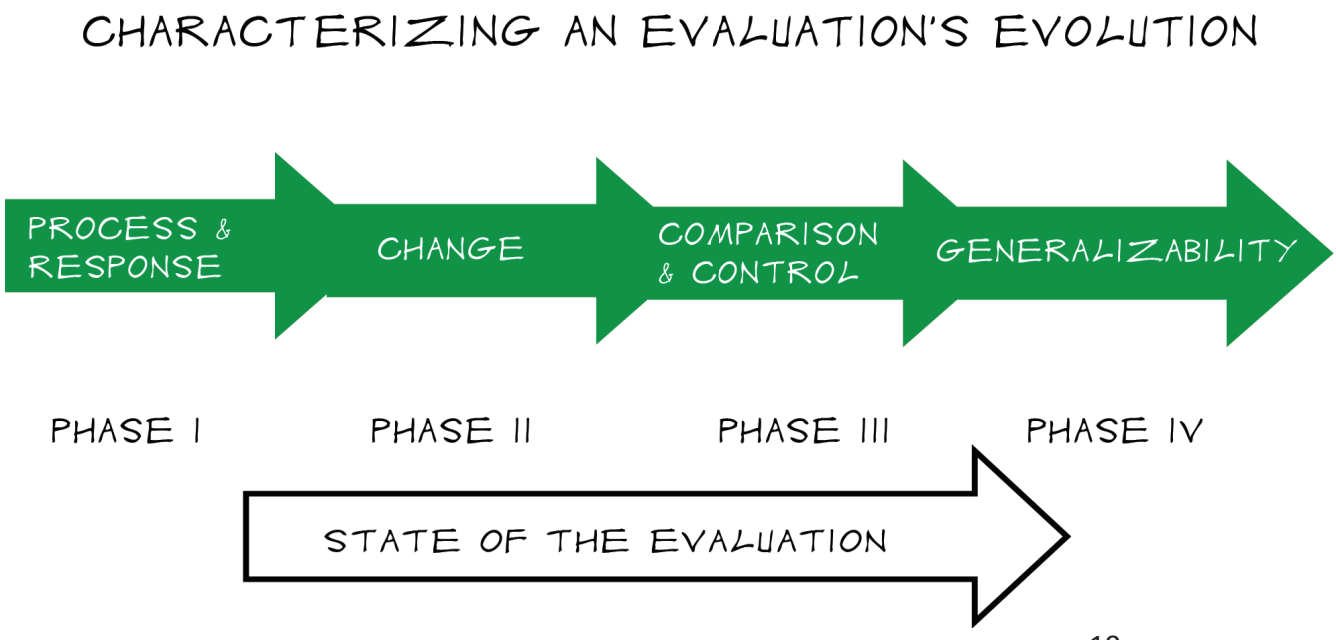
Each iteration of a program is related to the program's prior history but is also shaped by decisions based on new information about how and how well the program works, about what is needed by the target audiences or community, and by purely external factors like funding availability. The process of evolution involves learning, changing, and ultimately strengthening the larger system as a program is run, evaluated, revised, and re-run over time. Use *Program History and Lifecycle Worksheet* (**Workbook pages 27-28**) to plot your own program's lifecycle.

The process of program evolution through lifecycle phases is driven, in part, by evaluation. Information gathered through evaluation can be used to make positive changes to a program's implementation and scope, pushing the program forward - and sometimes backward - through the lifecycle stages. Accompanying and supporting the program's evolution is a similarly evolving pattern of evaluation activities. A key tenet of the SEP is that there are lifecycles in evaluation as well as the program, and that for any given program lifecycle phase or state of the program, there is an appropriate evaluation lifecycle phase. These lifecycle phases are defined in more detail below.

Figure 6 provides an image for evaluation lifecycle phases, analogous to the program lifecycle phases above. The "State of the Evaluation" is a synthesis of the multiple dimensions of a program evaluation. Movements from left to right in this figure correspond to potential increases in the scope and/or intensity of the evaluation effort. We distinguish four broad phases, delineated according to their basic purpose or goals. Early phase evaluations focus on how well the program is being implemented and how participants are responding to it; evaluations in the next phase assess change associated with program participation; evaluations with more elaborate comparison and control group designs allow for examination of causality, and the fourth phase examines how generalizable the program's results are likely to be to other contexts and settings. Use the Evaluation HIstory and Lifecycle Worksheet (**Workbook pages 29-30**) to diagram your program's evaluation lifecycle.

If you use the Netway for Evaluation Planning, you may record your program lifecycle in the "Information" section of your program, and your evaluation lifcycle in the "Evaluation Plan" section of your program.

FIGURE 6. PHASES IN EVALUATION LIFECYCLE



Activity: Evaluation History

This activity parallels the previous activity. Participants will reflect on the evaluation that has been done on this program over time, on what they know about the program and how they know it, and the factors that have shaped the evaluation decisions. To the extent that the underlying program may have moved forward or backward among lifecycle phases at various points in time, it may be that the evaluation as well has moved iteratively back and forth among evaluation phases along the continuum in Figure 6. Exploring a program's history of evaluation illuminates both the evolution of the state of evaluation, and also the factors shaping evaluation decisions.

The prompt questions below are designed to help the working group think through the evaluation history.
What types of evaluation have been done on this program, and when? (Include both formal and informal evaluations and feedback, whether or not it was officially recognized as "evaluation".)

- ***What have been the primary purposes of the evaluations that have been done?***
- ***How have results been used? Have the results influenced decisions about the program?***
- ***Who have the results been shared with or reported to?***
- ***Does your funder require a certain type of evaluation?***
- ***What have been the sources of information that have been used? (Documents, responses from individuals on surveys or in interviews, etc.)***
- ***Have there been evaluations that are descriptive of the program (e.g. interviews with program participants or leaders)?***
- ***Have there been formal evaluations of the program that included structured data collection (qualitative or quantitative data)?***
- ***Have there been evaluations that used comparison groups or control groups?***

Following the discussion, the group should try to capture the program's evaluation history in a schematic way similar to the program history (See Workbook Pages 29-30). If there are several working groups engaged at the same time, use the opportunity to have working groups take turns sharing and describing their illustration and the program's evaluation history with the larger group. These brief presentations should include a description of how the program has been evaluated over time and a brief description of the factors that have influenced the decisions about evaluation (stakeholder priorities, evaluation funding changes, program challenges, and so on.)

Alignment

Alignment between program and evaluation lifecycle phases is essential for ensuring that programs obtain the kind of information that is most needed at any given program lifecycle phase, and that program and evaluation resources are used efficiently. New programs should generally be doing process and implementation evaluations and basic satisfaction surveys rather than more controlled pre-post assessments. These programs are still changing a great deal, and need basic rapid feedback that can be incorporated into the next round of implementation.

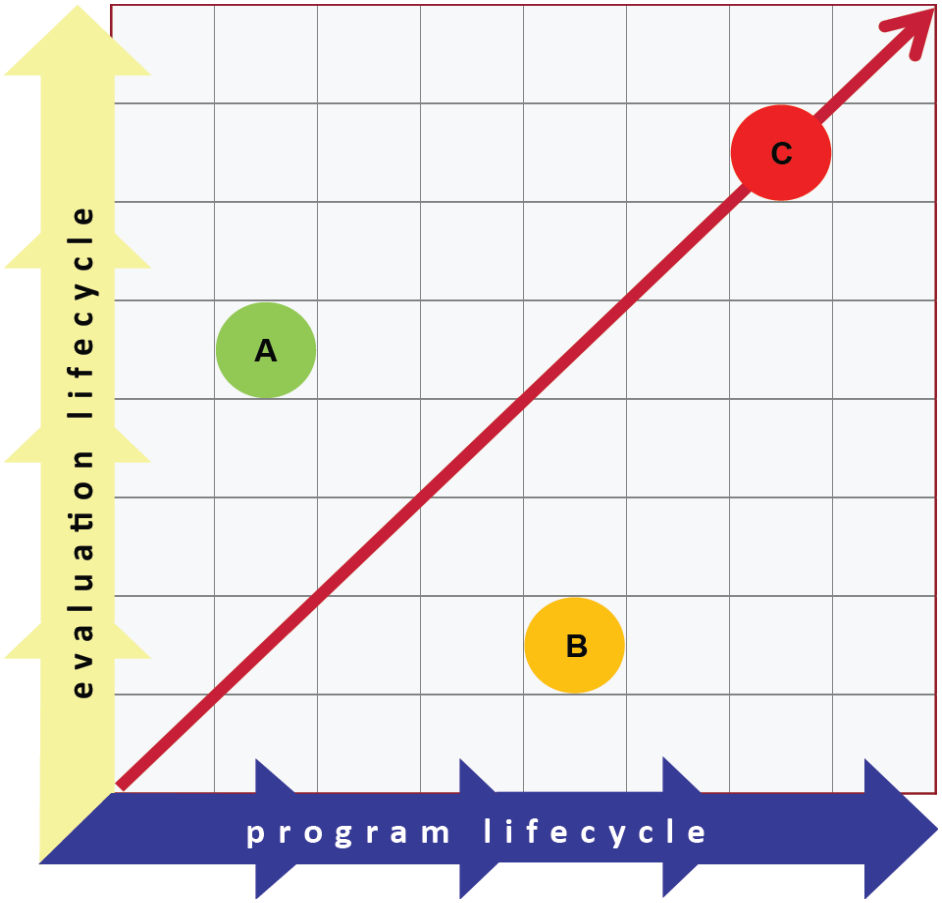
Using sophisticated outcome evaluation strategies on a program that's still in an early lifecycle phase is more than just a waste of resources. Outcome evaluation for an early phase program might happen to yield favorable results, but since the program is still changing considerably this seemingly favorable outcome might not hold up in subsequent rounds of the program, and could lead to an over investment in something that has not yet stabilized. The opposite risk is also significant: early outcome evaluations might show poor results and lead to the premature cancellation of a program that actually has great promise but needs to have some basic weaknesses resolved.

On the other hand, managers of a mature, consistently-presented and well-received program typically need to make decisions about whether to re-commit or even expand the resources being devoted to it. At that phase it is critical to evaluate program outcomes, and obtain evidence of change associated with or possibly caused by the program. This program might not really be attaining its intended outcomes and should be retired or substantially revised in order to meet a community need; alternatively it may be an extremely valuable program that is not being disseminated as widely as it should be, because it cannot build a strong enough case to funders. Without appropriate evaluation, program resources will not be allocated as well as they could be. Questions of participant or facilitator satisfaction alone simply would not serve the program well.

Figure 7 offers a simple representation of what alignment and non-alignment of program and evaluation lifecycles looks like, using the sequential lifecycle phases we have established above. In this illustration, "Program C" lies on a 45-degree line indicating ideal alignment between its program and evaluation lifecycle phases.

In practice it is very common to have programs whose program and evaluation lifecycles are not aligned. Early phase programs, such as Program A in **Figure 7,** may face stakeholder pressure to evaluate effectiveness through sophisticated outcome evaluations even though the program is still changing considerably from one round to the next. As described above, this pressured over-investment in evaluation may be using more resources than are warranted at this stage, and raises the risk of poor program decisions. Program B may have been in place and stable for some time, but is using simple end-of-session satisfaction surveys. This under-investment in evaluation *also* raises

FIGURE 7. LIFECYCLE ALIGNMENT



the risk of poor program decisions. This non-alignment may simply be the consequence of being trapped in a rut of familiar routines, or it may be that the program is constrained with insufficient resources (knowledge or funding). These are common and realistic scenarios, but misalignment can be costly and increases the risk of poor decisions about programs. Moving toward alignment should be treated as a key goal of the evaluation plan. Using the lifecycle definitions on page 23 (also in **Workbook on page 23**) and the *Lifecycle Identification and Alignment Worksheet* (**Workbook pages 31-32**), assess your lifecycles' alignment and discuss its implications for your program.

Program Lifecycle Definitions

Every program has its own unique development. Any given program might move forward and backward through and between the phases, although the general tendency will be to progress through the phases over successive implementations. We have briefly touched on the lifecycle phases, the following section presents more information on each phase. We define four broad phases, each of which is then more finely separated into two sub-parts.

1) Program Phase I: Initiation

A Phase I program is typically a new program that is just starting up or an existing program that has

been overhauled and revised considerably and is being piloted in its new form. Programs in this initiation phase will almost inevitably go through revisions.

a) Phase IA programs are in their initial implementation(s) either as a newly conceived program or as an existing program adapted from another context or from basic research.

b) Phase IB programs have been through initial trials but are still relatively new and are still going through substantial changes or revisions to major parts of the program.

2) Program Phase II: Development

Programs are considered to be in the development phase when they have been implemented successfully and are still undergoing some refinement, but, compared to those in Phase I, the scope and pace of change are much smaller.

a) Phase IIA programs are going through significant changes but some program elements have settled into consistent patterns.

b) Phase IIB programs are still going through change in some components, but most program elements are being implemented consistently.

In the Course of a Lifetime - Ontogeny



An evolutionary systems evaluation perspective leads us to think differently about programs. For instance, the idea of ontogeny in evolutionary theory is concerned with the origin and the development of an organism from inception through maturity. In human organisms ontogeny refers to the different phases of development from infancy to old age. Instead of thinking of our programs as static entities, this notion encourages us to think of each program as continuously evolving through different phases in a lifecycle, much like any organism does. While this lifecycle will manifest itself differently for each program, much as different people develop at different rates at various times in their lives, we can sketch out a hypothetical sequence that would likely fit many programs into multiple phases like the initiation, development, maturity or stability, and implementation or dissemination phase.

These stages aren't meant to be a strait-jacket or an inflexible taxonomy. For any given program, the progression may not be sequential. Like some people, a program may be precocious. It may for instance quickly evolve through the development phase and become stabilized or routinized. Or, a program can revert to an earlier stage, much like the young adult that temporarily reverts to juvenile behavior before resuming more mature development. At any phase, we may decide whether to continue the program or not. Sometimes it is apparent even early in a program's development that it is not able to be implemented well or that it has a fundamental flaw in its conception or structure.

This notion of a program lifecycle has practical implications for evaluation. How should a program be evaluated at each stage of its lifecycle? In organizations that are simultaneously running multiple programs – and most organizations do this routinely – what are the advantages of thinking about the collection of programs as constituting a type of portfolio and encouraging variation of programs at different stages of development? What role can evaluators play in helping program administrators and organizations assess where their programs are in their development and in encouraging them to think about when and how they will evolve their programs to their next phase?

In many of our program contexts, we become committed to the program as it currently exists. The program evolves up to a point and then we get a type of "lock-in" where we seemingly get stuck in a phase and are unable to move any further. Program decisions turn into a struggle between program preservationists who fear change and the potential loss of their familiar context, resources, or even their jobs, and program critics who push for ever-extending demonstrable results and emphasize ever-shrinking funding and resources.

An evolutionary perspective on programs and the idea of ontogeny emphasize program change as something to be expected and embraced. Instead of the commitment to preserving the program as it is, they encourage the idea that programs have a limited life-span, that they should not be assumed to live forever, that it is normal to see them as part of an ongoing trial-and-error learning process, and that the abandonment of an older program and the development of new ones is part of the normal cycle-of-life. From the beginning of the program, and throughout its evolution, the focus is on where the program is in its development and how to move it to the next phase. In effect the idea of a lifecycle creates system pressure to move programs along and not allow them to become static.

PROGRAM AND EVALUATION LIFECYCLE DEFINITIONS AND ALIGNMENT

PROGRAM LIFECYCLE		EVALUATION LIFECYCLE
Phase I Initiation		**Phase I Process & Response**
Program is in initial implementation(s), either as a brand new program or as an adaptation of an existing program.	PHASE IA	*Examines implementation, participant and facilitator satisfaction. Uses process and participant documentation and assessment and post-only evaluation of reactions and satisfaction.*
Program still undergoing rapid or substantial change or revision, after initial trials.	PHASE IB	*Focuses on implementation, and increasingly on presence or absence of selected outcomes. Evaluation is post-only; outcome measures are under development with attention to internal consistency (reliability).*
Phase II Development		**Phase II Change**
Scale and scope of revisions are smaller; most program elements are still developing while a few may be implemented consistently.	PHASE IIA	*Examines program's association with change in group outcomes, for these participants in this context. Uses unmatched pre- and post-test of outcomes, quantitative/ qualitative assessment of change, assessment of measure reliability and validity.*
Most program elements are implemented consistently; minor changes may still take place as some elements may still be developing.	PHASE IIB	*Examines program's association with change in group (and/ or individual) outcomes, for these participants in this context. Uses matched pre- and post-test of outcomes, quantitative/ qualitative assessment of change, verifying measure reliability and validity.*
Phase III Stability		**Phase III Comparison & Control**
Program is implemented consistently; participant experience from one implementation to the next is relatively stable (formal lessons or curricula exist).	PHASE IIIA	*Assesses eFFectiveness using design and statistical controls and comparisons (control groups, control variables or statistical controls).*
Program has formal written procedures/protocol and can be implemented consistently by new facilitators.	PHASE IIIB	*Assesses eFFectiveness using controlled experiments or quasi-experiments (randomized experiment; regression-discontinuity.)*
Phase IV Dissemination		**Phase IV Generalizability**
Program is being implemented in multiple sites; adaptations to new contexts have been made.	PHASE IVA	*Examines outcome effectiveness across wider range of contexts. Multi-site analysis of integrated large data sets over multiple waves of program implementation.*
Program is fully protocolized and is being widely distributed	PHASE IVB	*Formal assessment across multiple program implementations that enable general assertions about this program in a wide variety of contexts (e.g., meta-analysis).*

FIGURE 8. PROGRAM AND EVALUATION LIFECYCLE DEFINITIONS AND ALIGNMENT

3) Program Phase III: Stability

Programs are considered to be in the stability phase when they are being implemented consistently. Program planners and providers know what can be expected in implementing the program; there are relatively few surprises. Participant experiences are relatively consistent from one session to the next. The program has finalized its procedures and protocols.

a) Phase IIIA programs are being implemented consistently and have lesson plans or curricula to guide facilitators.

b) Phase IIIB programs have formal written procedures or protocols that make it possible for new facilitators working in that context to deliver the program consistently.

4) Program Phase IV: Dissemination

The dissemination phase is a period when the program is adapted for wider implementation while still adhering to the essentials of the program model. Logistical issues regarding support of the program over a broader range of circumstances are addressed. In short, dissemination phase programs are run at multiple new locations with new and diverse sites, staff and participants.

a) Phase IVA programs are being implemented in multiple sites in different contexts; adaptations to new contexts have been made in order to maintain the essential meaning of the program.

b) Phase IVB programs are in wide distribution, well beyond the initial context in which it was developed and used.

Most programs do not progress all the way through to the dissemination phase. In many cases, experience and evaluation will show that a program is not sustainable for various reasons. It is possible for evaluation to reveal that a program is not achieving the desired outcomes or that there are negative consequences. Perhaps the funding stream has dried up, or participation was too low to maintain the program. Many programs will be retired then revamped in order to try another approach, thus facing another cycle of growth and starting the process over again.

Evaluation Lifecycle Definitions

Evaluation lifecycle phases are distinguished according to the kinds of claims one would be interested in making, the corresponding evaluation methodology or design, and the quality of measures. **Figure 8** (page 23) provides a concise version of these evaluation phase definitions.

1) Evaluation Phase I: Process and Response

Evaluation in this phase emphasizes implementation and process assessment in order to provide rapid feedback that will be used to refine the program model, "debug" the program procedures, identify barriers to high-quality adoption, and assess participant response to the program.

a) Phase IA evaluations examine program implementation or process, and participant and facilitator satisfaction. These typically use documentation strategies, and post-only evaluation of reactions and satisfaction. Evaluations may rely more heavily on qualitative measures, such as open-ended questions, but quantitative measures are also used.

b) Phase IB evaluations are also typically process, implementation or satisfaction assessments, but they extend the evaluation scope to examine the extent to which selected outcomes are present or absent. Evaluations are post-only, quantitative or qualitative outcome measures are under development or are being adapted from other uses and their reliability is being established.

2) Evaluation Phase II: Change

This phase of evaluation emphasizes the assessment of changes in outcomes (e.g., knowledge, skills, attitude, behavior, performance) that occur in association with the program. The major distinction between the two sub-phases is where the change is being measured – within groups or within individuals.

a) Phase IIA evaluations typically involve unmatched pretests and posttests of outcomes and assessment of consistency (reliability) and validity of measurement. Change is assessed within groups, and may use quantitative or qualitative methods. Results tend to be utilized for management and accountability of the program.

b) Phase IIB evaluations typically consist of a pretest and posttest of outcomes matched at

the level of the individual, using quantitative or qualitative methods. The matching allows for more precise analysis of patterns of change that may be occurring, and enables exploration of reliability and validity of measures. Because participant identification is necessary to match pre and post outcomes and results are increasingly used for public accountability, the required level of participant protection increases. Human subjects review and protection (informed consent, anonymity or confidentiality) is typically undertaken here and increasingly formalized.

3) Evaluation Phase III: Comparison and Control

The emphasis in this phase is on evaluating effectiveness – that is, whether the program is responsible for causing the observed changes in outcomes. Here, evaluation involves the use of comparison groups or variables and statistical controls for adjusting for uncontrolled factors. Phase III evaluation designs typically call for use of more sophisticated statistical analysis, so programs using Phase III evaluations may need the assistance of a data analyst or statistician.

a) Phase IIIA evaluations use design and statistical controls and comparisons (control groups, control variables or statistical controls).

b) Phase IIIB evaluations use controlled experimental or quasi-experimental designs (randomized experiment; regression-discontinuity) for assessing the effectiveness of the program.

4) Evaluation Phase IV: Generalizability

These in-depth and extensive program evaluations focus on how well programs dependably display consistent outcomes over an increasingly broad range of circumstances. Evaluations at this phase may include meta-analysis or synthesis across multiple sites and implementations, investigation of regional/national effects, and/or assessing program "generalizability." Phase IV evaluation designs call for more sophisticated use of statistical analysis, so programs using Phase IV evaluations may need the assistance of a data analyst or statistician.

a) Phase IVA evaluations are multi-site integrated assessments yielding large data sets over multiple waves of program implementation.

b) Phase IVB evaluations present a formal assessment across multiple program implementations

Activity: Lifecycle Alignment Review

The purpose of this activity is to provide a structured framework for assigning lifecycle phases to both a program and its evaluation. This lays the foundation for a discussion of the program's alignment.

First, the working group should review the Program and Evaluation Lifecycle Definitions and Alignment handout (see Workbook Page 23) corresponding to Figure 8. Based on the discussions and diagrams created in the previous two Activities, the working group should review the phase definitions in the handout and decide which program and evaluation lifecycle phases most closely fit the program in its current form. Record these on the worksheet on Workbook page 31.

With regard to program lifecycle, the group may recognize that some "parts" of their program are undergoing more or less change than others. In assigning an overall program lifecycle, this variation across parts of the program should be taken into account. (The distinctions between Phases IB, IIA, and IIB are specifically about this kind of variation in scale and scope of change.) With regard to evaluation lifecycle, if the components of the current evaluation fall into different lifecycle phases it may be difficult to assign a single evaluation lifecycle phase to the entire program. If possible, base the assignment on an identification of the primary purpose of the current evaluation and the methodology that is being used to serve that primary purpose. If this is not possible, then make note of the fact that there are multiple evaluation phases underway and record which aspects of the program are being evaluated in which way.

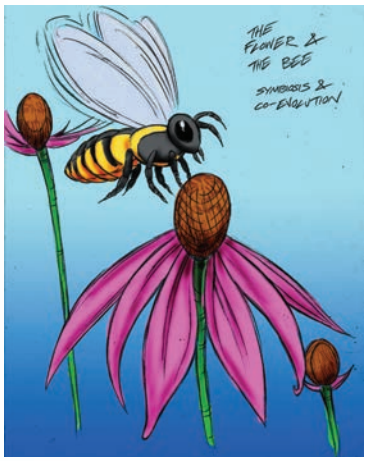
Once the lifecycle analysis for each the program and the program evaluation are complete, use the Lifecycles Alignment Chart from the Workbook (page 32) to plot the location of the program relative to the 45-degree line. In practice, since lifecycle considerations are not the only factor influencing evaluation decisions, it is quite common for program and evaluation lifecycles to not be aligned. If this is the case for your program, discuss what other factors are in play for this program, and what the consequences of non-alignment are. (Refer to the discussion of points A, B, and C in Figure 7 , or "Importance of Lifcycle Alignment" on pages 25-26 of the Workbook for examples of this kind of alignment review.) Refer back to the Map of Stakeholders to identify stakeholder-driven reporting requirements or guidelines, and explore whether these are appropriate for the lifecycle phase of the program.

that enable general assertions about a program in a wide variety of contexts (e.g., meta-analysis).

Lifecycle Application

Figure 7 showed how the phases of these two lifecycles ideally are synchronized. That is, an evaluation should be appropriate for the lifecycle phase of a program. However, it is one thing to present these "ideal" phases as synchronized, and it is another thing entirely to make these phases "fit" what is occurring in a real-world program context. Your job as the Evaluation Champion is to facilitate the discussions - first regarding the program's current lifecycle phase, and the current evaluation methods being used - then how best to work toward alignment and continue the

The Flower and the Bee - Symbiosis and Co-Evolution



The ideas of symbiosis and co-evolution are critically important in evolutionary biology. One of the most familiar examples of this phenomenon is the relationship of the flower and the bee. Each provides something to the other. The flower provides nectar that is produced into honey, and the bee acts as the vehicle for plant sexual reproduction by moving pollen from one flower to another. Both benefit from the exchange. Neither participates in this exchange consciously. Flowers didn't strategize one day that they needed bees as a vehicle for reproduction. And bees didn't decide that flowers would be good vessels for honey production. They co-evolved over millennia in a manner that makes them co-dependent.

There are several ways that symbiosis and co-evolution are important for evaluation. First, if all programs evolve through different stages over time, then we must recognize that the evaluation approaches we use at each stage need to differ throughout the life of the program. That is, the way we would evaluate a program during its initiation stage would not likely be appropriate for evaluating it during its growth stage, and so on. In effect, the evaluation of a program has its own lifecycle and one of the major tasks of systems evaluation is to encourage the symbiotic or co-evolutionary relationship between program and evaluation lifecycles. In the initiation phase an evaluation needs to be dynamic and flexible, providing rapid feedback about implementation and process. In many program evaluations this is accomplished with simple monitoring or post-only feedback forms, unstructured observation, qualitative methods, informal debriefing and feedback, and through communications systems. In the development phase of an evaluation, the focus tends to shift to the observation and assessment of change and we focus on things like designing observational procedures and measures of key outcomes, assessing the consistency and construct validity of measures, looking at pre-post differences and examining the relationships among different observations, qualitative or quantitative. The mature phase of an evaluation tends to emphasize the idea of control. At this point the program is routinized and stable enough to compare performance of participants with some standard expectation of performance or with outcomes of people who participate in alternative programs or none at all. This is the realm of experimental and quasi-experimental designs and of more structured and comparative qualitative approaches. The translation or dissemination phase in evaluation is typically concerned with generalizability or external validity. It examines the consistency of outcomes across different settings, populations or program variations. This is the realm of secondary and meta-analysis and of program review approaches that seek general inferences about the transferability of the program. Encouraging a symbiotic relationship between the evaluation approach and the program lifecycle is a critically important systems evaluation process.

Second, the ideas of symbiosis and co-evolution also have important practical implications for the level of support people have for evaluation. In many evaluation contexts, one hears a series of laments about how unmotivated people are to evaluate or their resistance to doing evaluation. For instance, the evaluator asks "Why don't these program people just cooperate when I ask them for data?" Program implementers ask "Why don't these evaluations address something that would be useful for us?" Program participants want to know "Why do they keep bugging us for data? We don't get anything from this." In the ideal, we would want the situation to be a co-evolutionary one where program participants are providing information naturally as part of their participation, where program administrators are getting what they want from the provided data, and where evaluation happens almost transparently as an integrated aspect of program implementation. That is, the ideal is the flower and the bee. This is a difficult ideal to achieve in practice. It requires that the evaluation systems be engineered in such a way that each stakeholder group's incentive to participate in the evaluation is well understood.

developmental progression of the program. (See Activity - Lifecycle Alignment Review on page 25). Alignment of program and evaluation lifecycles will not necessarily occur in just one cycle of evaluation (particularly if there is initially a large discrepancy between the program and evaluation lifecycles). Decisions about how quickly to work toward alignment will have to be weighed against the feasibility of different approaches as well as external pressures (i.e., funder mandates).

2.05 Logic Model

The goal of this step is to help the working group generate an initial logic model that captures the program assumptions, context, inputs, activities, outputs, and short-, medium-, and long-term outcomes. Use the Workbook document *Getting Started with your Logic Model,* as well as the *Blank Logic Model Template(***Workbook pages 35-37, 41***)* for this step. The following section is a description of the components of the logic model. An example is shown in **Figure 9**. In addition to the worksheets mentioned above, the programs we have worked with have also found it helpful to be aware of the key aspects of a good logic model, and therefore we created the *Logic Model Template and Guidance* and the *Logic Model Feedback Form* (see **Workbook pages 39, 43-45***)* to explain what should be in each section of the model.

Although modeling may all be done on paper, many programs begin their logic modeling on paper, and later move to entering the logic model into the Netway (in the section titled "Logic Model"), which makes editing and printing easier.

Inputs
Inputs should include a brief, clear, informative, and complete list of key resources such as staff, curriculum, teaching materials, outside partners and facilities. Often staff inputs are described in fractional "Full Time Equivalent" (FTE) increments. The inputs should give the reader an "at a glance" idea of the program's size and scale. The list does not need to be lengthy but should describe succinctly the resources needed to implement the program.

Activities
The activities list should be complete and understandable, as it is the basis for the remainder of the Logic Model. Based on the boundary discussion, staff should be able to determine what gets included as an activity. Activities should be well defined so that someone unfamiliar with the program can understand the activity titles. For example, instead of naming an activity "Green Garden" you could use a more descriptive (yet short and succinct) title such as "Green Garden Weekend Workshop" or "Green Garden 1-Day Conference" so that the reader gets a better idea about what is happening. The activity titles should be written so that they can be understood by a general non-specialist reader.

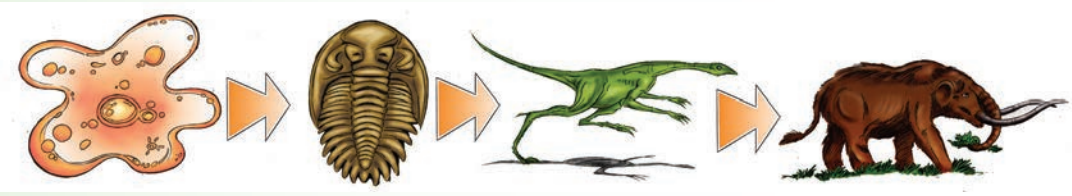
It is sometimes difficult to decide how narrowly or broadly to define an activity. For example: for a program consisting of multiple workshops, should the activity list include each workshop separately, or should they be bundled together into a single activity referring to the combined workshop series? The choice will depend on the nature of the program and on the kind of detail in the rest of the program model. If the individual workshops are expected to have distinct outcomes (one focuses on basic knowledge, one on how to use equipment safely, etc.) and if the resulting short-term outcomes are important to distinguish, then it may be useful to list the individual workshops as separate activities. However if you are working at a higher level of generality, and particularly if the workshops are really interwoven, then it might work better to have the workshop series as a single activity, with more comprehensive, broader outcomes.

The activities list should only include activities that reach people who participate or who are targeted and should be consistent with the program as it is described in the program description. It should typically NOT include administrative, marketing or other activities carried out by program staff. (Recruitment, follow-up and a host of other administrative activities generally do NOT belong in a logic model. However, there are exceptions. For example, it would be legitimate to have "recruitment" in a logic model if an important outcome was to increase program diversity, and the program was working to diversify the organization's volunteer pool.)

Outputs
Outputs are the by-products of activities. These could include certificates of attendance or completion and objects that were created as part of the program (such as products generated by participants in activities). If something is created by a participant it should be listed in the outputs

The Survival of Programs with "Fitness" - Evolution and Evaluation

The theory of evolution is one of the most important achievements in the history of science. Darwin's Origin of Species and his articulation of the theory of natural selection forms the foundation of virtually all of the life sciences and continues to have profound effects in the social sciences, arts, humanities and, as we all know, in the political and religious realms. The theory of evolution is essentially a systems theory in that it describes how different systems interact and develop over time. This systems theory has a profound effect on how we think about evaluation. To give you an idea of how different an evolutionary systems evaluation might be, consider how the basic idea of evolution sounds when framed in terms of programs and evaluation:



Every program can be viewed as an organism in a population of similar programs that constitutes its species. Program theories, whether stated explicitly or not, make up the essential instructions of the program. Programs have variations within each species of program. Programs have unique characteristics: the people who implement them, the activities that constitute them, the setting and assumptions that guide them, the participants who take part in them. This program variation is essential for their evolution.

Program variations are implemented, have consequences, and are selected for in subsequent program generations. Some programs and their characteristics and theories survive over time; most become extinct. Programs and program theories get selected and survive because of the fitness of their characteristics to a specific environmental or ecological niche. While most of us probably hope or believe that programs are selected for using rational criteria to yield specific desirable characteristics or outcomes, in many situations they probably survive because people like them, get used to them, or because there are institutional, political and economic forces that favor their survival.

Over time, programs and their theories evolve. This evolution is based on the same principle of natural selection that underlies all evolution in life. The process of consciously developing and evolving programs is a type of artificial selection, a special subtype of natural selection. Artificial selection is to natural selection as plant or animal breeding is to natural reproduction. Evaluation can play a key role in that artificial selection, both in encouraging and enhancing variability and in providing feedback and influencing selection. As in evolution generally, it's not clear where program evolution is heading or whether any adaptation can be said to constitute 'progress.' Slight program variations and adaptations can survive that subsequently make little apparent sense. Program features may exist today that were adaptive in the past but remain largely as residuals, long beyond their original adaptive genesis.

Just as with other organisms in nature, in addition to their participation in a broader species, each program has its own individual life (ontogeny), a unique life course that moves through various phases. Programs are born or initiated. They grow and change as they are implemented and revised. They mature and may reach a relatively stable state sometimes becoming routinized and standardized. And, they regenerate, die, are translated and disseminated, and so on, starting new cycles of program instances.

This is simply a restatement in terms of programs and program theory of the theory of evolution generally. It incorporates the ideas from evolution of the life-course of the individual organism (the ontogeny) and the tree-like descent of multiple generations of organisms from ancestors (phylogeny). Like the theory of evolution it is simple in conception and readily communicated. And, like that theory it has behind it a world of complexity and implications which have implications for evaluation practice.

section. If something is an effect on a participant, then it should be listed in the outcomes section. The outputs should be clearly described and included for activities that are likely to generate outputs.

Outcomes

The difference between short-term, medium-term, and long-term outcomes is a matter of relationship to each other, not just the passage of calendar time. Short-term outcomes are the earliest outcomes directly linked to the activities, medium-term outcomes stem from the short-term outcomes and connect to the long-term outcomes, and long-term outcomes are generally the furthest out from the activities.

Outcomes should not be limited to things that can be "measured". The goal in the logic model is to portray a full picture of the program and what it is expected to lead to, even if some of the outcomes are impossible to measure and even if some (long-term outcomes especially) are expected to emerge long past the time of any evaluations that may be done.

In general, the descriptions of outcomes should be clear, informative, and appropriately reflect the program description. They should be written in a general way (i.e., not context specific) and should be phrased as effects on, or changes in, participants and/or their communities or society.

Short-term Outcomes

Short-term outcomes should describe results and effects on participants that are logically connected to activities. These outcomes could include things like changes in awareness, knowledge, attitudes,

FIGURE 9
SAMPLE LOGIC MODEL

INPUTS	ACTIVITIES	OUTPUTS	SHORT TERM OUTCOMES	MEDIUM TERM OUTCOMES	LONG TERM OUTCOMES
VOLUNTEER TEACHERS (GRADUATE STUDENTS. FACULTY	CLASSROOM SESSIONS ON MODULE. FACILITATED BY EXPERT VOLUNTEERS	COMPLETED ACTIVITY SHEETS	INCREASE TEACHERS' ACCESS TO QUALITY RESEARCH-BASED MATERIALS	INCREASE STUDENT EXCITEMENT ABOUT SCIENCE AND ENGINEERING	INCREASE POOL OF INDIVIDUALS CHOOSING SCIENCE AND ENGINEERING CAREERS
MODULE SCIENCE KITS (HANDS-ON MATERIALS. LESSON PLANS	RECRUITMENT OF GRADUATE STUDENT VOLUNTEERS (ESP. FROM UNDER-REPRESENTED GROUPS)	CONTACTLIST OF AFFILIATED GRADUATE STUDENTS	GRADUATE STUDENTS INCREASE ENJOYMENT IN TEACHING YOUTH	K-12 STUDENTS INCREASE KNOWLEDGE OF SCIENCE AND ENGINEERING	INCREASE DIVERSITY AMONG SCIENTISTS AND ENGINEERS
.5 FTE MANAGER	PLAN AND COORDINATE SESSIONS WITH SCHOOL TEACHERS AND VOLUNTEERS	RECORD OF COMPLETED CLASSROOM SESSIONS	YOUTH PARTICIPANTS LEARN NEW SCIENCE CONTENT RELATED TO MODULE OBJECTIVES	SCHOOL DISTRICT EXPANDS SCIENCE TEACHING RESOURCES THROUGH USE OF MODEULES	INCREASE SCIENCE LITERACY IN GENERAL PUBLIC
			FAMILIES LEARN AND TALK ABOUT SCIENCE TOPICS TOGETHER	GRADUATE STUDENTS GAIN SKILLS IN SCIENCE TEACHING	
			YOUTHE INTEREST IN SCIENCE INCREASES		

ASSUMPTIONS - SCHOOL DISTRICT SCIENCE TEACHERS ARE RECEPTIVE TO LESSONS BY OUTSIDE EXPERT VOLUNTEERS. SCIENCE MODULES MEET STATEWIDE CURRICULUM REQUIREMENTS. SCHOOL DISTRICT LACKS ADEQUATE UP-TO-DATE SCIENCE TEACHING MATERIALS. STUDENTS IN GRADES 3-5 LEARN WELL WITH HANDS-ON SCIENCE PROJECTS.

CONTEXT - THIS PROGRAM IS DESIGNED FOR STUDENTS IN GRADES 3-5. AS THERE IS A RELATIVE SHORTAGE OF FUN HANDS-ON TEACHING RESOURCES FOR THIS AGE GROUP. THE STUDENT POPULATION IN THIS SCHOOL DISTRICT IS DIVERSE. AND THERE IS A PARTICULAR INTEREST IN INCREASING THE INTEREST OF UNDER-REPRESENTED GROUPS IN SCIENCE CAREERS. THE RECRUITMENT OF GRADUATE STUDENT VOLUNTEERS FROM UNDER-REPRESENTED GROUPS IS INTENDED TO CONNECT THE ELEMENTARY SCHOOL STUDENTS WITH DIVERSE POSITIVE ROLE MODELS IN SCIENCE.

skills, opinions, aspirations, and motivations. Place outcomes in the short-term column if the "first glimmer" of their appearance occurs early on. For instance, the skills of participants may continue to expand over time, but they should be listed as an outcome during the time frame when generally those skills manifest for the first time. Remember that sometimes there is a chain reaction of short-term outcomes, and that even if one outcome generally arrives before another outcome, they can still both be listed in the short-term column. For instance, a student might learn a skill that leads to an increase in their self-esteem. Even though one came before the other, they may both be considered short-term outcomes.

Medium-term Outcomes

Medium-term outcomes should describe effects on participants that logically connect short-term outcomes with long-term outcomes. They tend to be follow-on effects on participants in the form of deeper or more sustained changes in short-term outcomes, and/or the spread of those effects from individual participants to their families or larger groups. Medium-term outcomes could include new behaviors or changes in behavior, practice, decision-making, policies, social action, awareness, knowledge, attitudes, skills, opinions, aspirations, and motivations.

Long-term Outcomes

Long-term outcomes describe the ultimate effects logically connected to medium-term outcomes. These are generally written as occurring beyond the individual or personal level and are likely to include things like social, economic, civic, or environmental effects. The subject of long-term outcomes is generally "the community" or even broader terms such as "a reduction in HIV/AIDS infection".

Assumptions and Context

The assumptions and context sections of logic models are sometimes overlooked, but in fact they provide information that is essential for understanding how and why a program works the way the rest of the logic model suggests that it does. The Workbook provides worksheets to assist the team in working through these items (**Workbook pages 41-54**), the *"Uncovering Buried Assumptions"* and the *Describing Program Context* in particular.

Assumptions

Assumptions are the beliefs and thought patterns about how and why a program is expected to succeed. In order to help identify assumptions, ask staff what things might occur that would prevent the program from achieving its long-term outcomes. Assumptions may include the idea that the program will be funded through the next funding cycle, or that the program will have access to space and other resources that may be provided by partners or funders upon whom they rely, or that the training methodology is appropriate for the intended audience, and so on.

Context

Context is the environment (including the social, cultural and physical context) in which a program will take place. For example, context could be within an afterschool program, at a farmer's market, or within the Spanish-speaking community. Context also includes factors that affect the need for the program, or the community history that might make the program particularly relevant. (HINT for both assumptions and context: If these details have been mentioned in the program description they do not need to be repeated at length here, but should still be mentioned. Both of these sections should be brief.)

2.06 Pathway Model

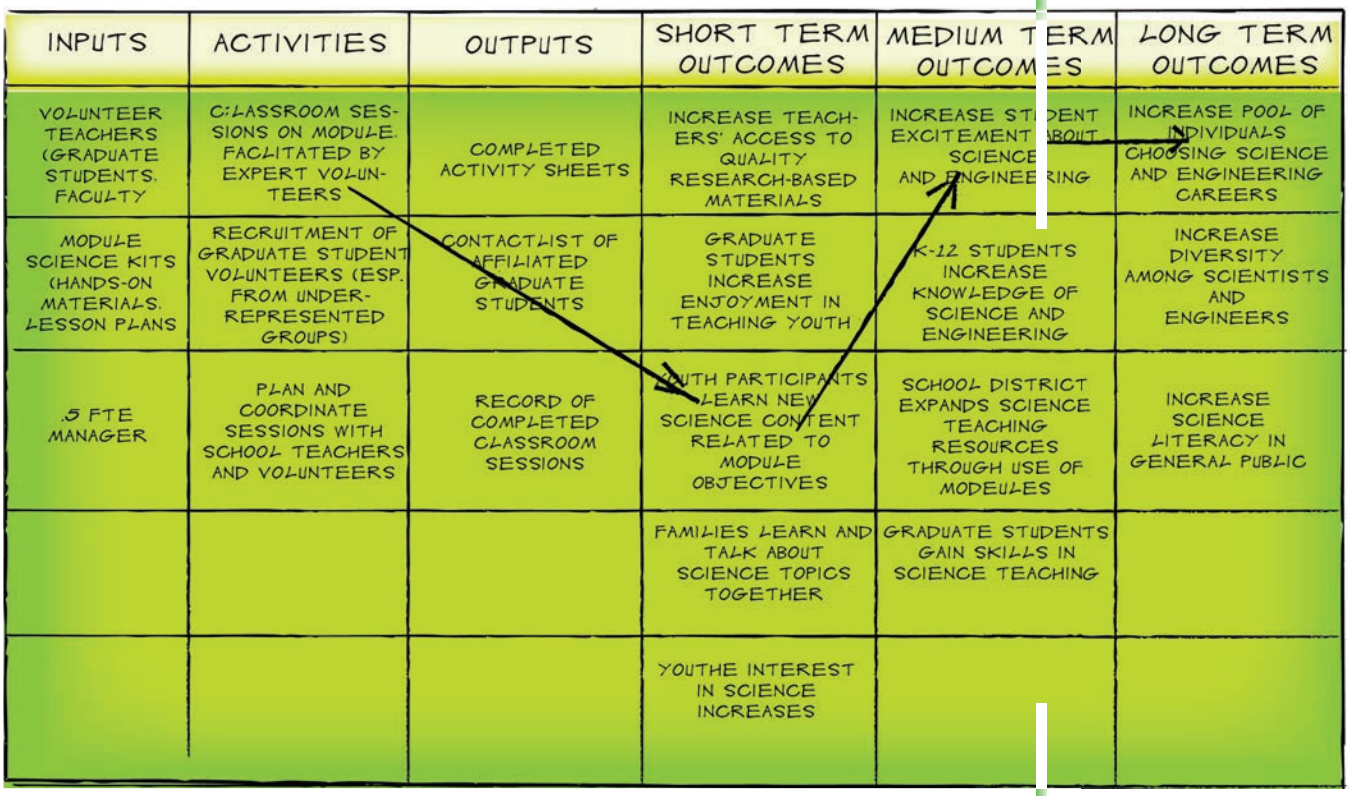
A Pathway Model is a visual program model closely related to the columnar logic model. A pathway model overlaps with a logic model in that it is based on the activities and outcomes that are in the logic model (See *Relationship between Logic and Pathway Models*, **Workbook page 67**). Unlike the logic model, it does not include the details of inputs, assumptions, or context. However, it adds a significant element by incorporating the logical connections that lead from an activity to one or more short-term outcomes, and from there to medium-term outcomes, and ultimately to long-term outcomes. It tells the story of how the program works, in a way that can't be captured in the columnal logic model. Together, the logic model and pathway model present a very informative,

concise picture of the program and how it is believed to work.

The pathway model can be developed using the logic model as a basis for articulating clear and direct linkages between program activities and outcomes. If it has not been done already, discussion should help program staff begin to identify key pathways or "through-lines" that connect the activities to outputs and outcomes in their logic model. One way to think of this is to explain:

[ACTIVITY "A"] leads to [SHORT-TERM OUTCOME "X"],
which leads to [MEDIUM-TERM OUTCOME "Y"],
which leads to [LONG-TERM OUTCOME "Z"].

FIGURE 10
SAMPLE PATH FROM ACTIVITY TO OUTCOME



For example, using the sample logic model in **Figure 10**, program staff may believe that the activity "Classroom session on module, facilitated by expert volunteers" leads to the short-term outcome "Youth participants learn new science content related to module objectives", the learning leads to the medium-term outcome "Increase youth excitement about science and engineering", and that leads to the long-term outcome "Increase pool of individuals choosing science and engineering careers."

A single pathway may look something like that in **Figure 11** and a more complete pathway model like that in **Figure 12.**

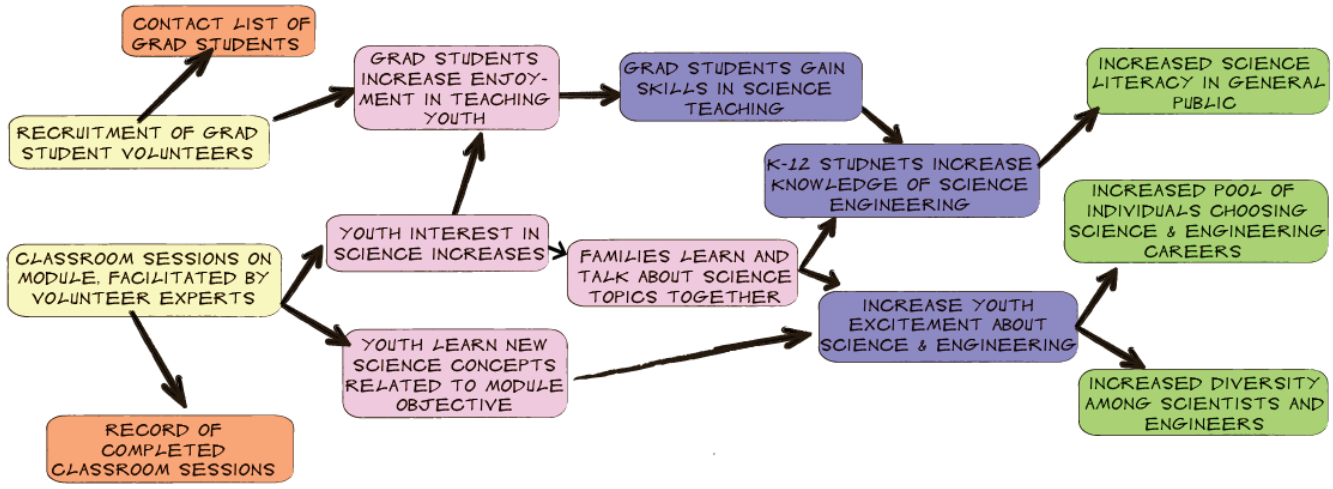
The working group should continue to create these pathway links until "how the program works" has been described to their satisfaction. The pathway model is different from a logic model in that it shows how the components of the logic model are connected.

FIGURE 11
PATHWAY FROM ACTIVITY TO OUTCOME



FIGURE 12.
PROGRAM PATHWAY MODEL

The items in the completed pathway model should be consistent with the logic model. The connections between activities, short-, medium-, and long-term outcomes should be relevant, appropriate and logical. In general, short-term outcomes should rarely be connected directly to long-term outcomes. It is also important to check and make sure that there are not any "orphaned" items. This refers to activities that are not connected to any outcomes or outcomes that are not



Activity: Logic and Pathway Model Peer Review

Before finalizing the program models we encourage program staff to get reviews from others. Because program staff is typically characterizing their programs from an internal perspective, they may omit elements that they take for granted or know implicitly (it is easy to fill in pieces which may be missing from the model with logic inside one's own head). Peers typically have the benefit of being external to the program and can therefore better judge the ability of the model to communicate a clear picture of the program's logic. Peer reviews also benefit the reviewer. As reviewers begin to identify weaknesses in a peer's model, they may become better able to identify similar problems in their own model. This type of feedback can be especially valuable to the logic and pathway model development process when programs reviewing each other either share similar missions or are from the same organization.

For program staff who do have the opportunity for in-person feedback, this activity is designed to be used with poster-sized pathway models displayed on the wall (using either poster paper with index cards listing each logic model elements taped in place, and then drawing connecting lines, or a large-format printout 42-inch poster of a digitally imaged pathway model). Each program's description should also be displayed next to the model. Reviewers use a structured guide (see "Review Guide for Pathway Models" on page 69 of the Workbook) and a set of stickers pre-printed with specific images (a leaping frog, a fence, a cloud, etc.) to identify common pitfalls of model development. Peer reviewers are also encouraged to write open-ended comments as necessary, and discuss all of their feedback with the recipient at the end of the activity. (We frequently use the same poster for this activity and the next one. The trick is to put a mark in the two top corners of the poster, cover the poster with a sheet of clear overlay, and use a permanent marker to make the same corner marks on the overlay. Participants can take the posters and the feedback home and quickly re-align them.)

For program staff who do not have the opportunity for in-person feedback, they may elect to partner with another program team and provide feedback to each other by using the "Feedback Summary for Logic & Pathway Models" (Workbook pages 43-46). As the Evaluation Champion you should encourage programs to partner up to do a review. There is a lot to be learned from reviewing another program's model, as well as getting outsider input on one's own model. At the minimum, as Evaluation Champion you should take some time to provide feedback to each program team you are working with.

connected to any activities.

It is also important to review the pathways or explanatory "through-lines" that connect specific activities and outcomes. Make sure that the pathways clearly communicate the "story" or "program logic" and do not dead-end at short- or medium-term outcomes.

In general, the pathway model should efficiently communicate the program logic and should be easily readable (e.g., the number of connections should not be excessive).

Look at the completed pathway model and ask:

- Are there any activities that are not connected to any outcomes?
  If yes, why may these gaps exist? Was something simply left out of the model? Or, is there a program activity that does not really address the program goals? In the case of orphaned activities, ask whether the activity is serving an intended purpose or whether, for example, the activity is included because it has always been done.
- Are there any outcomes that are not connected to any activities?
  If yes, why may these gaps exist? Was something simply left out of the model? Or, is the program expected to lead to a particular outcome that is not supported by your current activities? In the case of orphaned outcomes, ask whether there are adequate activities to address the outcome.
- Are there any pathways that dead-end at short- or medium-term outcomes?

If you find that any of these situations arise, it is important to revisit the program boundary discussion and the logic model. We use the *Review Guide for Pathway Models* (**Workbook page 69**) to help structure an analysis of the Pathway Model, as well as the *Pathway Model Feedback Form* on **Workbook page 46**.

2.07 Determining Evaluation Scope

Typically evaluation changes over time as a program evolves. In any given evaluation cycle (e.g., annual evaluation) we usually don't try to assess every activity and outcome on a logic or pathway model, we make choices about what is most important, relevant and feasible. That is, we determine what the "scope" of the evaluation will be at this point, which elements of the program we will focus on. In most cases, 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- and medium-term outcomes.

This step in the Protocol represents a change in focus. The Evaluation Champion and the working group will now move from thinking purely about the program, its boundaries, logic and stakeholders move toward thinking about evaluation. This shift should be acknowledged, and the group may want to discuss any learning and other benefits that have come out of their work so far.

Discuss Feasibility, Credibility, Accuracy and Usefulness

One of the Evaluation Champion's responsibilities is to help the working group identify parts on the models where focused evaluation efforts will be most valuable for them. There are many points to consider and discuss during this step. In order to help guide this decision-making process, we have developed a goal statement which may serve as a touchstone for evaluation planners as they move through the rest of the steps of the Protocol:

> *The goal of evaluation is to obtain accurate, useful insights about the answers to evaluation questions in a manner that is feasible, is credible to relevant stakeholders, makes strategic use of limited time and resources, and contributes to our general knowledge, to future evaluations and to program evolution.*

Evaluation time and resources are 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. Therefore, it is important to be strategic by recognizing what is feasible in the coming

Evaluation Goal:

To obtain accurate, useful insights about the answers to evaluation questions in a manner that is feasible, is credible to relevant stakeholders, makes strategic use of limited time and resources, and contributes to our general knowledge, to future evaluations and to program evolution.

The Netway makes Pathway Modeling simple. Just click and drag logic model elements onto each other to connect them. Built-in tutorials will help you learn how to easily edit your pathway model (also see Workbook pages 63-65.)

In fact, the original purpose of the Netway was to do pathway models, and the term "Netway" comes from the concept of "Networked Pathways."

evaluation cycle. Feasibility discussions should take account of program activity, time constraints, funding constraints, and/or reporting constraints. Use caution not to select too large a scope.

Accuracy refers to reliability and validity of a given evaluation strategy and set of tools. Program staff should choose an evaluation strategy that will provide the best accuracy (which will lead to credibility and usefulness) within their contextual feasibility restraints.

As the group works through the sections of the evaluation plan, the building blocks established in the program modeling phase of the Protocol – stakeholder analysis, lifecycle analysis, program logic and pathway models – will come together to help establish strategies and priorities for achieving the evaluation goal in the coming evaluation cycle.

Examine the Logic and Pathway Models

The pathway model and logic model are the primary tools you will use to assess and decide upon the scope of evaluation for a particular evaluation cycle. The program models represent a broad overview of everything that is done and assumed in a program and all of the expectations for the program's effect. It can help program staff to think of the pathway model and logic model as a blueprint for their program over the next "X" number of years. Keep in mind that it may not be possible to address every aspect of this model in the next year, or maybe even in the next five years. The "Evaluation Scope" encompasses the components of the program that will be focused on in the upcoming evaluation cycle. To guide your thoughts on the following concepts, consider using our *Mining the Model Worksheet,* in the **Workbook, page 71**) .

And so on, and so on… - Causal Pathways



The idea of cause and effect is central to systems thinking. The field of System Dynamics, for instance, develops cause-effect models or "causal" chains and uses them to think about the way causes produce effects throughout the system and the different types of feedback loops that result and can often lead to unanticipated outcomes. In effect, you are describing the chain of thinking in a system: "X leads to Y which leads to Z, and so on, and so on…"

The notion of causality is critically important in systems evaluation. It is central to theory of change approaches to evaluation, to path analysis, to theory-driven approaches, and to much more. In program logic models there is a general idea of causality – activities are expected to produce outputs and to lead to short-, medium-, and long-term outcomes and ultimately impacts. However, one problem with traditional logic models is that they are "columnar" in nature. The entire set of program activities, or outputs for each phase of outcomes are typically treated as a whole. That is, in traditional logic modeling while we expect that program activities produce outcomes, we usually do not specify which activities are expected to produce which outcomes. In other words, traditional logic models do not spell out the specific cause-effect relationships that are expected.

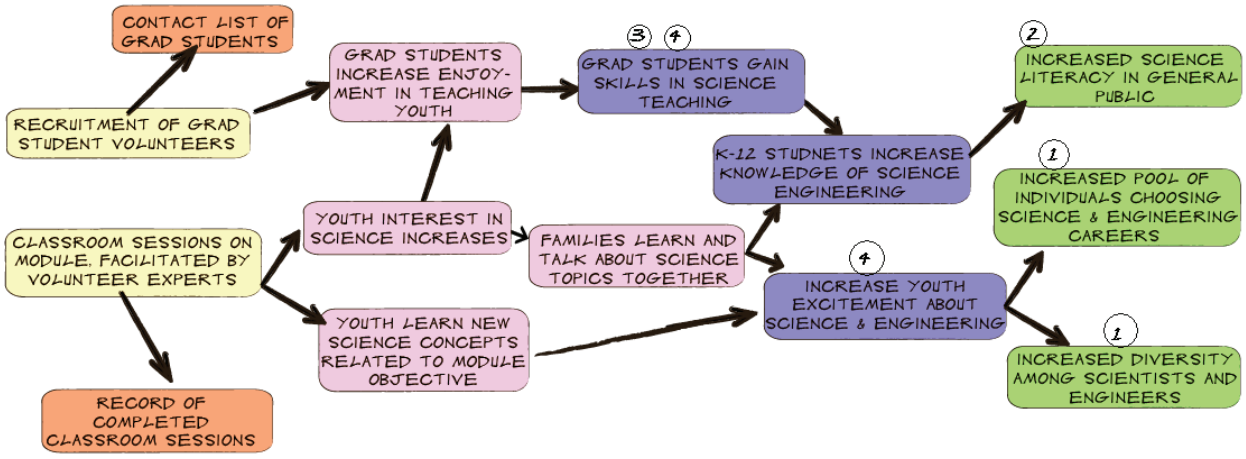
Because systems thinking suggests that distinguishing different cause-effect chains can be important, we prefer program logic models that describe the specific causal pathways involved in programs. For example, typical programs usually involve multiple activities, outputs and outcomes. In a pathway approach, you would specify each connection that you think might be relevant. You might specify that activity A affects short-term outcomes A and C, which in turn affect medium-term outcomes E and F, and long-term outcomes A and D. You might also expect that there will be feedback loops in your model. For instance, changing the results of a short-term outcome could trigger a change in another short-term outcome that then reverberates in or feeds back to the first outcome.

This kind of causal pathway model is useful in telling the story of the program and is essential in developing a high-quality evaluation of it. A program model is likely to have many pathways from activities to outcomes. Drawing pictures of the pathway model enables you to understand better how you think your program should operate. It is especially useful to trace the "through-lines" of your program, the major causal paths through the model of your program. The through-line points out program activities that may not lead to any outcomes and helps you to identify key outcomes that should be measured.

Revisit Stakeholders

An evaluation scope discussion should include a re-examination of the Map of Stakeholders. Are there any outcomes or 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 get precedence over others? Once again, visual clues are helpful. Transfer the pathway model onto paper or a drawing board. Create a numbered list of the key stakeholders and place each stakeholder's number onto the pathway model next to the outcomes and pathways that that stakeholder would be most interested in (see **Figure 13**).

FIGURE 13. STAKEHOLDERS LOCATED ON PROGRAM PATHWAY MODEL



Revisit Lifecycle

This is an opportune time to reconsider the lifecycle of the program. There may be a desire to focus on a medium-term outcome but if the program is still in its initiation phase it might be more appropriate to focus on demonstrating effects on short-term outcomes first.

Finding the Key Pathway Markers

In any pathway model there are some paths that are more important than others, and some outcomes that are more central to the model. The following questions can help guide thinking about which pathways to focus on in the current evaluation cycle.

Required: Is there a particular activity, outcome, pathway (or set of pathways) on which the program is required to report? If so, then evaluation should definitely include this in its scope.

Easiest: The primary (or most obvious) pathway may be the best focus for the evaluation. Is the program already collecting information that addresses some aspect of the model? Is there an "easy" or obvious point on which an evaluation could focus? We refer to this as the "low-hanging

Activity: Mining the Model – Part 1

As in the Pathway Model Peer Review, this Mining the Model activity also uses stickers and a worksheet (See the "Mining the Model Worksheet" in the Workbook, Page 71-72.)

Starting with a clean model (preferably updated after feedback from the previous Logic and Pathway Model Peer Review Activity (page 32), and working on their own poster this time, participants will be able to narrow the focus of the evaluation and develop evaluation priorities. The pre-printed stickers may use stars (main activities and outcomes), keys (key links and pathways), and letter stickers (where each letter represents a stakeholder priority) or they can draw stars and use highlighters. Complete other activities on the worksheet - including comments for lifecycle considerations.

Once placed, these marks will provide visual cues that indicate the priorities of the current evaluation, and will guide formation of the Evaluation Purpose Statement. Sometimes the visual cues support the participants' previous thinking about what is most important, sometimes they yield surprises.

fruit."

Hub (Grand Central): Like Grand Central Station in New York City, on your model there may be a "central point" through which many of the pathways pass. By evaluating this outcome, the program may potentially be able to address many aspects of its model simultaneously.

Once again, visual cues are helpful for many people. On the pathway model circle hubs, highlight key pathways, and so forth. (See **Mining the Model - Part 1** Activity on page 35 of this Guide.)

Determine the Scope for this Evaluation Cycle

Once the key markers have been identified and stakeholders have been placed on the pathway it is time to determine the scope for the evaluation. What we mean by scope is essentially how far into the pathway model (from left to right) the group intends to evaluate over a certain period of time (this evaluation cycle). For instance, if someone is determining the scope for an evaluation for the upcoming year, and they have determined the key markers as described above, they will want to set the scope for the evaluation to include as many of the key markers as can reasonably be accommodated within the year. It is useful to show the scope of an evaluation graphically. You can do this by drawing a line that encloses the pathways that will be included in your evaluation within it as shown in **Figure 14**.

Keep in mind that the evaluation scope will generally change from year to year and will generally move from a focus on activities, outputs and short-term outcomes in earlier evaluations to one that includes medium- or long-term outcomes later on.

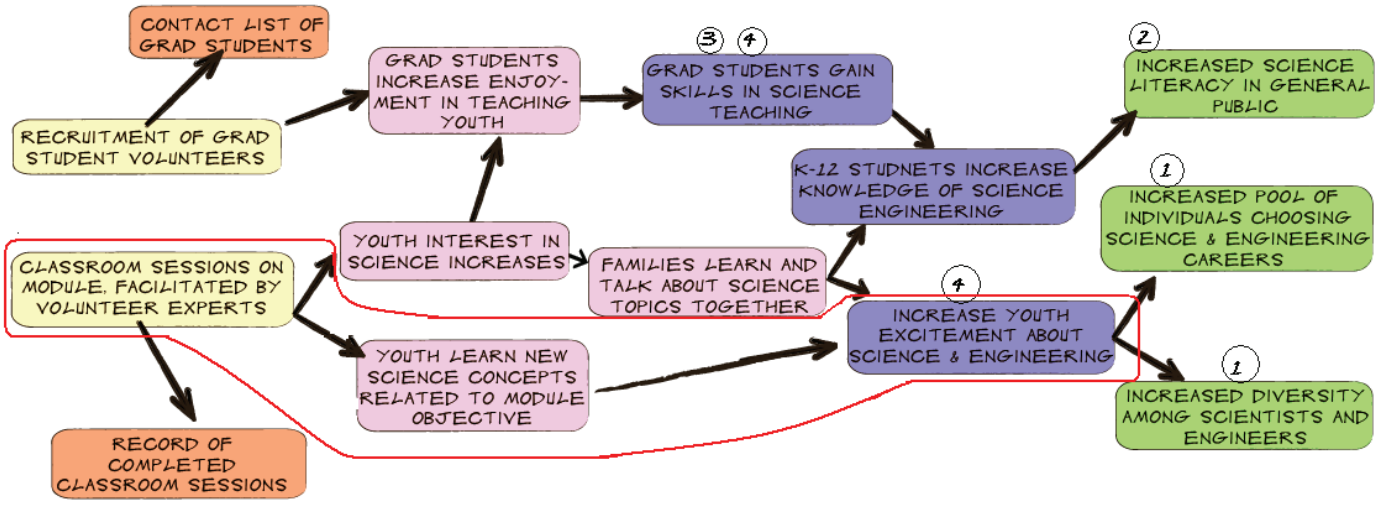
2.08 Program-System Links

No program is an island. That is, it is a rare program that is so unique that it cannot learn or benefit from knowing how others view the program and have addressed similar challenges in the past.

At some point each program should examine what others are doing that might be related to or inform the program and its evaluation, and integrate this information into the program's pathways. This step involves turning the working group's attention to other programs and to the research literature. These sources may suggest measures that could be used for the current evaluation or provide evidence that could help support the logic conveyed by the links in the pathway model.

Ask the working group if there are other programs like this one. These may be in the same organization or in another local organization, or may be physically distant. Are there other programs with similar or shared outcomes, even if the program activities are very different? What evaluation tools are used for comparable programs, and are they available for this program to use? What research supports the logic represented in the pathway model? Are there measures in the research literature that may be useful for this evaluation? To find similar programs, draw upon professional

FIGURE 14.
EVALUATION SCOPE



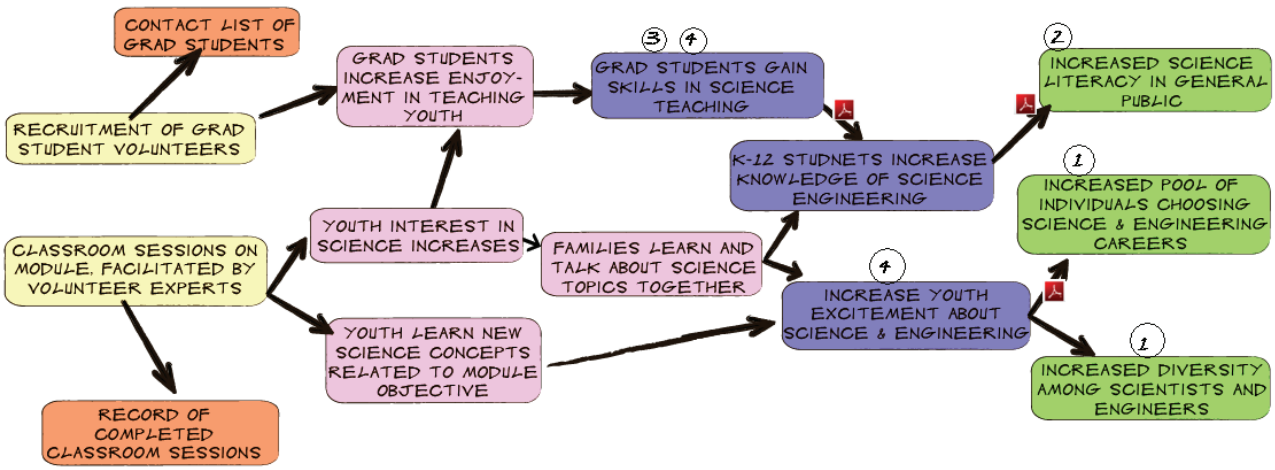
connections as well as links or resources that may be available from the larger systems that this program is part of – national organizations, professional associations, and so on.

Don't rely only on similar programs. The working group should search the scientific literature for current research that is being conducted in their general program area. This can be a time-consuming task, but it is also an important one. Having an understanding of the evidence that already exists can help the program to focus its evaluation efforts and to identify well-researched and widely used measures to use. Depending on the program's resources and capacity, as well as the interests of its stakeholders, organizations may choose to devote more or less time to this step.

For example, consider a course that teaches youth how to reason scientifically. Suppose that a short-term outcome for this course is that youth will be able to reason scientifically about everyday problems. An example of evidence for this causal relationship could be a research article that shows that courses in scientific reasoning methods that utilize relevant everyday topics may enhance the transfer of scientific reasoning to other problems/situations (e.g. Williams, et al., (2004). Thinking Like A Scientist About Real-World Problems: The Cornell Institute for Research on Children Science Education Program. *Journal of Applied Developmental Psychology*, 25:1, 107-126).

Once the working group has surveyed resources from other programs and the literature, it is useful to visually depict this on the pathway model (see **Figure 15**). Relevant literature (evidence) that relates to the causal relations can be "mapped" onto the pathway model (we call this evidence mapping, and it is summarized in our published article: Urban, J. B., & Trochim, W. (2009). *The Role of Evaluation in Research-Practice Integration: Working Toward the "Golden Spike"*. American Journal of Evaluation, 30(4), 538-553. For a short summary see **Workbook page 73** for *The Golden Spike - Linking Evidence to Practice.* Similarly, relevant measures that relate to specific outcomes can also be "mapped" onto the pathway model (we call this measurement mapping). See the **Activity: Mining the Model - Part 2** (below) for an example of how to conduct the mapping exercise. Results from the evidence mapping and measurement mapping may lead the working group to want to

FIGURE 15. LITERATURE MAPPING (GOLDEN SPIKE)



Activity: Mining the Model – Part 2 - Linking Evidence to Practice

As part of thinking about program-system links, participants may find it useful to return to their model for an additional "mining the model" activity to help them visualize exactly where on the model these links may exist. As in the previous two activities, this activity may use stickers. We usually have them place the stickers on the same poster/overlay sheet as in Mining the Model - Part 1.

System symbol stickers (e.g., a microscope for STEM education) = system priority/interest
Coat hanger stickers = there is prior research to support or "hang on" that link
Measure stickers (e.g., an image of a checklist) = there is a measure that can be used for that outcome/activity

Once all of the key areas of the pathway model have been identified using the stickers in both Mining the Model Activities, the participants should be able to use this version of their model to help (re)define the scope for the current evaluation cycle based upon their priorities, interests and available resources.

Activity: Stakeholder Interview

This is a "homework" activity. In order to encourage participants to delve deeper into understanding stakeholder perspectives, this activity asks each participant to interview 2-4 different key stakeholders about the program, its model, and future evaluation work. Drawing from their stakeholder list or map, they need to identify key stakeholder whom they would be interested in interviewing, and then agree on a plan for how to capture stakeholder comments and input. A workgroup member should contact each potential participating stakeholder, provide an explanation of this activity, and state that the stakeholder input would be appreciated. If you plan to record the interview the stakeholder should be informed of this, and if agreeable establish a time and date for the interview. The working group should then create a list of questions for the interview based on the list provided in Workbook pages 9-10). Sharing your program description, logic model and pathway model during the interview can provide a helpful starting point for the interview, but you may wish to ask some questions about their perceptions of the program before sharing these with them.

revise the scope of the current evaluation; particularly if there is pressure to report on longer-term outcomes. For example, if the working group is able to identify strong research evidence that already links one of their medium-term outcomes (e.g., "increase student excitement about science and engineering topics") to one of their long-term outcomes (e.g., "increase the number of individuals who choose science and engineering careers"), they may decide to focus the current evaluation on demonstrating that program activities lead to that medium-term outcome. If the evaluation can provide evidence that program participation increases student excitement about science, they can then logically argue (based on the appropriate research evidence) that the long-term outcome should follow.

The working group should keep in mind that they may not be able to find evidence or measures for all of the relations or elements that have been articulated in their pathway model.

2.09 Reflection and Synthesis

The purpose of this step is to take stock of where the working group is in their thinking about the program and its evaluation, to review and revise the work done so far, and, if possible, to interview stakeholders so that their opinions can also be integrated.

It would be a rare program indeed that goes through the steps in the Protocol in sequential order one time and gets it "right" or to a level with which they are satisfied. If this process is working correctly, when they get to this synthesis step the working group should immediately see things from earlier steps that they want to go back and revise or rework. In addition, when possible, interviewing stakeholders can shed new light on the program and can inform revisions to the established program boundaries and model (you may wish to revisit *Stakeholder Interview Guide,* **Workbook pages 9-10**). However, this process is not always feasible. Therefore, it is important that the working group continue to keep stakeholder opinions in mind throughout the revision process.

The synthesis step actually occurs throughout the entire planning process, but we place the step here so that you deliberately stop at this point to assess the entire program modeling picture. Another way to think of this synthesis step is to contrast a step-by-step process with a more dynamic and adaptive one. If you look at the steps in the planning stage of the Protocol you might get the impression that the way to do evaluation planning is to take each step in order and then be done. In fact, what we want to encourage is a more dynamic approach to planning that cycles through these steps several times – and not even necessarily in the same order – until the working group is satisfied that they have developed a high-quality model and evaluation scope appropriate for their purposes.

When does this process end? In one sense it never does. In another sense, we have to be practical and recognize that the goal is to get the best model for our purposes within the time that we have available. So, ideally the working group will decide to end the process at some point that makes sense for their work – and recognize that this is a process they should revisit from time to time.

When the working group has settled on a program model that they will use going forward in this process, they may choose to do a more formal review of their own, using the structured feedback document (*Logic Model and Pathway Model Feedback Form*, **Workbook pages 43-46**). Structured feedback can help to provide an "outside" perspective on the program logic and pathway models.

One of the key products of this synthesis step is a revised program description that reflects new thinking about program boundaries, stakeholders and outcomes. The thinking done at this step in the Protocol will also contribute to a concise evaluation purpose statement, which is the second step in the evaluation planning phase.

At this point the working group will have the compiled set of materials that reflect the work done in this stage of the protocol, including the stakeholder map, mission and description, program logic model, pathway model, boundary analysis, lifecycle analysis from the perspectives of key stakeholders, evaluation scope and system links.

Stage 3: Evaluation Plan Creation

Introduction

Now that the program models have been refined, the working group can move from talking about the program to talking about evaluation. This stage of the Protocol involves several steps that build on the prior stage to aid in the creation of an evaluation plan that will guide the implementation of the evaluation. This stage includes: Introduction to Evaluation Plans, Development of an Evaluation Purpose Statement, Development of Evaluation Questions, Development of Measures and Measurement Strategies, Development of a Sampling Plan, Identification or Selection of an Evaluation Design, Development of an Analysis Plan, Development of an Evaluation Reporting Plan, Development of Evaluation Schedule and Implementation Plan, and the Finalization of the Evaluation Plan.

One of the key aspects of systems evaluation planning that distinguishes it from traditional evaluation planning is the concept of the program lifecycle as described earlier. When you are developing your evaluation plan, it is crucial to consider the lifecycle phase of your program. That is, the evaluation questions, sampling approach, measurement choices, design, analysis, reporting and use will be different for programs at different phases of development. In the sections below we will emphasize this concept and how it is used in developing your evaluation plan.

3.01 Evaluation Plan Overview

The goal of this step is to present and discuss the purpose and components of an evaluation plan. The working group should understand the work that is entailed, and the importance both internally and externally of having a fully articulated and written plan.

A written evaluation plan captures important program and evaluation information and guides the implementation of the evaluation. A written plan can be especially important if there is new staff that will be doing some of the work. In addition, having a written evaluation plan can help to establish credibility with stakeholders – especially funders – because it conveys information to them and is a demonstration of your commitment to quality evaluation.

The written evaluation plan you develop will likely include section headings that map onto many of the steps in the Protocol. Several of the outputs already created from Stage 2 will be used to develop the written evaluation plan. Typically, we include the program mission statement and program description, the program logic model and pathway model, and stakeholder map, and evaluation scope. The plan will also include new products– the evaluation purpose statement, the evaluation questions, a description of the proposed sample, the proposed measures, the proposed design, a data collection and management plan, the plans for data analysis, plans for evaluation reporting and utilization, and an evaluation timeline. These new components are the classic elements of any good research plan, covering the objectives and methodology to be used. In addition to the written sections of the plan, a complete document will include, as appendices, the actual measures you will be using (if available). Sometimes the logic model, pathway model and stakeholder diagram are found in the appendices.

A key distinct concept in evaluation planning is differentiating between an evaluation cycle and the evaluation of the program over its entire life. An evaluation plan is typically written in terms of how the staff will evaluate the program during the next evaluation cycle. An evaluation "cycle" is the period of time over which a single iteration of evaluation occurs. Many organizations use an annual evaluation cycle, others may be more frequent if the program has a shorter cycle or if they require rapid feedback to aid program development. The endpoint of an evaluation cycle is often determined by when the organization and program are doing major reports. Again, many organizations find it convenient to do annual reports, which is a tip-off to the notion that they are on an annual evaluation cycle. When the working group developed the evaluation scope in Step 2.08, they were essentially determining the extent of the program evaluation they were going to take on in the next evaluation cycle (e.g., over the next year).

We have already seen that programs evolve through phases over time. Evaluation needs to change and adapt so that it is always appropriate (that is, symbiotically linked) with where the program is in its development. This means that over the life of a program there are likely to be multiple evaluation cycles and the evaluation plan for each cycle will change. We will keep coming back to this idea in the sections below. The **Workbook (pages 75-92)** offers both *Tips for Writing Evaluation Plans,* and the *Feedback Form for Evaluation Plans* as resources for preparing and assessing your plan.

3.02 Evaluation Purpose

The Evaluation Purpose section of the evaluation plan provides a short description of the reason for conducting the current evaluation and what the working group hopes to achieve by it. Although we touched on this during the Mining the Model activities, writing the Evaluation Purpose section is an important part of any evaluation effort and must be revisited. It helps to define the goals and boundaries of the current evaluation effort, and can serve as a "big picture" reference point for the working group as they work through the details of developing an evaluation methodology. (For *Guidance for Evaluation Purpose Statement, see* **Workbook page 95.***)*

The Evaluation Purpose section appears at the beginning of the evaluation plan document and should describe what is and what is not being evaluated and the overall goal or purpose of the evaluation. You might even think of it as a kind of "mission statement" for the evaluation (rather than for the program). It sets boundaries by identifying the program elements being considered for the evaluation, which goals or objectives are of most interest, and what the intended uses of the evaluation results are. As the working group completes other steps in the Protocol and makes decisions about aspects of the evaluation strategies, the Purpose section should be revised or updated to include a brief description of the methodology choices that have been made and the reasoning behind these decisions.

Since one of the major functions of the Evaluation Purpose section is to provide the overall rationale to guide the specific detailing of the evaluation plan, the Purpose Statement will tend to vary systematically with the program lifecycle. For instance, in an early phase evaluation the major purpose might be to provide a thorough description of the process of delivering the program or of its implementation so that the program might be further developed or improved in subsequent iterations. A Phase II purpose might emphasize that the evaluation is conducted to assess the degree to which the program appears to be associated with key outcomes of interest. In Phase III, the purpose might stress the desire for the evaluation to demonstrate the effectiveness of the program on key outcomes.

It is important to keep in mind that this section should be brief. It is a good idea to begin this stage by drafting an initial Evaluation Purpose section (recognizing that it will be incomplete at this point in the process). As you complete each of the subsequent steps in Stage 3, revisit the evaluation purpose statement and add to and revise the statement.

3.03 Evaluation Questions

Evaluation questions are the broad questions asked about the program in general – "Is the program being implemented well?" or "Does our program have an effect on outcomes A, B, or C?" They will focus the evaluation planning and anchor all subsequent sections of the Plan. Note that evaluation questions in this context are not the same thing as questions that would be included in a questionnaire or interview for participants (those are measurement questions). Evaluation questions are the major questions your evaluation is trying to address in a specific evaluation cycle. Every question will have at least one associated measure in place by the time the evaluation plan is complete.

It is important for the working group to keep in mind the lifecycle phase of the program and its relationship to the evaluation lifecycle. Each stage of the program lifecycle emphasizes different types of evaluation questions. Remember that the goal is to eventually align the program lifecycle and evaluation lifecycle. Depending on where the program is starting from, this may not be feasible

in just one evaluation cycle - it may take several evaluation cycles for the program and evaluation lifecycles to be aligned. When developing evaluation questions, it is important to consider what can be done to move the lifecycle and evaluation phases into better alignment.

In addition to maintaining alignment with the program and evaluation lifecycles, fidelity to the program's pathway model and purpose statement, language and format have to be carefully considered while the working group continues to brainstorm, revise and edit evaluation questions. Each evaluation question should clearly identify the program element(s) that is/are being examined (one or more activities, the overall program, specific outcomes, etc.) The phrasing of evaluation questions matters. Consider the difference between asking whether a program is "associated with a change" as opposed to "causes a change." The difference could mean a much more (or less) controlled evaluation strategy, and enable a much different claim that could be made at the end of the evaluation cycle.

When writing the evaluation questions, make sure the questions are clearly worded and understandable to someone who is unfamiliar with the program. Be sure to consider (and possibly revisit and revise) the evaluation purpose statement so that the questions make sense given the overall purpose of the evaluation.

The following framework will help to craft evaluation questions for the selected scope of the evaluation. Keep in mind that these are only examples of questions and that there are many "correct" ways to word questions (though some are more appropriate for certain programs than others). To develop an evaluation question, start by identifying a program element (activity or set of activities, outcome, or link) that falls within the evaluation scope in your pathway model. For instance, examples of a Phase 1 evaluation lifecycle question might be:

How well is [ACTIVITY "A"] implemented?

To what extent are participants satisfied with [Activity "B"]?

How useful were the program handouts?

To what extent do participants demonstrate [OUTCOME "X"]?

Phase 2 evaluation questions are typically focused on assessing changes in outcomes associated with participation in the program. Examples of Phase 2 evaluation lifecycle questions might be:

Is participation in the program associated with a change in [OUTCOME "X"]?

How do participants' levels of [OUTCOME "Y"] compare to non-participants' levels of [OUTCOME "Y"]?

Phase 3 evaluation questions will typically explore issues of causality or "effectiveness" of the program. These require the use of evaluation designs that are more carefully controlled and/or longitudinal (i.e., follow participants over a longer period of time). Many programs will not reach a Stage 3 evaluation in terms of their program lifecycle phase. Either their purpose is not to demonstrate the effectiveness of the program, or evidence from earlier phase evaluations will cause the organization to abandon the program and try an alternative approach to the problem. For those programs where a Phase 3 evaluation lifecycle question would be appropriate, some examples of questions might be:

To what extent does participation in the program cause [OUTCOME "Y"]?

To what extent does participation in the program cause changes in [OUTCOME "Z"]?

Phase 4 evaluation questions are focused on understanding the generalizability of the program to other contexts, settings, and participants and are typically used when the goal is to broadly

Once again, alhough these steps may all be done on paper or in your word processing document, these steps may be documented in the Netway (under the Evaluation Plan section of your program). Use the worksheet section to do your drafts, and the plan section (above the worksheet section) to enter your actual plan. This can facilitate aligning your plan by evaluation questions, as well as printing your final models and plan.

disseminate a program that has already clearly demonstrated its effectiveness through rigorous prior evaluations. Examples of Phase 4 questions might be:

> Does participation in the program have the same effect on participants of [different ages, different races, living in different contexts (e.g., urban, rural, suburban)]?
>
> Can the program be replicated with similar results across multiple settings?

In general, if there are multiple activities that lead to one crucial outcome --or as is more often the case, multiple outcomes that arise from an individual activity-- you may certainly use a semi- colon and bullet points to collapse more than one evaluation question into a single sentence, such as:

> What is the effect of the program on: [OUTCOME "X"]; [OUTCOME "Y"]; [OUTCOME "Z"]?

Keep in mind that when we are developing evaluation questions we are generally only looking for a few key overarching questions that can guide the evaluation during its next cycle. Imagine we have a program with five activities. We could develop five separate evaluation questions that ask "How well are we implementing Activity 1?", "How well are we implementing Activity 2?" and so on, or we could shorten this to "How well are we implementing the program?" When stating evaluation questions, use fewer and more general questions if possible.

Typically you want to keep the number of evaluation questions manageable. Three or four evaluation questions will in most cases be sufficient to accomplish a reasonable evaluation in any given evaluation cycle. Use the *Developing Evaluation Questions* worksheet, and the resource *Evaluation Questions and Program Lifecycle* (**Workbook pages 99-101**) to help you develop your questions. The questions will provide a basis for the rest of the evaluation plan.

Activity: Question-Claim Match-Up, Stakeholder Review
Once the working group has a set of draft evaluation questions developed, there are a couple of quick activities that can help narrow the list down, and ensure that the questions are appropriate to program priorities and desired claims.

- ***The first activity is a "question-claim match-up." This activity simply consists of brainstorming the possible claims one could make after collecting data to address the question at hand. For example, if the question were "Is the Earth round?" the claim you would hope to make is that either A) the Earth is round or B) the Earth is not round. This exercise should help working group members recognize how the language and phrasing of their question may affect their methodological choices and the overall usefulness and appropriateness of their evaluation effort. This activity should be done as a group, because brainstorming about claims may bring up differences in perspectives and opinions that should be settled before moving forward with the evaluation plan. The "Getting to Measures" worksheet in the Workbook (page 109) can guide this discussion.***
- ***The second activity is a peer or stakeholder review. If possible, invite stakeholders and/or peer program implementers to take a look at the developed evaluation questions. The following prompts may help guide their feedback:***
  - ***What sort of claims would be possible, if the evaluation yielded favorable evidence for this evaluation question?***
  - ***Does this evaluation question fit the Evaluation Purpose Statement? How/how not? Any suggestions?***
  - ***Comment on the alignment between this evaluation question and the program's lifecycle stage; does this evaluation question make sense for "where" this program "is"?***
  - ***Does this evaluation question clearly relate to the Program Model? If so, how? If not, what's missing?***
  - ***Does this evaluation question make clear exactly what is being measured (the key constructs)? The group(s) to whom the question applies? What the basis for comparison will be (if appropriate)?***

3.04 Measurement and Measures

Measurement is a very complex topic. Measurement approaches need to be tailored to the specific program and lifecycle phase of program development. This makes it very difficult to specify simple rules for how to develop a measurement plan for any given study. There is a huge literature on how to apply the great variety of measurement approaches that one might use in an evaluation. It is important that the Evaluation Champion be familiar with the scope of strategies and methodologies available. There are widely available resources that would be useful when thinking about measurement planning (e.g.: http://www.socialresearchmethods.net/kb/measure.php).

When writing the measurement plan, consider each evaluation question, identify the focal construct, and describe in detail the measures that will be used. Be sure to include a description of the measure type (e.g., survey, observation, interview, etc.), identify the origin of the measure (cite the source if the measure was found in the literature or describe the development of the measure if you are creating a new measure), and discuss the reliability and validity of the measure (if the reliability and/or validity has not been tested, be sure to state this as well).

This Protocol step is broken into two parts: "Defining Variables and Measurement Strategies" and "Identifying Measures." Evaluation Champions may find that it makes the most sense to think about the first part, which builds from work on evaluation questions, directly after developing the questions. However, it may be advisable to look ahead to sampling and analysis before returning to work on identifying measures, which can be an intensive research and development process. Though the entire Protocol is designed to be iterative, you may find that the steps in this stage are particularly difficult to think about separately. This is because decisions about measures and measurement, sampling, design and analysis all depend on one another, and therefore must be considered simultaneously.

1. Defining Variables and Measurement Strategies

Before you can really begin to think about measurement, you first need to clearly define what it is that you are trying to measure. Begin by reviewing the evaluation questions. The variables are the things that you think influence or are influenced by something else. For example, if the evaluation question is whether participation in the program is related to an increase in science knowledge, then the variables are participation in the program and science knowledge. When a variable is an abstract idea that is derived from empirical evidence but may not be directly observable (such as behaviors, attitudes, or knowledge) it is called a construct. It's important to clarify the variables and construct(s) so you know what it is that you are trying to measure.

After the variables have been identified, they must be defined. Often, what "science knowledge" means to one stakeholder may be different from what it means to another. This is another step in the Protocol when it is critical to bring in multiple stakeholder perspectives. Ideally, the working group will invite stakeholders to participate in a discussion that will establish internal working definitions for the variables identified in each of the evaluation questions.

In addition to developing definitions, the working group will need to brainstorm indicators of the presence of the variable they define. This can be particularly tricky when trying to think about indicators of constructs. By their very nature, constructs are not directly observable and so the working group needs to think about how they would "know" the construct exists. This may be best described as brainstorming what a construct "looks like." For example, an indicator of "science knowledge" could be the ability to explain a science concept to a peer, or adequate performance on a science test, while an indicator of "interest in science" could be signing up for additional science courses, or engaging in science-based hobbies.

Once the variables are clear, you can begin thinking about measurement. There are many different strategies that you can use to measure variables. Some strategies will be more appropriate than others depending on the variable you are trying to measure and the context. Examples of common measurement strategies are surveys, observations, interviews, and focus groups.

Using a table like the one shown below can help working group members think through these questions systematically. Refer to **Workbook pages 103-108** for some examples, and utilize the *Getting to Measures Worksheet,* and *Key Constructs and Measurement* (**Workbook pages 109-111**) to do this for yourself.

Activity: Peer Evaluation Plan Brainstorm

One of the best ways to come up with an appropriate, creative and efficient evaluation plan strategy is to brainstorm as many strategies as possible and then narrow them down based on your priorities. Often, peers and stakeholders outside of the working group can provide a useful and creative contribution to this brainstorming process.

First, the working group will need to identify peer program implementers and/or stakeholders who are willing to participate in a plan brainstorming session. Then, using the "Identifying Key Constructs and Measurement" (see Workbook Page 111), the working group should fill in the left-most column with each of their final evaluation questions. Each peer or stakeholder is then asked to fill out the remaining columns of the table, offering ideas about constructs, their definitions, measurement strategies and tools.

If the working group chooses to conduct this activity later in the planning process, they may also include columns for sampling, design and analysis strategies. With these columns included, the working group can also use the brainstorm guide to develop a "rehearsal" evaluation plan for an imagined or well-known program. Brainstorming an evaluation strategy for a neutral program will help working group members better understand the consequences of their choices and the challenges facing them as they develop their own evaluation strategy.

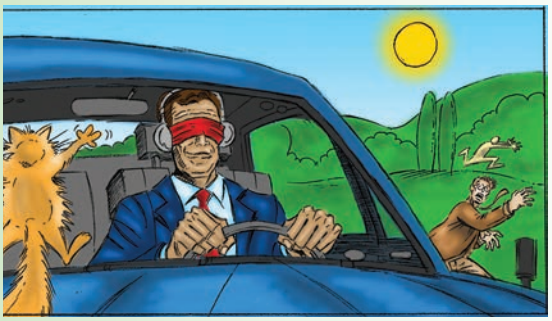
Evaluation Question	*Constructs/ Variables*	*How is it defined?*	*What does it look like?*	*How might it be measured?*
Is program participation related to an increase in science knowledge?	Science knowledge	Understanding of science concepts	Ability to explain science concepts to a peer	Track and record peer-to-peer explanations using video and an observational checklist

The process of identifying variables, defining them and brainstorming measurement strategies should be repeated for each question. When this process is complete, the working group should be able to easily identify the number and types of measures that they will need. Often this process will lead to the realization that one measure may be able to address (or collect data that will address) more than one evaluation question. Once this list of desired measures has been created, the working group should reconsider the following about their proposed strategies before moving onto identifying measures:

- Program Lifecycle (is this type of measurement appropriate to the program's developmental stage?)
- Feasibility (Do you have resources (staff, budget) to collect data in this method, or to analyze the data? Is the methodology suited to the program context (time required, setting/locale of the program, age appropriate)?
- Accuracy (Does this methodology get you as close to the source of actual data as possible? For example, teacher reports of student skills may provide different results than direct student observation. Does the methodology present confounding factors (does a written test actually measure knowledge, or does it measure reading skill?)
- Credibility (Will program stakeholders find this kind of evidence credible?)
- Usefulness (Will the results provide information that is useful to program practitioners or program development?)

Driving With Your Eyes Open - Feedback

In dynamic living systems different parts of the system continually interact with and influence each other over time. An essential aspect of this involves the role of feedback. If we didn't have feedback we simply would not be able to function as living beings. Imagine trying to drive a car down the road without being able to see anything or hear anything, without the basic sense mechanisms that provide information about what we are experiencing. That's essentially what feedback means, and it is an essential concept in systems thinking.



At a macro level you might think of evaluation itself as a feedback mechanism for our society. Without evaluation we are "driving blind" in trying to operate our programs (an unfortunate state of affairs that occurs far too often). In addition, without feedback, learning will not occur. While an individual program evaluation can tell us something useful about a program, multiple evaluations of many similar programs over time can provide us with the cumulative feedback needed to begin making sense of what works – but only if we have systems for accumulating and synthesizing the individual evaluations.

The idea of feedback is also critical within the evaluation of a program. We gather input from multiple participants, to get their feedback about what they think the program is and what it is trying to do. We feed back the results of an evaluation so that others can see what is happening and learn from it. Feedback is essential even throughout the process of conducting the evaluation. When we are collecting data, we monitor whether we are getting adequate responses and how respondents and others in the organization are reacting to the process.



So, feedback is everywhere in an evaluation. One of the purposes of systems evaluation is to encourage us to become more conscious of the informal feedback that is already occurring and to make it more coherent and structured so that it can function even more effectively and we don't drive off the road.

Unobtrusive or Nonreactive Measures

One of the biggest practical challenges in evaluation is motivating program participants to engage in providing data. When analyzed from a systems perspective, it's typically the case that both the participants and the evaluators want something out of the program. Participants usually want the program itself, or what it might do for them potentially. Evaluators want data or information about how the program affects the participants. And, each has negative motivators. Participants usually don't like the burdens that formal measures, especially tests, impose on them. And evaluators don't like being in the position of having to impose on the participants (and often the program staff). From a systems point of view, the ideal potential solution often amounts to looking for a symbiosis between the interests of the evaluator and the program participant.

Now consider how this might inform how we approach measurement. Let's say that in an informal science education program designed to teach children about how to use a microscope you would like to assess their knowledge of the material conveyed in the program. You could construct a paper-and-pencil test that the children would complete at the end of the program. But that would be obtrusive (not to say a drag!) and it would be better if you could assess knowledge more symbiotically and less obtrusively. How might this be done? One approach would be to collect data about knowledge of microscope use in the natural course of their doing the program. You might do this by observing how they try to use the microscope initially and how they perform with it in the last task of the program (a type of before-after assessment). This could be done by observing them directly or by rating or scoring the results of what they record in the natural course of using the microscope. It may even be that the "program" doesn't involve a real microscope but is one simulated through a computer program. In this case, measures of performance could be unobtrusively built into the software itself. In this example of measurement symbiosis, both parties get what they want. The children who participate get to take part in an engaging (we hope) program without doing any burdensome tests and the evaluator gets data on knowledge as reflected in their performance without having to cajole them to take a test (or impose that requirement on the program staff).

These kinds of measurement approaches are known as unobtrusive or nonreactive measures. We have several favorites that we like to cite. For instance, when an evaluator wanted to assess peoples' radio station preferences, instead of doing a survey research study they came up with the clever solution of having auto mechanics in the area note what radio stations were on in the cars that were brought in. Or, when museum evaluators wanted to assess which exhibit paintings people were most interested in they took the creative approach of replacing selected floor tiles in front of each painting and then making careful measures of wear-and-tear at the end of the exhibit. There are lots of potential ways to conduct measurement unobtrusively: direct observation and coding, photography, video, use of archival research, and so on. In fact, one of the most fertile sources of short-term outcome measures is likely to be program outputs – products that are naturally generated in the course of participating in the program. When we take these outputs and code, rate or score them we are in effect turning them into measures that might reflect things like performance, knowledge or even satisfaction or interest. Although unobtrusive measures may require considerable forethought and preparation, they can also be fun to create and integrate into a program. This helps everyone get what they want and encourages a greater symbiosis between the program and its evaluation.

Precisely because these types of measures can be implemented without the participants' knowledge, it is especially important with these strategies to be careful with issues of privacy and protection of human subjects more broadly. If in doubt about the ethics of a particular strategy, be sure to consult with an expert in this area. Universities have Institutional Review Boards or other entities charged with ensuring protection of human subjects in research and evaluation. Outside universities, there are experts and consultants who specialize in this.

2. Identifying Measures

Choosing appropriate measures for each evaluation question will likely be one of the central challenges in writing and executing the evaluation plan. Measures will generally fall into one of

three categories:

1. Demographic or descriptive measures - measures that track (simply count) events and/or participants and, if relevant, their characteristics
2. Process measures - measures that capture the type or quality of the program event or interaction
3. Outcome measures - measures that capture effects of the program including associated change for a group, significant change for an individual, causal relationships between activities and outcomes.

Many programs will use more than one of these types to address their evaluation questions, and each of these types could be formatted in a variety of different ways.

There are three main strategies that can be used when attempting to identify a measure that is appropriate for a given program and question. These include:

- Using existing local measures (those already in use within the organization)
- Locating measures in the literature (those that have been used in similar settings and are discussed in peer-reviewed literature)
- Developing measures from "scratch" or by modifying a pre-existing measure to fit your program)

In order to maximize accuracy, as well as credibility to many stakeholders, it is often desirable to use an established and validated measure, but at times modifying or creating an existing measure is necessary. The following questions/steps can help locate or identify potential measures, and are listed in sequence of how we approach identifying measures.

Using Existing Local Measures (already in the office)

Begin by looking at the measures currently being used by the program. Perhaps program practitioners have developed and used their own measure for quite some time, or maybe they already have a measure that fits their needs.

- What measures are currently used in the program?
- Are there specific measures that have been mandated by a funding agency?
- Are the current measures something that have been tried and true, and have literature to support them? (This would be ideal, and if the measures don't fit this category there may be a tradeoff between resources available to locate an appropriate established measure and using the existing one.)
- Do these measures match the evaluation questions and are they appropriate given the evaluation lifecycle? (This is a critical issue – existing internal measures may have been developed for a different evaluation purpose and for different evaluation questions. Be sure they really will serve your current purpose and questions.)

Locating Measures in the Literature

A second option is to locate measures that are already developed and are supported in the literature. It may make sense to try and locate measures in the literature if you have evaluation questions that do not have measures already in place, or for evaluation questions that are using measures that have not been validated. Measures obtained from other sources will need to be cited appropriately by the program, and some may even have fees associated with them. Networking with colleagues can also help to broaden your measures library. Do you have colleagues in another department, office, or geographic location who may be measuring the same activity or outcome of interest? Have they found a measure that is working well? Sometimes you may find a measure in the literature that has several subscales (portions of the measure that each addresses a different construct.) You may not be interested in all of the subscales included in the measure, but you may find that one or more of the subscales addresses your outcome of interest. Sometimes these relevant subscales can be found in measures that, on the surface or taken as a whole, seem unrelated to your target outcome. The following are some questions to consider when thinking about looking for measures

in the literature:

- Which evaluation questions do not yet have an identified measure?
- Which evaluation questions may already have an existing measure, but the working group would prefer to have a measure with stronger support in the literature?
- What measures are colleagues using to measure similar outcomes, and are they available for this program?
- What does a literature search for related programs, activities or outcomes reveal? A validated measure may be identified either in a research article or through a related article that is cited in the bibliography.
- Is there a larger measure that has a sub-scale that could be used to measure the outcome of interest?
- Is there a measure that can address more than one evaluation question?
- Is there an existing measure that can be modified to fit the program's needs by changing a word or two?

Developing Measures

Developing a measure is likely to be appropriate particularly for newer programs where there is no history of prior measurement, or in fields where there are limited measures already available. Any time a new measure is created, people will question how good the measure actually is at measuring what it is supposed to measure (validity) and whether it does so consistently and dependably (reliability). Eventually, new measures will be expected to undergo testing to demonstrate their reliability and validity (measures that you find in the literature have typically been tested for reliability and validity and a good measure will report just how reliable and valid it is. This is an advantage of using a measure that you find in the literature).

Creating new measures is disadvantageous in that it will not allow you to cite evidence of reliability and validity for the current evaluation, nor compare results to those obtained by others. However, creating a new measure may be the only option available for many programs, and creating a new measure, pilot testing it, and assessing and refining it can, over time, be the foundation for a good new measure. (See other resources, such as http://socialresearchmethods.net, for information on validity and reliability.)

Key Questions Regarding Measures:

When making decisions about measures, be sure to use your evaluation questions as a guide. When thinking about which measures to use, review the following questions:

- **Identify outcomes to be measured**. What exactly are you trying to measure? Does the measure you have actually measure the outcome of interest? For example, if the outcome of interest is self-esteem, make sure you have a measure of self-esteem and not self-concept or some other similar construct. These are different things. If the evaluation tool is a broad measure, does it address all of the outcomes you want to measure? Does it cover "too much"? That is, does it collect data you do not need and won't use? If so, try to pare it down in order to not waste your or your participants' valuable time and attention.
- **Determine the measurement strategy.** Which measurement strategy is most appropriate given the outcome that you are trying to measure and the context in which the program is taking place? Surveys are a common measurement strategy; however, they are not the only

New Programs			Mature Programs
Informal Open Ended Pilot testing Exploring measurement approaches	Reliability Consistency Validity	Standard protocol of measurement instruments	Formal Structured Quantitative Identifying existing measures

possible strategy. Consider whether using interviews, observations, content analysis, etc. might be appropriate.

- **Match measure to sample**. The sampling plan may target adults or youth, each having its own literacy level, etc. Is this measure appropriate for the sample?
- **Consider lifecycle.** Does the measure fit the stage of the program and evaluation lifecycles? Newer programs are probably looking for rapid feedback from participants about their reactions to the program - which might be met with simple satisfaction surveys; whereas more mature programs will be looking to show cause and effect relationships and might use more established and tested measures. Below is a tentative chart of the intentions of evaluation at each stage of development.

Assess quality of the measure. This is the "bird in the hand vs. two in the bush" decision. A program may have a choice to make between using what is on hand already (which may be ready to go, and may even have data from past years giving evaluators the opportunity to compare results), or trying to find a "better" existing measure. A "better" measure in this case might mean one that has been tested in careful studies for validity and reliability, has the credibility of having been used in additional research papers, and for which large-scale study results are available to which results can be compared.

Assess feasibility. There's no point in listing a measure in an evaluation plan if it is simply not realistic that program staff will be able to find it, afford it, modify it appropriately, test it, use it, analyze it, and/or report on it. Will staff have time available to use this measure?

Assess strategic Value. If time and resources are limited then efforts should be focused on the opportunities that have the highest "payoff". Consulting with stakeholders or advisory groups is recommended in order to be sure that the choice is made well.

Identify references. If using established or named measures, are they properly referenced? Measures developed and field tested by others should be cited in the evaluation plan and in any other writing where the measure is mentioned.

At the risk of repeating ourselves, remember that decisions about measures don't occur in a vacuum – they are related to lifecycle, sampling, design and analysis issues, that is, they will both affect and be affected by these other topics. For help with this decision making process, see **Workbook pages 103-134** for several resources to assist this, including: *Getting to Measures Worksheet; Key Constructs and Measurement; Introduction to Measurement and Measures; Obtaining a Measure - Find, Modify or Write; Measure Checklist; Finding Measures;* and *Survey Review Form.*

3.05 Sampling Plan

After the evaluation questions have been identified, the working group needs to describe the source of the evaluation data. Sampling is the process of selecting units (e.g., a subset of people, things, documents, events, organizations, or groups) from a population (the entire set of people, things, events, documents , organizations, or groups) of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen. The Evaluation Champion should be familiar with at least the general ideas behind sampling, including external validity issues, and the distinction between nonprobability and probability sampling. There are many resources available on sampling, and while these topics will be covered in brief here this discussion should be supplemented with external resources including the *Developing a Sampling Plan* document in the **Workbook pages 137-141**, and other sources such as http://www.socialresearchmethods.net/kb/sampling.php. Sampling will also be affected by measurement and evaluation design, so once again these steps should occur in a parallel, dynamic and interactive process rather than in sequential fashion.

Key Concepts in Sampling

Unit of Analysis: In evaluation, sometimes we focus on individuals and sometimes we focus on groups. The level on which the evaluation is focused (e.g., individuals, families, classrooms, schools, etc.) is called the unit of analysis and will depend on the focus of the evaluation question. It's essential that the selection of the unit of analysis is done consciously because the unit of analysis selected for data collection must be the same as what we use to draw conclusions. For example, imagine that an evaluator collects data from individual adolescents on the amount of time they spend engaged in the after-school program and on their risk-taking behaviors. The

evaluator analyzes the relationship between after-school program involvement and risk-taking to see if adolescents who are more engaged in the after-school program have lower levels of risk-taking. The data describes the individuals, the conclusions drawn are about the individuals, and the individuals are the unit of analysis.

In some evaluations, groups are the unit of analysis but data are collected from individuals. In other words, the unit of analysis may not be the same as the unit of observation (the cases about which measures are actually obtained in a sample). For example, imagine that an evaluator hypothesizes that classes that use the Jolly Phonics reading program will have higher reading proficiency scores than classes that do not use the program. Reading proficiency is measured by giving each of the individual students a test. However, the individual student test scores are averaged together to create a classroom average score for each classroom. Then, classrooms that use Jolly Phonics are compared to those that do not use Jolly Phonics. It is the differences in the classroom average scores that are used to explain variation in reading proficiency between classes. In this example, the unit of observation and the unit of analysis are not the same. The unit of observation was the individual students and the unit of analysis was the classrooms. In the previous example (after-school program participation and risk-taking) the unit of observation and the unit of analysis were the same (individuals).

Generalizability: You will want to begin by identifying who or what you want to be able to say something about. Imagine that we are interested in evaluating "no tolerance" drug policies in high schools in the United States. Ultimately, we want to be able to say something about all high schools in the United States so our **population of interest** is all US high schools. Next, we need to identify our **sampling frame**. This is a list of all elements in the population. In our example, the sampling frame would be a list of all high schools in the United States. The **sample** (the subset of the population) is drawn from the sampling frame. Evaluators generalize from samples to populations if the sample is representative of the population. Depending on the sampling technique used, we can be more or less confident in the representativeness of the sample (see sampling strategies section below). In some circumstances it may be feasible to avoid the issue of generalizability by conducting a census (studying the entire population of interest). For example, if you are only interested in generalizing the results of your evaluation to the actual participants in your program, you could conduct a census by including all program participants in your sample.

Sampling Strategies: Remember that our primary objective when selecting a sample is to try and make it representative of the population to which we are interested in generalizing our results. If we want to be able to say something about all participants in the program, but we cannot actually study all program participants, we want to make sure that the sample is representative of all program participants. There are 2 major sampling strategies: Probability and nonprobability sampling techniques. In general, **probability sampling** allows us to be most sure that our sample is representative of the population. Probability methods rely on a random selection method so that the probability of being selected for the sample is known. **Nonprobability methods** do not rely on random selection and the probability of being selected for the sample is unknown. A few common probability and non-probability sampling approaches are reviewed here, but the reader is encouraged to explore outside sources for additional information.

Probability Sampling Strategies: Simple Random Sampling is a technique that gives every element in the sampling frame the same probability of being selected for the sample. For example, if I wanted to draw a simple random sample from the population of program participants, I might assign all participants a number and randomly select some subset of participants for the sample using a random number generator (e.g., using a random number function in Excel). Many populations are made up of clusters within hierarchies. For example, the individuals who make up the population of 3rd graders are clustered within schools. Cluster Random Sampling makes use of these clusters to aid in sampling. First, the evaluator can randomly select the clusters and then, from within the selected clusters, randomly select the sample. Note that in order to be truly representative of the population, cluster random sampling requires that the process of selection must be random at each stage of selection. A non-probability approach to cluster sampling can be used (see section

on Hierarchies below), however the results are not as broadly generalizable.

Non-probability Sampling Strategies: Convenience Sampling is a technique whereby the sample is selected based on convenience and ease of access rather than based on representativeness. Convenience sampling is appropriate for early lifecycle evaluations where the goal is not to achieve generalizability beyond the participants included in the evaluation. Purposive Sampling is a technique in which the sample is selected deliberately (though not randomly) because the participants have some very specific characteristic of interest. This approach makes the most sense when the evaluator has a great deal of knowledge about the population of interest. Purposive sampling does not produce a sample that represents some larger population but it can be exactly what is needed for earlier lifecycle evaluations where the interest is less in generalizability than in getting initial evidence about how the program performs with a specific group.

Determining the Sample

The working group should consider: "Who will participate in the evaluation?" Guide the working group to focus exclusively on whom or what will answer the evaluation question(s) and can be measured. Do not fall into the trap of broadly describing the population served by the program. Focus specifically on the population and sample that is relevant for the evaluation question(s). For instance, imagine that there is a program for mothers of premature babies. The evaluation question is "does the program improve the height and weight of the babies?" The primary measure for this evaluation question is the height and weight of the babies at the end of the program. The sample should describe the babies since they are the focus for this evaluation question (not the moms).

The program description section should include a rough estimate of the number of participants predicted for the coming year. The sample section should describe whether some or all of the participants will be included in the evaluation (e.g., the % of participants who will be "sampled"). This will allow readers to determine to what degree the results are generalizable to those who were involved in the program. For instance, if the program expects to have 1,000 participants, yet staff only plans to sample 20 of them, they might have a difficult time generalizing the results to all participants.

As with other aspects of the evaluation plan, sampling changes over the life course of a program. Programs in the Initiation lifecycle phase will probably select their sample based on availability and convenience in order to generate rapid feedback. More mature programs that are trying to make stronger assertions based on their evaluation will have to more formally address internal validity issues and generalizability (and may therefore need to use a probability sampling technique).

Phase I	Phase II	Phase III	Phase IV
Convenience .....Formalized Opportunistic..... Deliberate / Planned Non-Probability .....Probability			

Hierarchies: One issue to keep in mind when sampling is that there may be hierarchies or multiple levels at which different types of sampling take place even within a single evaluation. For instance, if you are conducting an educational program, you might sample school districts, schools within those, grades within those, classrooms within those and students within those. At each level you might use different approaches. For instance, your choice of which school districts you can work with may be predetermined and entirely opportunistic – you'll work with whichever district is close and willing to participate. However, you might be able to select different schools or classes within school districts in a systematic way. In this example, you could only generalize the results of the evaluation to the selected school district. As programs mature, sampling plans also tend to mature and become more structured and complex.

When writing the sampling plan for the evaluation, consider each evaluation question and describe in detail the population of interest, who will participate, approximately how big the sample will

be, how the sample will be recruited, whether there are multiple levels and/or types of sampling strategies employed, and how participants are selected at each level (i.e., sampling strategy) (see **Workbook page 143,** *"Sampling Plan Worksheet"*). And, keep in mind that decisions made here affect and are affected by decisions made in other parts of the evaluation plan.

3.06 Evaluation Design

An evaluation design shows how the evaluation is structured with respect to measurement, administration of the program, sampling and any comparison groups that are included. It provides an important schematic that can be used to guide the choice of data analysis. Simplified general research designs are described below, but selecting a design will vary depending on Evaluation Champion and working group preferences. Once again, we refer you to the literature for more in-depth information on design, including http://www.socialresearchmethods.net/kb/design.php.

Relationship between Designs and Claims

The kinds of claims that you can make based on the results of the evaluation vary depending upon the kind of design you choose to use. For example, if you want to be able to state that participation in the program is related to a change in some outcome, you need to use a design that assesses change. Not all designs are created equal. Some designs are better than others at addressing the kind of claim we want to make. When considering which kind of design to use, it is important to think about what kind of claim you want to make and select a design that can provide evidence for that claim. It is also important to consider the feasibility of the design as well as whether or not it is appropriate given the lifecycle phase of the program. It is possible that after reviewing different design options, the working group may decide to revise the evaluation questions.

In addition to considering the kinds of claims you want to make, it is also important to take note of the kind of language that is used in the evaluation question. For example, if the evaluation question asks "Does participation in the program cause outcome X", this implies that a particular type of design that can assess causality is used. The strongest design for assessing a cause/effect relationship is a Randomized Controlled Trial (RCT; a pre-post-test with random assignment to groups). This type of design is considered a Phase 3 (Comparison and Control) Evaluation Lifecycle design and is most appropriate for a Phase 3 (Stability) Program Lifecycle program. On the other hand, when you are doing first-time implementation of a new program an RCT would not be appropriate and you might be advised to choose something like a post-only case study design. The evaluation questions may need to be revised accordingly (in this example, you may opt for something like "do participants in the program show outcome X?").

Criteria to Consider when Selecting a Design: There are several criteria that should be considered when selecting a design: (1) Time order, (2) Covariation, (3) Rules out other possible causes, and (4) Shows change. In order to demonstrate time order, we need to use a design that clearly demonstrates that the "cause" or the program happened before the "effect" or the outcome that we are interested in assessing. Covariation means that changes in the "cause" or the program are related to changes in the "effect" or the outcome of interest. In order to demonstrate covariation, we need a design that shows that when the program occurs the outcome of interest occurs and that when the program does not occur the outcome of interest does not occur. Typically, this is demonstrated by using a design that includes at least two groups. One group receives the program (and hopefully exhibits the outcome of interest) and one group does not receive the program (and hopefully does not exhibit the outcome of interest). In order to rule out other possible causes, we need a design that demonstrates that the program (the presumed "cause") is the only reasonable explanation for the "effect" or outcome of interest. This is typically an extremely difficult criterion to meet. Any number of factors other than the program could "cause" the outcome of interest. In order to demonstrate that change occurred, a design that includes a "before and after" or pre- and post-test is needed.

The strength of the claims we can make depends on how well the design addresses these criteria. In other words, does the design we select allow us to make the desired claims? The chart below provides examples of some of the more commonly used designs and the associated claims that can typically be made.

Aligning Claims with Designs

Associated Claim	*Design where X=program O=observation*	*Time Order?*	*Covariation?*	*Rules out other possible causes?*	*Shows change?*	*Program lifecycle phase it may be appropriate for*
After program, these participants show desired levels of outcome Z in this setting and context.	X O (post-only)	Yes	No	No	No	IB
According to these participants, in this setting and context, the program is associated with a change on outcome Z.	X O_{post}/O_{pre} (retrospective "post- then pre-")	No	No	No	Yes	IIA
Participation in the program is associated with a change in outcome Z in this setting and context, with these participants.	O X O (simple pre-post)	Yes	No	No	Yes	IIB
The program is effective in this setting and context, with these participants.	O X O O O (pre-post with comparison group)	Yes	Yes	Somewhat	Yes	IIIA
The program is effective in this setting and context, with these participants. It may also be effective in other settings and contexts, with other participants.	R O X O R O O (pre-/post- with random assignment)	Yes	Yes	Mostly	Yes	IIIB

For more information on the criteria described above and designs see *Overview of Evaluation Design* and *Alignment with Design* (**Workbook Pages 147-149)**, as well as http://www.socialresearchmethods.net/kb/desdes.php

Design Notation

We often describe a design using a concise notation that enables us to summarize a complex design structure efficiently. If two or more of the same kind of elements function the same way in a design (e.g., all measures are given to all participants at the same time) then a single symbol may be used to represent the entire set; if they function differently (e.g., some measures are pre-post and some are post-only) then you can use subscripts to differentiate them.

Observations or Measures are symbolized by an 'O'. Distinguish among specific measures, with subscripts, as in O1, O2, and so on.

The Activity or Program is symbolized with an 'X'. As with observations, use subscripts to distinguish different activities or program variations.

Groups are given their own line in the design structure. Samples are divided into groups that do or do not participate in the activity. If the design notation has three lines, there are three functionally distinct groups in the design. Group type – such as "random" (R), or "non-equivalent" (N) - is designated by a letter at the beginning of each line (i.e., group).

Time moves from left to right.

For example:

O X O Represents a pre-test before and a post-test after the activity

and

N O X O Represents a pre-post group with a non-equivalent comparison

N O O group that didn't participate in the activity

Notice that the design notation tells something about how the participants are organized or grouped in an evaluation (this relates to sampling) and it shows how measures are sequenced or organized (this relates to measurement). And, the structure of a design will usually circumscribe what will be done in analyzing the data collected. So, design is a fairly central topic in evaluation planning.

As always, it is important to keep the evaluation questions in mind when thinking through the various aspects of evaluation planning. If this is not done, there is the danger of developing a nice evaluation design that doesn't actually help to answer the focal questions.

Much like the measures section, there are a few key questions to consider once the design has been outlined:

- Is there a clear connection between the evaluation questions, chosen measures and the resulting design?
- Is the design appropriate given the claims that you would like to be able to make?
- Is this design appropriate for this program's lifecycle?
- Is this design feasible given the program resources and organizational capacity?
- Is this design feasible given the duration and setting of the program? For example, a short 30-minute activity does not lend itself to an elaborate pre-post measure.

It's important to link design issues to the lifecycle of the program. As you learned in the Lifecycle Analysis step we believe that the ultimate goal is for the evaluation lifecycle to be aligned with the program lifecycle. Different evaluation designs are more or less appropriate depending on the program lifecycle phase.

In conclusion, the design plan should address each of the evaluation questions, be described in detail (e.g., post-only, pre-post, pre-post with comparison group, etc.), be appropriate given the lifecycle stage of the program, and be appropriate for generating evidence for the desired claims.

3.07 Data Management and Analysis

The goal of this step is to succinctly articulate and put into writing the plan for managing and analyzing the evaluation data. Data management requires attention to details, and knowledge of data collection, data entry, and data storage. More information on this will be provided in the Evaluation Implementation phase, but for now you should be at least thinking about how to get this plan in place.

As for Analysis, information on how to analyze data is vast, and careers are based on this skill alone. This should not discourage or cause fear in the average program practitioner because most programs' needs can be met through relatively simple analysis methods. Also, programs that don't have the capacity for their analysis needs frequently have resources available to address this. Having an analysis plan and tools in place, in addition to working with an experienced evaluation facilitator or statistician, can help allay concerns that program staff may have. For an introduction to qualitative, quantitative and mixed methods strategies for manipulating and synthesizing data see http://www.socialresearchmethods.net/kb/analysis.php.

First, step back and think about how the data you plan on collecting will be used to answer the evaluation question(s). The analysis plan should explain how each variable (and corresponding measure) will be used. If you aren't planning on using the information, then why are you collecting

it? Consider what kind of data you will have and how the information could be summarized. For example, if you are using surveys, can you calculate individuals' scores on those surveys? If you are conducting observations, are you using a coding sheet and can numerical scores be calculated? What would a high score indicate? What would a low score indicate? Could you compute average scores on the surveys? If you are comparing groups, could you compare the average scores for one group to the average scores for the other group? If you are doing a pre- and post-test could you compare the average post-test scores to the average pre-test scores? If you are collecting qualitative data through interviews or focus groups, can you look for specific themes that are relevant to the evaluation question(s)? In the analysis plan section, describe how the data you plan on collecting could be used to address the evaluation question(s).

Next, think about how the data will be managed. Describe your plan for data collection (how measures will be administered), how the data will be handled and stored, and how the data will be organized in preparation for analysis. This includes thinking about what software program if any will be used for data storage, how the data will be coded, and how any sensitive data will be kept secure.

Finally, describe in detail the data analysis strategies that will be used to address each question, making sure that the analysis strategies are appropriate for generating evidence to answer the questions. As with the previous sections, look at the draft analysis section and ask:

- Is there a clear connection between evaluation questions, measures, sample, design and analysis?
- Is this analysis strategy appropriate for this program's design which is appropriately connected to the stage of development (lifecycle)?
- Will the analysis answer the evaluation questions?
- Is the selected analysis feasible given the program resources and organizational capacity? If not, how will the organization attain either the assistance or professional development necessary?

3.08 Evaluation Reporting Plan

Most programs already have basic reporting requirements, both internally and externally. A good evaluation plan contains a clear plan for how the results of the evaluation will be reported and utilized. Your reporting plan should be consistent with the stakeholder analysis done earlier. For each evaluation question, examine the key stakeholders and determine whether they would be interested in this question and its results. This will help ensure you will keep the actual report concise and easy to read, and illustrate a clear "throughline" that starts from the evaluation questions and continues through the measures, sample, design and analysis. Your plan should address:

- What reporting is required? What other reporting opportunities do you have?
- Does this fit your evaluation purpose?
- Have evaluation questions been answered or addressed? If so what form should "answers" take?
- What is the report's audience? (Include internal and external audiences.)
- What type of reporting will you do? An evaluation summary? Informal ongoing reports to be generated and distributed at every staff meeting?
- When do reports occur? Monthly? Quarterly? End of fiscal year?

Reporting needs are likely to change depending upon the program lifecycle. For instance, Phase I program reporting tends to be more internally focused and more private, whereas later phase reporting moves toward becoming more public. Also, earlier lifecycle phase reporting tends to be less formal while later lifecycle phase reporting tends to be more formal.

It is worth keeping in mind that good evaluation opens up lots of opportunities for communication. "Reporting" tends to sound and feel obligatory, but "communication" suggests inclusion and a positive constructive response to feedback. As Evaluation Champion you may want to encourage

the working group to think about what their communication opportunities may be, beyond the mandated reporting. This positive perspective can help reinforce an understanding of the value of evaluation.

When writing the reporting plan section of the evaluation plan, be sure the plan addresses each of the evaluation questions, fits the purpose of the current evaluation, and clearly describes how the evaluation results will be utilized. It is also important to make sure that the plans for utilizing evaluation results are appropriate given the evaluation purpose (you may want to revisit/revise the evaluation purpose statement at this point) and that the plans for utilization are appropriate given the program's current lifecycle phase.

3.09 Implementation Plan and Schedule

The goal of this step is to develop your evaluation timeline (schedule) identifying key implementation milestones. The measures, sampling, design, analysis and reporting efforts should all be covered in this section. For instance, if a measure is still "in development", the timeline section should include the date by which the measures will be either located and/or developed internally. If these measures are not put into place at the beginning of the evaluation cycle, program staff will be missing data from a significant portion of the year described by the evaluation plan. If existing measures are used in "waves" periodically throughout a season or within a workshop series, those time periods should also be listed in the timeline section. Either program-specific time periods (week one of our six week series) or external times and dates (the start/stop dates of the local school year) are acceptable.

The Netway makes printing out your models and plan simple. From your program page, click on "Build a Report" and you can select which items you want in your report.

Here are some questions to ask when creating an evaluation timeline:

- Does the timeline indicate the entire time period with a start date AND end date?
- If relevant, are there dates associated with each measure listed in the design section?
- Does the timeline include time for sample identification and development of contact information, if needed?
- Has it been specified when data entry and analysis will be performed?
- Are specific times for reports included?
- Is the timeline appropriate and practically feasible?
- Does it address when you will review your plan and move into the next evaluation cycle?

3.10 Finalize Evaluation Plan

At this point, go back and revisit step 3.01 (*page 39 – Introduction to Evaluation Planning*). Remind yourself of the sections of your evaluation plan, then print it out and bind it for easy reference. Holding the actual evaluation plan in your hand is a significant accomplishment.

Activity: Finalize Evaluation Plan

The goal of this activity is to review, finalize, and prepare to share the evaluation plan with leaders in the organization, as well as to prepare to move on to the Evaluation Implementation phase.

Using the "Evaluation Plan Feedback Form" (see Workbook pages 87-92) feedback may be provided to each working group about their entire plan. When possible, it is useful to receive feedback from peers and colleagues, as well as the Evaluation Champion. After incorporating the feedback, we recommend printing and binding the entire plan for easy reference, as well as distributing both hard copies and electronic formats to the Evaluation Champion, working group, program leader, external evaluator (if used) and organization director. We have also learned that the work done on the modeling for evaluation purposes is quite useful to organizational and program planners and developers, especially for presenting the program and showing evaluation intentions to funders and other stakeholders.

It may seem like the plan is never quite complete and ready for printing, but there comes a time when you have to stop planning and begin implementing the evaluation plan and is not practical to keep fine tuning it. Those edits can be reserved for the next evaluation cycle.

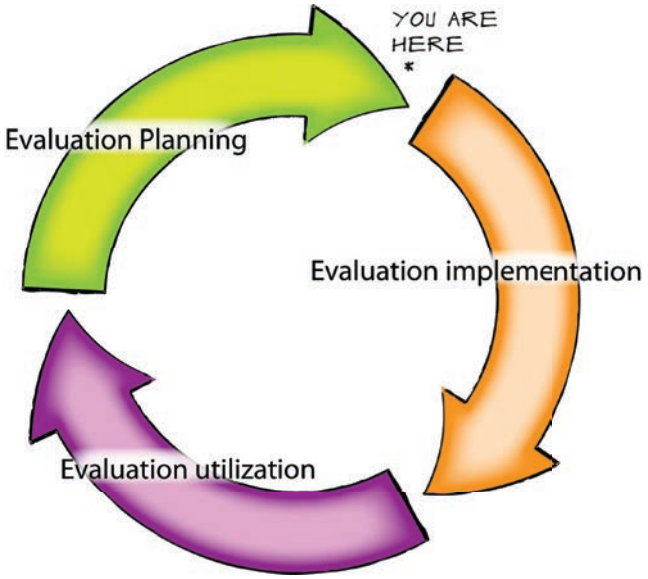
Evaluation Planning Summary

It has been a lot of work up to this point, but, "Well begun is half done" goes the proverb often attributed to Aristotle. Perhaps in this case we are even more than half done. Let's take a step back and look at where we have come from, where we are, and where we are going.

We began by stating that evaluation is ideally viewed as having three phases – planning, implementation, and utilization. With the completion of the evaluation plan, the program has moved through the first phase of the process, and certainly through the most thought-intensive part of the process. In addition, there should be a timeline in place for the implementation and a plan for reporting and utilization, so planning for the entire 3-phase process has been completed. You are now ready to move on to the implementation phase.

A survey of organization directors and program leaders and staff who have gone through this process has shown that participants believed the Evaluation Partnership (EP) process using the Systems Evaluation Protocol helped them by facilitating clearer thinking about program goals and how program activities connect to those goals, and afforded them a greater understanding of evaluation. Organization Directors responding to an open-ended question on benefits of the EP reported that it increased the organization's ability to communicate plans and results to funders and other stakeholders.

At this point we hope you, too, agree that the concept of a Systems Evaluation Protocol offers a useful approach to evaluation. You should be more aware of each program as a dynamic and evolving organism, and be cognizant of how these dynamics influence the program's evaluation needs. A complex nested system of organizations may initially muddle the conception of systems program evaluation, but this approach emphasizes that evaluation can be presented as much more than an end-of-program judgment of the work which comprises the staff's work life. Evaluation has a role in every phase of program planning and management. Evaluation Partnerships encourage collaboration among individuals with many different perspectives and priorities, and increase everyone's valuation of each other's work.



Glossary

Activities	*active pieces of the program that reach program participants*	
Analysis	*the process of deriving order and meaning from data*	
Assumptions	*beliefs and thought patterns about how and why a program is expected to succeed, or that would prevent a program from succeeding which are not otherwise explicitly stated*	
Boundary analysis	*the process of defining the structure and elements of a program; specifically what is considered part of the program as opposed to activities or elements that may be outside the program*	
Cleaning (data)	*after data collection but before transforming or analyzing data evaluatiors must screen the data for accuracy - allowing you to go back and clarify problems or errors due to incomplete, missed or illegible answers.*	*Trochim (2006)*
Coding (data)	*Coding is a process for both categorizing qualitative data and for describing the implications and details of these categories. Initially one does open coding, considering the data in minute detail while developing some initial categories. Later, one moves to more selective coding where one systematically codes with respect to a core concept.*	*Trochim (2006)*
Comparison group	*a sample or population who are comparable to the evaluation sample, but who participated in the program in a different way or did not participate in the program*	
Construct	*an abstract or general idea inferred or derived from empirical evidence*	
Control group	*a group as closely as possible equivalent to the treatment group but does not receive the treatment*	
Correlation	*the degree to which two or more sets of measurements vary together; e.g., a positive correlation exists when high values on one scale are associated with high values on another; a negative correlation exists when high values on one scale are associated with low values on another*	*JCSEE (2003)*
Cyberinfrastructure	*an interactive website*	
Data	*material gathered during the course of an evaluation which serves as the basis for information, discussion, and inference*	*JCSEE (1994)*
Data management	*the ongoing process of recording, documenting, tracking, protecting, and organizing data*	
Design (evaluation)	*an evaluation design structures the research, to show how all of the major parts of the research project -- the samples or groups, measures, treatments or programs, and methods of assignment -- work together to try to address the central research questions*	*Trochim (2006)*
Effectiveness	*the ability to produce an effect*	*Wikipedia*
Efficacy	*the capacity for beneficial change*	*Wikipedia*
Evaluation capacity	*the potential to plan, implement and utilize effective, useful, and professional evaluation practice*	
Evaluation capacity building	*involves the design and implementation of teaching and learning strategies to help individuals, groups, and organizations learn about what constitutes effective, useful, and professional evaluation practice*	*Preskill & Boyle, 2008*
Evaluation cycle	*the iterative process of planning, implementing, and utilizing an evaluation*	
Evaluation Partnership	*building evaluation capacity through a relationship between the evaluation facilitator and the partnering program, organization or system*	

Evaluation Champion	*person in the leadership role who will facilitate the protocol and guide the evaluation process within in the evaluation partnership, allowing the program partner(s) to remain the program expert(s), the EC also serves as a driving force for addressing contextual factors of both the organization and the larger systems within which the program is embedded*	
Evaluation plan	*a guide for the implementation of an evaluation*	
Evaluation purpose statement	*the introduction to the evaluation plan, which summarizes the context, both current and historical, in which the current evaluation exists, the key assumptions of the program and its evaluation, the current evaluation questions and methods, and the evaluation questions that will be addressed in the future*	
Evaluation questions	*the broad questions about the program that the evaluation seeks to address*	
Evaluation system	*to the comprehensive and integrated set of capabilities, resources, activities and support mechanisms for conducting evaluation work*	
Evolutionary systems perspective	*programs, as well as evaluation, are constantly evolving as a result of interacting with their complex, interconnected environments*	
Facilitation (of evaluation)	*the process of partnering with and training program, organization or system staff, as an evaluation expert*	
Implementation	*the process of carrying out an evaluation plan; including activities such as data collection, management and analysis*	
Interview	*a series of orally-delivered questions designed to elicit responses concerning attitudes, information, interests, knowledge, and opinions*	*Wheeler, et. al. (1992)*
Lifecycle (evaluation)		
Lifecycle (program)	*the individual course a program takes as it evolves, changes, or remains the same, over time*	
Lifecycle analysis	*the process of identifying and describing a program's current lifecycle phase and the programs's evaluation lifecycle phase, and assessing for alignment*	
Logic Model	*an outline of a program's inputs, activities, outputs, outcomes, context and assumptions*	
Long-term outcomes	*the ultimate impact logically connected to earlier term outcomes in a logic or pathway model, likely to include broader social, economic, or environmental effects*	
Measure	*a method or tool used to collect information*	
Measurement	*the process of using a method or tool to collect information*	
Medium-term outcomes	*describe effects on participants that logically connect short-term outcomes to long-term outcomes*	
Mixed methods	*using both qualitative and quantitative evaluation strategies to address the same evaluation question*	
MOU	*stands for memorandum of understanding, which is the written agreement between the participating program, its parent organization, and the Evaluation Champion*	
Organization	*an office or institution that consists of one or more programs*	
Outputs	*the by-products of activities that may serve as evidence that the activity was completed, but not part of the larger flow of logic which explains the effects on participants*	
Pathway model	*a graphical representation of the activities and outcomes that make up a program, and how they are interrelated*	
Post-only	*an evaluation design in which the observation or measurement takes place exclusively after the program*	

Pre-post	*an evaluation design in which observations/measurements take place both before and after the program*	
Program	*a series of activities conducted with the intention of producing some effect (outcomes) on participants*	
Program boundary	*an imagined "line" (artificial construct) between the activities, outputs and outcomes considered part of a program and those outside a program,*	
Program description	*a summary of the basic components and characteristics of a program*	
Protocol (evaluation)	*a step-by-step guide*	
Qualitative analysis	*the process of deriving order and meaning from data using non-numerical methods*	
Quantitative analysis	*the process of deriving order and meaning from data using numerical representations and statistical methods*	
Quasi-Experiment	*a study in which the subjects to be observed (sample) are not randomly assigned to different groups, but grouped according to a characteristic that they already possess*	*Wikipedia*
Random assignment	*units in the sample are randomly assigned to different groups or treatments in the study*	*Trochim (2006)*
Reliability	*the degree to which an evaluation consistently produces the same result*	
Reporting	*the process of communicating results and recommendations to internal or external stakeholders*	
Sample	*part of a population*	*Wheeler, et. al. (1992)*
Sample size	*the number of observations that constitute the sample*	*Wikipedia*
Scope (evaluation)	*how much of the pathway model the program staff intend to evaluate in a given evaluation cycle*	
SEP	*stands for Systems Evaluation Protocol, which is a standardized protocol designed to enable any program to develop a uniquely tailored evaluation for that program*	
Short-term outcomes	*describe effects on program participants that are logically and directly connected with the activities*	
Stakeholder analysis	*the process of identifying and describing the perspectives of all of the potential people and/or organizations that have a stake in the program and its evaluation*	
Stakeholders	*any person or organization having an interest in the program, and are therefore may be involved in or affected by the evaluation .*	
STEM	*science, technology, engineering and mathematics*	
Subscale	*a smaller set of items on a measure that have shared validity and reliability independent of the larger measure*	
Survey	*a method for collecting quantifiable information about a population*	
System	*a parent organization to one or more offices or institutions, in which complex interactions between programs, organizations and their contexts take place*	
Systems evaluation	*the assessment of the functions, products, outcomes and impacts of a system (set of programs, activities or interventions). Systems evaluation is an approach to conducting program evaluation that considers the complex factors that are inherent within the larger "structure" or "system" within which the program is embedded*	
Systems perspective	*taking into account the larger contextual and environmental factors around a program, organization, or system, including the complex interactions between each*	

Timeline	***a calendar or list of dates showing the evaluation stages and activities, and indicating the dates by which they should be implemented and be completed***	***Wheeler, et. al. (1992)***
Treatment group	***a sample sub-group that is exposed to the program, project, or instructional material as well as all of the conditions of the investigation***	
Utilization (evaluation)	***the process of using an evaluation, and its results, to make decisions about program change***	
Validity	***the extent to which the test scores or responses measure the attribute(s) that they were designed to measure***	***Wheeler, et. al. (1992)***
Variable	***a characteristic that can take on different values***	***JCSEE (1994)***

Glossary References

Health Services Research and Development Service . (2000). Organizational Change. Boston, MA. HSRDS (2000)

Joint Committee on Standards for Educational Evaluation. (1994). Standards for Evaluations of Educational Programs, Projects and Materials. New York: McGraw Hill. JCSEE (1994)

Joint Committee on Standards for Educational Evaluation. (2003). The Student Evaluation Standards. Thousand Oaks, CA: Corwin Press JCSEE (2003)

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. American Journal of Evaluation. 29 (4) pp 443-459. Preskill & Boyle (2008)

Wheeler, Patricia , Haertel, Geneva D. and Scriven, Michael. (1992) Teacher Evaluation Glossary. Western Michigan University. Wheeler, et. al (1992)

Appendices

#	Step	Title
I	1.02	Memorandum of Understanding Template
II	1.05	Organization Evaluation Capacity Survey
III	1.05	Program Evaluation Capacity Survey

Appendix I
Memorandum of Understanding Template

[YEAR] Evaluation Partnership

This Memorandum of Understanding (MOU) is meant to be a vehicle to plan the year's evaluation planning and capacity building activities and to help assure that all parties are clear about what is planned and expected for the new Evaluation Partnerships beginning in [YEAR]. It is not intended to be a formal legal document.

The sections below describe the background and goals for this initiative, and the roles and responsibilities of the participants. A preliminary timeline of project activities and completion is on the final page.

1. **Background**

The Cornell Office for Research on Evaluation (CORE), under the leadership of Professor William Trochim, created this "Evaluation Partnership" and "Systems Evaluation Protocol" approach to building evaluation capacity among participating partner organizations. Funding for this effort has come from the National Science Foundation and Cornell Cooperative Extension. More on the history of the project and CORE's systems approach to evaluation are at http://core.human.cornell.edu/outreach/evaluation/index.cfm.

[Here, put a few lines of text to describe the background of how your organization or program got involved in using this approach, and of any institutional partnerships which are involved in your use of this approach.]

2. **Goals**

Based on preliminary discussions, the Evaluation Partnerships between the [EVALUATION CHAMPION] and [ORGANIZATION] will cover [#] "programs": [LIST PROGRAMS]. For purposes of the Evaluation Partnership, each of these programs will have a working group consisting of one or more individuals.

The goals of the first year of this Evaluation Partnership are to:

- Develop a Logic Model, Pathway Model, and Evaluation Plan for the selected program;
- Build on and expand participants' shared understanding of all programs, their commonalities and distinct features, and how they relate to other programs;
- Enhance the evaluation skills and evaluation leadership within [ORGANIZATION], and expand internal capacity for using this evaluation method with additional programs.

Evaluation Partnership work may extend into future years, if mutually desired, under terms that would be established in separate future MOUs.

3. Project Preparation

Establishing a shared understanding

The Evaluation Partnership begins with discussions and communication between [EVALUATION CHAMPION] and the leadership and members of [ORGANIZATION] to establish a shared understanding of interests, responsibilities and commitments. This MOU should reflect that shared understanding, and will be adjusted as needed. Participants on both sides are encouraged to raise questions, discuss options,

and seek mutually satisfying plans. When details are clarified, participants will be asked to complete the signature page and return the MOU to [EVALUATION CHAMPION].

4. Roles and Responsibilities

[EVALUATION CHAMPION] Roles [and staff if appropriate]

The evaluation facilitator is:

- Names and Titles
-

The Evaluation Partnerships follow a step-wise evaluation planning process known as the "Systems Evaluation Protocol" (see the link provided earlier for more on the systems approach used by CORE.) [EVALUATION CHAMPION] will schedule and conduct in-person and audio/web-conference trainings which may include other partners; and will communicate regularly through e-mail, telephone, and various web resources with EP participants in support of their evaluation efforts throughout the project duration. [If there is more than one person on the facilitation team, identify roles]

[ORGANIZATION] Program Roles
Individuals in leadership positions in [ORGANIZATION] have important roles to play in ensuring the success of this evaluation planning effort. These roles include

- Setting realistic expectations for evaluation; providing resource support for participating staff; staying abreast of progress; supporting deadlines;
- Preparing staff for their participation in the EP, adjusting staff work-loads and position descriptions as needed in order to incorporate the commitment to evaluation;
- Clearly and consistently supporting the commitment to evaluation and recognizing its value to the Team's work and to the system more generally;
- Demonstrating positive use of evaluation results, including in communications to internal and external stakeholders (such as key funders, peers in participating counties, and so on).
-

Specific next steps to be completed by the [ORGANIZATION] Leaders are the following:

- In consultation with [EVALUATION CHAMPION], confirm the selection of the programs that will be the focus of this Evaluation Partnership. These will be the "EP Programs", and should be listed on the signature page of this MOU.
- For each of these EP Programs, identify the team member(s) who will be working on the evaluation planning for that program, and indicate who will serve as the main contact (Evaluation Project Manager, or EPM) for each program. (Please attach this information as requested on page # of this MOU.
- Review this MOU with the team members working on the EP. Obtain signatures as needed on the signature page of this MOU, and return it to [EVALUATION CHAMPION] by [DATE] to indicate an understanding of and agreement with the scope of the proposed work. Please contact [name and contact info] to discuss any questions or concerns you may have prior to signing.
- Complete the Organizational Evaluation Capacity Survey and return it to [EVALUATION CHAMPION] by [DATE]

Working Group members

A key feature that has facilitated prior Evaluation Partnership work has been to have one person designated as the Evaluation Project Manager (EPM) for each EP program. The EPM together with other staff members working on the evaluation planning for an EP program will be considered the "working group" for that program. Collaborative work among working groups is welcomed and encouraged. The particular roles are described below:

- The EPM will take the lead on ensuring that the stages of the evaluation planning process are completed, that deadlines are met and materials are developed and sent to [EVALUATION CHAMPION] as directed. The EPM is expected to participate in all in-person and audio/web-conference trainings and meetings over the course of the Partnership (or to designate a substitute from his/her office if needed.)

- The EPM (in consultation with other staff as desired) will complete an initial "Program Evaluation Capacity Survey" and return it to [EVALUATION CHAMPION] by [DATE] - sent in advance of the first in-person training session.
- Working Group members will be expected to contribute to the work of program modeling and evaluation planning, and are strongly encouraged to participate in all trainings and audio/web-conferences for the Partnership.
- EPMs and Working Group members are expected to join the Partnership listserv that will be used over the course of the Evaluation Partnership, and active participation is highly encouraged.

All Participants

- Team Leaders, EPMs, and Working Group members may be asked to participate in surveys or other mechanisms that [EVALUATION CHAMPION] may request as part of its own assessments of progress and impact. In addition, [EVALUATION CHAMPION] may request permission to make video or audio tapes of training sessions and meetings. These recordings will be used for internal purposes, and/or may be used as a basis for future training materials. Requests for permission will be specific as to future use.

5. Project Implementation and Tentative Timeline

The work of the Evaluation Partnership will take place through a combination of in-person training meetings, audio- and web-conferences, and on-going support and consultations by phone, by e-mail, and through the EP listserv. A Preliminary Timeline is included as an Attachment at the end of this MOU. This will be adjusted and dates established more firmly following discussions with the [ORGANIZATION] Team to determine what will work best for all sides.

In addition to the specific training activities, resources will be made available electronically, and the Evaluation Team at [EVALUATION CHAMPION] will be available by phone and by e-mail. An additional source of support comes from the network of Evaluation Partners from prior cohorts who continue to participate on the EP listserv for on-going questions and discussion.

Deliverables:

At the end of this year of the Evaluation Partnership, each EP Program will have a well-articulated Logic Model, Pathway Model and Evaluation Plan. Individual staff members involved in the Evaluation Partnership will have received focused evaluation support, and will be in a position to take an ongoing leadership role in terms of evaluation planning at their offices and within their teams. The teams as a whole, and these individual staff members, will be part of an expanding and increasingly active Evaluation Network that can serve as a valuable resource for supporting and expanding evaluation capacity.

Costs:

The [ORGANIZATION] partners will be responsible for all local costs associated with this project, including staff time, staff travel, printing, mail, and computing. [EVALUATION CHAMPION] will be responsible for the teleconferencing fees, on-site costs of hosting the training meetings, and for its own related travel costs.

6. Program Selection and Staff Participation

Please attach a list of the programs selected by the [ORGANIZATION] Team for inclusion in the Evaluation Partnership. For each program please give its title, the names and contact information of the staff members who will be working on this program for the EP process, and indicate who will serve as EPM.

7. Signatures indicating review and acceptance:

Please complete this signature page for each of the selected programs, filling in the program title and highlighted EPM information each time.

Program Name: _____

_____ _____
(Organization Leader Signature and Title) Date

_____ _____
(Program/Evaluation Leader Signature and Title) Date

_____ _____
(Evaluation Champion/Facilitator Signature and Title) Date

Attachment: Preliminary Timeline

The exact timing of the steps is still to be determined, taking into consideration the deadlines and workloads of the [ORGANIZATION] team members, [EVALUATION CHAMPION]'s commitments, and the priorities of effective and efficient trainings. Further details on this timeline will be set once [EVALUATION CHAMPION] and the [ORGANIZATION] teams have discussed options. The preliminary version included here is intended to provide more detail about time commitments and the training steps and sequence.

September/October (Preparation Stage)

- Review and sign MOU
- Gather contact information, start initial communications, set up listserv for Evaluation Partners
- Complete Organization Evaluation Capacity Survey ([ORGANIZATION] Leader)
- Complete one Program Survey for each program (EPMs, with colleagues as needed)
- Web-conferences (dates tba) to cover:
    - o Initial program stakeholder maps and guidance for follow-up interviews
    - o Program "boundary analysis" and lifecycle analysis; program description

November/December (Program Modeling)

- Web- conferences (dates to be confirmed) to cover:
- Two-day in-person meeting at [LOCATION], including:
    - o Sharing results of stakeholder analyses
    - o Full development of Logic and Pathway Models, with peer reviews and [EVALUATION CHAMPION] support and feedback
    - o "Mine the Model" session to draw out program insights and use program models, stakeholder analysis, and program lifecycle to identify desired evaluation scope, initial evaluation questions
- Revise Logic and Pathway models. [EVALUATION CHAMPION] will review models

Winter/Spring (Evaluation Planning)

- Final revisions to Logic Model and Pathway Model as needed
- Web-conference to discuss linkages between the overall [ORGANIZATION] goals and the individual programs.
- Web- conferences and possibly one in-person meeting to support evaluation plan development (Evaluation Questions, Sampling Plan, Measurement, Evaluation Design, Data Management and Analysis, Reporting Plan, and Implementation Timeline.)

May/June (Finalization)

- Draft Evaluation Plans due [TBA]
- [EVALUATION CHAMPION] reviews of evaluation plans returned to teams by [TBA]
- Final Reports (final Evaluation Plans plus Stakeholder Maps, Logic and Pathway Models, and proposed measures) due [TBA].

Appendix II

Assessing Evaluation Capacity

Measures of evaluation capacity are easier to find than they used to be. Evaluation Capacity Building is becoming more popular, and it can be helpful to assess any changes in an organization's and program's capacity to do evaluation. Think about what changes you're trying to make in their capacity, and consider measuring it. For example, using the following guiding questions, we were able to see changes in evaluation policies, support & training, staff attitudes, and evaluation use after going through the SEP. As well, collecting this type of data prior to working through the SEP allows the Evaluation Champion to anticipate needs for training and support.

Some Considerations when Assessing Evaluation Capacity

At the Organization Level

1) How many FTE are employed?
2) How many FTE are dedicated to evaluation?
3) How many Staff have formal assignments for evaluation specified in their job descriptions?
4) How often are evaluators external to the organization utilized?
5) What kind of training do staff have for evaluation?
6) What resources/support has the organization put into evaluation training?
7) Do budgets plan for evaluation? If yes, what % of the budget, if no, how is evaluation funded?
8) Do funders require your programs to conduct evaluation? Explain.
9) What IT resources are available for evalution? - Database and analysis software? Survey tools?
10) What are the attitudes toward evaluation within the organization?
11) How ofter are programs changed based on evaluation?
12) Are there incentives offered to staff to encourage participation in the evaluation process?, if Yes, what?
13) What supports and barriers are there at the Organization level for evaluation?
14) Among staff doing evaluation, what sort of planning do they do in advance to prepare for the evaluation?
15) What, if any, formal or informal guidelines or policies are there to guide evaluation decisions?

At the Program Level

1) Is there a program description, as well as goals and/or a mission statement?
2) How is the program funded, and what requirements are there by the funder for evaluation?
3) What funding is available for evaluation?
4) Is there already a logic model for the program? Does it list inputs, activities, outcomes, context and assumptions?
5) Has evaluation been conducted and utilized?
6) What guides evaluation decisions?
7) What training or experience do staff have on evaluation?
8) Is there an IRB or other oversight committee, and has it been used to review previous evaluations?