# TABLE OF CONTENTS

Table of Contents ......................................................................................................2  
Preface .......................................................................................................................3 
Acknowledgements ...................................................................................................4 
I. Introduction to Systems Evaluation ...................................................................5 
   A Systems Perspective on Programs and Program Evaluation .......................7 
II. The Systems Evaluation Protocol .....................................................................10  
   Phase 1: Evaluation Planning ............................................................................11  
      Stage 1: Preparation ....................................................................................11  
      1.01 Enter the System ..................................................................................11  
         1.02 Develop Memorandum of Understanding ..................................12  
         1.03 Identify Internal Stakeholders .....................................................13  
         1.04 Identify the Working Group(s) ....................................................13  
         1.05 Assess Evaluation Capacity ..........................................................14  
         1.06 Launch Evaluation Partnership ....................................................15  
      Stage 2: Program Model Development .......................................................15  
         2.01 Stakeholder Analysis .......................................................................15  
         2.02 Introduction to Program Model Development Process ...............16  
         2.03 Program Review .............................................................................17  
         2.04 Program Boundary Analysis ..........................................................18  
         2.05 Lifecycle Analysis ...........................................................................21  
         2.06 Logic Model ....................................................................................27  
         2.07 Pathway Model ...............................................................................29  
         2.08 Determining Evaluation Scope ......................................................31  
         2.09 Program-System Links ....................................................................33  
         2.10 Program-Logic Model Synthesis .....................................................34  
      Stage 3: Evaluation Plan Creation ..................................................................35  
         3.01 Introduce the concept of an Evaluation Plan ..................................35  
         3.02 Develop Evaluation Questions ......................................................36  
         3.03 Develop Sampling Plan .....................................................................37  
         3.04 Identify or Develop Measures .......................................................38  
         3.05 Develop Evaluation Design .............................................................41  
         3.06 Develop Analysis Plan .......................................................................43  
         3.07 Develop Evaluation Reporting Plan ................................................43  
         3.08 Develop Implementation Plan and Schedule ...................................44  
         3.09 Finalize Evaluation Plan ...................................................................44  
      Summary .........................................................................................................45  
Glossary ....................................................................................................................46  
Appendices ...............................................................................................................47
This Systems Evaluation Protocol is intended to serve several purposes: it is first and foremost designed to be a step-by-step guide for evaluation professionals who wish to integrate a systems evaluation perspective into their work with organizations in order to enhance the quality and appropriateness of program evaluations. This Protocol was specifically created in the context of education and outreach programs generally and specifically for programs in Science, Technology, Engineering and Mathematics (STEM) education and outreach sponsored by the 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 and we hope that a broader audience will find it useful.

The Introduction below goes beyond simply laying the groundwork for the steps of this Protocol, and it is here that we address a second goal – that of providing an overview of the “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.

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

The appendices offer sample materials and worksheets for many of the outputs in the Protocol.

Finally, this guide is meant to be a snapshot of where we are now in the development of this approach in the Fall of 2008. We are presenting it at this time as our original pilot study comes to a close, but anticipate that our insights and understanding will continue to be shaped and influenced by our experiences in the future. This publication focuses on the first phase of Evaluation – Evaluation Planning. Future publications are planned for similarly constructed protocols for Evaluation Implementation and Evaluation Utilization, as well as revised versions to the current Protocol.

We anticipate that this Protocol will continue to undergo changes as our understanding of the systems interactions becomes more succinct. Likewise, materials in the Appendix will undergo similar evolutions. We have recently received a 5-year grant (beginning August 2008) to continue this work and to begin work on the Implementation and Utilization Protocols. Our expectations are to make our materials available after each evaluation cycle (annually), and we encourage feedback and discussion on systems evaluation.

For comments or questions please contact:

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
120 Martha Van Rensselaer Hall
Cornell University
Ithaca, NY 14853
ACKNOWLEDGEMENTS

Many people worked on this project in 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 leaders on key projects.

Cornell Office for Research on Evaluation (CORE):

The Principle Investigator leading the research and Protocol development is Dr. William M. K. Trochim. He is a professor in the Department of Policy and Analysis and Management at Cornell University.

The position of Manager of Evaluation for Extension and Outreach was filled at different times by Cathleen Kane (also Evaluation Project Manager in CUCE NYC) and Monica Hargraves (also Evaluation Project Manager for CCE Tompkins County.)

Graduate Research Assistants working on this project include Jennifer Brown Urban (Department of Human Development), Sarah Hertzog (Department of Human Development) and Derek Cabrera (Policy Analysis and Management, who also worked as a post-doc.)

Extension Associate Laura Colosi (Policy Analysis and Management).

CORE staff: Claire Hebbard, Claire Lobdell, Debra Perlmutter.

Undergraduate Research Assistants were vital for literature searches, finding measures and other organizational tasks, and included: Timothy Au, Amanda Bates, Lindsey Bober, Haixin Dang, Stacey Elbaum, Anna Hays, En Chi (Angela) Hsu, Oluwarantimi Damilola Oluwasegun, Sabrina Rahman, Marissa J. Steinmetz, Helen Zaitsev and Anthony Zhu.

Co-PI: Stephen Hamilton (Associate Provost for Outreach, Cornell University).

Executive Directors and Evaluation Project Managers from Cornell Cooperative Extension: Keith Severson, Kim Eaton, Bekky Hargrave (Chenango County); Richard Halpin, Stephanie Graf (Jefferson County); Paul O’Connor, Kelly Somerlot (Onondaga County); Clive Chambers, Patrick Ames (St. Lawrence County); Kenneth Schlather, Monica Hargraves (Tompkins County); Lydia Reidy, Roxanne Churchill (Ulster County); and Don Tobias, Cathleen Kane and Michelle Currie (New York City).

Director and Evaluation Project Managers at the Museum of the Earth at the Paleontological Research Institution: Rob Ross and Samantha Sands.

Director and Evaluation Project Managers at the Cornell Center for Materials Research (an NSF MRSEC program): Melissa Hines, Nevjinder Singhota, Kevin Dilley, Jane Earle.

Database developers at Gorges Websites (Ithaca, NY): Chris Grant and Matthew Clark.

Collaborator: Nancy Trautman (Senior Extension Associate, Cornell University).

Independent STEM Consultant: 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 constructs have guided our work including complexity theory, evolutionary theory and natural selection, general systems theory, ecology, system dynamics, and ideas on research-practice integration (including evidence-based practice and translational research).

The field of evaluation is 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, whereby we refer to 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 enhance 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 projects in order to assure that projects that use it will incorporate such principles when developing project pathway models and identifying key pathways and nodes (outputs and outcomes), determining the boundary conditions for project models, assessing project lifecycles, and selecting evaluation designs that are appropriate to project evolution.

It was the examination of some of our beliefs about evaluation that led to our work on 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 project or organizational improvement (see sidebar “Driving with your Eyes Open,” page 41).
- 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 below provides a pictorial representation of the Phases of Evaluation.
- The Evaluation Planning phase is a critical step for systems evaluation in that this is where introducing the systems perspective will shape how Evaluation Facilitators and the 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 just doing evaluations of individual programs.
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 of 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 6). 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 curriculum, 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. A dynamic system is necessarily composed of evolving relationships and programs. (see sidebar “The Rock and The Bird”, page 8).

- 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 24). Following our previous example, Ms. Smith’s third grade class will be Mr. Brown’s fourth grade class, and so on through to graduation.

- 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 the complexity of a system.

- 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 14).

- Each stakeholder of a system has specific expertise, and brings a distinct perspective and motivation for evaluation. This comprehensive set of stakeholders should be identified and included in the evaluation design or and/or process and be viewed as partners in the evaluation. Placing the program participants of 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 and participants in the center. A stakeholder within the organization would have a local perspective, and one that is placed at the outermost ring would have a global perspective (see sidebar “The Local and the Global,” page 9).

- The role of the evaluation expert is as an Evaluation Facilitator - this allows the program staff to retain their expert stance on their program. The Evaluation Facilitator and the client organization together form what we refer to as an Evaluation Partnership (EP). The role of Evaluation Facilitator is to educate and inform the team members about the process of evaluation, as well as to facilitate the group’s discussions and encourage them to examine thoughtfully the organization’s and program’s needs, stakeholder issues, and the program itself.
Terminology

Before proceeding further we would like to clarify some of our hierarchical 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 parent organization). Within that there are assumed to be multiple organizations (one or more of which you will be working with). Each organization may have one or more program areas, and each program area may have one or more programs. Programs in turn are made of up multiple components, including inputs, activities, outputs and outcomes (see figure 2).

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 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 be able to identify how our evaluation approaches need to evolve over time both to encourage program evolution and to provide feedback and learning about it.

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.
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).

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 at another. 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. 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.
II. The Systems Evaluation Protocol

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. The SEP specifically articulates the unique facilitation or partnering role that the evaluation professional necessarily plays when conducting systems evaluation. The process will consist of many collaborative meetings that will seemingly spiral through several focal points over time. This process is essential to the nature of the SEP! It is through these discussions that the organization and its programs will develop a new outlook upon their program that will change both their understanding of how the program stakeholders perceive the program and 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 called the Systems Evaluation Protocol presents the Protocol 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 10). 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.

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. 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 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 that use the acronym ABC: check for a clear AIRWAY; make sure the person is BREATHING; check for 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.
Phase 1: Evaluation Planning

The purpose of the Evaluation Planning Phase is to create an evaluation plan and make it operational.

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 an evaluation plan that will guide the subsequent Implementation and Utilization Phases of a complete evaluation effort.

Stage 1: Preparation

The Preparation Stage is intended to help you, the Evaluation Facilitator, enter into an organization, establish the Evaluation Partnership (EP), identify people’s key roles, acquaint the participants with the SEP process and establish baseline information regarding the target program(s) and organization. For purposes of discussion here, we will talk in terms of there being a contact person in the organization who will work closely with you. We call this person the Evaluation Project Manager (EPM). How this contact person (or perhaps group of people) is selected and the scope of his/her/their responsibilities should be determined early on.

1.01 Enter the System

The SEP always begins with the initial contact between the Evaluation Facilitator and the organization’s key decision makers. Sometimes it will be the key decision maker who asks you to assist in a program evaluation, but more often it will be the program leader or staff person who first contacts you. Ultimately, in this step it is important to make contact with the key decision maker(s) to lay the foundation of a working relationship with the person responsible for approving the Evaluation Partnership.
1.02 Develop Memorandum of Understanding

After you have made initial contact with the decision maker(s) you will need to communicate with those decision makers to negotiate and outline the responsibilities and expectations of the Evaluation Partnership members, and put the finalized commitments into a written Memorandum of Understanding (MOU). This step may take a couple of weeks, depending upon the number of people involved and your experience with moving through the steps. The discussion (and document) should detail the expectations for the Evaluation Facilitator, organization administrators, EPM and organization staff. There are several key points to negotiate at this time:

- **Time commitments.**
  - Present a general timeline of the expected process of a quality evaluation. It may take 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. There will be at least one meeting with stakeholders, and several meetings with the EPM and key staff as you learn about the program’s priorities. There will also be a need to educate the participating 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.
  - Policy regarding evaluation. Although many organizations have some sort of policy stating that evaluation is a requirement, most organizations don’t think about the amount of time that this will take. Ten to fifteen percent of an employee’s working time is not unreasonable. We want to encourage organizations to make their commitment more explicit by writing it down.

- **Other resource requirements** – what are the facilitator costs (your fees) as well as the planning costs – travel, time for meetings, phone calls, room rental, fees for licenses, printing, etc.? What about equipment – laptops, projectors, copiers, etc.?

- **Who will be the EPM(s), and what are their time commitments?** This person (or persons) will be working with staff in planning and implementation on a daily basis and then meet with you to discuss issues as they arise. The Evaluation Facilitator can work with one or more EPMs to train them on evaluation so that they become an evaluation resource to the organization when the Facilitator is not available. This simplifies communication between the organization and the Evaluation Facilitator, and builds the evaluation capacity of the organization.

- **Responsibilities and roles of the EP members.** How will the key decision maker(s) support the EPM? What expectations do the key decision makers have of the EPM reporting back to them? What can they expect from you and your availability? How should communication between you and the EPM occur?

- **What is the timeline of 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 was of necessity for the development process of the SEP. You may wish to address the entire evaluation process in a single MOU. But it is very important to be clear on expectations about the length of time each phase takes. Some programs have moved rather quickly through these discussions (4 months), while others may take a year or more. We often begin by asking them about their annual cycles – when are reports due? When is the next program being held? The answers to these will determine your timeline.
At a minimum the MOU is meant to be a vehicle to plan the 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 made available for easy reference. A sample MOU can be found in Appendix I. We have traditionally worked through this step via teleconferencing and email.

1.03 Identify Internal Stakeholders

An important first step in the launch of the EP is to bring together the internal stakeholders of this organization and its programs in order to convene an introductory or launch meeting. This list typically includes directors, administrative staff, and staff from this and other programs in the organization. The launch will help the organization understand the process of what’s going on. Not all the people attending this launch meeting will participate in each and every step of the process, but it is particularly important that this first meeting be inclusive.

1.04 Identify the Working Group(s)

The working group will consist of the people who should be involved in a conversation about the program and the program evaluation. 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 programs, and why?
- Is there anyone who might be upset to later find out that they were not included in a conversation about evaluation?
- 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 Survey? This is usually the EPM, program leader and/or staff.

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

See our example Contact List in Appendix II. When you work with many offices in the same system or organization you may want to share the contact list with all participants so they can contact each other easily and begin building an evaluation network.
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.

1.05 Assess Evaluation Capacity

Assess Organization Evaluation Capacity

Request that the appropriate person identified in the previous step complete and return the Organization Evaluation Capacity Survey. This instrument only needs to be completed once for each organization, even if you are doing several program evaluations. The purpose of this step is to identify: a) the resources available within the organization (personnel, budget, and technology), b) any organizational evaluation policies, and c) the current evaluation capacity of the staff.

A potential secondary purpose will be to revisit this survey later, and to review it after the organization has implemented its first evaluation. This could be a motivational tool for the organization upon review of the progress and changes they have made. (See the Organization Evaluation Capacity Survey in Appendix III.)

Assess Program Evaluation Capacity

Ask an appropriate individual or group to complete and return the Program Evaluation Capacity Survey. This should be completed for each of the programs that will be evaluated. 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. 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. (See the Program Evaluation Capacity Survey in Appendix IV.)
1.06 Launch Evaluation Partnership

The purpose of this meeting and presentation is to introduce evaluation, the Evaluation Partnership, and the Protocol and its resources (computer systems and/or software, exercises, instruments, tools) to the organization as a whole. This can also be a general discussion for addressing the organization’s questions (see Launch Presentation in Appendix V.) Everyone in the organization should be invited. The goal is to inform everyone within the organization about evaluation and its role within the organization – as well as how it will affect each individual. Frequently we also hold workgroup-specific meetings at the launch to further delve into the subsequent steps.

Stage 2: Program Model Development

Introduction

This stage is intended to further acquaint a program working group with the SEP process, enhance their knowledge of evaluation concepts and the use of systems approaches for analyzing programs, and develop and fine tune each program’s Logic Model and Pathway Model. In the previous stage we were working with the organization as a whole. During this stage we are working primarily at the level of the program, but there may be some overlap between programs. Of course, because we take a systems view, we will be considering the broader organizational stakeholder context. But in this part of the Protocol we primarily work with staff of each individual program.

This stage involves several steps: Stakeholder Analysis, Introduction to Program Model Development, Program Review, Program Boundary Analysis, Lifecycle Analysis, Logic Model development, Pathway Model development, Determination of Evaluation Scope, Program/System Links Examination, and Program/Logic Model Synthesis. You will likely hold multiple meetings and it may take several weeks or months to complete this stage. At times you may involve members from all programs (when they can benefit from seeing what other groups are saying regarding stakeholders), at other times you may decide to work only with members of a single program or working group. As Facilitator you should present these materials in a way that matches your own style and fits the needs of the organization and each program. Many of the steps may be performed in a different sequence or in tandem, and will also 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 Post-It® 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 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.

Create a “Map of Stakeholders” - a visual depiction of the stakeholders and their relationship to each other (see Figure 4: A 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 at the center of the graph, 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 the stakeholders and ensure that everyone in the working group is comfortable with the map. There is a blank map in Appendix VI. It is likely you will have to go through several iterations to produce a map that everyone is comfortable with.

Guide the group to consider which aspects of the program each stakeholder is most interested in learning about. It can be helpful for later steps if you record the identified interests of each stakeholder.

2.02 Introduction to Program Model Development Process

Now it is a good time to teach more in-depth on the topic of Logic Models.

Although some people may argue that this step should be placed ahead of the Map of Stakeholders, we have identified it as the second step because the stakeholder analysis immediately engages the working group and gets them thinking about the context of their program before they start thinking about what the evaluation will involve. Keep in mind that you may prefer to move it even further back to just before the Logic Model step 2.06 – depending on your local needs.
2.03 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 the benchmark initial description given in the earlier Evaluation Capacity Surveys. 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 surveys and stakeholder map. It is also appropriate to examine any curricula, guidelines, manuals or evaluations that might exist in relation to the program.

The following questions may help guide the discussions:

- Why does the program exist?
- What is the program’s mission/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 they actually do – brainstorm the activities of the program. It is essential to build engagement with the working group early on, before getting into the details of their mission or vision statements. If you start by asking about the program’s mission or vision is it almost certain to kill the conversation. You have to start with where the people live – and that is in what they do. They get excited talking about their activities and this conversation will help build your team.

You should approach this step knowing that some of these questions can probably never be completely answered to the satisfaction of everyone. 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 what the program consists of. This discussion will necessarily cause some areas of unexpected confusion, but this is an important discussion to have before moving on to the next step, which will be to set some boundaries to 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.

They shouldn’t get ahead of themselves — don’t start drawing the logic or pathways models yet. Stay focused on encouraging brainstorming. The next few steps will give shape to what is generated in this step. The result will be a tangled mess that shows the level of detail in their 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.

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

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 divergently.

The Program Boundary Analysis exercise is one of 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. 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. The output should be an improved program description, improved understanding of the program, and the 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 their program. Below are several guidelines for thinking about what the program is, and developing a comprehensive program description.

1. Ask the program staff exactly what it is that they do, and ask other stakeholders 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 the program.
2. What are the “elevator stories” of the program – if someone were to describe their program in two sentences during a short elevator ride, what would they say?
3. Refer to the first question of the Program Evaluation Capacity Survey: “Is there a formal definition of the program?” If there is such a definition this can be the working program definition and boundary. If not, do they have any informal descriptions of it? Did you come up with a better definition during the previous step (2.03 Program Review?) Examples might include program descriptions on their website, or in 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.
4. 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.
5. 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, assuming that the staff come to the program adequately trained to carry out the program (note, this would be one of your assumptions for later on in the Logic Model).
6. Look at the Map of Stakeholders. How would the 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 take 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 their understanding of their
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 differing views in a meaningful way?

7. 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?

8. 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? Keep in mind that it may be easier to demonstrate impact for intangible goals, such as increasing youth interest in science, if they are aggregated at a higher level.

9. Look back at the Program Evaluation Capacity Survey. 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?

10. 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 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 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 story – instead we chose to focus on the porridge (rather than on the chairs or the beds, or on the Bears’ return).
As one last example of different possible boundaries consider how does someone define “school”? Does school begin when the first class begins, or does it begin when a student walks through the door 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 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 consists only of the activities done with the trainers or does it include 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 the typical program participants?

The output of this whole step should be a newly revised and even more precise program description that the working group has consensus on. At this point they should agree on what activities are considered to be the program, and which ones are not.

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

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 programs, 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 in 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 part of what we call the “program”? What exactly is the boundary of the program? The same 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.
2.05 Lifecycle Analysis

Every program - like an organism - has its own lifecycle and proceeds through different phases or stages. Integrating principles associated from both systems theories and evolutionary theories, the SEP was explicitly designed to identify where a program is in its lifecycle (see Figure 5) and to encourage a progression of evaluation (see Figure 6) appropriate to each program lifecycle phase. Like its program, the evaluation of the program moves through different phases and has identifiable lifecycle phases as well.

The program and evaluation lifecycles are idealized sequences of phases through which a program and its evaluation typically progress, and are ideally synchronized with each other. Very often in practice they aren’t aligned, and some aspects of a program may be in one phase, and other aspects may be in a different phase, resulting in boundaries between phases not being clear.

For example, new programs should generally be doing basic satisfaction surveys rather than sophisticated pre/post tests. These programs are still changing a lot, and it’s better to get basic rapid feedback that can be incorporated into the next round, rather than devote extensive resources that won’t be pertinent to the changing program. On the other hand, you may encounter a mature (and consistently presented) program that is still using satisfaction surveys when pre-post measures would be more useful, informative and reliable. It may take a couple of cycles to bring disparate program and evaluation lifecycles closer together.

Program Lifecycles

Each program will follow its own unique path through a lifecycle. Any given program might move forward and backward through and between the phases as needed, although the general tendency will be to progress through the phases over successive implementations.

There are four program lifecycle phases:

1. **Program Phase I: Initiation.**
   a. A Phase I program is a new program just starting up, and will likely face some revisions.
      i. 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.
      ii. Phase IB programs are new programs that have been revised and are being reimplemented. These programs have procedures that tend to change from one implementation to the next.

2. **Program Phase II: Development.**
   a. Programs move into the development phase when they have been implemented successfully one or more times with empirical evidence of participant response. The development phase emphasizes consistent implementation and the development of standardized program procedures or protocols that can produce detectable change in outcomes.
      i. Phase IIA programs are implemented accurately with respect to their model and presented consistently over multiple implementations. Program activities are identifiable from observation or measurement of implementation. The program model incorporates needed adaptations to local context and setting without requiring fundamental change to the model. Participant experience from one implementation to another is relatively stable (e.g. formal lessons or curricula).
ii. **Phase IIB** programs have formalized written procedures (e.g., protocol or implementation plan) outlining how to implement them. The procedures can be followed consistently by new implementers and yet yield a similar experience for participants.

3. **Program Phase III: Maturity.**
   a. Programs move into the *maturity phase* when they meet the requirements of a Phase II program AND show evidence of being associated with changes in outcomes. A program is not mature if it is still significantly changing. In the maturity phase, program planners and providers know what can be expected in implementing the program; there are relatively few surprises. There are relatively few programs that reach this level of maturity in terms of consistent outcomes or effectiveness. A mature program is one that can be considered “evidence-based” because claims about program outcomes are substantiated through formal evaluation.
   i. **Phase IIIA** programs are ones that are associated with consistent change in relevant outcomes. They are aware of other similar programs, much like the child who is becoming aware of the world around them and are conscious that there are others like them
   ii. **Phase IIB** programs have been shown to be effective when compared with an alternative program or no program.

4. **Program Phase IV: Implementation and Dissemination.**
   a. The *implementation and 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, *implementation and dissemination phase* programs are run at more multiple new locations with new and diverse sites, staff and participants.
   i. **Phase IVa** programs are being implemented in multiple sites for the purpose of assuring that effectiveness holds in different places, settings and times.
   ii. **Phase IVb** programs are in wide distribution under the assumption that evaluation evidence shows that they can be implemented widely with fidelity and with comparable effects.

Most programs are not expected to survive through a complete maturity process. 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 Lifecycles

1. Evaluation Phase I: Process and Response
   a. Initially evaluation on new programs should emphasize 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 assesses participant response to the program. Phase I programs benefit most from exploratory rapid feedback.
      i. Phase IA evaluations are typically post-only assessments of participant reactions to the program and their satisfaction from participating.
      ii. Phase IB evaluations are also typically post-only assessments, but focus on substantive outcomes, and are used for program development and assessment of internal consistency (reliability) of outcome measures.

2. Evaluation Phase II: Change
   a. 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 difference between the two sub-phases is where the change is being measured – within groups or within individuals.
      i. Phase IIA evaluations typically use unmatched pretest and posttest of outcomes and assessment of consistency (reliability) and validity of measurement. Results tend to be utilized for management and internal accountability of the program.
      ii. Phase IIB evaluations typically consist of a matched pretest and posttest of outcomes. Verify reliability and validity of change. Because participant identification is necessary to match pre and post outcomes and results are increasingly used for public accountability, the level of protection of participants increases and human subjects review and protection (informed consent, anonymity or confidentiality) is increasingly formalized.

3. Evaluation Phase III: Comparison and Control
   a. The emphasis in this phase is on evaluating whether the program is responsible for causing the observed changes in outcomes (effectiveness). Here, evaluation involves the use of comparison groups or variables and statistical controls for adjusting for uncontrolled factors.
      i. Phase IIIA evaluations use design and statistical controls and comparisons (control groups, control variables or statistical controls).
      ii. 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 and Synthesis
   a. These in-depth program evaluations focus on how well programs dependably have 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.”
      i. Phase IVA evaluations are multi-site integrated assessments yielding large data sets over multiple waves of program implementation.
      ii. Phase IVB evaluations present a formal assessment across multiple program implementations that enable general assertions about a program in a wide variety of contexts (e.g., meta-analysis).
## 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. 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.

The first stage in this hypothetical program lifecycle might be termed the initiation phase and spans from the initial conceptualization through design and initial piloting. The first time you implement a program – even one that was well established elsewhere – you will likely be dealing with the usual issues of initiation: identifying and training program staff, localizing the program to the immediate context, reacting to the unanticipated problems that arise, and so on. The second lifecycle stage might be labeled the growth or development phase and would encompass the process of successive revision of the program as it gets implemented repeatedly over time. With programs in this stage, you are getting used to the program and how it plays out in practice. You have some idea of what to anticipate. You might still be surprised by what occurs, and are still adapting the program as you learn, but you are also increasingly able to anticipate problems before they arise, and you are developing a storehouse of experience in how to deal with them. In the third stage, which we might term the maturity or stability phase, the program has been carried out at least several times with some degree of implementation success. By this point you probably have a good idea of what to expect. You have stable processes for implementation, and the program is routinized and often features a written protocol or process guide. The program is no longer dependent on particular individuals for implementation. If something happens to the initial implementers of the program it can still be carried out by others with high fidelity. The fourth stage might be termed the translation or dissemination phase. The primary focus of this phase is on extending the program to other settings or populations of interest, pushing the ecological boundaries of the program as originally conceived into new niches or applications.

These stages aren’t meant to be a straight-jacket or an inflexible taxonomy. For any given program, the progression may not be sequential. 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. In the best of all worlds probably only a small minority of programs will or should survive to the translation or dissemination phase.

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 or even their jobs, and program critics who push for ever-extending waves of program implementation.

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.

### Evaluation Lifecycle

<table>
<thead>
<tr>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
<th>Phase IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process &amp; Response</td>
<td>Generalizability</td>
<td>Conservation &amp; Control</td>
<td>Process &amp; Response</td>
</tr>
<tr>
<td>Phase IA</td>
<td>Phase IB</td>
<td>Phase IA</td>
<td>Phase IB</td>
</tr>
<tr>
<td>Post-only assessment of outcomes, implementation assessment, outcome measurement evaluation and assessment of internal consistency (reliability).</td>
<td>Controls and comparisons (control groups, control variables or statistical controls).</td>
<td>Controlled experiments or quasi-experiments (randomized experiment; regression-discontinuity) for assessing the program effectiveness.</td>
<td>Multi-site analysis of integrated large data sets over multiple waves of program implementation.</td>
</tr>
<tr>
<td>Unmatched pretest and posttest of outcomes, qualitative assessment of change, and assessment of reliability and validity of measurement.</td>
<td>Matched pretest and posttest of outcomes. Verify reliability and validity of change. Human subjects review.</td>
<td>Formal assessment across multiple program implementations that enable general assertions about this program in a wide variety of contexts (e.g., meta-analysis).</td>
<td></td>
</tr>
</tbody>
</table>
Lifecycle Application

Figure 7 shows 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 is to facilitate the discussions - first regarding their program’s current lifecycle phase, and then around the current evaluation methods and its lifecycle phase. Don’t agonize over these descriptions, approximations will provide the necessary information. Programs can move from one phase to another, or may have traits of more than one phase for each.

<table>
<thead>
<tr>
<th>Program Lifecycle</th>
<th>Evaluation Lifecycle</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phase I</strong></td>
<td></td>
</tr>
<tr>
<td>Initiation</td>
<td>Process assessment and post-only evaluation of participant reactions and satisfaction.</td>
</tr>
<tr>
<td>Is program in initial implementation(s)?</td>
<td>Phase IIA</td>
</tr>
<tr>
<td>Is program in revision or reimplementation?</td>
<td>Phase IB</td>
</tr>
<tr>
<td><strong>Phase II</strong></td>
<td></td>
</tr>
<tr>
<td>Development</td>
<td>Post-only assessment of outcomes, implementation assessment, outcome measurement development and assessment of internal consistency (reliability).</td>
</tr>
<tr>
<td>Is program being implemented consistently?</td>
<td>Phase IIA</td>
</tr>
<tr>
<td>Does program have formal written procedures/protocol?</td>
<td>Phase IIB</td>
</tr>
<tr>
<td><strong>Phase III</strong></td>
<td></td>
</tr>
<tr>
<td>Is program associated with change in outcomes?</td>
<td>Phase IIA</td>
</tr>
<tr>
<td>Does program have evidence of effectiveness?</td>
<td>Phase IIB</td>
</tr>
<tr>
<td><strong>Phase IV</strong></td>
<td></td>
</tr>
<tr>
<td>Dissemination</td>
<td>Controls and comparisons (control groups, control variables or statistical controls).</td>
</tr>
<tr>
<td>Is effective program being implemented in multiple sites?</td>
<td>Phase IVA</td>
</tr>
<tr>
<td>Is evidence-based program being widely distributed?</td>
<td>Phase IVB</td>
</tr>
</tbody>
</table>

Produce a copy of the Program Lifecycle Decision Tree worksheet (in the Appendix) to the working group and have them complete it after a discussion of the following questions:

- When did you begin this program?
- Was it something you developed in-house, or did you use an already established curriculum?
- Do you make changes each time you present the program, or do you consistently present it?
- Do you typically implemented your program in similar or different contexts or places?
- Is there a set curriculum that would allow any qualified person to teach this the same way every time?
- Have you collected evidence or data that shows that your program is effective at attaining its goals?
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 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.

After the group has a consensus of where their program is in its developmental cycle you can provide them with a copy of the Evaluation Lifecycle Worksheet (Appendix VIIb). Bring out the Organization Evaluation Capacity Survey because you may want to consult the information in it while making these decisions. For each evaluation that is used for a program there will be a corresponding evaluation lifecycle phase. The following questions are designed to help program staff determine the lifecycles of their evaluations.

- Does your program include informal evaluation or feedback?
- What types of evaluation have you done?
- How long has evaluation been a formal part of your program?
- Have you ever done a study that is descriptive of your program (e.g. interviews with program participants or leaders)?
- Have you ever done a formal study of your program that included structured data collection?
- Does your funder require a certain type of evaluation?
- How are you using the results of evaluation?

Finally, once the lifecycle analyses for both the program and the program evaluation are complete, use the Lifecycles Map in Appendix VIIc to plot the phase of each. Is the program lifecycle appropriately coupled with the lifecycle of its evaluation? If not, have a conversation about why this is so, and how it can be changed. Refer back to the Map of Stakeholders to explore if reporting requirements or guidelines are appropriate for the lifecycle phase of the program.
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 lifecourse 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 with cast implications for evaluation practice.

2.06 Logic Model

The goal of this step is to help program leaders generate an initial logic model that outlines their assumptions, context, inputs, activities, outputs, and short-, medium-, and long-term outcomes. Use the Logic Model Worksheet provided in Appendix VIII for this step. The following is a description of the components of the logic model (LM) and an example is shown in Figure 8. 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 our Logic Model Rubric (see Appendix IX) for distribution at this point. Later you may use this rubric to provide feedback to them on the quality of their modeling.

Inputs

Inputs should include a BRIEF and complete list of key resources such as staff, curriculum, outside partners and facilities. 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 or exhaustive.

Activities

The activities list should be complete, as it is the basis for the remainder of the Logic Model. Arising from the boundary discussion, staff should be able to determine what constitutes an activity. Activities should be well defined so that someone unfamiliar with the program can understand the activity titles. For example, instead of calling it “Green Garden” you could list 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 activities list should only include activities that reach people who participate or who are targeted. It should 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 if the program wants to evaluate organizational projects or initiatives that are not just confined to education and outreach - in which case some “administrative” sounding things would be a relevant activity. For example, it would be legitimate to have “recruitment” in a logic model if the program was working to diversify the organization’s volunteer pool.)

**Outputs**

Outputs are the tangible by-products of activities. These could include certificates of attendance or completion and objects that were created as part of the program (such as a birdhouse). If something is done by a participant it should be listed in the outputs section. If something has an effect on a participant, then it should be listed in your outcomes section.

**Outcomes**

The difference between short-term, medium term, and long-term outcomes is a matter of relationship to each other. 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. Frequently, a connection is made to literature to show support for the long-term outcome, and staff are not likely to see these long-term fruits of their labor.

**Short-term Outcomes**

Short-term outcomes should describe results and learning (effects on participants) that are logically connected to activities. These outcomes could include changes in awareness, knowledge, attitudes, skills, opinions, aspirations, and motivations. Place outcomes in the short term column at the “first glimmer” of their appearance. 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 short term outcomes.

**Medium-term Outcomes**

Medium-term outcomes should describe mid-term effects on participants that logically connect either activities or short-term outcomes with long-term outcomes. Medium-term outcomes could include new 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 impact logically connected to either short or medium-term outcomes. These are generally written as occurring beyond the individual or personal level and are likely to include 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”.

---

Phase 1: Evaluation Planning
Stage 2: Program Model Development

28
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.

Context

Context is the environment in which a program will take place. For example, context could be within an afterschool program, at a farmer's market, or the within the Spanish-speaking community. (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. Both of these sections should be brief.)

2.07 Pathway Model

A 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:

\[ \text{[ACTIVITY “A”] leads to [OUTPUT “Q”] which leads to [SHORT-TERM OUTCOME “X”] which leads to [MEDIUM-TERM OUTCOME “Y”] which leads to [LONG-TERM OUTCOME “Z”].} \]

For example, using the sample Logic Model in Figure 8, Figure 9 shows that program staff may believe that the activity “facilitate module with students” may lead to the output “completed activity sheets”, which leads to the short-term outcome “make special materials available to teachers and students”, the medium-term outcome “increase student excitement about science and engineering topics”, and the long-term outcome “increase the number of individuals who choose science and engineering careers.”
Figure 10 shows this sequence as a single pathway.

The working group should continue to create these pathway links until “how the program works” has been fully described. This is different from a logic model in that it shows how the components of the logic model are connected. Look at the completed Pathway Model and ask:

- Are there any activities that are not connected to any outcomes?
- Are there any outcomes that are not connected to any activities?
- If yes, try to determine why these gaps may exist.
  - Did they simply forget to include something in their model?
  - Or, are they doing an activity which does not really address their program goals?
- Are they expecting a particular outcome, but not including an activity that would be expected to result in that outcome?

Using assistive technology/software – such as a flow-chart software (available in PowerPoint or other software) - the pathways model may look something like that in Figure 11.
2.08 Determining Evaluation Scope

A role of the Evaluation Facilitator is to help the program staff determine where evaluation efforts can be most valuable for them. There are many points to consider and discuss during this step. The logic model and its pathway model represent a broad overview of everything that is done in a program and all of the expectations for the program's impact. 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. It can help program staff to think of their logic model and pathway model as a blueprint for their program over the next “X” number of years. The “Evaluation Scope” consists of the components of the pathway model that will be focused on in the upcoming evaluation cycle. There will be no need to craft an evaluation question for every outcome listed in the models. You will want to concentrate on those outcomes that are within your current scope, and many factors can influence the selected scope.

Discuss Feasibility

It is important to be strategic and to determine what is appropriate and feasible in the coming evaluation cycle based on program activity, time constraints, funding constraints, and/or reporting constraints. Keep this concept in mind as you continue, and use caution about selecting too large a scope.

Revisit Stakeholders

A pathway analysis 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 all of the stakeholders and place each stakeholder’s number onto the Pathways Model next to the outcomes and pathways that each stakeholder would be most interested in (see Figure 12).
**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 would be more appropriate to focus on a short-term outcome.

**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 pathway (or set of pathways) on which the program is required to report? If so, then evaluation should definitely cover this one.

**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 their model? Is there an “easy” or obvious point on which an evaluation could focus? We refer to this as the “low-hanging fruit.”

**Hub (Grand Central):** Like Grand Central station, there may be a “central point” through which many of the pathways pass. By measuring this outcome, the program may potentially be able to address many aspects of its model simultaneously.

Once again, visual clues are helpful for many people. Transfer the Pathway Model onto paper or a drawing board. Create a numbered list of all of the stakeholders and place each stakeholder’s number onto the Pathway Model next to the outcomes and pathways that each stakeholder would be most interested in. Circle key nodes, highlight pathways, and so forth.

**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 much of the pathway model 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 13.
Keep in mind that the evaluation scope will generally change from year to year and will generally move from a focus on outputs and short-term outcomes in earlier evaluations to one that includes medium- or long-term outcomes later on.

2.09 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 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 onto the program’s pathways. This step involves turning program staff’s attention to other programs and to the research literature.

Ask the working group if there are other programs like this one. These may be in the same organization or in another organization – and 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 literature supports these measures?

But don’t rely only on similar programs. Show the working group how to search the scientific literature for current research that is being conducted in the program area. This can be a time-consuming task, but it is also an important one. (In our case, we assisted in this step by having undergraduate students work on conducting literature searches, and we are making the aggregate results available to many similar programs.) Having an understanding of the evidence that already exists can help the program to focus its evaluation efforts and to identify well-researched 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 the 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 they have located and reviewed relevant literature (evidence) that relates to the causal relations have them indicate this publication visually (such as with a PDF icon or author and year, see Figure 14) on the line connecting the activities and outcomes. Help them keep in mind that they may not be able to find evidence for all of the relations or elements that have been articulated in their pathway model.

### 2.10 Program-Logic Model Synthesis

The purpose of this process is to take stock of where the working group is in their thinking about their program and its evaluation.

You should encourage them to review all of the products from the previous steps in the Protocol. It would be a rare program indeed that goes through the steps in the Protocol in order one time and gets it “right” or to a level that they are satisfied with. If this process is working correctly, when they get to this synthesis step the working group should immediately see things that they want to go back and revise or rework from earlier steps. For instance, they may now have a greater understanding of who their stakeholders are and which stakeholders are most important for the scope of the evaluation they intend to undertake.

This activity actually occurs throughout the entire process so far, but we place the point here so that you deliberately stop at this point to assess the entire 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 this 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 scope 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.

One of the key products of this synthesis step would be a concise evaluation purpose statement. Why is evaluation being conducted? The evaluation purpose statement should be a short paragraph that explains what the evaluation is intended to accomplish, what the main focus will be, and what the results will be useful for.

The product or output of this synthesis step is the compiled set of materials that reflect the work done in this stage of the protocol, including the stakeholder map, program logic model, pathway model, boundary analysis, and lifecycle analysis.
from the perspectives of key stakeholders, evaluation scope and system links. It is at this point that we offer feedback to each program on their logic models, using the logic model rubric mentioned before during the introduction to logic models.

## Stage 3: Evaluation Plan Creation

### Introduction

Now that the program models have been firmed up, the working group can move from talking about the program to talking about evaluation. The third stage, “Evaluation Plan Creation” involves several steps that aid in the creation of an evaluation plan that will serve as a guide for the implementation of the evaluation. This stage includes: Introduction to Evaluation Plans, Development of Evaluation Questions, Development of a Sampling Plan, Identification or Development of Measures, 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, considering what phase your program is at is crucial. That is, the evaluation questions, sampling approach, measurement issues, design and analysis and 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 Introduce the concept of an Evaluation Plan

The goal of this step is to present and discuss the components of an evaluation plan. You should help the working group to understand the importance of having a fully articulated and written plan. Your discussions will change from being about understanding the program and its goals to now talking about evaluation methods.

There are many resources that outline the contents of an evaluation plan, as well as many uses of the evaluation plan. Having an evaluation plan can help to establish credibility with stakeholders – especially funders – because it can aid communication to others about the value of the program. Having a strong understanding of the importance of the evaluation plan will help the group to continue in their commitment to evaluation.

Several of the outputs of Stage 2 will be used to develop the written evaluation plan. Typically we include the program mission and description, the program logic model and pathways model, and decisions about evaluation scope from the previous stage. The plan will also include outputs from this stage - the evaluation questions, a description of the proposed sample, the proposed measures, the program design, and the plans for data analysis. It will probably be useful to review both the program phase and evaluation scope as described earlier because both of these concepts are critical in developing the evaluation plan.

A key distinction in evaluation planning is differentiating between an evaluation cycle and the evaluation of the program over its entire life. An evaluation “cycle” is the evaluation that you do over a specific period of time. Many organizations use an annual evaluation cycle. 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 group developed the evaluation scope in Step 2.08, they were essentially determining how much evaluation of the program they were going to take on in the next evaluation cycle.
(e.g., over the next year). An evaluation plan is always described in terms of the next evaluation cycle. An evaluation plan describes, in detail, how the staff will evaluate their program next year (or the next implementation of the program).

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) for 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.

### 3.02 Develop Evaluation Questions

Evaluation questions are the broad questions asked about the program in general – “Is the program being implemented well?” or “Do our activities have an effect on our desired outcomes?” They will function as the focus of the evaluation planning. Please note that evaluation questions in this context are not the same thing as questions that would be included in a questionnaire for participants; those are measurement questions. Evaluation questions are the major questions your evaluation is trying to address in a specific evaluation cycle (that is, in an evaluation of a specific scope).

Every question should be associated with one or more activities and/or their outcome(s) as defined within the “Scope of the Evaluation,” (from Stage 2) and every question will have at least one associated measure in place by the time the Evaluation Plan is complete.

Also, remind the group of the lifecycle phase of the program, and the planned evaluation lifecycle. Each stage of the lifecycle emphasizes different types of evaluation questions.

<table>
<thead>
<tr>
<th>Program Phase</th>
<th>Type of Questions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1</td>
<td>Description and Implementation</td>
<td>How well is the program implemented?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How do program participants respond to the program?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How do program participants perform on outcome x?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>What is the participant satisfaction level with our activities?</td>
</tr>
<tr>
<td>Phase 2</td>
<td>Relationship and Change</td>
<td>Is participating in the program related to change in a specific outcome?</td>
</tr>
<tr>
<td>Phase 3</td>
<td>Cause and Effect</td>
<td>Does participation in the program cause a change in outcome x?</td>
</tr>
<tr>
<td>Phase 4</td>
<td>Generalizability and Translation</td>
<td>Does the program work effectively and consistently in this setting or context?</td>
</tr>
</tbody>
</table>

The following framework will help to craft evaluation questions for the selected scope of the evaluation. To develop evaluation questions, start by identifying a connected segment or link that falls within the evaluation scope in your pathway model. For instance, for a new program in Phase 1 we might ask an evaluation question like:

**How well is [ACTIVITY “A"] implemented?**

**How do participants perform on Outcome X?**

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.
For a program in Phase 3 you might select an activity and an associated outcome within the scope of the evaluation and substitute them in the question template below:

**What is the effect of [ACTIVITY “A”] on [OUTCOME “Y”]?**

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 [ACTIVITY “A”] on: [OUTCOME “X”]; [OUTCOME “Y”]; [OUTCOME “Z”]?**

For example going back to the pathway model in Figure 9,
Q1: “What is the effect of this curriculum on student test scores?”
Q2: “What is the effect of our volunteer recruitment on students’ excitement about science as a career?”

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. The remainder of the Evaluation Plan is constructed to address these questions.

### 3.03 Develop Sampling Plan

After the 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., people, organizations) from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen. The facilitator should be familiar with at least the general scope of this literature (a good research methods course would be minimal appropriate training for this), including external validity issues, and nonprobability and probability sampling. There are many resources available on this, such as http://www.socialresearchmethods.net/kb/sampling.php. Sampling will also be affected by measurement and design, so once again these steps will occur in a parallel process rather than a sequential process.

Ask the staff: “Who will participate in your evaluation?” Guide your clients to focus exclusively on who or what will answer their evaluation questions and can be measured. Do not fall into the trap of broadly describing the population served by the program. For instance, if a program was for mothers with premature babies, and the only measure/evaluation at the end of the program was the height and weight of the babies, your sample would exclusively describe the babies, not the moms, even though the program was meant for them.

The program description section should have included 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 lifecourse of a program. Newer programs 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 external validity issues and random sampling.

<table>
<thead>
<tr>
<th>Phase 1</th>
<th>Phase 2</th>
<th>Phase 3</th>
<th>Phase 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convenience</td>
<td>Formalized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opportunistic</td>
<td>Deliberate/planned</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonprobability</td>
<td>Probability</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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 doing 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. As programs mature, sampling plans also tend to mature and become more structured and complex.

**Units and Level of Scale:** Another thing to keep in mind is that you are likely to be measuring at different levels or “units” of generality even in a single evaluation. For example in an educational program you may want to measure an outcome for a classroom as a whole (e.g., class activity level or “climate”) as well as at the level of individual students within the class (e.g., student’s subject matter knowledge or interest in science). In this case, you would have a multi-level sampling approach where you are measuring at different levels of scale in the same evaluation. This shows that sampling is related to measurement, design and analysis.

When writing the sampling plan for the evaluation, describe in detail who will participate, whether there are multiple levels of sampling and how participants are selected at each level. And, keep in mind that decisions made here affect and are affected by decisions made in other parts of the evaluation plan.

### 3.04 Identify or Develop Measures

Measurement is a very complex topic. Measurement approaches need to be tailored to the specific program and 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 do the great variety of measurement approaches that one might use in an evaluation. The facilitator should be familiar with at least the general scope of this literature (a good research methods course would be minimal appropriate training for this). You might consult widely accessible resources when thinking about measurement planning (e.g., [http://www.socialresearchmethods.net/kb/measure.php](http://www.socialresearchmethods.net/kb/measure.php)).

Choosing appropriate measures for each evaluation question will likely be one of the central challenges in writing and executing the evaluation plan. Measures will usually take roughly three forms: 1) demographic or descriptive measures, 2) process measures, and 3) outcome measures. Most program leaders will need a great deal of guidance on how to identify and select measures that are appropriate to the program and evaluation lifecycles, and the complexity of programs being in more than one phase of development may cause confusion and uncertainty.

If available, it is best to use an established and validated measure, but at times modifying or creating an existing measure is necessary. The following questions/
steps can help identify measures, and are listed in sequence of how we approach identifying measures.

**Existing measures (already in the office)**

Begin by looking at the measures currently being used in the program. Perhaps they 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 their measures don’t fit this category there may be a trade off between resources available to locate an appropriate established measure and using the existing one.
- Do these measures match the evaluation questions and appropriate evaluation lifecycle?

**Locating Measures in Literature**

The next best option is to locate measures that are already developed and are supported in literature. Situations where this may occur are with evaluation questions that do not have measures already in place, or for evaluation questions that are using rough or un-proven measures. These measures will need to be cited appropriately by the program, and some may even have fees associated with them. Networking with colleagues can also help them to broaden their measures library, and sometimes utilizing a subscale of a non-related measure can address a specific outcome of intent.

- Which evaluation questions do not yet have an identified measure?
- Which evaluation questions may already have an existing measure, but that staff would prefer to have a measure with stronger support in the literature?
- What measures are colleagues using for similar programs, and are they available for this program?
- What does a literature search for related programs, activities or outcomes reveal? A proven measure may be identified either in the article or through a related article in the bibliography.
- Is there a larger measure that has a sub-scale that could be taken from it?
- Is there a larger measure that can address more than one evaluation question, and possibly replace a current un-supported measure for one of them?
- Is there an existing measure that can be modified to fit the program by changing a word or two?

**Developing measures**

Creating new measures is disadvantageous in that it will not allow the program to cite evidence of reliability and validity for their 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 new measure, particularly for newer programs that call for simpler measures. These measures should be undertaken with care to minimize threats to construct validity and reliability. (See other resources, such as [http://socialresearchmethods.net](http://socialresearchmethods.net), for information on validity and reliability.)
Key Questions Regarding Measures:

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

- **ID Outcomes measured.** Does the measure answer the evaluation question? If the evaluation tool is a broad measure, does it address all outcomes being measured? (For example, the ERS forms used by many Nutrition Programs measure multiple outcomes such as knowledge, skill and behavior.)

- **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 appropriate stage of the program and evaluation lifecycles?

- **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.

- **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?

- **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.

- **References**—If using established or named measures, are they properly referenced? Measures developed and field tested by others require the Program to formally list its source.

Once a measure has passed this review it should be included in the written evaluation plan.

How does measurement change over the lifecourse of a program? Newer programs are probably looking for rapid feedback which can be met with simple satisfaction surveys, whereas more mature programs will be looking to show cause and effect. Below is a tentative chart of the intentions of evaluation at each stage of development.

<table>
<thead>
<tr>
<th>New Programs</th>
<th>Mature Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informal</td>
<td>Reliability</td>
</tr>
<tr>
<td>Open-ended</td>
<td>Standard protocol of measurement</td>
</tr>
<tr>
<td>Pilot testing</td>
<td>Consistency</td>
</tr>
<tr>
<td>Exploring measurement</td>
<td>Structured</td>
</tr>
<tr>
<td>approaches</td>
<td>Validity</td>
</tr>
<tr>
<td>Identifying existing measures</td>
<td>Quantitative</td>
</tr>
</tbody>
</table>

At the risk of repeating ourselves, remember that decisions about measures don’t occur in a vacuum – they are related to sampling, design and analysis issues, that is, they will both affect and be affected by these other topics.
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 system 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.

### 3.05 Develop Evaluation Design

An evaluation design shows how the evaluation will be conducted, as it relates to program implementation, sampling and data collection. The design links sampling, measurement and analysis and shows how all of the major parts of the research project work together to try to address the central research questions. Simplified general research designs are described below, but selecting a design will vary depending up facilitator or EPM preferences. Once again, we refer you to the literature for more in-depth information on design, including [http://www.socialresearchmethods.net/kb/design.php](http://www.socialresearchmethods.net/kb/design.php).

We often describe a design using a concise notation that enables us to summarize a complex design structure efficiently.

- **Observations or Measures** are symbolized by an ‘O’. Distinguish among specific measures, with subscripts, as in \( O_1, O_2 \), and so on.
- **The Activity** is symbolized with an ‘X’. As with observations, use subscripts to distinguish different activities or activity 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 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:

\[
\begin{array}{ccc}
O & X & O \\
N & O & XO \\
N & O & O
\end{array}
\]

Represents a pretest and a posttest before and after the activity

\[
\begin{array}{ccc}
N & O & XO \\
N & O & O
\end{array}
\]

Represents a pre/post group with a non-equivalent comparison 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. One other thing. It is important to always keep the evaluation questions in mind when doing the various parts of evaluation planning. If this is not done, there is the danger of developing a nice evaluation design that doesn’t help answer the questions that were being focused on. This may seem obvious, but sometimes the working group’s enthusiasm from thinking about all the possible outcomes can cause a drift away from the evaluation questions.

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

- Is there a clear connection between the evaluation questions, chosen measures and the resulting design?
- 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 evaluation lifecycle should match up with the program lifecycle, and that different evaluation designs should be used for different evaluation lifecycle phases. As a reminder, these are laid out in the table below.

<table>
<thead>
<tr>
<th>Program Phase</th>
<th>Evaluation Phase</th>
<th>Typical Evaluation Designs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Initiation</td>
<td>1a. Process</td>
<td>Post-only assessment of reaction and satisfaction</td>
</tr>
<tr>
<td></td>
<td>1b. Response</td>
<td>Post-only assessment focused on outcomes</td>
</tr>
<tr>
<td>2. Development</td>
<td>2a. Change at the group level</td>
<td>Unmatched pretest and posttest of outcomes, assess reliability and validity of the measure</td>
</tr>
<tr>
<td></td>
<td>2b. Change at the individual level</td>
<td>Matched pretest and posttest of outcomes, assess reliability and validity of the change.</td>
</tr>
<tr>
<td>3. Maturity</td>
<td>3a. Comparison</td>
<td>Control groups, controls variables or statistical controls</td>
</tr>
<tr>
<td></td>
<td>3b. Control</td>
<td>Controlled experimental or quasi-experimental designs</td>
</tr>
<tr>
<td>4. Implementation and</td>
<td>4a. Generalizability</td>
<td>Multi-site integrated assessment</td>
</tr>
<tr>
<td>Dissemination</td>
<td>4b. Synthesis</td>
<td>Meta-analysis – multi site with multiple contexts</td>
</tr>
</tbody>
</table>
3.06 Develop Analysis Plan

Information on how to analyze data is huge and careers are based on this skill alone. This should not discourage or cause abject fear in the average program manager because most programs’ needs can be met through the relatively simple analysis methods. Also, programs that don’t have the capacity for their analysis needs frequently have resources available to them for addressing this. Having a plan (and tools) in place for analyzing evaluation data, in addition to working with an experienced evaluation facilitator, can help allay any concerns that program staff may have. As that facilitator, you should be familiar with qualitative, quantitative and mixed methods methodologies for manipulating and synthesizing data (also see http://www.socialresearchmethods.net/kb/analysis.php).

The goal of this step is to articulate and put into writing the plan for analyzing the evaluation data once it has been collected. This section can be quite succinct. As with the previous sections, look at the draft analysis section and ask:

- Is there a clear connection between evaluation questions, measures, 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?

We help program staff to install and use the Analysis ToolPak that comes with Excel (see Appendix XI). This Analysis ToolPak is a free program available to anyone with Microsoft Excel, and is simple enough for most people to use. In addition to various resources available from http://www.socialresearchmethods.net/kb/, additional free online software and tools can be found at http://statpages.org/.

3.07 Develop Evaluation Reporting Plan

A 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. The reporting section should also be concise and understandable, and illustrate a clear “throughline” that starts from the evaluation questions and continues through the measures, design and analysis. Most programs currently have basic reporting requirements both internally and externally. The working group could start this section by listing the required reporting for the program, and then continue by looking at the plan for analysis and asking:

- Have evaluation questions been answered or addressed? If so what form should “answers” take?
- Does the reporting section indicate all audiences the organization will be reporting to? (Include internal and external audiences.)
- Are both frequency and type of reporting identified? (Frequency: Monthly? Quarterly? End of fiscal year? Type: An evaluation summary? Informal ongoing reports to be generated and distributed at every staff meeting?)

Reporting needs are likely to change depending upon the program lifecycle. For instance, Phase 1 program reporting tends to be more internally focused and more private, whereas later phase reporting moves toward becoming more public. Also, earlier phase reporting tends to be less formal while later phase 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 Facilitator 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.

3.08 Develop Implementation Plan and Schedule

The goal of this step is to develop the Timeline (schedule) for the evaluation and key implementation milestones. The measures, design, analysis and reporting sections 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 in the beginning of the year 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.

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?
- 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?

3.09 Finalize Evaluation Plan

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

Using the evaluation plan rubric (see Appendix X) we provide feedback to each working groups about their entire plan. If necessary or desirable they may continue to edit their plan until they are happy with it, at least for this evaluation cycle. We recommend printing and binding the entire plan so that it is available to the program leader, evaluator 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 as they move to the future. A good evaluation plan helps them to present their programs 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 it is time to conduct the evaluation and is not practical to keep fine tuning it. Those edits can be reserved for the next evaluation cycle.
Summary

At this point take a step back and celebrate the program’s progress. 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 going, and where we are.

We had suggested that evaluation is ideally viewed as having three stages – planning, implementation, and utilization. With the completion of the evaluation plan, the program has moved through the first stage of the process, and certainly through the most thought-intensive part of the process. In addition, they should have a timeline in place for their implementation and a plan for reporting, so planning for the entire 3-phase process has been completed. You are now ready to move on to the implementation phase (for which we are currently working on drafting a similar protocol utilizing a systems perspective).

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) 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. 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 are hoping that we have succeeded in demonstrating that the concept of a Systems Evaluation Protocol offers a useful approach to evaluation. You should be more aware of the dynamic and evolving organism that each program is, and be cognizant of how these dynamics influence the program’s evaluation needs. The complex nested systems 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.
The glossary is in process, and we hope to have definitions for the next version of the protocol. If there are other terms you would like defined, please let us know.

Analysis
Cyberinfrastructure
Design
Evaluation Capacity
Evaluation Cycle
Evaluation Implementation
Evaluation Partnership (EP)
Evaluation Plan
Evaluation Planning
Evaluation Professional, Evaluation Facilitator
Evaluation Questions
Evaluation Scope
Evaluation Utilization
Evolutionary Systems Perspective
Facilitation
Lifecycle
Logic Model
Measure
Organization
Pathway Model
Program
Program Boundary
Program Managers
Program/Program (inconsistent use in text and titles of materials)
Protocol
Sample
Stages of Evaluation – Planning, Implementation, Utilization
Stakeholder
STEM
Systems Evaluation
Systems Evaluation Protocol (SEP)
Systems Perspective
Timeline
Working Agreement (MOU)
Appendices

Appendix I - Sample MOU

Appendix II - Systems Evaluation Protocol Contact List

Appendix III - Organization Evaluation Capacity Survey

Appendix IV - Program Evaluation Capacity Survey

Appendix V - Launch Presentation

Appendix VI - Stakeholder Mapping
   A - List of Stakeholders
   B - Stakeholder Incentives Worksheet
   C - Stakeholder Map Worksheet

Appendix VII - Lifecycles
   A - Program Lifecycles
   B - Evaluation Lifecycles
   C - Lifecycle Alignment

Appendix VIII - Logic Model Worksheet

Appendix IX - Logic Model Rubric

Appendix X - Evaluation Plan Rubric

Appendix XI - Statistical Analysis Guidelines for Phase I-B Projects
Appendix I - Sample MOU

This Memorandum of Understanding (MOU) describes the roles and responsibilities of the participants in the [Organization] Evaluation Planning Partnerships for the 2008 calendar year. The document details the expectations for the Evaluation Team and county Extension administrators and staff, and provides a list of deliverables, a timeline and budgeting considerations. This MOU is not intended to be a formal legal document; it is meant to be a vehicle to plan the year’s activities and help assure that all parties are clear about what is intended and can feasibly be accomplished.

1. Background

Cornell Cooperative Extension (CCE) launched a new effort in 2006 to develop a process for building evaluation capacity throughout the Extension system, and for linking participating counties into a new statewide Extension Evaluation Network. The work in the first year focused on CUCE-NYC. Six additional County Associations were invited to join the project in 2007. The six new Associations each selected four programs to include in their pilot effort, and identified a staff person to serve as “Evaluation Project Manager” (EPM). With in-house coordination from the EPM, the selected programs used the tools and processes developed by the Evaluation Team in the previous year, and were provided with a common template and web-based software for the development of Program Logic Models and Evaluation Plans. Final Reports – including Program Descriptions, Logic Models, Pathway Models, Stakeholder Charts, and Evaluation Plans, were finalized and reviewed at a year-end meeting of the EPMs and the Evaluation Team in December 2007.

The goals for the Evaluation Planning Partnerships in 2008 are to:

- Support revisions to the Evaluation Plans that were developed at the end of 2007 to incorporate any re-assessments of what will be feasible, and any new priorities on the part of the Programs, Associations, or other stakeholders
- Complete the provision of research-based input for the priority Measures that were requested in 2007 by EPMs (including actual evaluation instruments where possible, and/or evaluation instruments that can be refined appropriately)
- Further develop the evaluation planning capacity of participating Associations through provision of materials and information/training sessions (via teleconference or video conference) that will allow Association program staff to implement their Evaluation Plans; find and assess measures as called for in their Evaluation Plans; analyze evaluation data; and use evaluation results effectively for internal feedback (program improvement) and more formal reporting to internal and external stakeholders
- Organize and make available on the Evaluation Team website the EPM training materials and tools including “Tips” documents, presentation notes, templates, etc. so that these are readily accessible to participating Associations
- Work together with participating Associations to identify and develop internal policies, procedures, and communication strategies that promote a positive culture and practice of evaluation
- Continue to build the connections and network among EPMs and EDs in the participating Associations as a means of developing mutually beneficial evaluation capacities
- Strengthen the statewide Extension capacity for evaluation of 4-H Youth Development Programs by hosting a one-day conference in May 2008
2. Evaluation Team Responsibilities

Overall Role
The Evaluation Team will work with all CCE Associations selected for participation in 2008. The Evaluation Team currently includes:

- Evaluation Director (Name and Title)
- Evaluation Manager (Name and Title)
- Evaluation Graduate Research Assistant (Name and Title)
- CCE Assistant Director for Program and Professional Development (Name and Title)

The team will provide technical assistance and support, conduct training and information sessions, and facilitate ongoing communication among participating Associations. All parties will be expected to utilize video conferencing, the web, and conference calls wherever possible to achieve efficiencies.

Project Preparation
The Evaluation Team will launch the 2008 EPP work by conducting an introduction and planning teleconference with all EPMs, and a subsequent teleconference with participating Executive Directors (EDs). These teleconferences will review the staff changes that have taken place on the Evaluation Team, and will lay the foundation for the year ahead. On the call with the EDs, this MOU will be reviewed.

Evaluation Implementation
With new (and possibly revised) Evaluation Plans in hand, the critical task for EPP Associations in 2008 is implementation. To provide guidance and support during this phase, the Evaluation Team will conduct collaborative development/training sessions for the Evaluation Project Managers (EPMs). These sessions will be held via teleconference, videoconference, or on-site. Additional support will be provided through on-going phone and e-mail consultations with the Evaluation Team. The primary contact for the EPMs will be the Evaluation Manager.

The interactions during the evaluation implementation phase will enable the participants to address the following tasks:

- Refine Program Logic Models and Evaluation Plans to make program planning and evaluation more effective, and more feasible
- Identify and use high-quality, appropriate evaluation instruments
- Manage, analyze, and interpret the data
- Use the Netway software to create basic reports, and communicate effectively with internal and external program stakeholders
- Make recommendations for future analysis and reporting

Mid-Year Meeting
The Evaluation Team will lead a half-day meeting, to coincide with the 4-H Youth Development Evaluation Conference in May. At this meeting, to be held on May 19th, participating EPMs will be asked to share perspectives on their evaluation planning and implementation, challenges and barriers identified, and any new tools or approaches developed. The group will consider opportunities and challenges for the development of evaluation capacity system-wide in future years, and will discuss how they can most effectively continue to work together as an evaluation network to continue to develop evaluation capacity for their offices and for the Extension system.
3. Partner Responsibilities

The Role of the Executive Director

1) Overall Role
The Executive Director’s role is essential to the success of this effort to develop evaluation capacity and embed evaluation practice within their Association. This may take many forms, including:

- Clear statements to staff regarding the Association’s commitment to program planning that facilitates effective evaluation
- Continued use of evaluation results, whether in communications to internal stakeholders such as the Board of Directors or external stakeholders such as key funders
- Setting realistic expectations and providing resource support for the EPM and project staff to build sustainable capacity and practice

2) Project Preparation

- On the initial teleconference with the Evaluation Team the ED will review this Memorandum of Understanding and should gain an understanding of: expectations and planned outcomes for 2008; what this might entail in terms of additional staff time and resources; and the support available from the Evaluation Team
- The ED is asked to review this Memorandum of Understanding and sign it by April 1, 2008. This will signify that he/she understands and agrees to the scope of the proposed work
- The ED is expected to identify a staff member who will serve as the as EPM. It is assumed that the EPM for 2007 will continue to serve as EPM for 2008, but if this is not the case the ED should inform the Evaluation Team of the change. Key criteria to consider in choosing an EPM might include:
  - Communications and consultation skills
  - Positional authority and respect among colleagues
  - Technical background (e.g., familiarity with evaluation, general computing skills)
  - Interest and motivation
  - Time availability

3) Review of Evaluation Plans
As a first step in the Evaluation Implementation effort, EDs are asked to review the existing Evaluation Plans with their EPM and Program Leaders for the participating programs in order to:

- Become familiar with what has been learned over the course of the first year of work with the EPP
- Assess what still needs to be done in terms of staff training and support on evaluation
- Determine, with the EPM, how best to support and encourage the use of the evaluation protocol and tools that have been provided through the EPP
- Adjust staff work loads and job descriptions to incorporate the commitment to evaluation

4) Support the EPM
The ED’s support of the EPM is essential for the EPM’s success. Supporting the EPM in the content, scheduling and coordination of the deliverables is a key role for the ED.
The Role of Evaluation Project Manager

1) Overall Role

As in 2007, the Evaluation Project Manager will be the primary contact between the Evaluation Team and the CCE Association office. The EPM will work closely with program staff from all participating Programs to support the effective implementation of their Evaluation Plans. He/She will be responsible for disseminating the evaluation training and planning strategies recommended by the Evaluation Team and the ED.

2) Project Preparation

The EPM should hold a meeting with the ED to review Evaluation Plans; coordinate strategies for ensuring follow-through on Evaluation Plan revisions and implementation; and make sure that there are sufficient staff resources and staff time allocated to Evaluation Plan implementation. As part of the forward-looking planning, the EPM is asked to assemble a “master calendar” of evaluation plan deadlines and requirements and oversee the execution of all crucial internal deadlines and deliverables.

3) Evaluation Implementation

The EPM, will be responsible for the content, scheduling and coordination of the deliverables for their CCE office evaluation project. As such, the role of the EPM includes but is not limited to responsibility for the following:

- Schedule, coordinate and host all teleconferencing and remote meetings and/or trainings pertinent to their site
- Schedule and lead internal meetings as needed in order to support and promote the successful implementation of the Evaluation Plans
- Present mid-year results at the meeting for all participating counties on May 19, 2008
- Participate in the 4-H Youth Development Evaluation Conference on May 20, 2008 as part of the EPM Panel discussion of their experiences in evaluation planning through the EPP project
- Coordinate and complete follow-up internal evaluations designed to assess progress made as a result of participation in the EPP project
- Participate in ongoing communication among all the EPMs and the Evaluation Team.
- Support the growth of the CCE/Cornell University evaluation network
- Submit a written year-end report, using a template to be provided by the Evaluation Team, of overall outcomes of the EPP effort

While the amount of time spent on revisions and evaluation plan implementation may vary from week to week, it is expected that for weeks with no in-services, video conferences or projects/assignments the EPM will likely spend up to 5% of his/her time on their CCE county evaluation objectives during this implementation year.

4. Deliverables

Resource Archive of Evaluation Measures, Training Materials, and Templates

Staff in participating Associations will have access to the Evaluation Team website where these resources will be housed and expanded over time.
Strengthened Internal Capacity

Educated and experienced EDs, EPMs and program staff will increase each Association’s ability to evaluate their programs. It is expected that these individuals will express their valuation of the evaluation planning process, and extend this opportunity to all colleagues at the participating CCE offices. The EPM will be encouraged to take an ongoing leadership role in terms of evaluation planning and the use of the Netway at his/her office.

Extension Evaluation Network

External evaluation capacities will also be increased in the form of an expanding Cornell Extension Evaluation Network. Participating counties will make lasting connections, gain skills, and improve CCE county infrastructure using the Netway, our website/listserv, and the opportunity to attend statewide conferences and intra-office meetings. Staff from non-participating counties will attend the 4-H Youth Development Evaluation Conference in May and be introduced to the concept of evaluation planning.

5. Timeline

The following timeline is approximate and is intended to provide an overview of the process. It will likely change and adapt to the needs constraints of the participating office and the Evaluation Team.

- **JANUARY – MARCH 2008**
  - Finalize staff changes on Evaluation Team
  - Meet with EPMs via teleconference to provide and gather updates, launch this implementation year
  - Meet with all participating EDs via videoconference
  - EDs review and sign MOU
  - Develop internal Association calendars for project year to track key program and Evaluation Plan deadlines
  - Conduct first collaborative development/training session, to focus on “Finding and Assessing Potential Evaluation Measures”

- **APRIL - AUGUST 2008**
  - Conduct half-day mid-year meeting for EPMs (afternoon of May 19, in Ithaca)
  - Host 4-H Youth Development Evaluation Conference (May 20, in Ithaca)
  - Conduct second collaborative development/training session, to focus on “Managing and Analyzing Evaluation Data and Responses”

- **SEPTEMBER – DECEMBER 2008**
  - Conduct third collaborative development/training session, to focus on “Using Evaluation Results Effectively in Feedback and Reporting”
  - Schedule and hold fourth collaborative development/training session if needed, on a topic to be determined
  - EPMs submit Year-End Final Reports on overall evaluation capacity, and Evaluation Plan Implementation
6. **Budgetary Considerations**

The participating county office will be responsible for all local costs associated with this project, including telephone, printing, mail, computing, meeting refreshments, teleconferencing and video-conferencing (if used) etc.

The remainder of the cost to each county will consist of travel for EPMs to attend the mid-year meeting and conference in May.

Note: In addition to the investments being made by EPP Associations, we would like EDs to know that we estimate that Extension Administration is committing approximately $XXXX per year and per Association to the support of the Evaluation Planning Partnerships.

7. **Changes to the Evaluation Agreement**

At any time, the partnership site and/or Evaluation Team may call for a review of the Memorandum of Understanding. Any changes to the MOU must be agreed upon by both groups.

8. **Agreement**

Accepted by:

_________________________________________________________
Executive Director

_________________________________________________________
Date

_________________________________________________________
[Name of Facilitation Team Director]

_________________________________________________________
Date

9. **Contact Information**

Please provide the name and contact information for either the Executive Director or the designated initial primary contact for this project.

Partner Site: ____________________________________________________________

Contact Information: ______________________________________________________
## Appendix II: System Evaluation Protocol Contact List

### Systems Evaluation Protocol Worksheet

<table>
<thead>
<tr>
<th>Name</th>
<th>Title/Role</th>
<th>Contact Information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Organizational Evaluation Capacity Survey

For the 2007 Evaluation Planning Partnership Launch

<table>
<thead>
<tr>
<th>Organization:</th>
<th>Date:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of EPM:</td>
<td></td>
</tr>
</tbody>
</table>

This checklist is designed to help us prepare for our evaluation planning partnership. It is intended to be completed with the ED and/or the EPM for the organization. It is meant to assess organization-wide evaluation capability, so only one such form needs to be completed per organization.

## I. Resources

### PERSONNEL

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Do you have staff with formal or informal training in evaluation or related areas (e.g., measurement, research methods, statistics, and qualitative analysis)?
  - **If yes, please provide details**

- Do you have staff with ongoing formal assignments to conduct evaluation in relation to any of your programs?
  - **If yes, please provide details**

- Do you have contractors, volunteers or other persons outside of your staff that work with your organization and have skills or experience in evaluation and related areas?
  - **If yes, please provide details that describe the types of capabilities such people have.**

- Do you routinely use external contractors, volunteers or other persons outside your organization to assist you in accomplishing evaluation?
  - **If yes, please provide details that describe the types of evaluation work they provide**

### BUDGET

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Do you have an organizational budget line explicitly devoted to evaluation?
  - **If yes, please provide the dollar value and percent of the total budget allocated to**
**Evaluation:**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| □ | □ | Do you use general administrative or project budget lines to cover evaluation costs?  
If yes, please describe: |

How much do you estimate your organization spends annually on all evaluation-related expenses?  
**dollars per year**

**II. TRAINING**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| □ | □ | Do you have regular in-service training on evaluation or related topics?  
If yes, please describe: |

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| □ | □ | Over the past year have you provided support in-service training or for employees to take part in professional training on evaluation or related topics?  
If yes, please describe: |

How much do you estimate your organization spends annually on all evaluation training and staff development efforts?  
**dollars per year**

**III. INFORMATION TECHNOLOGY RESOURCES**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| □ | □ | Do you have an organization-wide computer network?  
If yes, please describe the network and how it is managed: |

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| □ | □ | Do you have organizational policies for document sharing and version control?  
If yes, please describe the policy: |

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| □ | □ | Do you have staff trained in database development and management?  
If yes, please describe their training and how this is managed: |

**SOFTWARE RESOURCES**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>□</td>
<td>□</td>
<td>Do you have Microsoft Office (e.g., Word, Excel, Powerpoint)?</td>
<td></td>
</tr>
</tbody>
</table>
|   |   | Do you have database software in house (e.g., Access, SQL Server)?  
|   |   | Do you have statistical software in house (e.g., MINITAB, SPSS, SAS)?  
|   |   | Do you have qualitative software in house (e.g., Atlas TI)?  
|   |   | Do you have staff who are trained to operate evaluation-relevant statistical or qualitative software?  
|   |   | If yes, please describe capability:  
|   |   | IV. Evaluation Policies  
| Yes | No | Do you have an evaluation requirement for all programs in your organization?  
|   |   | If yes, please describe frequency (e.g., annual) and nature of requirement:  
|   |   | Do you require that each program have a written evaluation plan?  
|   |   | If yes, please describe what type of plan is required? Note especially any methods or measures that you require:  
|   |   | Do you require that each program explicitly budget for their evaluation work?  
|   |   | Please describe what issues around budgeting:  
|   |   | Do you require that each program report evaluation results on a regular basis (outside of annual report requirements)?  
|   |   | If yes, please describe frequency (e.g., annual) and nature of requirement:  
|   |   | Do you include evaluation responsibilities in formal job descriptions for any of your staff?  
|   |   | If yes, please describe:  
|   |   | Do you base staff review or promotions in part on evaluation data?  
|   |   | If yes, please describe the issues involved:  

THANK YOU! Please fill this out electronically.  
If you have any questions contact
Appendix IV: Program Evaluation Capacity Survey

Program Evaluation Capacity Survey
For the 2007 Evaluation Planning Partnership Launch

<table>
<thead>
<tr>
<th>County Name:</th>
<th>Date:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of EPM:</td>
<td></td>
</tr>
</tbody>
</table>

This checklist was designed to help us prepare for our evaluation planning partnership. Please fill out as much as you can and/or have the information ready so we can review the checklist together in team meetings during the upcoming launch.

Please prepare a separate checklist for each of the four Programs you have chosen as pilots for the year one of our evaluation planning partnership.

<table>
<thead>
<tr>
<th>Program Area Name:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program Name:</td>
</tr>
<tr>
<td>Staff Names (and Titles):</td>
</tr>
</tbody>
</table>

**Part I. Program Information**

**DESCRIPTION**

Yes | No
---|---

- Is there a formal description of the Program?
  - **If yes, please include Program description in the space below.**
  - (A short paragraph from a proposal or your website is OK.)

- Do you have formal goals/mission for this Program?
  - **If yes, please include a description of the Program’s goals/mission in the space below.**

**INPUTS**

Yes | No
---|---

- Please check the inputs or resources that go into this program. (full/part time staff, curricula, etc.)

- **Staff?** (Please list by total % FTE, not by name.)
### Curricula? (Please list by name)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Other? (Please list)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### What is the annual budget for Program?

### PARTICIPANTS

**What is the estimated number of participants per year in this Program?**

**Please briefly describe your intended participant population.** (Primarily mothers? Persons on public assistance? Students interested in Veterinary science?)

(Please describe)

**Age?** Please indicate the normal age range of Program participants.

(Please list)

**Sex?** What is the estimated ratio of female to male participants in this Program?

- % Estimated *female* participants
- % Estimated *male* participants

**Race/Ethnicity?** Please check the race/ethnicity categories that constitute a significant % of participants served. (We do not have a specific % in mind that would qualify as “significant”, so please use your best judgment.) Space is provided at the bottom if you would like to include additional information or details.

- African American or Black
- American Indian/Alaskan Native
- Asian
- Hawaiian Native/Other Pacific Islander
- Hispanic/Latino
- White
- Other (“Other” is reserved both for multi-ethnic participants and for participants who identify as ethnicities not listed above.)
### Education?
Please check the education categories that constitute a significant % of participants served.
(Space is provided at the bottom if you would like to include additional information or details.)

- [ ] 8th grade or less
- [ ] 9th-11th grade
- [ ] 12th grade or GED
- [ ] Beyond high school
- [ ] Some college
- [ ] 2 year college grad
- [ ] 4 year college grad
- [ ] Post 4 year college grad

### Monthly Household Income?
Please check the monthly household income levels that constitute a significant % of participants served.
(Space is provided at the bottom if you would like to include additional information or details.)

- [ ] $0-$500
- [ ] $501-$1,000
- [ ] $1,001-$1,500
- [ ] $1,501-$2,000
- [ ] $2,001-2,500
- [ ] over $2,501

### Region/Locale?
Is there a region or locale in your county where this Program primarily takes place, such as a specific town or area? For instance, some urban gardening projects in NYC have primarily taken place in certain neighborhoods or boroughs. If so, please describe below.

### ACTIVITIES

What types of activities are conducted with participants in this Program?
Please check all that apply.

- [ ] Workshops? (one stand alone class)
- [ ] Classes? (series of classes)
- [ ] Training? (in the field)
- [ ] Training? (at the CUCE office)
- [ ] Demonstrations? (e.g. a cooking demonstration at a health fair)
<table>
<thead>
<tr>
<th>OUTCOMES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>
| □       | □       | Do you have immediate outputs expected as a result of this Program? (These are normally tangible outputs, such as a certificate received directly after the final activity.)  
**If yes, please list expected immediate outputs.**  
Certificates for some programs |
| □       | □       | Do you have short-term outcomes expected as a result of this Program?  
**If yes, please list expected short-term outcomes.**  
Participation in the County Fair |
| □       | □       | Do you have mid-term outcomes expected as a result of this Program?  
**If yes, please list expected mid-term outcomes.** |
| □       | □       | Do you have long-term outcomes expected as a result of this Program?  
**If yes, please list expected long-term outcomes.** |
| □       | □       | **Other?** (Please list:) |
## CONTEXT

Please describe the "real life" circumstances in which this Program takes place. (In the space below, please provide us with a bulleted list of a few of the most important external issues or factors affecting the staff, activities and participants in this Program. Examples of context would include a science class held in a juvenile hall (participants’ context), if a campaign to halt obesity is part of a new national initiative (programming context), or if a grant is coming to a close (funding context).

Please briefly outline your Program context here:

## ASSUMPTIONS

In the space below please describe a few of the crucial assumptions central to the work of this Program. Examples of Program assumptions include, a curriculum developed by a faculty PI must remain unchanged, (input assumptions) primary participant recruitment comes from agency partners at Head Start, (participant assumptions), or by the use of the “multiplier effect” trainees will reach over 25 additional persons in the year following Program completion (outcome assumption).

Please briefly outline your Program assumptions here:

## Part II. Evaluation Information

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
<th><strong>Is the Program currently being evaluated?</strong></th>
</tr>
</thead>
</table>

If yes, by whom? (check all that apply)
- [ ] Internal staff?
- [ ] External parties?

Please include a short description here:

**If the Program is evaluated, how are the results shared?**
- [ ] Formally? If so, how?
- [ ] Informally? If so, how?

**If evaluation results are shared, how are they currently used and by whom?**

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
<th><strong>Are Cornell faculty directly involved with the Program?</strong></th>
</tr>
</thead>
</table>

If yes, please briefly describe the role of the faculty:

**Who funds the Program?**

What kind of evaluation reporting do the funders require?

Please describe:

**What specific evaluation instruments have been used in the past/present?** (satisfaction surveys, pre-post tests, etc.) Please list the existing evaluation instruments for this Program here:

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>

**Have evaluation questions or hypotheses been developed?**

If yes, please list them here:

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>

**Is evaluation data currently being collected?**
<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>If yes, how is the data collected?</td>
<td></td>
</tr>
<tr>
<td><strong>What methods are used?</strong> (if mixed, click both)</td>
<td></td>
</tr>
<tr>
<td>□ Qualitative?</td>
<td></td>
</tr>
<tr>
<td>□ Quantitative?</td>
<td></td>
</tr>
<tr>
<td><strong>Is current evaluation data being analyzed?</strong></td>
<td></td>
</tr>
<tr>
<td>If yes, by whom?</td>
<td></td>
</tr>
<tr>
<td>□ □</td>
<td></td>
</tr>
<tr>
<td><strong>Is there a plan for reporting the evaluation results?</strong></td>
<td></td>
</tr>
<tr>
<td>When? By whom to whom?</td>
<td></td>
</tr>
<tr>
<td>□ □</td>
<td></td>
</tr>
<tr>
<td>Has the evaluation of this program been formally reviewed and approved by the University Committee on Human Subjects?</td>
<td>□ □</td>
</tr>
</tbody>
</table>

**THANK YOU! Please fill this out electronically.**

**If you have any questions contact**
Appendix V: Launch Presentation
Evaluation Partnerships
EPP Goals

I. Enhance the evaluation skills and leadership within each EP office

II. Complete Logic Model and an Evaluation Plan for the selected programs

III. Provide access to evaluation templates, tools and processes for evaluation planning
Evaluation Planning Partnerships

- Each office has selected up to four programs
- Evaluation consultant team:
What is evaluation?

- Evaluation is the systematic assessment of the worth or merit of some object.
- Evaluation is the systematic acquisition and assessment of information to provide useful feedback about some object.
Steps in Evaluation:

1. Identify & Engage Stakeholders
2. Develop Model
3. Develop Evaluation Plan
4. Gather Data
5. Assess Data
6. Utilize Results
Step 1: Identify & Engage Stakeholders

1. Engage Stakeholders

- Participants
- Community
- Organizational Administrators
- Program Staff
- The Public
- Researchers
- Policymakers
- Funders
### Step 2: Develop Model

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Activities</th>
<th>Outputs</th>
<th>Short</th>
<th>Medium</th>
<th>Long</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Context / Assumptions*
<table>
<thead>
<tr>
<th>Logic Model Report</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Organization</strong>:</td>
</tr>
<tr>
<td><strong>Office</strong>: NYS CCE</td>
</tr>
<tr>
<td><strong>Program Area</strong>: Family and Youth Development</td>
</tr>
<tr>
<td><strong>4-H After-School Clubs and Special Interest Projects</strong></td>
</tr>
</tbody>
</table>

| Today's Date | January 9, 2007 |

### 4-H After-School Clubs and Special Interest Projects

<table>
<thead>
<tr>
<th><strong>Family and Youth Development</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>New York Office</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Activities</strong>:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent/Science &amp; Tech</td>
</tr>
<tr>
<td>CE Coordinator</td>
</tr>
<tr>
<td>Public Presentation</td>
</tr>
<tr>
<td>Program Coordination &amp; Training</td>
</tr>
<tr>
<td>Youth Leadership Academy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Volunteers</strong>:</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE Educators</td>
</tr>
<tr>
<td>Staff</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Short-Term Outcomes</strong>:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased interest and participation in 4-H</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Mid-Term Outcomes</strong>:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased leadership skills</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Long-Term Outcome</strong>:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased interest in higher education</td>
</tr>
</tbody>
</table>

**Assumption 1**: Assuming that 4-H offers the right opportunities for urban youth

**Assumption 2**: Assuming that 4-H offers the right opportunities for rural youth

**Assumption 3**: Staff is proficient in training and the trainer skills

**Context**: 4-H is changing to meet the needs of urban youth

**Contact**: 2 Fiscal & Organization Changes
Pathway Logic Model

Program Information

Office Name: New York City
Program Area Name: Family and Youth Development
Program Name: 4-H After-School Clubs and Special Interest Projects

Pathways Diagram

Legend: Activity, Immediate Outcome, Short-Term Outcome, Medium-Term Outcome, Long-Term Outcome
Step 3: Develop Evaluation Plan

- **Program**
  - Description
  - Questions
  - Activities
- **Methods**
  - Sampling
  - Measures
  - Design
  - Analysis
- **Reporting**
Step 4: Gather Data

- Document analysis
- Interviews
- Photos/videos
- Tests & Exams
- Observations
- Review panel
- Group interviews
- Surveys
- Simulations
Step 5: Assess Data

- Data Preparation
- Descriptive Analysis
- Inferential Analysis
**Step 6: Utilize Results**

**Feedback & Reporting**

- Feedback to Participants
- Feedback to Staff
- Formal report
- Executive summary
- Visual summary
- Impact statement (Issue/Response/Results)
- News release
- Oral summary
- Discussion guide
Timeline: Planning to Implementation

1. Create Plan
   - Build internal capacity and begin to address the following:
     - Logic Model
     - Evaluation Plan
       - Questions
       - Sample (population)
       - Measurement(s)
       - Design
       - Analysis
       - Reporting

2. Operationalize Plan
   - Translate plan into operations:
     - Review existing tools
     - Create new tools
     - Investigate on-site logistics and partners
     - Commit the plan to the program calendar
     - Revisit reporting requirements

3. Implement Plan
   - Execute plan in the field:
     - Monitor and improve evaluation
     - Collect outcome data
     - Enter data
     - Analyze
     - Internal Reporting
     - External Reporting

4. Utilize Plan
   - Use results internally and externally:
     - Program improvement
     - Program decisions
     - Dissemination
     - Contribute to evidence base
     - Connect to other programs and evidence

Phase I

Phase II
EP Resources

- Templates and Tools
- Evaluation Network
- The Netway
- Website
- Listserv
- In-Service Training
Online Resources

- Research Methods Knowledge Base
  http://www.socialresearchmethods.net

- Penn State Cooperative Extension (website) Extension and Outreach Program Evaluation:
  http://www.extension.psu.edu/evaluation/

- University of Wisconsin Extension (online course) Enhancing Program Performance through Logic Models:
  http://www1.uwex.edu/ces/lmcourse/

- Cornell Website:
  http://evaluation.cce.cornell.edu
Appendix VI - Stakeholder Mapping

A - List of Stakeholders
Systems Evaluation Worksheet

Evaluand: _____________________________

Make a list of all of the stakeholders.

List of Stakeholders
Evaluand: ______________________________________

List each of the stakeholders from your List of Stakeholders Worksheet in the first column. In the “Motivated By” column to the left, brainstorm individually or in groups the types of things that motivate the stakeholder. In the “Possible Incentives” column, brainstorm individually or in groups the types of incentives that could be used. Leave the “Incentive Chosen” column blank until after you have selected your methods.

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>Motivated By</th>
<th>Possible Incentives</th>
<th>Incentive Chosen</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix VI - Stakeholder Mapping

C - Stakeholders Map Worksheet

Using the list of stakeholders that you generated on the previous page, create a “Stakeholder Map”. Consider levels of scale and how this relates to the stakeholders. Imagine that the Evaluand is at the center of the circle. Place the stakeholders in the surrounding circles based on how “near” or “far” they are from the Evaluand. The further away the stakeholder is placed, the less they interact with the Evaluand. Keep in mind that level of interaction does not necessarily correlate with level of impact. For example, the National Science Foundation (NSF) may be one of the stakeholders on your list. NSF may have minimal involvement with the Evaluand which means that it would be in a circle that is further away from the center (the Evaluand). However, NSF may have the greatest impact on your program - Without their financial support the Evaluand would not exist. Note these points of conflict.
Program Lifecycle Assessment

**Instructions:** Read through each question and think about it in reference to your program. The object of this assessment is to classify your program into the phase and sub-phase that best matches its current level of development. Place a check mark in the box next to the phase that seems to you most appropriate.

**Notes:** Some things to think about when making your choice:
- There is no "right" or "wrong" answer.
- One phase is not "better" than another.
- Many programs are multi-faceted. Some components may be more developed than others. If so, select the earlier phase.

Program __________________________
Organization _______________________
Comments: ___________________________
# Evaluation Lifecycle Assessment

**INSTRUCTIONS:** The object of this assessment is to classify your program's evaluation approach into the phase and sub-phase that best matches its current level of development. Read through each evaluation description and think about it in reference to your program. Place a check mark in the box next to the phase that seems to you most appropriate.

**NOTES:** Some things to think about when making your choice:
- There is no "right" or "wrong" answer.
- One phase is not "better" than another.
- Many programs are multi-faceted. Some components may be more developed than others. If so, select the earlier phase.

## Program:
___________________________

## Organization: ______________________

## Comments:

<table>
<thead>
<tr>
<th>Phase</th>
<th>Process &amp; Response</th>
<th>Evaluation Lifecycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase IA</td>
<td>Process assessment and post-only evaluation of participant reactions and satisfaction.</td>
<td></td>
</tr>
<tr>
<td>Phase IB</td>
<td>Post-only assessment of outcomes, implementation assessment, outcome measurement development and assessment of internal consistency (reliability).</td>
<td></td>
</tr>
<tr>
<td>Phase IIA</td>
<td>Unmatched pretest and posttest of outcomes, qualitative assessment of change, and assessment of reliability and validity of measurement.</td>
<td></td>
</tr>
<tr>
<td>Phase IIIB</td>
<td>Matched pretest and posttest of outcomes. Verify reliability and validity of change. Human subjects review.</td>
<td></td>
</tr>
<tr>
<td>Phase IIIA</td>
<td>Controls and comparisons (control groups, control variables or statistical controls).</td>
<td></td>
</tr>
<tr>
<td>Phase IIIB</td>
<td>Controlled experiments or quasi-experiments (randomized experiment; regression-discontinuity) for assessing the program effectiveness.</td>
<td></td>
</tr>
<tr>
<td>Phase IVA</td>
<td>Multi-site analysis of integrated large data sets over multiple waves of program implementation.</td>
<td></td>
</tr>
<tr>
<td>Phase IVB</td>
<td>Formal assessment across multiple program implementations that enable general assertions about this program in a wide variety of contexts (e.g., meta-analysis).</td>
<td></td>
</tr>
</tbody>
</table>
Appendix VII: Lifecycles

C - Lifecycle Alignment
List each of the activities, short-, medium-, and long-term outcomes, in the appropriate columns. Draw a square around each of the activities and a circle around each of the outcomes. Connect activities and outcomes with lines (see Analysis of Causal Relations Activity).
## Appendix IX: Logic Model Rubric

Feedback Summary for Logic Models
This rubric is intended to provide constructive feedback on the Logic Models.

<table>
<thead>
<tr>
<th>Office and Program Names:</th>
<th>Category</th>
<th>Max. Score</th>
<th>Current Score</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Inputs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Appears to be a complete list</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Includes brief list of key resources,(eg: staff, curriculum, outside partners/facilities)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Specifies % FTEs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Does not mention staff by name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Gives reader at a glance idea of program size/scale</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Activities</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Complete list</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Well bounded (have put thought into the boundary issue; e.g., train the trainer versus direct service)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Defined (eg: one would not have to be intimately familiar with the program in order to understand the Activity titles)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Only includes activities that reach people who participate or who are targeted (does not include administrative or marketing activities carried out by program staff)</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Outputs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Tangibles or by-products of the Activities (e.g., Certificate of Completion, a birdhouse)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Must be measurable (e.g., attendance record could be an Output)</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Should not include effect on a participant (that would be an Outcome)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Short-term Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Should be explicitly related to Activities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Describes short term results and learning logically connected to Activities</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mid-term Outcomes</strong></td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>----</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Should be explicitly related to either Activities or Short-term Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Describes medium term results and action logically connected to either Activities or Short-term Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Long-term Outcomes</strong></th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Describes ultimate impact and conditions logically connected to either Short or Mid-term Outcomes</td>
<td></td>
</tr>
<tr>
<td>• Generally written as occurring beyond the individual or personal level (likely to include social, economic, civic, environmental effects)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Assumptions</strong></th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Beliefs and thinking about the program and how it will occur</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Context</strong></th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Environment in which the program is taking place</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Formatting</strong></th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Fonts and spacing consistent</td>
<td></td>
</tr>
</tbody>
</table>

| **Total** | 100 |

**External Reviewer Comments:**
## Appendix X: Evaluation Plan Rubric

### Feedback Summary for Logic Evaluation Plans

This rubric is intended to provide constructive feedback on the Evaluation Plan.

### Office and Program Names:

<table>
<thead>
<tr>
<th>Category</th>
<th>Max. Score</th>
<th>Current Score</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mission</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Concise, understandable</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Concise, understandable</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Estimated # of participants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Evaluation Questions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Well crafted and measurable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Feasible</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Clearly based on activities and outcomes outlined in the Logic Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sample</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Specific data sources described</td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>• Includes number of participants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• How representative is sample of program participants?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Is sample large enough to reach conclusion?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Existing vs. in development described</td>
<td></td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>• One measure per outcome listed in the evaluation questions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Appropriate for Lifecycle stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Clear description of the measurement type or procedure (measure clearly listed)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Quantitative/qualitative specified</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Referenced where appropriate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Mandate specified where appropriate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Design</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Concise, understandable</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Text description where notation used</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Clear connection between evaluation questions, measures, and design</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Quantitative/qualitative specified</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Appropriate for Lifecycle stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Feasible</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clear connection between evaluation questions, measures, and design</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative/qualitative specified</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate for Lifecycle stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feasible</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reporting</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback loop to staff and stakeholders specified</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concise, understandable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation questions addressed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative/qualitative specified</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All audiences for report listed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency and type of reporting listed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timeline</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific time periods listed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dates provided for specific measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dates provided for data entry and analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Times for reports included</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate and feasible</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formatting</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fonts and spacing consistent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete sentences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**External Reviewer Comments:**
Who Should Use This Guideline?

This guideline should be used by projects that are in Phase I-B. These are projects that meet the following criteria:

- In the initiation phase of development
- Have basic quantitative measures of knowledge, attitudes or behavior
- Have measured participants pre and post program
- Have _not_ matched pre-post scores

It is assumed that the user has a basic familiarity with Microsoft Excel, including how to enter data into columns, how to format a worksheet, and how to do simple formulas.

Introduction

These guidelines describe how Microsoft Excel can be used to accomplish a basic analysis of the pre-post quantitative data for a project where it is assumed the pretest and posttest scores are not matched for participants.

The guidelines describe how to:

- set up Excel for basic statistical analysis
- enter basic data into an Excel spreadsheet
- obtain descriptive statistics
- draw a histogram for each variable
- conduct a statistical test to assess change

Installing the Analysis ToolPak

Before you can do the analyses described here, you have to first install the Analysis ToolPak that comes with Excel. How do you know if you already have it installed? Open Excel, select the Tools menu and see if you have a menu item called Data Analysis. If you do, then the Analysis ToolPak has already been added to your version of Excel. If you do not see the menu item Data Analysis, then you need to add the Analysis ToolPak. Follow these steps:

1. Select the Tools menu, then the Add Ins… item:
2. Check the Analysis ToolPak item and click OK

To verify that you have added the Analysis ToolPak correctly, select the Tools menu again. You should now see a Data Analysis… menu item.
Enter Data Into Spreadsheet

These guidelines only show how to analyze the pretest and posttest data for your project. You may have other data that you would also like to analyze. For instance, you may want to look at descriptive statistics for each response on a measure, explore demographics or your respondents, or do a qualitative analysis of open-ended responses. These tasks are not addressed here. These guidelines simply show you how to analyze your pretest and posttest total scores for your project.

Before entering the data into the spreadsheet we assume that you have already cleaned the data, assigned each respondent an ID number (and written it on their instrument), and computed the total pretest score and posttest score for each respondent. For instance, let’s say you give a 10-item test of knowledge relevant to your program. We assume that you have scored the answers and have obtained a total score for each person (e.g., a score between 0 and 10).

You will only enter data for respondents from whom you have completed data. If a respondent did not complete the test, you will not enter their data into the spreadsheet.

To enter your data into the spreadsheet, do the following:

1. Enter in a label for Pre ID, Pretest, Post ID and Posttest in the first row of the first four columns.
The ID column will be used to enter in the ID number for each respondent. You should make sure this ID number corresponds with the number on their instrument. By doing this, you will assure that you will always be able to track a number in the Excel spreadsheet directly back to the instrument itself.

2. For each respondent, enter in their data. For example, let’s say the first respondent (ID=1) got a score of 7 out of 10 on the pretest. Put their ID number in the Pre ID and their score in the Pretest columns as shown below. In the example, the second respondent got a score of 5 on the pretest. Continue entering all of the pretest scores.
Imagine that you had pretest scores for 20 participants. Your spreadsheet might look like the following:
Now you can enter in the posttest scores in the next two columns. For instance, you might only have 15 posttest respondents (you may have lost a few people over the course of the program). The complete data might look like the following:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pret ID</td>
<td>Pretest</td>
<td>Post ID</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>13</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>14</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>15</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>16</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>17</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>18</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>19</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>20</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
That’s all you need to do to enter the data. If you haven’t already, you should save the data. Go to File → Save and give the Excel spreadsheet a filename.

**Compute Descriptive Statistics**

The first step in analyzing the data is to obtain basic descriptive statistics on both the pretest and posttest. Let’s begin with the Pretest variable first. To get descriptive statistics for the pretest, do the following:

1. Select the Tools menu and the Data Analysis… item. A Data Analysis selection box will pop up. Click on Descriptive Statistics and then click OK.
2. On the Descriptive Statistics pop-up window, we have to first specify the input we want analyzed – the pretest column. To do this click on the little spreadsheet icon just to the right of the Input Range box as shown below.

The Descriptive Statistics window will reduce. You can now click one on the top of the first column (on the letter A) and then hit the enter key. You should once again see the Descriptive Statistics window, but now it will have the Input Range filled in.
Now, fill in the rest of the box as shown above. Check the ‘Labels in First Row’ box, select the ‘New Worksheet Ply’ button and type in the name ‘Pretest Descriptives’ for the worksheet, and click on the Summary Statistics box. When you have done all this, click OK. You will see a new worksheet pop up with your descriptive statistics on it. It will look something like the following:
I’ve made the first column (column A) wider by dragging the vertical divider between column A and B to the right a bit. That way you can see the complete label for each statistic. In this example, we see that for the pretest the Mean (or Average) was 4.75, there were 20 participants (Count = 20) and the data ranged from a Minimum of 2 to a Maximum of 7.

Notice the tabs at the bottom of the screen. The currently selected tab is the one labeled Pretest Descriptives. Where did your data go? They are on the tab labeled Sheet 1. To get back to your data, click on the Sheet 1 tab.

3. Now, let’s do the descriptive analysis for the Posttest scores. Be sure you are on the Sheet 1 tab and can see your data. Also, you may want to save your Excel file again (go to File ➔ Save).

The steps for the posttest are identical to the pretest steps, except that you will specify the posttest column for the input range, and name the worksheet ply ‘Posttest Descriptives.’ Select Tools ➔ Data Analysis… and click on the Descriptives Statistics box and click OK.

After you select the posttest column for the Input Range, and rename the worksheet
ply to say ‘Posttest Descriptives’ your screen should look like the following.

Once you have confirmed that the specifications are correct, hit OK. Your results should look something like the following:
For the posttest, we see that the Mean = 7.8, the range is from a Minimum of 5 to a Maximum of 10 and there were 15 respondents (Count = 15). Again, save your workbook (File → Save). To get back to your data, click on the Sheet 1 tab at the bottom of the workbook.

**Compute Histograms**

The next step in the analysis is to compute a histogram or bar chart for the pretest and posttest variables. The histogram shows the distribution of each variable – the number of respondents who scored within each range of possible scores.

Before we can compute the histogram, we have to set up a table that will tell the program how we would like the bars displayed. If you haven’t already done so, click on the Sheet 1 tab to get back to the sheet that has the raw data on it. We know that both the pretest and posttest in our example could have scores that range from 0 (none correct) to 10 (all ten questions correct). Let’s construct a histogram that has one bar for each possible score. Now, do the following to get a histogram:

1. Create the information for the bars of the histogram. As in the figure below, in column F type the word ‘Score’ into the first row, and then type the numbers 0 through 10 in the next 11 rows.
2. Now, select the Tools menu, Data Analysis... item and select histogram from the pop-up menu and click OK as in the figure below.

![Data Analysis dialog box](image)

3. Now select the input range for the Pretest variable just like you did in the descriptive statistics. Next, select the Bin range – that’s the new column you set up in column F with the possible scores shown. Be sure to check the ‘Labels’ box, select ‘New Worksheet Ply’ and enter in the title ‘Pretest Histogram’, and check the box for ‘Chart Output’. Your screen should look like the one below. When you’ve confirmed
that it does, click OK.

The histogram will show up in a new worksheet and look something like this:

Notice that your score column is shown on the left side of the sheet and that the number of participants who received each score is shown in the second column. The histogram or bar chart is shown next to the table. You can alter the look of the chart to make it easier to read. Just click on the chart and resize it as you would any
graphics object. Here, I've stretched it vertically so we can see the bars more clearly.

4. Repeat steps 2 and 3 above to get the histogram for the posttest. To get back to the data, remember to click on the Sheet 1 tab at the bottom of your screen. We'll use the exact same Bin Range in this histogram so our two graphs will be comparable. Here’s what my histogram specification looks like when I applied it to the posttest variable and named the new worksheet appropriately:
And here’s what the histogram looks like, after I adjusted it to make it a little taller:

Test For Change

The last step in our analysis is to test whether there is a significant change in scores from pretest to posttest. We will use the t-test to accomplish this analysis. The t-test compares the averages of the pretest and posttest to see if they are statistically different from each other. Do the following steps:

1. Select the Tools menu, Data Analysis… item, select the item ‘t-Test Two-Sample Assuming Unequal Variances’ and click OK, as shown in the following figure:
We are using the Unequal Variance version of the t-test because we saw in the descriptive statistics that the variance of the pretest was 1.67 and the variance of the posttest was 2.17. If these two were closer, we could use the Equal Variance version of the t-test.

2. Specify the t-test input. As in the earlier analyses, you have to specify the input range. But this time, we are comparing the two variables, the pretest and posttest, so we have to specify both. For Variable 1 Range, specify the Posttest column. Then for Variable 2 Range, specify the pretest.

For the Hypothesized Mean Difference, enter 0 (this is called the ‘null’ hypothesis – we predict that if the program does not work there will be no difference between the pretest and posttest). Click on the Labels box, select New Worksheet Ply and enter the name ‘t-test’.

When you click OK, you will get the following results:
I made column A wider so that we can read the labels for each result. The results show that the pretest average was 4.75 while the posttest average was 7.8. The t-value is 6.38 with a df=28. This t-value is much higher than the critical value of 2.04 for a .05 level of significance (what we specified for Alpha in the t-test screen above). We would interpret this to indicate that there was a statistically significant increase in the means from pretest to posttest in our project.

Using the Results

The results that you have generated can easily be copied and pasted into a word processing program, can be edited as needed, and can be incorporated into reports.