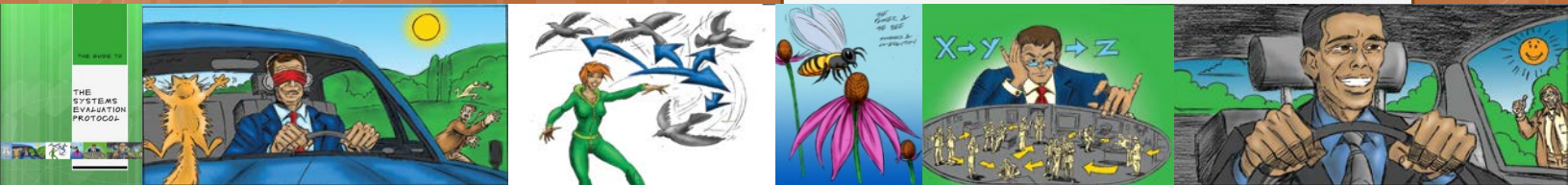WORKBOOK FOR

# THE SYSTEMS EVALUATION PROTOCOL

## PHASE I: PLANNING

V 1.1

# Workbook For The Systems Evaluation Protocol

## Phase I: Evaluation Planning

Monica J. Hargraves and Jane C. Buckley, Editors

*Cornell Office for Research on Evaluation (CORE)*

*Cornell University*

# TABLE OF CONTENTS

# Preface

The materials offered in this workbook were originally developed for STEM program staff engaged in an Evaluation Partnership with the Cornell Office for Research on Evaluation (CORE) between 2006 and 2014. The Evaluation Partnerships were designed to build evaluation capacity among participants and to support evaluation planning and implementation for selected education and outreach programs.

These Evaluation Partnerships provided a context for refining and testing CORE's Systems Evaluation Protocol (SEP) and the supporting web-based software called the Netway. The SEP uses a systems approach to evaluation and takes evaluators or program staff through a series of steps to develop an evaluation uniquely tailored to that program. The National Science Foundation supported CORE's research as part of its effort to develop evaluation strategies and infrastructure that would strengthen the development and evolution of STEM education programs (NSF grants #0535492 and #0814364). Cornell Cooperative Extension (CCE) provided additional funding as part of its effort to build evaluation capacity in CCE.

Participants in CORE's Evaluation Partnerships have largely been educators and program managers running non-formal education programs in a wide array of STEM topic areas. The programs involved in our research and development effort came from two systems: Cornell Cooperative Extension (CCE) in New York State and a nation-wide network of university-based Materials Research Science and Engineering Centers (MRSECs). These contexts are reflected in the examples used in some of the materials here. Nevertheless is bears emphasizing that the SEP steps and materials are relevant for any program, not just those in the STEM arena. It is our expectation that users of this Workbook will adapt our materials, when necessary, to fit the culture and context of the program(s) at hand.

These materials are publically available as web content and individual pdf's online via the Netway (www.evaluationnetway.com). The Netway is an online software system which provides tools for completing the SEP steps (stakeholder mapping, logic modeling, pathway modeling, evaluation plan development, etc.), and contains a database of program models and evaluation plans developed by Netway users, as well as tutorials and written support materials in addition to those printed here.

We would like to thank our colleagues at CORE who reviewed, developed, or co-developed these materials with and for us, particularly: Tom Archibald, Claire Hebbard, Margaret Johnson, Cathleen Kane, and Jennifer Urban. We owe particular thanks to CORE's Director and the originator of the Systems Evaluation Protocol and the Netway, Professor William Trochim. In addition we would like to thank the many colleagues in the CCE and MRSEC systems who entered into Evaluation Partnerships with CORE. Their work, insights, feedback, and willingness to explore the frontiers of systems evaluation with us have contributed significantly to the development and refinement of our approach to evaluation and these materials.

Although this workbook was developed for the Systems Evaluation Protocol, the content offers instruction and support for anyone engaged in the evaluation of programs or other initiatives intended to achieve particular outcomes, using a wide range of approaches to evaluation. We wish you well in your evaluation and program development efforts.

Monica J. Hargraves and Jane C. Buckley

Editors

# I. Introduction
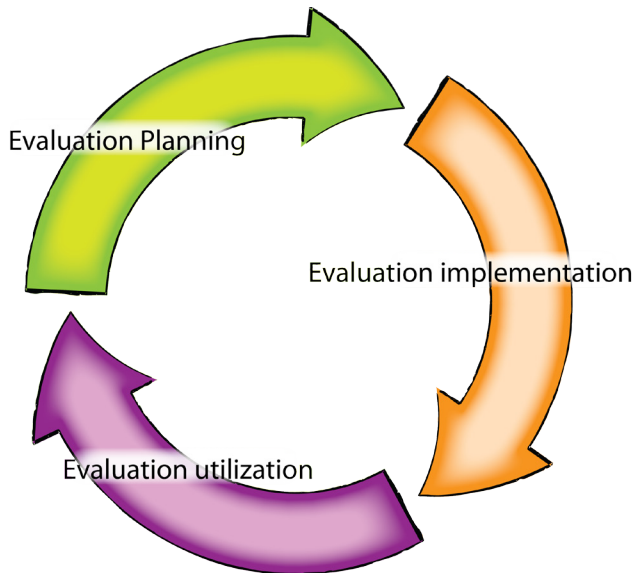
There are three phases of evaluation :

Phase I Evaluation Planning

Phase II Evaluation Implementation

Phase III Evaluation Utilization.

This workbook focuses on the first phase: Evaluation Planning.

The Planning Phase of evaluation involves more than just selecting a measurement tool and scheduling data collection. It includes the pre-thinking and analysis that are needed in order to ensure credible, accurate, useful, and feasible evaluation.  The Planning Phase of the Systems Evaluation Protocol (SEP) therefore includes three stages:

**Preparation** - Establish a working group, assess organizational resources, and identify issues that affect the program and its evaluation.

**Modeling** -  Analyze critical aspects of the program (definition, boundaries, stakeholders, history) and develop a structured model of what the program does, what it is thought to achieve, and how change is believed to occur.

**Evaluation Plan Development** - Using the program model and related analyses as a foundation, identify evaluation priorities (evaluation questions and purpose statement) and develop appropriate strategies for sampling, data collection, analysis, and work planning.

The "At a Glance…" section on the next page lists all the steps within these stages of the Planning Phase of the Systems Evaluation Protocol.

At the end of the Planning Phase, you will have a written document that: reflects careful, strategic thinking; will guide the practical implementation of the evaluation; and ensures that all the individuals involved in conducting the evaluation have a shared understanding of the goals and reasoning behind the various steps. You will also be prepared to incorporate the results that emerge from this evaluation into subsequent evaluation planning, so that evaluation continues to guide decisions about the program's evolution.

This Workbook follows the steps of the SEP, focusing on the Modeling and Evaluation Plan Development stages, and offers practical support material to assist evaluators in understanding and completing the Protocol steps. As such, it is a companion to the Guide to the Systems Evaluation Protocol. [1]

[1] Trochim, W., Urban, J.B., Hargraves, M., Hebbard, C., Buckley, J., Archibald, T., Johnson, M., and Burgermaster, M. (2012) The Guide to the Systems Evaluation Protocol (V2.2). Ithaca, NY.

# At a Glance:
## The Systems Evaluation Protocol Phase I: Evaluation Planning

### Stage 1 - Preparation

The Preparation stage is intended to acquaint the participants with the Systems Evaluation Protocol (SEP) and the evaluation project arrangements, and identify current evaluation resources. The Preparation stage involves the following steps:

- **Enter the System:** Connect with key decision makers of the organization to discuss commitment to evaluation planning and the evaluation project.

- **Develop Memorandum of Understanding:** Work with key decision makers in the organization to create a written document that: describes the roles and responsibilities of participants in the evaluation project; details the expectations for the evaluation consulting team, partner site administrators and staff; and provides a timeline of project activities and completion.

- **Identify Internal Stakeholders:** Identify people in the program/organization who should be involved or consulted in evaluation planning.

- **Identify the Working Group:** Identify those who will play a key role in developing the evaluation plan.

- **Assess Evaluation Capacity:** Identify the resources available within the organization and within the program - the degree of evaluation training the staff has already received; information technology resources; and evaluation policies.

### Stage 2 - Modeling

The Modeling stage is intended to enhance participant knowledge of evaluation concepts, and identify how their program "works". The Modeling stage involves the following steps:

- **Stakeholder Analysis:** Determine all of the potential people and/or organizations that may have a stake in the program.

- **Program Review:** Gain a firm understanding of the components and characteristics of the program including how it operates and whom it serves.

- **Program Boundary Analysis:** Determine the conceptual limits of the program; what is "in" and what is "out" when defining the program.

- **Lifecycle Analysis:** Determine where the program and its evaluation are in their respective lifecycle or evolutionary phases; assess implications for this evaluation cycle.

- **Logic Model:** Generate an initial logic model including the assumptions, context, inputs, activities, outputs, short-, mid-, and long-term outcomes.

- **Pathway Model:** Develop a pathway model incorporating logic model elements and the program's theory of change to illustrate linkages between program activities and outcomes.

- **Evaluation Scope:** Determine the specific components of the pathway model that will be the focus in the upcoming evaluation cycle.

- **Program-System Links:** Introduce tools and strategies for finding similar programs and shared outcomes; develop research support by drawing on literature and on resources in the systems within

which the program exists.

• **Reflection and Synthesis:** Finalize the logic and pathway models by assessing the models from the perspectives of key stakeholders, reviewing the Program Boundary Analysis, reviewing the Program and Evaluation Lifecycle Analyses, and revising the models as needed. This step also involves integrating relevant research literature as it relates to the causal pathways that have been articulated in the Pathway Model.

## Stage 3: Evaluation Plan Creation

The third stage, "Evaluation Plan Creation," focuses on the creation of an evaluation plan that will guide the implementation of the evaluation. The Evaluation Plan Creation stage involves the following steps:

• **Introduce the concept of an Evaluation Plan:** Present and discuss the components of an evaluation plan.

• **Evaluation Purpose:** Develop a summary of the evaluation's focus and goals.

• **Evaluation Questions:** Develop evaluation questions based on the logic and pathway models, lifecycle analysis, stakeholder analysis, and systems insights. The evaluation questions will function as the core determinants of all the evaluation plan components.

• **Identify or Develop Measures:** Identify measures already being used in evaluating the program and assess them for quality and feasibility; identify other existing measures that might fit the program evaluation needs; and/or develop any new measures that are needed.

• **Develop Sampling Plan:** Define the population of interest, sampling frame and sample and describe the source(s) of the evaluation data.

• **Develop Evaluation Design:** Describe how the samples, interventions, and measures will be coordinated over time.

• **Develop Data Management and Analysis Plan:** Describe how data will be managed as it comes in and how it will be stored. Articulate the plan for analyzing the evaluation data. Include information on how data will be managed.

• **Develop Evaluation Reporting and Utilization Plan:** Detail the plan for utilizing and reporting the results of the evaluation to key stakeholders identified earlier.

• **Develop Implementation Plan and Timeline:** Develop the schedule for the evaluation and key implementation milestones.

# Stakeholder Analysis

Stakeholder analysis is the process of identifying and describing the perspectives of all of the potential people and/or organizations that have a stake in the program and its evaluation.

## FAQs

### How will my stakeholder analysis affect my evaluation plan?

Identifying and considering stakeholders is important throughout almost every step of the evaluation planning process. Considering the perspective of diverse stakeholders will help you with program definition and boundary issues, with the creation or revision of logic and pathway models, with the selection of the evaluation scope, and with the articulation of the evaluation's purpose. Explicitly or implicitly, stakeholders will have different ideas about what types of evaluation questions, measurement strategies, evaluation designs, and analysis and reporting plans are credible or useful. For example, a funder might be enamored with experimental or quasi-experimental designs, while program participants are interested in telling their stories, and program implementers want to know if participants enjoyed the program and learned something. Thus, by taking all of these (divergent) perspectives into account, your evaluation plan will more likely be well received by key stakeholders. Going further to actually involve some of them in the evaluation process makes it more likely that the evaluation results will eventually be used; if stakeholders have had no input in and relation to the evaluation plan, they may not see the results as relevant, and might simply ignore them.

### How should I go about brainstorming stakeholders?

When brainstorming stakeholders, think about all of the major groups that have or could have an interest in your program. These may include: participants, implementers, funders, partner organizations, community members, tax payers, etc. Once you have brainstormed these larger categories, it may be easier to identify individuals or sub-groups.

## Guiding Documents

- Stakeholder Worksheet: Blank Map Template
- Stakeholder Perspectives Worksheet
- Stakeholder Interview Guide

# *Stakeholder Worksheet: Blank Map Template*

**Goal**: Brainstorm as many stakeholders for your program as you can. Indicate their proximity to the program using the concentric circle diagram below.

**Program name:**

Program

Netway (www.evaluationnetway.com)
Document available at: http://www.evaluationnetway.com/guide/evaluation-guidance/stakeholder-analysis
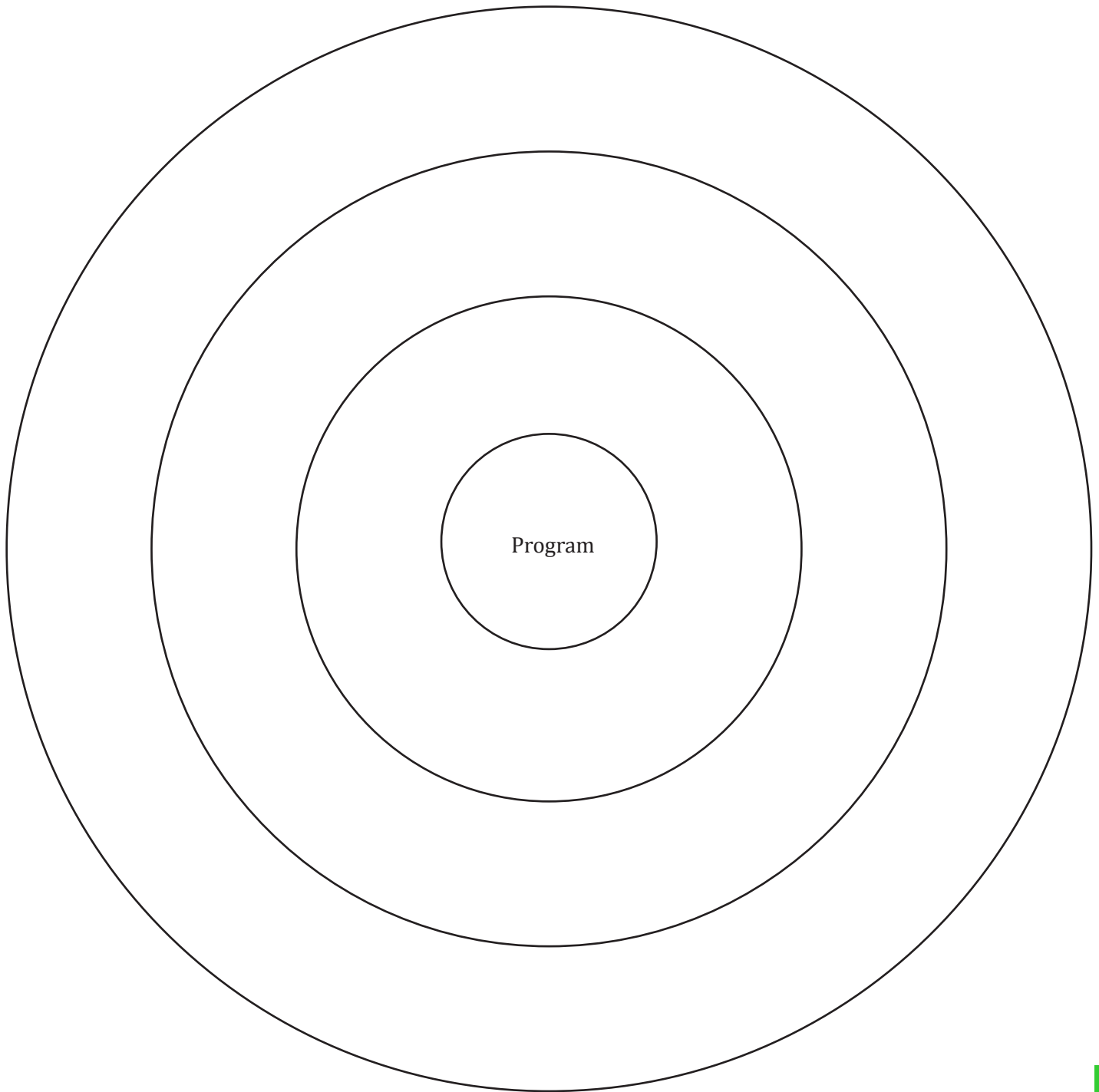
# *Stakeholder Perspectives Worksheet*

Goal: articulate the different ways your stakeholders might view the program (use these insights to enhance your program description).

| **Stakeholder** (start with the "close in" ones, and move "out"): | **Phrases this stakeholder might use to describe your program** (what would they care most about? What would jump out at them or make them care about it?): |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |

**To Do Later:** Revisit the stakeholder questions, ideally in the form of group discussions with various stakeholders included (colleagues, participants, volunteers, collaborators, etc.) Add to the insights about different perspectives in the right-hand column above, and take note of which parts of the program various stakeholders are primarily interested in.

# *Stakeholder Interview Guide*

**Purpose**: to gather insights and perspective from key stakeholders of the program being evaluated, in order to get a more full view of :

- what the program is or could be

- what outcomes or impact it does or could have

- relevant aspects of the context in which the program operates

- (possibly) key assumptions that should be articulated in the program model

- stakeholder priorities that could affect program development and evaluation

**Use**: This input will be used in several ways:

- To sharpen and enrich the Program Description

- To create or improve the Program Models

- As guidance for focusing the evaluation later

**Steps**:

1. Identify 2-4 key stakeholders, particularly those who are likely to have perspectives different from those of program staff. Any of the following might be useful:

   - Participants

   - Staff who've worked on this program or in this area before

   - Funders

   - Community leaders

   - County legislators

   - Faculty

   - Collaborators

   - Industry representatives, suppliers, end-users, etc.

2. Contact your selected stakeholders to explain what you are interested in, why you selected them, why you value their input and how it will be used, and how you will be capturing the information (if you are planning to record the conversation, get permission to do so). Set up time for a brief interview.

3. Review the candidate questions on the next page and select the ones you will use, adapting as needed for this particular stakeholder, your relationship with him/her, etc. Keep a record of the questions you ask

4. Prepare for how you will capture their comments (if you intend to quote them, ask permission).

Netway (www.evaluationnetway.com)
Document available at: http://www.evaluationnetway.com/guide/evaluation-guidance/stakeholder-analysis .

# *Candidate Stakeholder Interview Questions*

This list is suggestive, but not exhaustive. Some of these are more appropriate for programs that are just being developed; others are more relevant for programs that have already been operating. Use what fits.

1. How would you describe this program?

2. How do you perceive your relationship to this program?

3. Who do you think are the primary beneficiaries?

4. What do you think motivates/will motivate participants to do the program?

5. How do participants get involved?

6. How does this program work/how would you imagine it working?

7. What do you see as the main value or contribution of this program?

8. What are the desired outcomes for participants?

9. What other outcomes do you see or foresee?

10. What do you think are the most important outcomes?

11. What are key elements of context that influence success or failure of the program? (community or industry needs or challenges, strengths or vulnerabilities, relevant history, political or regulatory issues, ...)

12. Are there similar programs being offered elsewhere that we should be aware of?

13. Are there other programs or organizations working on this in this area? What is this program's niche?

14. Looking ahead to evaluation, what do you think will be the most important questions to explore?

# Program Definition and Boundaries

Program definition and boundary analysis involve defining the structure and elements of a program; specifically, what is considered to be part of the program as opposed to activities or elements that may be just outside the definition of the program.

## FAQs

### Should I include administrative activities in my program model?

These background activities - like recruitment, program marketing, training of instructors, or fundraising - are essential for program implementation. However the question of whether they should be included in your program model is complicated. In reality, all programs have some kind of admin effort in the background. The question here is, if you had 1 minute to describe to someone the essence of your program and how it works, would these be among the activities you would talk about? Most of the time the answer is likely to be "no". There are exceptions though, which is why there is not a hard and fast rule about this. Sometimes an activity that sounds administrative is actually directly related to achieving certain program outcomes. For example, marketing and participant recruitment are often routine (though important!) efforts. But for a program whose success depends on participation by a very specific demographic mix of participants, perhaps including hard-to-reach individuals, the outreach and recruitment effort becomes central to the program in a distinctive way.

### How do I determine what my program boundaries should be?

Determining a program's boundaries is not as simple as it may seem. Consider the example of public school as a program. You may think that everyone interprets "school" the same way. But if you look closely you will find that, like all programs, school shares a fuzzy boundary with things that could be defined as "just outside" the boundary. For example: after-school programs, riding the school bus, teachers working on lesson plans outside of school hours, and school food service are all arguably part of "school." Depending on your perspective on a program and its goals, these things may be included or not. It is important to have a rich discussion with your colleagues about how to define the boundaries of a program for the purpose of the program model. There truly is no single "right" answer. However, you may be able to agree upon a set of simple rules for deciding what is "in" and what is "out." Perhaps you will agree to include only the activities that you have control over, and/or that directly relate to the outcomes you want to include. For example, if you want to explore outcomes relating to or affected by parental involvement, then your boundary on school would include activities beyond what happens in the classroom. Whatever you decide about the boundary at this stage, it is likely to be a question that comes up again as you get deeper into the process of program modeling and evaluation planning.

### How long should a program description be?

A program description should be relatively brief, though long enough to cover the details relevant to key stakeholders (often 1-3 paragraphs is about right). Try to imagine describing your program, completely and accurately, to someone riding next to you on an elevator...and you're not in the Empire State Building.

### What should be included in a program description?

A program description should provide a very brief overview of the program, so that a reader would understand what the program entails (main activities, context), who participates in it (target audiences, actual audiences), why someone might want to take it (intended outcomes), and where the program "comes from" (history, context). It is important that the program description match the program models in terms of the activities and outcomes included and those that are not (program boundaries). In addition to what is included in the program model, the program description provides some supplemental information about the program's target participants, the scale of the program, and its history and context.

# Guiding Documents

- Guidance for Boundary Analysis
- Program Boundary Worksheet
- Guidance for Program Descriptions

# *Guidance for Boundary Analysis*

*Some questions to consider:*

1.  What is the essence of this program? (activities and outcomes) What is not?

2.  Who are the participants? Who are not?

3.  How is the program being delivered, and is that an essential part of the program? (For example, is this an After-School Science education program, or a Science education program that at the moment happens to be offered in an after-school setting but might be offered in different settings equally well? Is this a program that could be delivered in lecture format, or is the "hands-on learning" a key aspect?)

4.  How much of the possible span of activities have to be included in order to have a full view of this program? (For example, is it important for this program to include things that might be considered prep work, recruitment, training, etc., or can those be assumed or taken as given, or inherently separate from the working of your program?)

5.  How much of the possible span of outcomes have to be included in order to get a full view of this program? (For example, would you ever want to make statements about your program's relevance to community well-being, or national resource issues? Having those within the vision of your program does NOT mean you will be evaluating those kinds of outcomes, but if they are an important part of what shapes or motivates the work of the program then the connections matter.)

**Tip**: Think about who (individuals, or broad stakeholder groups) might have different views on these questions, or who might raise different questions.

**Tip**: Try to be broad enough to make this a complete picture of your program, and yet accurate enough so that the pieces in there are essential to the view of the program. Keep in mind some practical considerations:

*   What kinds of statements or claims will you want to be able to say at some point? (We do NOT expect or want you to evaluate everything within your program. But we do want your program view and your program model to be comprehensive enough so that you can make good choices within the program boundaries about what to evaluate.)

*   Who are your audiences for program discussions? (Are key stakeholders interested in a broad and extensive view of your program, or are they focused on what might be considered sub-pieces?)

**Tip**: Things that seem relevant to your program but in the end do not "belong" inside the boundaries of your program definition may have a place in the Assumptions or Context sections of your Logic Model. That is, if they help paint a picture of your program and offer important information about what makes it work, then there is still a place for this information even if it isn't central enough to merit being in the heart of the program model.

# *Program Boundary Worksheet*

**Program Name:** _____

**Goal**: to establish program boundaries by articulating specific elements within broad sub-parts (if any) that make up the program, and start to establish elements for the Logic Model.

**Instructions**: Identify the sub-parts (if any) of the program and organize the more specific activities, participants groups, and anticipated outcomes into these parts. For example, some programs may depend on volunteers to implement activities with participants. In this case, program staff may think of the program as have two main sub-parts: volunteer activities and outcomes; and participant activities and outcomes.

| Program sub-part: | | |
| --- | --- | --- |
| | **With Whom?**<br><br>(participants, ultimate target audiences) | **What?**<br><br>(activities or projects, using short titles) | **Why?**<br><br>(what will these activities achieve?) |
| IN | | | |
| OUT | | | |

# *Guidance for Program Descriptions*

*Program Descriptions, like Logic Models, are basically communication devices. Logic Models provide more detail than Program Descriptions, and provide more information about the implicit program theory. Program Descriptions on the other hand are more "readable" for most audiences, and can serve a range of purposes (marketing, website information, etc.) The following suggestions are not intended to define a Program Description rigidly, but are meant to offer general guidance and suggestions that can be applied to and adapted for individual programs.*

**Generally**, *a program description should...*

- Be in prose (not bullets, at least not exclusively).
- Be what you consider complete (it should stand alone) but not be excessively long.
- Convey a true sense of the program (staff and participants would recognize it).
- Take into account the perspectives articulated in stakeholder analysis.
- Align with boundaries set forth in program boundary analysis.
- Be clear, readable and understandable to an outside audience (does not contain jargon or references to things internal to the organization).

*A program description should cover at least some of the items in the following groups, according to what is relevant for this particular program:*

1) **Program information,** such as:

- program capacity (average or expected N=?)
- main program activities (including general content and format: lecture, workshop, hands-on)
- who delivers program and their training (staff, volunteers, etc.)
- average program duration
- target population
- participant selection criteria (if applicable)

2) **Context**, such as:

- local community and programmatic context of program
- state and/or federal programmatic context of program setting (where and when program takes place)
- funding sources
- history of the program lifecycle stage of the program

3) **Goals**, including possibly:

- main outputs or "take-aways" for participants
- how activities are thought to lead to hoped-for outcomes
- the theoretical framework behind program approach (if there is one)

# Program Mission

The program mission statement is a description of the overarching or long-term goals specific to the program being described.

## FAQs

### What is the difference between a program's mission statement and description?

A program mission statement is a very short description of the ultimate goals of the program (help prepare students for college, increase energy efficiency of farms through education, etc.), while a program description is a concise yet complete description of the program itself.

### What should be included in a program mission statement?

A program mission statement is a statement about the major goals of the program. It should be specific to the program, but at the "big-picture" level. A program mission statement should be only a few sentences.

# Lifecycle Analysis

Lifecycle analysis refers to the process of characterizing the history and current phase of development of a program and its evaluation.

The program lifecycle is the individual course a program takes as it evolves, changes, or remains the same over time. Typically, programs earlier in the lifecycle are smaller in scale and more variable in their implementation. Later phase program often, though not always, become more stable and reach a wider audience. Understanding a program's lifecycle history is a critical part of planning for its evaluation.

The evaluation lifecycle exists in parallel with the program lifecycle. Early lifecycle evaluations are characterized by rapid feedback, basic designs, and less of a focus on external validity and reliability. Later lifecycle evaluations are typically more involved, build upon prior knowledge of the program, and allow for broader claims.

## FAQs

### What if my program and evaluation lifecycles are not aligned?

Misalignment is a very common problem. To address it well, it is important to understand more deeply what the consequences of misalignment are and how to communicate them to stakeholders. Alignment between program and evaluation lifecycle phases helps ensure that programs obtain the kind of information that is most needed at any given program lifecycle phase, and that program and evaluation resources are used efficiently.

New programs are still changing a great deal, and need basic rapid feedback about program process, satisfaction, etc. that can be incorporated into the next round of implementation. Focusing instead on longer-term outcome evaluation, for example, would use additional resources and introduces a risk of bad decisions: the outcome evaluation might happen to yield favorable results, but since the program is still changing considerably this seemingly favorable outcome might not hold up in subsequent rounds of the program, and could lead to an over-investment in something that has not yet stabilized. The opposite risk is also significant: early outcome evaluations might show poor results and lead to the premature cancellation of a program that actually has great promise but needs to have some basic weaknesses resolved.

On the other hand, for mature programs that are consistently-presented and well-received, additional and exclusive focus on basic feedback about program process or satisfaction would not serve the program well. These programs typically need evidence about effectiveness in order to make decisions about whether to re-commit or even expand the resources being devoted to it. Regardless of participant satisfaction and program stability, the program might or might not be achieving its intended outcomes. Without appropriate evaluation, program resources will not be allocated as well as they could be.

In cases of misalignment, you should work towards alignment as you develop this and future evaluation plans. How that should occur, and how long it will take, will depend on the particular program, reasons for misalignment, and stakeholder priorities.

### How does my program's evaluation lifecycle affect my current evaluation plan?

Knowing the evaluation lifecycle amounts to having an understanding of the "state of knowledge" of your program. This state of knowledge is an essential determinant of what kind of new knowledge the evaluation should attempt to build, in order to help the program evolve well.

For example, if there has been little or no evaluation done in the past on this program (it's in an early

evaluation lifecycle stage) then it will probably be best to begin building knowledge by conducting process-oriented and exploratory evaluations. If there have been years-worth of satisfaction surveys collected from participants, yet no evaluation of the association with desired outcomes, then it would make sense to expand the state of knowledge about the program by planning an evaluation that examines a few key outcomes.

### What should I do if my program doesn't fit well into any of the lifecycle phase definitions?

In practice, the boundary between one phase and the next can be fuzzy and difficult to pin down.  The value of this analysis is less about selecting the "right" box, and more about being thoughtful about where your program is in its evolution and where it needs to "go" next. Our general rule of thumb is to choose the lower phase if you really can't decide where your program belongs. We also recognize that programs have many parts, and sometimes some parts of a program are stable and well-established, while others may be quite new and untried.  It is difficult to assign a single lifecycle phase to the whole program in this case, so think in terms of assigning lifecycle phases to the individual parts.

### Can my program have an early lifecycle phase even if it has been around for a long time?

Yes it can. Program lifecycle phase is not just a matter of the passage of time. The definitions we use have to do with how much a program is changing from one round to the next.  So a program that has been around for 30 years but is currently undergoing some big revisions in how it gets delivered or what it covers would be considered an early phase program – it's in that new phase where you are shifting what works, trying new things, getting the "bugs" out.  Similarly, a program that has been around for many years but is "always changing" would be considered a relatively early phase program. This might be the case with a program whose name remains the same, but whose delivery method switches significantly on an on-going basis – this might be a very adaptive program that is constantly changing in response to audience needs or desires. It might be adapting in sensible ways, but by its nature it isn't settled enough to be considered stable and standardized.  Keep in mind that there is nothing inherently good or bad about being in one lifecycle phase or another!

### I don't know my program's history. How can I figure out what phase it's in?

It's useful to know a program's history in order to "see" where it is in its evolution, but you don't have to know the history to figure out its current lifecycle phase. The definitions of program lifecycle phases are based on how much they are (currently) changing and being adjusted from one time to the next. The changes may be in terms of the scope of what the program includes, the kinds of audiences that are being reached or targeted, the way it is being delivered (formats, settings), and so on. The magnitude of the changes is a factor in the program lifecycle phase – we distinguish between big changes and substantial revisions, versus smaller fine-tuning changes that tend to be used as the program is converging toward a steady form. We also distinguish between changes that cover large parts of the program (earlier phase), and changes that are focused in on smaller segments (later phase).

## Guiding Documents

- Lifecycle Definitions and Alignment

- Importance of Lifecycle Alignment

- Program History and Lifecycle Worksheet

- Evaluation History and Lifecycle Worksheet

- Lifecycle Identification and Alignment Worksheet

# Lifecycle Definitions and Alignment

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

# *Importance of Lifecycle Alignment*

Misalignment between program lifecycle stage and evaluation lifecycle stage is all too common, in practice. The purpose of this document is to explain and illustrate the costs and the risks of misalignment. This is intended to motivate an effort to work toward alignment when possible, and to support well-informed communication with stakeholders who may be pushing for a non-aligned evaluation plan. In addition, since misalignment may persist, this information is useful for clarifying how to be cautious in interpreting evaluation results.

Waste of Resources:

Being out of alignment – in either direction – amounts to a waste of resources. Evaluation should help programs evolve to their next stage. If the evaluation is misaligned, the program will not get the information it "needs".



### Program Phase > Evaluation Phase

Simple post-event satisfaction surveys are not really "enough" for ensuring good decisions about a long-established, consistently-implemented, and possibly large program.



### Evaluation Phase > Program Phase

Sophisticated evaluation strategies (perhaps with control groups and randomization) are more costly and take more time, and are not appropriate for evaluating programs that are newly developed, and still evolving rapidly.

Netway (www.evaluationnetway.com)
Document available at: http://www.evaluationnetway.com/guide/evaluation-guidance/lifecycle-analysis.

# *Risk of Bad Decisions:*

Using advanced outcome evaluation strategies on a program that's still in an early lifecycle phase (evolving rapidly and not yet stabilized in implementation) increases the likelihood risk of bad decisions. The program is inherently changeable, which introduces more random variability into evaluation results. In statistical terms, these are the risks of Type I and Type II errors (Type I = accepting something that's false; Type II = rejecting something that's true).



## Type I Error

Making a decision to significantly expand a brand new program based on favorable results from an initial outcome evaluation would be like basing a large construction investment on a beach erosion study that might have been done on what happened to be a particularly calm sunny day. The program is still evolving rapidly, and much more information is needed in order to make sound investment decisions.



## Type II Error

The opposite risk is important too: Abandoning a newly-developed program that had weak results on an initial outcome evaluation might eliminate a program that actually has a lot of potential but has not yet had a chance to develop and have the bugs worked out. The program is still evolving rapidly, and much more information is needed in order to ensure that you are not "throwing the baby out with the bathwater".

26

# *Program History and Lifecycle Worksheet*

**Name of Program** _____ **Date** _____

CHARACTERIZING A PROGRAM'S EVOLUTION
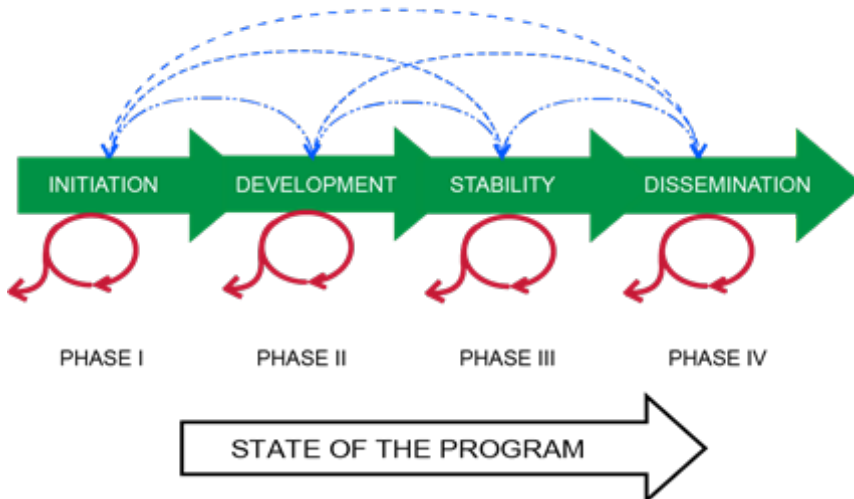
This graphic offers a way of characterizing a program's evolution. The "State of the Program" arrow emphasizes that it is not just the passage of time that marks a program's evolution. The blue dashed lines indicate that there is often iterative back-and-forth movement through the phases. The red circles and arrows remind us that programs may also stay in one phase or move incrementally within it for some time, and they may be retired at any point.

## The evolution of your program

On the graphic below, **sketch the path your program has followed, and mark where it is at the moment.** This is not meant to be a formal graph or time plot, just a schematic representation!

INITIATION    DEVELOPMENT    STABILITY    DISSEMINATION

Now, **describe briefly how your program has evolved so far, and where is it now.** (Bullet format is fine.) Think of yourself as a biographer working to tell the life story of your program. As you recall your program's history, specific considerations might include: how and why the program began; how it has changed over time and why; and the degree to which it continues to change.

- Where did your program "come from"? Has it evolved and changed? How?

- What do you know about your program (how it works, and what it does)?

- How do you know it? How certain are you?

28

# *Evaluation History and Lifecycle Worksheet*

**Name of Program**_____ **Date** _____

This worksheet is designed to help you reflect on the evaluation that has been done on this program over time, on what you know about the program and how you know it, and on the factors that have shaped previous evaluation decisions.

The graphic below provides an image for evaluation lifecycle phases, analogous to the program lifecycle phases in the Program History and Lifecycle Worksheet. The "State of the Evaluation" is a synthesis of the multiple dimensions of a program evaluation. Movements from left to right in this figure correspond to potential increases in the scope and/or intensity of the evaluation effort.
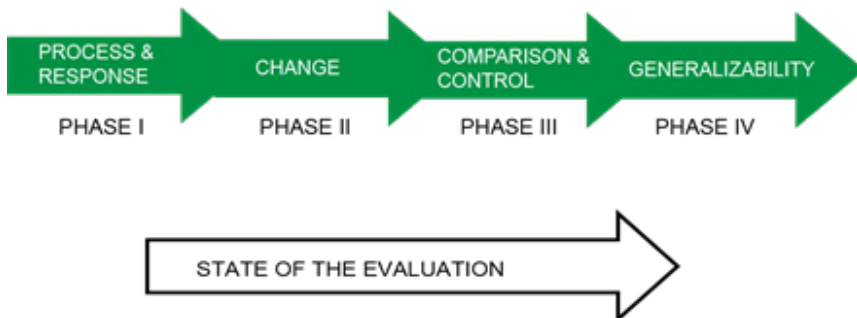


## The Evolution of your Evaluation

On the graphic below, **sketch the path your past evaluations have followed, and mark where the current state of evaluative knowledge on the program is at the moment.** This is not meant to be a formal graph or time plot, just a schematic representation!

Netway (www.evaluationnetway.com)
Document available at  http://www.evaluationnetway.com/guide/evaluation-guidance/lifecycle-analysis.

Now, **briefly describe how your program's evaluations have evolved so far, and where the current state of evaluative knowledge on the program is at the moment.** (Bullet format is fine.)

- What types of evaluation have been done on this program, and when? (Include both formal and informal evaluations and feedback.)

- What have been the primary purposes of the evaluations that have been done?

- How have results been used?  Have the results influenced decisions about the program?

- Who have the results been shared with or reported to?

- Does your funder require a certain type of evaluation?

- What have been the sources of information that have been used? (Documents, responses from individuals on surveys or in interviews, etc.)

- Have there been evaluations that are descriptive of the program (e.g. interviews with program participants or leaders)?

- Have there been formal evaluations of the program that included structured data collection (qualitative or quantitative data)?

- Have there been evaluations that used comparison groups or control groups?

# *Lifecycle Identification and Alignment Worksheet*

Program Name: _____     Date: _____

*To complete this worksheet, please refer to definitions of Lifecycle phases as given in the* **Lifecycle Definitions and Alignment** *handout.*

## I. Program Lifecycle Identification

The goal of this activity is to use your reflections about your program's history to select one of eight program lifecycle phases.

In practice, programs may be made up of components that are in different lifecycle stages. For example a program might combine an established curriculum or activities with some new initiatives, so that some parts of the program are well-understood and well-developed, while others are in a more pilot phase. If your program has components in different stages, pick the most representative phase overall, for now, but make note of the differences in program components.

What lifecycle phase is your program in? (IA, IIB, etc.) _____

Notes:

## II. Evaluation Lifecycle Identification

Now, use your reflections on the history of your evaluations of the program, again referring to the **Lifecycle Definitions and Alignment** handout, and select the evaluation lifecycle stage that best describes the current state of the evaluation. If your program has components in different evaluation lifecycle stages, pick the most representative phase overall, but make notes of and address the differences in your evaluation plan.

What lifecycle phase is your program's evaluation in? (IA, IIB, etc.) _____

Notes:

Netway (www.evaluationnetway.com)
Document available at: http://www.evaluationnetway.com/guide/evaluation-guidance/lifecycle-analysis

## III. Program and Evaluation Lifecycle Alignment

Alignment between program and evaluation lifecycle phases is essential for ensuring that programs obtain the kind of information that is most needed at any given program lifecycle phase, and that program and evaluation resources are used efficiently.

The figure bellow offers a simple representation of what alignment and non-alignment of program and evaluation lifecycle looks like. Program C in this illustration lies on a 45-degree line indicating ideal alignment between its program and evaluation lifecycle phases.
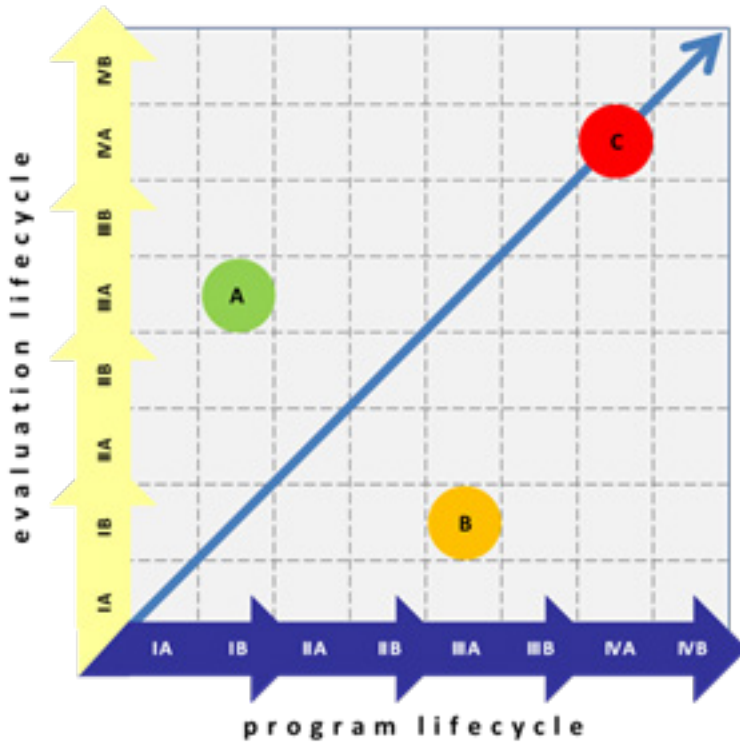
Some points to consider about alignment:

• In practice it is very common to have programs whose program and evaluation lifecycles are not aligned.

• Moving toward alignment should be treated as a key goal of evaluation planning.

• Alignment does not necessarily happen after one evaluation cycle.

• Evaluations and programs are developmental and grow over time.

Now, on the figure to the right, plot the point that shows your program lifecycle and evaluation lifecycle. If it is on the blue 45 degree line, your lifecycles are aligned. If not, use this chart and your program and evaluation histories to reflect on why your lifecycles are out of alignment.

If your lifecycles are misaligned, **take notes here** on why that might be:

*Lifecycle Identification and Alignment Worksheet page 2 of 2*
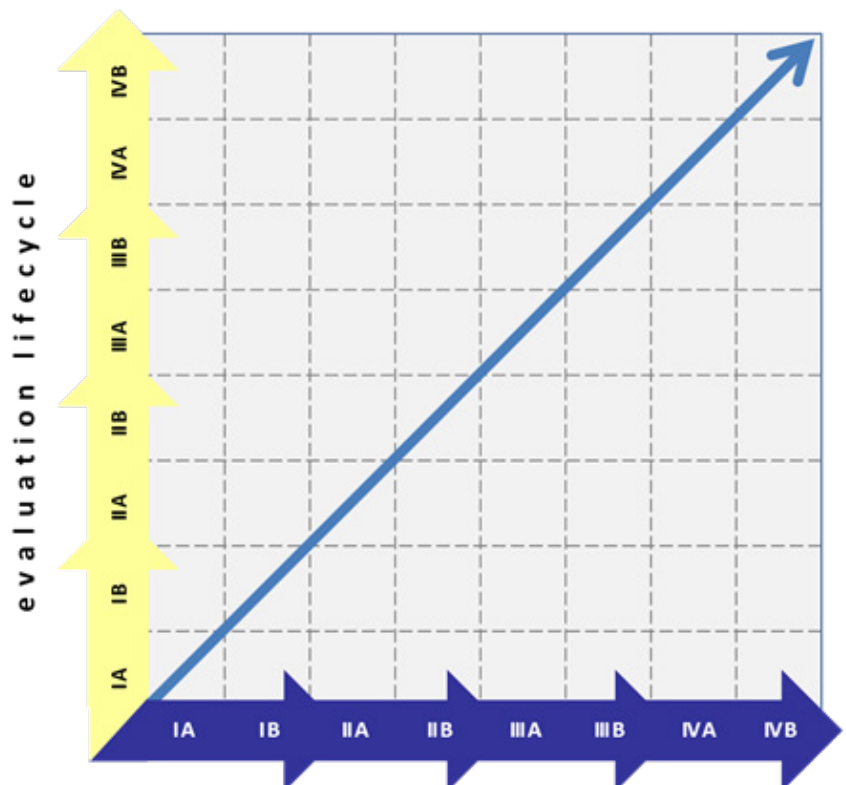
# Logic Model

A logic model is a representation of the thinking behind a program. Logic models come in many different formats. Some are more graphical than others. They all include the basic components of program activities and outcomes. Many also include program inputs and outputs. Some include program context and assumptions. Overall, logic models represent the theory of change behind a program - the ways in which the program planners imagine the program activities have an effect on the program's intended outcomes.

## FAQs

### Your logic models look different from other versions I've seen. Why is that?

There are several well-established methods/techniques for creating logic models. A columnar format is common to most versions, although there may be differences in the way the columns are labeled. The definition of what constitutes an "output" is probably the element that varies the most across different versions. Other differences sometimes show up in the context and assumption sections as well. Nevertheless, the diverse formats are all attempting to serve the same purpose: to convey succinctly the essential information about what a program entails, why it exists, and what kinds of outcomes are expected to result. The format we provide here allows us to address all the steps in the Protocol, and in particular it is essential for pathway modeling which is unique in articulating the detailed theory of change that underlies a program. Given the common features of most logic models, it is usually not too difficult to adapt a pre-existing one to the format and definitions for use on the Netway with the Protocol.

### What's the difference between an "indicator" and an "outcome"?

An outcome is the change you are expecting or hoping to see as a result of people's participation in your program. An indicator is how you might be able to tell that the outcome has been achieved. For example, an outcome of a nutrition education program might be "Families increase the variety of vegetables in their diet." An indicator of success might be "an increase in the weekly average number of different vegetables consumed at family mealtimes." Another indicator of that same outcome might be "an increase in the number of different vegetables observed in the family's grocery store purchases." The outcome should be about the underlying fundamental change that takes place, and that's what belongs in the logic model. The indicator for a particular outcome might be many different things, and is best decided a part of the evaluation and measurement strategy.

### Should I list every activity separately for the purpose of the logic model?

It depends. Activities that are lumped together can diminish the ability to pose more specific evaluation questions later on. For example, imagine a program that consists of many different workshops with different topics and formats. On the logic model, they are lumped together and labeled simply as "workshops." The logic model can now no longer show if and how the hands-on workshops have different outcomes than the online workshops, for example. On the other hand, if this same program lists every workshop as a separate activity, there may be diminishing returns. The model will become cluttered and more difficult to read. And, chances are, there will be a lot of repetitive connections to the same outcomes. In order to determine how much to "lump" or "split" activities for the purpose of the program model(s) it is essential to think about their related outcomes. A general guideline is: an activity should be listed separately if it has a unique outcome or set of outcomes compared to other activity(ies).

### Should I include administrative activities in my program model?

These background activities - like recruitment, program marketing, training of instructors, or fundraising - are essential for program implementation. However the question of whether they should be included in your program model is complicated. In reality, all programs have some kind of admin effort in the background. The question here is, if you had 1 minute to describe to someone the essence of your program and how it works, would these be among the activities you would talk about? Most of the time the answer is likely to

be "no". There are exceptions though, which is why there is not a hard and fast rule about this. Sometimes an activity that sounds administrative is actually directly related to achieving certain program outcomes. For example, marketing and participant recruitment are often routine (though important!) efforts. But for a program whose success depends on participation by a very specific demographic mix of participants, perhaps including hard-to-reach individuals, the outreach and recruitment effort becomes central to the program in a distinctive way.

### What is the difference between a logic model and a pathway model?

A logic model is a representation of the main components of a program and the theory of change underlying that program. Both the columnar model and the pathway model are types of logic models. These two types have a lot in common, but each also has unique features. Both models include activities, outputs and outcomes. The columnar logic model also includes inputs, context and assumptions. These components provide important information about what it takes to run this program, what environments it is designed for, etc. This information is useful to others who might want to replicate or learn from this program, and may be important aspects to evaluate. A pathway model does not include inputs, context, or assumptions but it has causal arrows between activities and outcomes, creating a graphical representation that highlights the theory of change in a much more specific and detailed way. These arrows make it possible to see and follow the story-lines that explain how the program seeks to achieve its goals. The graphical representation allows for detailed, shared understanding of what the program is and how it works – this forms an invaluable foundation for making good evaluation decisions.

### How do I know if an outcome is short-, mid-, or long-term?

There are no fixed time scale definitions for short- mid- and long-term outcomes. Rather, the general idea is that short-term outcomes arise soon after and as a direct result of an activity. Long-term outcomes describe the ultimate impacts of the program (within the bounds of the program definition). Mid-term outcomes then are the intervening changes that logically connect those immediate effects (short-term outcomes) to the ultimate impacts (long-term outcomes).

## Guiding Documents

- Getting Started with your Logic Model

- Logic Model Definitions and Guidance

- Blank Logic Model Template

- Review Checklist for Logic and Pathway Models

# *Getting Started with your Logic Model*

**Introduction:**

This document is meant to help you get started on developing your Program Logic Model (LM).

The "Before You Jump In" section walks you through some preliminary questions to help frame your Logic Model. The "Starting to Build" section covers where to start and how to work your way through the various components of the LM. Finally, we offer some general "Tips" in the last section.

**Before You Jump In:**

1.  Use what you have: It may help to review existing material that you already have about the program (written descriptions, marketing materials, annual reports, website content, etc.) If you are following the steps of the Systems Evaluation Protocol, then the Stakeholder Analysis, Program Review, Boundary Analysis, and Program Description will have given you a good base for developing the Logic Model. All of these can be helpful for identifying program activities, outcomes, context, and key assumptions.

2.  Find relevant examples: It can be helpful to look at examples of other logic models. Search the Netway for programs that similar to yours and review their logic models to get ideas.

3.  Think about "participants": There's no column in a Logic Model where you have to list participants explicitly, but it helps to have a clear idea about who your participants are. Your vision of who your participants are contributes to the Context section, where you will likely need to describe who's in the program and what it's important to know about them. Your vision of your participants also tends to affect the kinds of Outcomes you include, since it is through effects on participants and then the larger community that your program has impact.

The questions about who your participants are amount to a kind of "program boundary" issue, and it is sometimes not so obvious who should be viewed as a "participant". In a workshop setting it seems clear that "participants" are the people in the room. But if your program consists of workshops plus public events intended to raise awareness, then your program's participants are not just the ones in the room but also the attendees reached in some way at the public event. At a County Fair are you only thinking about the youth who are showing their projects? Or are you thinking about the impact this may be having on youth (or adults) who come to see the displays? If so, then they are "participants" in a sense as well.

4.  Think about Timeframe: How much time is involved between the beginning of the Activities and the ultimate Long-Term outcomes that you envision for your program? When you work on the outcomes within your LM you will have to distinguish between Short-, Mid-, and Long-Term.

    *   For Short-Term (ST) you should think about where you expect participants to be by the end of your time with them, or within a short time after that. If it's a one-time 3-hour event, then what do you expect them to have gained by the end of the afternoon? What will they be walking away with, or would be showing up within a few weeks? If it's a year-long after-school program then what would you expect to see during and by the end of that year? (That is, it's not really possible to give a calendar-time definition of "Short-Term", because the relevant time horizon varies with the program. The general principle is that Short-Term outcomes arise soon after and as a direct result of an activity.)

- For Mid-Term (MT) you should think about what tends to follow from those ST Outcomes. These might answer the "And then what?" questions following the ST effects. You might expect to see spill-over from direct participants to their family members or colleagues, and you might expect to see some initial behavior change that might follow from the knowledge and skill gains that would have occurred in the Short-Term.

- For Long-Term (LT) you should be thinking more broadly toward the ultimate goals of a Program, so this would tend to show up in terms of how things would unfold for the individual participants over a long time span (years, perhaps), and also of the cumulative effect on a community or group as the effects spread out. That is, if individual participants gain a greater interest in and capacity for science careers, then the community and country will (eventually) have the benefits of having greater numbers of scientists and a populace that is generally more versed in the tools and role of science.

**Starting to Build:**

There's no single "right" place to start with a Logic Model. Building backwards – starting from Long-Term outcomes – works well if you prefer to go from the big picture and ultimate goals and work back to the "how" part of what you have to do to get there. There's more of a program planning flavor in starting this way, so this may be particularly helpful for programs that are newly-developed or are still in the development phase. But, even for established programs, it may simply suit you to think of the big picture first and then work backwards.

It is also reasonable and may suit you better to start with Activities, since this means focusing first on what you DO as part of a Program. Inputs can be filled in at any time and are usually relatively straightforward. Outcomes follow from thinking through a series of "And then what?" questions. Context and Assumptions can be a bit difficult to bring to the surface and put into words, because although they are very important they are often just implicit or unstated in our minds. So it can sometimes work best to fill those in after the other pieces are written up.

**Assumptions** help fill in the explanations about why a program is expected to have these series of Outcomes (because, for example, you are assuming that young people who meet real scientists and have fun working with them are more likely to consider a career in that field than if they had never met a real live person who did that kind of work.) Assumptions can also hold a place for things that might or might not actually turn out to be true, but where a change would have serious implications for whether your program is likely to succeed or not. (For example, an assumption that there will continue to be funding for this program for the whole fiscal year, or that the prices of home heating fuels are going to continue to increase, etc.)

**Context** provides important information about the community being served by this program and the target participants, to give perspective to the Outcomes and Activities. For example, you might find it appropriate to explain that youth in this community have no other access to a safe after-school setting or program, or that language barriers make it important that this program be offered with Spanish-speaking volunteers and leaders, etc. The Context section should paint a picture of the setting within which the program operates, and the environment facing the participants.

*Tips*:

- *Use the "Logic Model Definitions and Guidance" handout* – read the guidance and examples for each LM section before filling it in; go through the checklist for each section after filling it in.

- *Keep your language as general as possible* when naming Inputs, Activities, Outputs, and Outcomes.

36

*For example:* *When naming Activities it is important to choose titles that are neither too broad nor too specific. Consider the following three options:*

1.  *"Health Education Workshop Series at JFK High School in the Bronx on October 25th" (too specific!)*

2.  *"Workshop" (too broad!)*

I*magine that you will be repeating this activity at different locations on different dates. What name would work for all of them?*

3.  *"Health Education Workshop for Teens" (just right…)*

- Be careful about "column drift" e.g. putting activities or outputs in the outcome column, etc. Refer to the definitions in the Logic Model Definitions and Guidance for clarification.

REALLY IMPORTANT:

- Be careful about mixing up "outcomes" and "targets" and "indicators": For the kind of Logic Model we use, and especially for evaluation planning purposes, outcomes should describe the changes that are expected to occur as a result of the program's activities, rather than being phrased in terms of goals or indicators of success.

For example, contrast the following three versions:

1.  *"10% of participants will pass the food safety certification exam"*

2.  *"Percentage of participants passing the food safety certification exam"*

3.  *"Increased knowledge of food safety principles"*

*Version #3 is what we would be looking for. It describes the underlying change that the program seeks, and it describes it in general terms and therefore allows flexibility to evaluate the hoped-for change in different ways. Version #1 would serve well as a target or goal, which could be useful as a benchmark of success for accountability purposes or funding proposals. But it is best viewed as a goal. Version #2 is more general but it is really just identifying an indicator of the underlying outcome that the program is designed to achieve.*

# Logic Model Definitions and Guidance

| Inputs | Activities | Outputs | Short-term (ST) outcomes | Mid-term (MT) outcomes | Long-term (LT) outcomes |
|---|---|---|---|---|---|
| ...are the key resources that go into making a program possible. | ... are the primary, intentional mechanisms by which program outcomes are achieved. They are often conducted or implemented by program staff. | ...are the artifacts or by-products created as a result of an activity taking place. | ...are changes directly connected to Activities, typically including awareness, knowledge, attitudes, and skills; - these are the first set of outcomes that might be observed. | ...are changes directly connected to Activities, Short- or other Mid-term outcomes, typically including behavior, or decision making; - these are a bridge between Short-term and Long-term outcomes. | ...are ultimate changes or impacts, directly connected to Mid- or other Long-term outcomes, typically including social, economic, civic, or environmental changes. |
| *examples*: - staff (incl. hours, or FTE) - outside partners - funding | *examples*: - Workshop on [topic] - Site Tour(s) - Newsletter "XYZ" on[topic(s)] - Materials development | *examples*: - certificates of completion - attendance record - record of sessions offered | *examples*: - increased knowledge of [topic] - improved skills - improved attitudes toward [topic] | *examples*: - participants apply knowledge to outside contexts - participants adopt and use new methods | *examples*: - change in knowledge of the broader population - increased economic stability |
| *checklist*: - does it include key resources? - does the list give a sense of your program's size and scale? | *checklist*: - do boundaries of Activities accord with your boundary analysis? - would Activity names be clear to an outsider? | *checklist*: - are your Outputs tangibles or by-products of participation in Activities? - do any of your Outputs look like they belong in Outcomes? | *checklist*: - does each ST Outcome describe short-term changes logically related to Activities or other ST Outcomes? - do boundaries of ST Outcomes accord with your boundary analysis? | *checklist*: - does each MT Outcome describe mid-term changes logically related to Activities or ST or other MT Outcomes? - do boundaries of MT Outcomes accord with your boundary analysis? | *checklist*: - does each LT Outcome describe long-term changes that logically relate to either MT or other LT Outcomes? - do boundaries of LT Outcomes accord with your boundary analysis? |

**Assumptions:** Beliefs and thought patterns about how and why a program is expected to succeed that are not otherwise explicitly stated; also, things that would prevent a program from achieving its long-term outcomes.

*Examples*: the approach to teaching used in the program is effective for this audience; the starting condition of participants is understood to be ... [x]; the program will have access to resources provided by partners or funders through the entire program cycle; participants are able to find transportation to program site, ...

**Context:** Information about the program setting or history; the environment in which a program will take place, and participants that paints a good but concise picture of the program. *Examples*: The program is the first effort in this community to provide a ...; ...is adapted from a long-established curriculum in order to attract a younger audience; ...is held within an afterschool program; ...at the university; ...within the Spanish-speaking community; ...for at-risk middle school youth...etc.

Netway (www.evaluationnetway.com)   Document available at: http://www.evaluationnetway.com/guide/evaluation-guidance/logic-model.

# *Blank Logic Model Template*

| Inputs | Activities | Outputs | ST outcomes | MT outcomes | LT outcomes |
|--------|-----------|---------|-------------|-------------|-------------|
|        |           |         |             |             |             |

**Assumptions:**

**Context**

# *Review Checklist for Logic and Pathway Models*

| Program Name: | |
|---|---|
| Completed by (Name and Title): | Date: |

Please check the appropriate box corresponding with either yes, somewhat or no for each line listed, and add any comments you may have. Reviewer feedback is crucial to the model revision process. Please write comments as needed to help explain your assessment and to offer further suggestions.

## <u>Logic Model</u>

**I.      Overall Assessment:**  Logic Models should describe a program accurately, concisely, and coherently. They should reflect the internal logic of the program. Please review this program's Logic Model (LM) and Program Description, and offer overall comments. For example, does the LM match the Description? Is the LM clear? Does it tell a reasonable and coherent "story" (is the logic of this program plausible and internally consistent)? Are the elements of the LM consistent with the implicit program "boundary"? Other comments?

**II.      Section by Section Assessment:**

| Objective/Category | Yes | Some what | No | Comments |
|---|---|---|---|---|
| **Inputs** | | | | |
| List of resources appears complete | | | | |
| Descriptions are informative | | | | |
| Gives reader accurate, at-a-glance idea of program size/scale (e.g., % FTE for staff; annual budget; avg # participants, etc. … as appropriate) | | | | |
| **Activities** | | | | |
| List of activities appears complete | | | | |
| Activity titles are clear and understandable | | | | |
| Activity titles written in a general way (not context-specific) | | | | |
| Activity list is consistent with program boundary (e.g. only includes activities that reach people who participate or who are targeted) | | | | |

Netway (www.evaluationnetway.com)
Document available at: http://www.evaluationnetway.com/guide/evaluation-guidance/logic-model.

| Objective/Category | Yes | Some what | No | Comments |
|---|---|---|---|---|
| **Outputs** | | | | |
| List of outputs appears complete | | | | |
| Outputs are all tangible or measurable evidence of program activities (depending on program, this might include attendance list, certif. of completion, projects completed, hrs of participation or contact, etc.) | | | | |
| Does not include effect on participants (outcomes) | | | | |
| **Outcomes in General** | | | | |
| Outcome titles are clear and informative | | | | |
| Outcomes written in a general way (not context-specific) | | | | |
| Outcomes are phrased as effects on, or changes in, participants and/or their communities or society (they are not actions, or objectives, or specific indicators) | | | | |
| **Short-term (ST) Outcomes** | | | | |
| List of ST outcomes appears complete | | | | |
| Outcome is in the correct column (i.e., it logically arises with or "soon after" the activity and is not an output or mid-term outcome, etc.) | | | | |
| **Mid-term (MT) Outcomes** | | | | |
| List of MT outcomes appears complete | | | | |
| Outcome is in the correct column (i.e., it plausibly arises beyond the ST, and reflects further outcomes for participants and/or spillover to others; is not a short-term or long-term outcome) | | | | |
| **Long-term (LT) Outcomes** | | | | |
| List of LT outcomes appears complete | | | | |
| Outcome is in the correct column (i.e., it reflects ultimate impacts on participants and/or their communities/society at large) | | | | |

44

| Objective/Category | Yes | Some what | No | Comments |
|---|---|---|---|---|
| **Assumptions** | | | | |
| Clearly describes beliefs and thinking about the program and how it will occur | | | | |
| List of assumptions appears complete | | | | |
| **Context** | | | | |
| Clearly describes environment (e.g., social, cultural, physical, etc.) in which the program is taking place | | | | |
| List of contextual factors appears complete (captures relevant factors) | | | | |
| Formatting | | | | |
| Format is consistent | | | | |
| Model is easy to read | | | | |

# Pathway Model

Note to reviewers: there's no prescriptive level of detail or generality that's "right" under all circumstances. Some intended audiences may need more or less detail than others. As an outside reader, you can't be sure of the specific audience. However, the overall goal should be to ensure that a pathway model is ***comprehensive and internally consistent for some presumed audience.*** Please comment accordingly and raise questions for the program modeler if needed.

| Objective/Category | Yes | Some what | No | Comments |
|---|---|---|---|---|
| **Items** | | | | |
| All activities and outcomes from logic model are represented in pathway diagram | | | | |
| Appropriate connections have been made to every item (no "orphaned" items) | | | | |
| **Connections** | | | | |
| All relevant connections appear to have been made | | | | |
| Connections are made to nearer term outcomes whenever possible (no "leap-frogging" or redundancy) | | | | |
| Big steps (as from a short to long term outcome) have been minimized | | | | |
| **Pathways** | | | | |
| Pathways clearly communicate the "story' or "program logic" that connects activities to long term outcomes | | | | |
| "Dead ends" only appear as important side-effects or outputs | | | | |
| **Representation** | | | | |
| Pathway diagram efficiently communicates program logic | | | | |
| Level of detail in both items and connections  is not excessive – model is  "readable" | | | | |
| **Overall Comments** | | | | |

46

## _Assumptions_

Program assumptions are beliefs about the program and how it will occur. Basic assumptions that are often made about programs include things like, "there is a need for this program," "there will continue to be interest in this program," "this program will be funded." However, program assumptions can, and should, be much more specific than this. For example, "We assume that hands-on activities engage young students most effectively." Assumptions like this are often so ingrained that they are hard to identify.

### FAQs

#### Why is it important to identify program assumptions?

Identifying program assumptions is important for several reasons. First, assumptions are an important part of the thinking behind any program. As such, it is important to identify them in order for outsiders to fully understand the program and why it is conducted the way that it is. Second, program assumptions are legitimate and potentially important candidates for evaluation. Providing evidence to support a program assumption helps build the foundation of evidence for the overall logic of the program. Finally, program assumptions can help account for evaluation results. For example if the results appear to be "negative" the explanation may be that one or more of the assumptions are not accurate.

### Guiding Documents

- Uncovering Buried Assumptions

# *Uncovering Buried Assumptions*

Sometimes, a program is based on assumptions that are "buried" – implied but not stated. For example, consider the following:

**In a nutritional program for youth, the pathway model has this link:**

**Teach proper nutrition ... -->... Youth make more healthy food choice*s***

This seems perfectly logical, and would reflect the program's intention to act positively to improve kids' diets. Now, pull out the unstated assumptions:

**#1: youth in the program's target population lack nutrition knowledge.**

**#2: schools and neighborhood stores offer healthy food choices**

**#3: youth participants have the ability to apply nutrition knowledge (self-discipline, resistance to peer pressure, etc.)**

It can be very important to bring these buried assumptions to light. They can affect your program planning, your evaluation questions, and eventually the interpretation of evaluation results.

**Action Steps:** To check for buried assumptions or presumed conditions, look at the pathway model in the program you are examining and focus on key pathways. Then ask:

- *What does this assume* about participants' needs and/or capacities?
- What does this assume about gaps or possibilities in the environment?
- What does this assume about other things that could get in the way?

As you identify buried assumptions, note them on the back of this page.

**Keep in mind:** These un-buried assumptions (and others you may come up with), might be important to evaluate. For the example above, you might ask:

*#1: Do youth in the program's target population lack nutrition knowledge?*

*#2: Do youth in the program's target population have healthy food choices available to them?*

*#3: Do youth in the program's target population have the self-discipline to apply nutrition knowledge?*

49

# *Context*

This section of a logic model provides a concise description of the elements of a program's context that are important for understanding the program. These elements may include the history of the program, needs that it is designed to address, information about participants, practical details about program location, or any other background information that would help a reader understand why the program exists and is designed the way it is.

## FAQs

### Why is it important to identify the program's context?

Just like program assumptions, context helps fill out the entire story of a program. It gives an outside reader a better understanding of the program, and helps an evaluation planner determine what is appropriate. For example, if the program participants are 3rd to 5th graders, it would be inappropriate to use an evaluation tool that requires a high school reading level. The context section of a logic model helps remind evaluation planners of program realities like this.

## Guiding Documents

- Describing Program Context

# *Describing Program Context*

**Goal**: Brainstorm a comprehensive list of contextual factors relevant to the program being evaluated.

Consider the following, and make notes in the table below:

- Participant specifics (age group, other demographic information, literacy, primary language, socioeconomic status if relevant, prior experience, selection criteria, attitudes, etc. as needed)

- Participant/program contact (type of contact or interaction, one-time or repeated, duration, etc.)

- Program setting (format, physical location, time of day, etc.)

| | Notes |
|---|---|
| Participants | |
| Contact with Participants | |
| Setting | |

53

## *Inputs*

Inputs are key resources such as staff, curriculum, partner organizations, and facilities that go into making a program possible. The list of inputs in a logic model gives an idea of what it takes to run the program and, by including quantities, indicates the scale on which the program operates.

### FAQs

#### What should be included in a list of inputs?

Typically the list includes things like % FTE for staff and/or volunteers, budget, and materials. In addition, there may be other inputs that are important to the "story" of your program such as collaborators, curricula, and so on.

## *Activities*

Activities are the active components of a program. They are often conducted or implemented by program staff. Activities are the primary, intentional mechanisms by which program outcomes are achieved. Examples of activities include: workshops, online forums, networking, agricultural field trials, and so on.

### FAQs

#### Should I list every activity separately for the purpose of the logic model?

It depends. Activities that are lumped together can diminish the ability to pose more specific evaluation questions later on. For example, imagine a program that consists of many different workshops with different topics and formats. On the logic model, they are lumped together and labeled simply as "workshops." The logic model can now no longer show if and how the hands-on workshops have different outcomes than the online workshops, for example. On the other hand, if this same program lists every workshop as a separate activity, there may be diminishing returns. The model will become cluttered and more difficult to read. And, chances are, there will be a lot of repetitive connections to the same outcomes. In order to determine how much to "lump" or "split" activities for the purpose of the program model(s) it is essential to think about their related outcomes. A general guideline is: an activity should be listed separately if it has a unique outcome or set of outcomes compared to other activity(ies).

#### Should I include administrative activities in my program model?

These background activities - like recruitment, program marketing, training of instructors, or fundraising - are essential for program implementation. However the question of whether they should be included in your program model is complicated. In reality, all programs have some kind of admin effort in the background. The question here is, if you had 1 minute to describe to someone the essence of your program and how it works, would these be among the activities you would talk about? Most of the time the answer is likely to be "no". There are exceptions though, which is why there is not a hard and fast rule about this. Sometimes an activity that sounds administrative is actually directly related to achieving certain program outcomes. For example, marketing and participant recruitment are often routine (though important!) efforts. But for a program whose success depends on participation by a very specific demographic mix of participants, perhaps including hard-to-reach individuals, the outreach and recruitment effort becomes central to the program in a distinctive way.

## *Outputs*

An output is a by-product or artifact that is created as a result of an activity taking place. It is different from an outcome, which is a description of a change that is thought to emerge as a result of the activity or program taking place. Outputs can be useful indicators of an activity having taken place, how often, and on what scale.

## FAQs

### How do outputs fit in to a pathway model?

Outputs result directly from activities. On a pathway model, arrows between activities and outputs show which activities result in the creation of which outputs. However because outputs are simply artifacts they do not, on their own, lead to outcomes. So there should not be arrows leading from outputs to outcomes.

### What is the role of outputs in an evaluation?

Outputs play an important role in evaluation. Because outputs are tangible artifacts of activities, connections between activities and outputs can be fertile ground for evaluation questions related to program implementation. For example, in the case of the model airplane program, one might ask "Did participation in the airplane model building workshop lead to the production of finished model airplanes by our participants?" This is essentially a question about the connection between an activity and an output. Answering this question would provide foundational evidence, typically for an early lifecycle program, about whether a program activity is working the way it is intended to.

### What is the difference between an output and an outcome?

An output is a tangible by-product of an activity (think artifact), while an outcome is an effect on a participant, the community, or society. In a logic model, outputs should be labeled as simple nouns ("photos of participants", "finished model airplanes", "contact list of participants"), whereas outcomes should be understood and written in terms of change ("participants' knowledge increases", "farm soil health improves").

## *Short-term Outcomes*

Short-term outcomes are early changes that arise during or very soon after an activity, and are logically and closely connected with program activities. For example, short-term outcomes may describe effects on participants' awareness, attitudes, skills or motivations.

## FAQs

### How do short-term outcomes connect to other parts of a logic/pathway model?

Short-term outcomes are closely connected to activities and either arise directly from an activity or from another short-term outcome. Short-term outcomes may lead to other short-term outcomes, or to mid-term outcomes. (They should not be connected directly to long-term outcomes. If you feel you want to connect a short-term to a long-term, pause and think about what mid-term outcome actually arises in between.)

### What defines "short-term" when talking about "short-term outcomes?"

It's not useful to define a standard time interval for short-, mid- or long-term outcomes in logic modeling, because the programs being modeled can be so different. If the program is a one-session workshop lasting 3 hours, then the timeframe for outcomes is likely to be much shorter than it would be for a program with multiple activities that lasts for a semester or a year. "Short-term" is a relative term in a logic model, referring to outcomes that occur by the end of the program or fairly soon thereafter, and are the initial changes that have to happen and lay a foundation for all subsequent changes.

### Should I include an outcome even if I can't imagine how it could ever be measured?

Yes. The logic model is just that – a model. It is supposed to convey a picture of how your program works and what it is expected to lead to or contribute to. Just because you can't count up or quantify or measure a particular part of that picture doesn't mean that it's not an important part of the picture. Even those ill-defined, intangible outcomes can be an important element of the program and can help an outsider understand the program better.

### What's the difference between an "indicator" and an "outcome"?

An outcome is the change you are expecting or hoping to see as a result of people's participation in your

program. An indicator is how you might be able to tell that the outcome has been achieved. For example, an outcome of a nutrition education program might be "Families increase the variety of vegetables in their diet." An indicator might be the weekly average number of different vegetables consumed at family mealtimes. Another indicator of that same outcome might be the number of different vegetables observed in the family's grocery store purchases. The outcome should be about the underlying fundamental change that takes place, and that's what belongs in the logic model. The indicator for a particular outcome might be many different things, and is best decided a part of the evaluation and measurement strategy.

### Should I include numbers in my outcomes?

Including numbers, such as "50 farmers will adopt new crop management practices," would specify a target for your program (and as such is completely appropriate for grant applications or program descriptions if the funder or audience expects that). However for purpose of wording outcomes in a the logic model, it is important to recognize that the logic model provides the basic logic of how your program works, and outcomes describe the changes you are expecting to see as a result of people's participation in the program. Stating specific numbers detracts from the articulation of the theory of change, by suggesting that the program is a mechanistic black box that "always" leads to 50 changes. So depending on your program situation, a more appropriate phrasing for this example outcome might be "Farmers adopt new crop management practices" (if you are just looking for them to change), or perhaps more specifically "Farmers adopt recommended best management practices for crop rotation" (if you are aiming for them to adopt a specific new practice).

## *Mid-Term Outcomes*

Mid-term outcomes are changes that logically and directly arise from short-term outcomes, or from other mid-term outcomes. For example, mid-term outcomes may describe changes in participants' behavior, decision making, or depth of understanding. They may also describe initial spillover effects on other individuals or parts of the community.

## FAQs

### What defines "mid-term" when talking about "mid-term outcomes?"

It's not useful to fix a calendar time interval for short-, mid- or long-term in logic modeling, because the programs being modeled can be so different. If the program is a one-session workshop lasting 3 hours, then the timeframe for outcomes is likely to be much shorter than it would be for a program with multiple activities that lasts for a semester or a year. "Mid-term" covers those outcomes that occur in between immediate or rapid short-term effects and the ultimate long-term changes that a program is aiming for.

### How do mid-term outcomes connect to other parts of a logic/pathway model?

The mid-term outcomes tend to answer the "and then what happens?" questions that take you from short-term outcomes out toward the larger and longer-term changes. They fill in the stepping stones of a change process. Mid-term outcomes may capture a deepening of the awareness and knowledge gains, and changes in behavior on the part of the participant. Mid-term outcomes also tend to include spillover effects – changes in the condition of surrounding people (family members, community members, etc.) as an individual participant's actions and behavior have wider effects.

### Should I include an outcome even if I can't imagine how it could ever be measured?

Yes. The logic model is just that – a model. It is supposed to convey a picture of how your program works and what it is expected to lead to or contribute to. Just because you can't count up or quantify or measure a particular part of that picture doesn't mean that it's not an important part of the picture. Even those ill-defined, intangible outcomes can be an important element of the program and can help an outsider understand the program better.

### What's the difference between an "indicator" and an "outcome"?

An outcome is the change you are expecting or hoping to see as a result of people's participation in your

program. An indicator is how you might be able to tell that the outcome has been achieved. For example, an outcome of a nutrition education program might be "Families increase the variety of vegetables in their diet." An indicator might be the weekly average number of different vegetables consumed at family mealtimes. Another indicator of that same outcome might be the number of different vegetables observed in the family's grocery store purchases. The outcome should be about the underlying fundamental change that takes place, and that's what belongs in the logic model. The indicator for a particular outcome might be many different things, and is best decided a part of the evaluation and measurement strategy.

### Should I include numbers in my outcomes?

Including numbers, such as "50 farmers will adopt new crop management practices," would specify a target for your program (and as such is completely appropriate for grant applications or program descriptions if the funder or audience expects that). However for purpose of wording outcomes in a the logic model, it is important to recognize that the logic model provides the basic logic of how your program works, and outcomes describe the changes you are expecting to see as a result of people's participation in the program. Stating specific numbers detracts from the articulation of the theory of change, by suggesting that the program is a mechanistic black box that "always" leads to 50 changes. So depending on your program situation, a more appropriate phrasing for this example outcome might be "Farmers adopt new crop management practices" (if you are just looking for them to change), or perhaps more specifically "Farmers adopt recommended best management practices for crop rotation" (if you are aiming for them to adopt a specific new practice).

## *Long-Term Outcomes*

Long-term outcomes are sustained or ultimate changes in program participants, their communities, or society, that arise logically from mid-term outcomes or from other long-term outcomes. For example, long-term outcomes may describe aggregate effects, changes in policy, and other big-picture impacts.

## FAQs

### Why should I include long-term outcomes that I will not be around to see?

A logic model is not supposed to just show things you can see or measure. It is meant to convey information and a vision of how the program works. Having a broad and full vision of what your program is about provides valuable information about what motivates the program, who might want to fund it, and what needs it is addressing. The long-term is an important part of that picture. Note also that including long-term, distant outcomes in your model does not imply that you are claiming that your individual program is solely capable of, or responsible for, huge outcomes like community well-being, poverty eradication, scientific breakthroughs, and so on. But you can claim how the program is contributing to broader changes and what you are working toward.

### How do long-term outcomes connect to other parts of a logic/pathway model?

Long-term outcomes should capture the ultimate goals of a program, either or both in terms of how things would unfold for an individual over a long time span (years, perhaps), or in terms of the cumulative effects on a community or society of having more and more participants with these experiences. Long term outcomes would be ones that mid-terms (or other long-term outcomes) lead to. It is likely also that at least some long-term outcomes would feature in the program description, or possibly in the mission statement, since those ought to provide information about why the program exists and what it is striving for.

### What defines "long-term" when talking about "long-term outcomes?"

It's not useful to fix a calendar time interval for short-, mid- or long-term in logic modeling, because the programs being modeled can be so different. In calendar time, long-term outcomes might occur within a year or two, or within twenty or even many more years depending on the nature and boundaries of the program being modeled.

### Should I include numbers in my outcomes?

Including numbers, such as "50 farmers will adopt new crop management practices," would specify a target for your program (and as such is completely appropriate for grant applications or program descriptions if the funder or audience expects that). However for purpose of wording outcomes in a the logic model, it is important to recognize that the logic model provides the basic logic of how your program works, and outcomes describe the changes you are expecting to see as a result of people's participation in the program. Stating specific numbers detracts from the articulation of the theory of change, by suggesting that the program is a mechanistic black box that "always" leads to 50 changes. So depending on your program situation, a more appropriate phrasing for this example outcome might be "Farmers adopt new crop management practices" (if you are just looking for them to change), or perhaps more specifically "Farmers adopt recommended best management practices for crop rotation" (if you are aiming for them to adopt a specific new practice).

# Pathway Model

A pathway model is a type of logic model. It is a graphical representation of the relationships between the activities, outputs, and outcomes that make up a program. Pathway models communicate the "story" or "theory of change" of a program. They are the essential foundation for determining the scope and questions that guide the evaluation of the program being modeled.

## FAQs

### What are the benefits of developing a pathway model aside from evaluation planning?

In addition to helping with evaluation planning, developing a pathway model is beneficial for program development and planning. The process of building a pathway model creates a unique opportunity for people to articulate their understanding of the program from their individual perspectives. Almost invariably, this process will lead to 'Aha!' moments, in which people come to understand their program in a new light. Another benefit comes from the fact that for many people the pathway model is an excellent communication tool. It offers a quick view of the entire program and helps "tell the story" of how the program works and what it hopes to achieve. The pathway model diagram can be used in accountability reporting and in efforts to secure new funding; and it can hang in your office to share with colleagues (and anyone who walks by) the thoughtful and purposeful design of your program.

### If two outcomes reinforce each other, how do I relate them in my pathway model?

There are times when two outcomes mutually reinforce each other. In cases like these, you can have arrows going both ways between the two outcomes in the pathway model. This representation stays true to your perception of how the program works. That particular area of the pathway model could be a good candidate for exploring in the evaluation planning process, exactly because of this interdependence. As you are deciding whether or not to use arrows going both ways, consider the impact on the overall readability and usefulness of the model. Does the importance of displaying that mutual reinforcement outweigh the risk of visual clutter?

### In a pathway model, can I have a short-term outcome lead to another short-term outcome?

Yes you can. In fact, it is often advisable. Doing so will let you present a more detailed and nuanced picture of the underlying process of change for the program. The resulting level of specificity is often necessary for the pathway model to be useful. Thinking ahead to the evaluation effort, since evaluating short-term outcomes is usually more feasible than evaluating more distant mid- or long-term outcomes, it can be advantageous to have early specific changes represented in detail.

### How will the pathway diagram be used for evaluation planning?

The pathway model will help guide your evaluation by revealing key outcomes and pathways that may serve as focal points for this evaluation cycle. This focusing is important because it is (usually) not feasible to evaluate the whole program in one year. The visual nature of the pathway model offers a uniquely powerful way to see the key "nodes" in the overall process of change, for example, outcomes that may have many arrows "going in" or "coming out", or both. (Ironically, although the pathway model is built from the knowledge held by the model-builders, it is sometimes only when they step back and look at the resulting visual display that these patterns can be recognized fully.) These model components become candidates for consideration, along with other factors such as stakeholder priorities, lifecycle considerations, and feasibility, in finalizing the evaluation questions.

A pathway model can also help with what is often a challenge for program staff, namely, that funders or other stakeholders want "evidence" about the program's impact on long-term outcomes which is simply not feasible for program staff to evaluate. In this case, the pathway model is useful for demonstrating the

conceptual linkage between shorter-term outcomes (which can be evaluated) and more distant outcomes. The burden of evaluation is reduced if you can then find published research to support these longer-term conceptual linkages.

### What is the difference between a logic model and a pathway model?

A logic model is a representation of the main components of a program and the theory of change underlying that program. Both the columnar model and the pathway model are types of logic models. These two types have a lot in common, but each also has unique features. Both models include activities, outputs and outcomes. The columnar logic model also includes inputs, context and assumptions. These components provide important information about what it takes to run this program, what environments it is designed for, etc. This information is useful to others who might want to replicate or learn from this program, and may be important aspects to evaluate. A pathway model does not include inputs, context, or assumptions but it has causal arrows between activities and outcomes, creating a graphical representation that highlights the theory of change in a much more specific and detailed way. These arrows make it possible to see and follow the story-lines that explain how the program seeks to achieve its goals. The graphical representation allows for detailed, shared understanding of what the program is and how it works – this forms an invaluable foundation for making good evaluation decisions.

## Guiding Documents

- Creating a Pathway Model from a Logic Model

- Relationship between Logic and Pathway Models

- Review Guide for Pathway Models

- "Mining the Model" Worksheet

- The "Golden Spike": Linking Evidence to Practice

# *Creating a Pathway Model from a Logic Model*

## 1. Start with the Logic Model

**Logic Model**

| Input: | Activities: | Outputs: | Short-Term Outcomes: | Mid-Term Outcomes: | Long-Term Outcomes: |
|---|---|---|---|---|---|
| Collaborators & Partners NYS Collaborators: NYS<br><br>Market Site Requirements Running Water Storage<br><br>Materials & Equipment Curriculum: CUCE-NYC Learning<br><br>Money/ Funding<br><br>Research Base<br><br>Staff<br> Community Educator Assistants (3 people, 21<br><br>Volunteers  25 - 30 students, volunteering 1-3 | Cooking Demonstrations/ Learning Stations at Farmers Market<br><br>Farmers Market Resource Awareness Activities at Collaborating Organizations | # of recipes distributed<br><br># of food samples taken<br><br># of Health Bucks distributed at CUCE-NYC N&H programs redeemed at NYC farmers markets<br><br># of referrals made to community food resources | Increased awareness of food-related budgeting and available resouces: FMNP Coupons, Food Stamps (EBT), WIC Checks, Health Bucks, Summer Meals, other<br><br>Increased knowledge of healthy food choices and benefits<br><br>Increased awareness of NYS growers, and locally grown produce<br><br>Increased knowledge/awareness of food safety when preparing foods<br><br>Increased exposure to preparation of healthy foods using simple, affordable recipes<br><br>Increase awareness of benefits of physical activity<br><br>Participants at community organizations go to their local farmers market<br><br>Improved skill in selection of quality produce<br><br>Improved attitudes about healthy lifestyles (i.e. eating, physical activity)<br><br>Organization staff and their participants have an increased awareness of their local farmers markets (locations, general benefits, etc.)<br><br>Organization staff and their participants are more aware of the food-related resources they can use to buy produce at farmers markets | Individuals view community farmers markets as an affordable resources for produce<br><br>Individuals access a variety of food-related resources<br><br>Shop regularly at farmers markets and use them as a source of fresh produce<br><br>Adoption of MyPyramid food & physical activity guidelines<br><br>Increased safe food preparation and practices at home<br><br>Impacted individuals share knowledge, attitude, skills with others in their communities<br><br>Communities experience an increased number of residents eating nutritiously on a limited budget using resources appropriately<br><br>Participants use food-related resources to buy more (amount and/or variety of) vegetables, fruits, and herbs (can be a ST outcome for those who arrive with EBT, Health Bucks etc.) | Communities/neighborhoods respond to consumers interest in and need for increased access to affordable, healthy foods<br><br>Fresh produce is easily accessed within neighborhoods<br><br>NYC families with limited resources can afford fresh produce<br><br>A secure NYC food system<br><br>Decreased risk factors for nutrition-related health problems and chronic disease in NYC neighborhoods |

## 2. Focus on the columns for Activities, Outputs, and Outcomes

**Logic Model**

| Input: | Activities: | Outputs: | Short-Term Outcomes: | Mid-Term Outcomes: | Long-Term Outcomes: |
|---|---|---|---|---|---|
| Collaborators & Partners NYS Collaborators: NYS<br><br>Market Site Requirements Running Water Storage<br><br>Materials & Equipment Curriculum: CUCE-NYC Learning<br><br>Money/ Funding<br><br>Research Base<br><br>Staff<br> Community Educator Assistants (3 people, 21<br><br>Volunteers  25 - 30 students, volunteering 1-3 | Cooking Demonstrations/ Learning Stations at Farmers Market<br><br>Farmers Market Resource Awareness Activities at Collaborating Organizations | # of recipes distributed<br><br># of food samples taken<br><br># of Health Bucks distributed at CUCE-NYC N&H programs redeemed at NYC farmers markets<br><br># of referrals made to community food resources | Increased awareness of food-related budgeting and available resouces: FMNP Coupons, Food Stamps (EBT), WIC Checks, Health Bucks, Summer Meals, other<br><br>Increased knowledge of healthy food choices and benefits<br><br>Increased awareness of NYS growers, and locally grown produce<br><br>Increased knowledge/awareness of food safety when preparing foods<br><br>Increased exposure to preparation of healthy foods using simple, affordable recipes<br><br>Increase awareness of benefits of physical activity<br><br>Participants at community organizations go to their local farmers market<br><br>Improved skill in selection of quality produce<br><br>Improved attitudes about healthy lifestyles (i.e. eating, physical activity)<br><br>Organization staff and their participants have an increased awareness of their local farmers markets (locations, general benefits, etc.)<br><br>Organization staff and their participants are more aware of the food-related resources they can use to buy produce at farmers markets | Individuals view community farmers markets as an affordable resources for produce<br><br>Individuals access a variety of food-related resources<br><br>Shop regularly at farmers markets and use them as a source of fresh produce<br><br>Adoption of MyPyramid food & physical activity guidelines<br><br>Increased safe food preparation and practices at home<br><br>Impacted individuals share knowledge, attitude, skills with others in their communities<br><br>Communities experience an increased number of residents eating nutritiously on a limited budget using resources appropriately<br><br>Participants use food-related resources to buy more (amount and/or variety of) vegetables, fruits, and herbs (can be a ST outcome for those who arrive with EBT, Health Bucks etc.) | Communities/neighborhoods respond to consumers interest in and need for increased access to affordable, healthy foods<br><br>Fresh produce is easily accessed within neighborhoods<br><br>NYC families with limited resources can afford fresh produce<br><br>A secure NYC food system<br><br>Decreased risk factors for nutrition-related health problems and chronic disease in NYC neighborhoods |

Netway (www.evaluationnetway.com)
Document available at: http://www.evaluationnetway.com/guide/evaluation-guidance/pathway-model

## 3. Make the connections: think about what each activity leads to or contributes to, and then what each outcomes leads to or contributes to, and so on.



| Activities: | Outputs: | Short-Term Outcomes: | Mid-Term Outcomes: | Long-Term Outcomes: |
|---|---|---|---|---|
| Cooking Demonstrations/ Learning Stations at Farmers Market<br><br>Farmers Market Resource Awareness Activities at Collaborating Organizations | # of recipes distributed<br><br># of food samples taken<br><br># of Health Bucks distributed at CUCE-NYC N&H programs redeemed at NYC farmers markets<br><br># of referrals made to community food resources | Increased awareness of food-related budgeting and available resouces: FMNP Coupons, Food Stamps (EBT), WIC Checks, Health Bucks, Summer Meals, other<br><br>Increased knowledge of healthy food choices and benefits<br><br>Increased awareness of NYS growers, and locally grown produce<br><br>Increased knowledge/awareness of food safety when preparing foods<br><br>Increased exposure to preparation of healthy foods using simple, affordable recipes<br><br>Increase awareness of benefits of physical activity<br><br>Participants at community organizations go to their local farmers market<br><br>Improved skill in selection of quality produce<br><br>Improved attitudes about healthy lifestyles (i.e. eating physical activity)<br><br>Organization staff and their participants have an increased awareness of their local farmers markets (locations, general benefits, etc.)<br><br>Organization staff and their participants are more aware of the food-related resources they can use to buy produce at farmers markets | Individuals view community farmers markets as an affordable resources for produce<br><br>Individuals access a variety of food-related resources<br><br>Shop regularly at farmers markets and use them as a source of fresh produce<br><br>Adoption of MyPyramid food & physical activity guidelines<br><br>Increased safe food preparation and practices at home<br><br>Impacted individuals share knowledge, attitude, skills with others in their communities<br><br>Communities experience an increased number of residents eating nutritiously on a limited budget using resources appropriately<br><br>Participants use food-related resources to buy more (amount and/or variety of) vegetables, fruits, and herbs (can be a ST outcome for those who arrive with EBT, Health Bucks etc.) | Communities/neighborhoods respond to consumers interest in and need for increased access to affordable, healthy foods<br><br>Fresh produce is easily accessed within neighborhoods<br><br>NYC families with limited resources can afford fresh produce<br><br>A secure NYC food system<br><br>Decreased risk factors for nutrition-related health problems and chronic disease in NYC neighborhoods |

## 4. Note, there may be …

- more than one arrow coming FROM an Activity or Outcome
- more than one arrow going INTO an Outcome
- arrows WITHIN a column (ST leading to other ST, MT to other MT, etc.)
- arrows in both directions between two Outcomes

There should NOT be …

- an Outcome with no arrow leading to it
- an Activity with no arrows leading from it
- dead ends: ST or MT outcomes with no arrows going out from them

Ideal Level of Detail? – it depends (on what you need to communicate, your audience's tolerance for detail, etc.)

## 5.    Full pathway model, using the Netway:

*Creating a Pathway Model from a Logic Model  page 3 of 3*
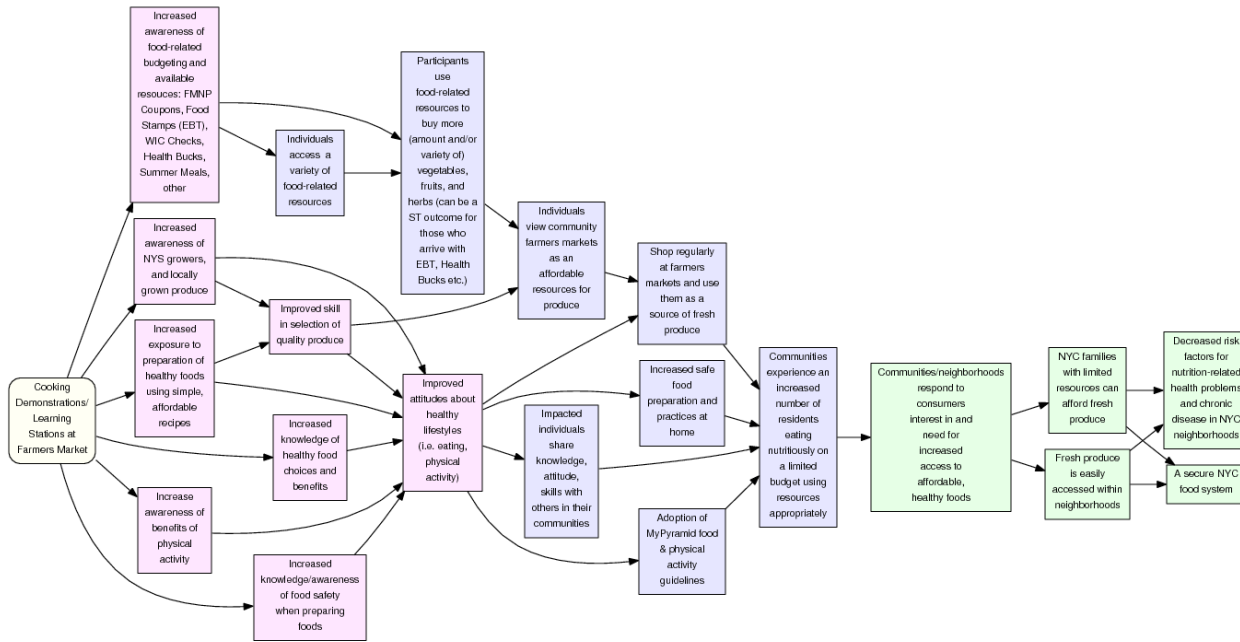
# *Relationship between Logic and Pathway Models*

**Columnar Logic Model**

**Pathway Model**



**Causality/ Program Theory**

**Shows "sub-programs"**

**Tells "story" graphically**

**Core Program Model Elements: Activities, Outputs, Outcomes**

**Inputs**

**Context**

**Assumptions**

**Mission**
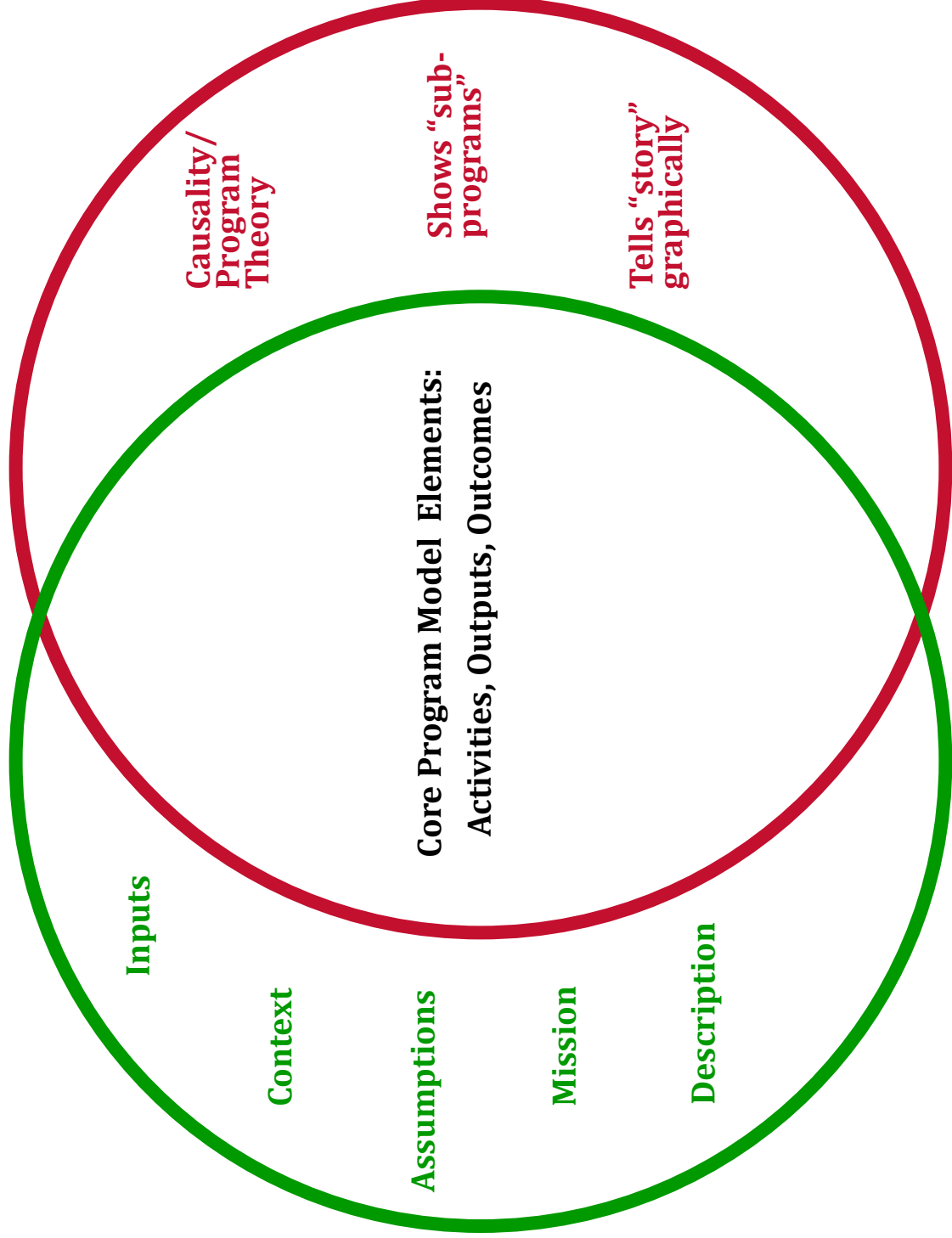
**Description**

# *Review Guide for Pathway Models*

*These steps are intended to guide a constructive review of a pathway model by an outside reviewer, and may also be used as a self-checklist for those who developed the pathway model.*

1.      Read the Program Description and then review the Pathway Model broadly. Do they match? If not, how are they mismatched?

2.      Look for good ideas and note/highlight them. These might include particularly good or novel outcomes, good links, whatever deserves recognition.

3.      If you see big leaps in logic, add a brief note with a suggestion if possible. (A big leap is where there's an arrow from an Activity all the way to a MT or LT outcome, or a ST all the way to a LT, etc.  It could also be a one-step arrow if there's a big leap of logic involved, so that it seems like something is being skipped over.)

4.      If you see something that makes you wonder about a boundary issue, add a brief note with a suggestion if possible.

5.      If you see something that is likely to be confusing to an outsider, or that could be worded more clearly, mark it and add a brief note with a suggestion if possible.

6.      From your own perspective and what you know of the key stakeholders' perspectives, think about whether the model captures a full view of the program. If necessary, propose an additional outcome or activity.

7.      Look for themes or common threads among outcomes and make a note of them.

8.      Step back and think about the model overall. Prepare some comments and observations to share as appropriate.

# *"Mining the Model" Worksheet*

*It's generally not feasible to evaluate all aspects of a program. Prioritization is essential. The visual pathway model creates a foundation for integrating program logic, internal and external stakeholder priorities, and program lifecycle considerations in order to hone in on strategically important evaluation questions. This worksheet walks you through a series of questions to identify the important considerations and annotate a printed pathway model.*

*Print a paper copy of your pathway model (if possible, print on 11x17 paper or larger poster paper). Use highlighters, colored pens, markers, or whatever you have at hand to mark up the model. The steps below guide you through a series of questions to identify the important considerations. When finished, step back and weigh the evaluation priorities that have emerged and assess what would be the best feasible purpose(s) for the coming evaluation cycle.*

### 1.	What does the Model show you about key program outcomes?

Use one color highlighter to circle **key outcomes** – ones that have a lot of arrows going into them or out of them, or both. For example, look for:

"Prime Destinations" (outcomes that have a lot of arrows going IN to them)

"Gateways" (outcomes with lots of arrows going FROM them)

"Hubs" (outcomes with lots of arrows going IN AND OUT)

### 2.	What are some Key Links?

Some things are important even if there aren't a lot of arrows going in or out. Each arrow, or link, represents some change that your program leads to or contributes to. Which ones do you think are important in the program? Mark these **key links** with a second color highlighter.

### 3.	What are the key pathways, or main storylines in the model?

Think about what's essential to the way your program works and succeeds – the main storylines. (If you had to strip down your model to just a few through-lines from Activities to LT outcomes, which story lines would you insist on keeping?) Mark one or two of these **key pathways** with a third color highlighter.

### 4.	Identify key external  stakeholders and their priorities:

List one to three key external stakeholders to your program (ones you are likely to report to in the coming year, or ones involved in important decisions about the program):

A. _____

B. _____

C. _____

Think about what each of these stakeholders cares most about in your program. Mark the **external stakeholder priorities** by writing the letter for each stakeholder next to the outcome/s or activity/ ies that they are most interested in.

Netway (www.evaluationnetway.com)
Document available at: http://www.evaluationnetway.com/guide/evaluation-guidance/pathway-model .

**5.        Mark important internal priorities, if any:**

Mark 1 or 2 outcomes that are of particular importance to you – these are **internal stakeholder priorities**. Put a star or asterisk next to these.

**6.        Lifecycle consideration:**

Consider what lifecycle stage your program is in, and direct your attention to the "area" of the pathway model indicated by **alignment of evaluation with the program's stage of evolution.** With a pencil, lightly draw a line around the "area" that seems most appropriate for this evaluation cycle. *(See the Lifecycle Analysis section of the Protocol and related worksheets. At the risk of oversimplifying, if your program is in an early developmental stage the most useful evaluation would likely tend to focus on activities and program process and/or assumptions, and possibly some ST outcomes; for more established programs where those early pieces of evaluation data are already established, it would likely be more appropriate to explore further out short-term or possibly some mid-term outcomes, or some deeper evaluation of outcomes that have already been explored.)*

**7.        Step back and review your annotated model, make note of emerging evaluation priorities.**

Look to see whether the mark-ups you have added to your model overlap and reinforce each other, or not. If they do (that is, if internal and stakeholder priorities coincide and also relate to key component(s) of the model, and align with lifecycle considerations) then this analysis points out a fairly clear direction for the program evaluation to take. The purpose of the evaluation would be fairly clear, and a key consideration becomes feasibility.

It may be, however, that the various considerations are not all pointing the same way (for example, two key external stakeholders are interested in very different outcomes, or stakeholders are pressing for more extended outcome evaluation than your program is ready for, etc.) If the emerging priorities don't overlap, weigh the usefulness or strategic value of the alternatives that have surfaced, and make decisions about what is feasible and most useful for the current evaluation effort.

In either case, note down your suggestions for a potential "Evaluation Purpose Statement" below, and use these notes for further discussion and consideration with colleagues and stakeholders if possible.

# The "Golden Spike"

## Linking Evidence to Practice

Pathway models can be extremely useful for linking what is (or can be) known through evaluation, to what is known from research. Mapping existing research onto a pathway model helps identify where evaluation is needed in order to fill gaps in the evidence base. In addition, this can reduce the burden on evaluation by clarifying how and where evaluation results connect to an existing evidence base.

The illustration below shows a key "through-line" in a pathway model for a youth science program. The program has a validated survey for measuring youth interest in science (indicated by the pdf icon attached to the short-term outcome), and it has the resources to conduct a pre-post evaluation to assess whether completing the engineering kits is associated with an increase in youth interest in science. It does not have the resources to follow up with participants in the future to see if they are more likely to go on to college. However, there is published research (indicated by the pdf icon) supporting the link on the right side of the model – namely, that increased interest in science education contributes to increased rates of higher education attainment.
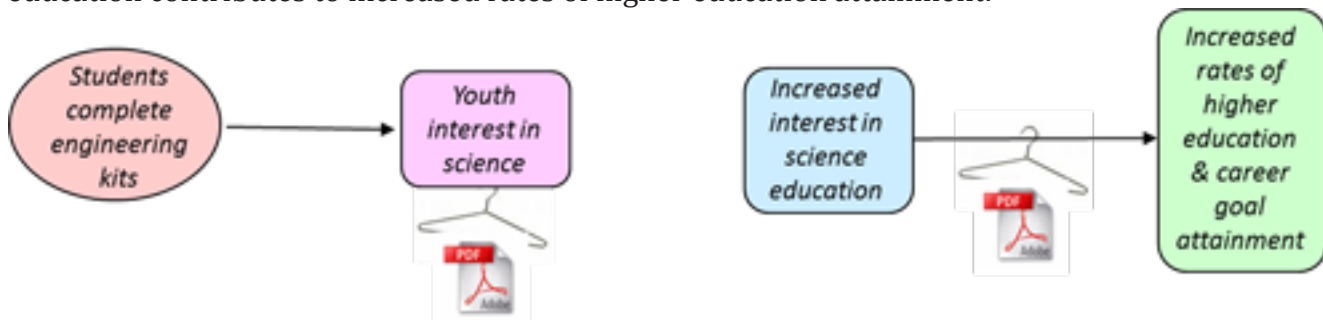


Harking back to the construction of the transcontinental railway in the US (in which a ceremonial golden spike was used on May 10, 1869 to make the final connection between the rail-lines coming from the east and west) the "Golden Spike" in evaluation refers to the point where evaluation and research meet.[1] In the example program here, this would be between the short- and mid-term outcomes linking interest in science to interest in science education.
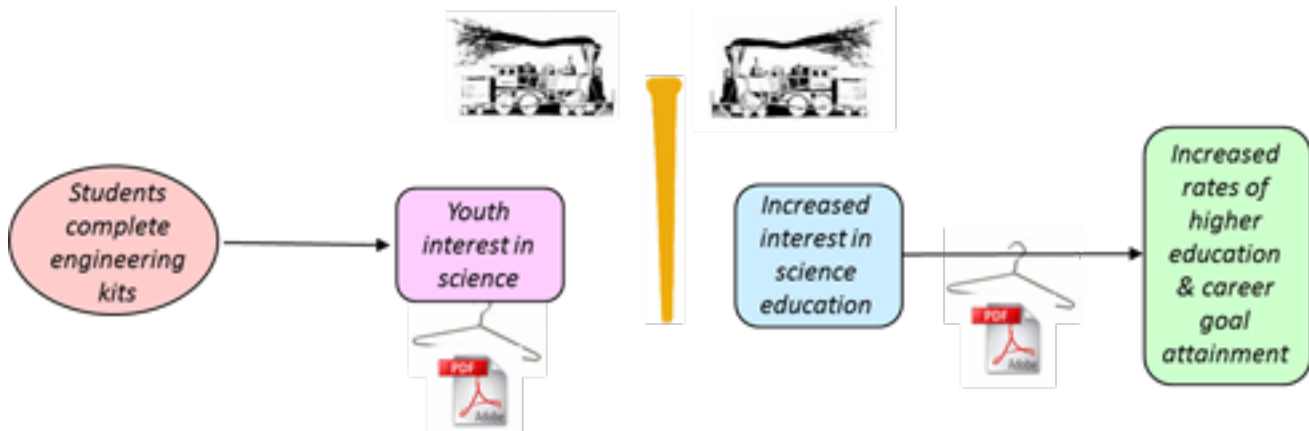


Evidence supporting this link would support a claim that this youth science program contributes to increased rates of higher education and career goal attainment.

[1]Urban, J.B., Trochim, W. (2009). The role of evaluation in research-practice integration: Working toward the golden spike. American Journal of Evaluation, 30, 535-553.

# Evaluation Plan Overview

An evaluation plan is a document that guides the implementation of an evaluation. It includes a description of the program, the program model(s), and a detailed description of the evaluation research strategy (sample, measurement, measures, design, analysis plan, etc.)

## FAQs

### How long should each section of the evaluation plan be?

Each section should be complete, yet concise. The length of a particular section can vary widely based on what that section is and based on the nature of the evaluation plan. Some sections, like Program Mission, are supposed to be very brief. Other sections, like Measurement or Analysis will be longer or shorter depending on the number of evaluation questions, the diversity of approaches used, and (perhaps) the lifecycle stage of the evaluation (later stage evaluations may require lengthier text to explain, though not necessarily). Overall, the goal is to make a plan that is thorough yet concise and is readable and understandable by someone external to your program.

### What should each section of the evaluation plan include?

Each section should include a complete, concise description of what you plan to do and why you chose that strategy. For the Measurement section, for example, you should include a description of the data collection method(s) you have chosen and which tool(s) you plan to use, along with an explanation for why those choices were made. (Otherwise, skeptical readers might wonder, "Why is an online survey the best choice here?" or "Why will ten interviews be sufficient?") Without going into too much detail, explain why the approach you have adopted is most likely to generate credible, accurate, useful insights about the program in a way that is feasible. Overall, the plan should hang together well, have internal consistency, and serve as a good guidance and communication tool.

### What is the purpose of having a written evaluation plan?

The written plan serves a number of different purposes. First, the process of developing and writing an evaluation plan helps you be more intentional and thorough in your approach to evaluation. In the absence of a formal planning process, it is easy to focus too much on simply finding or creating a measure. Writing out the plan helps you think through and clearly articulate all components (purpose, questions, design, analysis, etc.) of the evaluation, and helps ensure alignment between those various components and the larger needs that this evaluation should serve. In addition, there are benefits to having it as a written document. The written document increases institutional memory, both about the evaluation and about the program itself. The written plan can provide continuity in how evaluation is conducted. Since evaluation is an iterative, evolving process, the plan provides a history upon which future evaluation cycles can be built. Last but not least, the written plan (like the logic and pathway models) is an excellent communication tool for internal and especially external communication. The plan can be included in reports to funders, in applications for funding, or in presentations to legislatures or other important stakeholders. To these audiences, the written plan exemplifies your commitment to and expertise in evaluation.

## Guiding Documents

- Tips for Writing Evaluation Plans

- Feedback Form for Evaluation Plan

# *Tips for Writing Evaluation Plans*

## **Introduction**

This document offers guidance to help you complete the Evaluation Plan for your program. The sections below are standard components of any written evaluation plan format. In addition, they correspond, in order, to the sections that are included as part of the Evaluation Plan Report in the Netway software system1. This document describes the kind of content that belongs in each section, considerations related to the intended audiences, and formatting or data entry issues if relevant. Deeper explanations of the evaluation concepts and the decisions that have to be made are available from other support materials that are referenced in each section.

This document is organized into the following sections:

Overall Considerations

    A.  Intended Audiences

    B.  Practical Information about Entering and Editing Sections of an Evaluation Plan

    C.  Overall Formatting

Evaluation Plan Sections

    A.  Mission

    B.  Program Description

    C.  Activities

    D.  Evaluation Purpose Statement

    E.  Evaluation Questions

    F.  Sample

    G.  Measurement

    H.  Design

    I.   Data Management and Analysis

    J.   Reporting and Utilization

    K.  Timeline

[1] Cornell Office for Research on Evaluation. (2012). The Netway [Software for Evaluation Planning].

# Overall Considerations

### A.       Intended Audiences

Individual programs have a diverse array of stakeholders, some of whom might read and/or use the Evaluation Plan, others of whom will not. In drafting your Plan, consider who will be reading it and keep their needs and interests in mind. In an effort to cover the most standard uses, the instructions here are written for Evaluation Plans with two sets of primary potential audiences. The first are internal audiences or, more specifically, audiences who are or will be close to the program itself. These include staff who will be conducting the evaluation and who therefore will be guided by this plan; staff in other programs who might learn from the evaluation approach; and future staff who might be managing the program and implementing the evaluation at some future date. For all these groups, the Evaluation Plan can provide an invaluable record of key program characteristics, the evaluation choices that were made, and the reasoning behind them.

The second potential audience is more removed from the program and would include current or future stakeholders who are interested in how the program will be evaluated. These might include organizational stakeholders such as the Board of Directors, administrative leadership, etc., and external stakeholders such as current or prospective funders or collaborators. For those who will be receiving evaluation results, the Evaluation Plan should provide enough information to allow them to understand the evaluation and assess how it was conducted. Parts of the Plan will often provide useful material for a report of evaluation results, so that readers make an informed assessment about the reliability and value of the evaluation results. For potential funders, the Evaluation Plan can provide evidence of the care and skill devoted to evaluation of this program.

The content in each section should be written appropriately for these audiences. Particular audiences might value individual sections more or less than others, or might require a different tone or level of detail and you are free to write your Plan with them in mind, but those specific adaptations are not covered here.

### B.       Overall Formatting

*Fonts and White Space:* It is likely that the content in some of the sections of your Evaluation Plan will have been cut and pasted directly from existing text (e.g. the program Mission statement or Program Description might have come from your website or program brochures.) When you are entering text into the evaluation plan, please make sure you keep your fonts consistent, and keep an eye on the "white space" (space between paragraphs).

*Complete Sentences:* It is advisable to include a short introductory sentence for each section of the Evaluation Report. For instance: "The sample for the 4-H Club Program Evaluation will be…" Logic Models can rely on bullet points and strong phrases in the same manner as a resume or CV. In contrast, an Evaluation Plan should be written as a stand-alone report and should use complete sentences.

78

# Evaluation Plan Sections

### A.    Mission

Since an Evaluation Plan is written for an individual Program, the Mission statement should be for the specific Program in particular, not the Department/Program Area or the overall Organization. The Mission statement should be concise and easily understandable, and should convey the overarching "big picture" goals of the program.

### B.    Program Description

The Program Description should be concise and understandable and should provide readers with a good complete view of what the program involves and what it strives to achieve. The Program Description is usually written as part of the program modeling effort but it should still "stand alone" because readers will not necessarily have access to the logic model or other information in order to understand what the program is about. So there should be enough detail to allow them to understand the program and be able to tell how the proposed evaluation will fit the program.

### C.    Activities (relevant if using the Netway)

You have the option of including a list of the Activity titles used in the program's Logic Model, and (if desired) also including any Activity descriptions that were entered into the Netway while creating the Logic Model. The descriptions can be useful if the audiences for the Evaluation Plan are not all familiar with the details of the program. If so, be sure to revisit the Activity descriptions to review for accuracy.

### D.    Evaluation Purpose Statement

The preceding sections of the Evaluation Plan are all about the program; this section is the first one devoted to the evaluation. The Evaluation Purpose Statement is an important opportunity to make the case for your choice of evaluation scope – that is, it's an opportunity to be clear and positive about why you have chosen to focus on certain program elements in this round of evaluation (since you can't evaluate everything), and how the results will be used. It should provide a concise overall explanation of the evaluation priorities encompassed by your Plan. It should include a short description of the evaluation work that will be done. It should describe what is and is not being evaluated and the goal/purposes of the evaluation. It sets boundaries by identifying the program elements (outcomes, for example) and time frame being considered, what the results will be useful for, and which goals or objectives are of most interest.

### E.    Evaluation Questions

"Evaluation Questions" in this context are not the kind of questions you'd include on a questionnaire for participants,. The Evaluation Questions are questions you are asking about your program or activities within it, for example: *"Is the program being delivered as planned?"* or *"Is the program having an effect on its target outcomes?"* or *"What is the effect of the activity on participants'*

*knowledge?"* These function as the driving questions at the core of evaluation planning for each Program. The remainder of your Evaluation Plan is guided completely by these questions. Every Question should have at least one measure in place (and ideally more than one) to try to get at an "answer", and every measure should have a Sample, a Design, a form of Analysis, and Reporting associated with it.

Evaluation questions should be presented in the form of a question. Ideally, they will follow somewhat the standardized format that highlights the model elements that are being explored, for example:

> **EQ1:  What is the effect of [ACTIVITY "X"] on [OUTCOME "Y"]?**

For the Evaluation Question section, you may wish to introduce the questions with a sentence or two before listing them. It will be helpful also to number the questions with a number or with something like "EQ1, EQ2…" so that subsequent sections can refer to individual questions easily.

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 use a colon and bullet points to collapse more than one evaluation question into a single sentence, such as:

> **What is the effect of ACTIVITY "X" (e.g. "Youth Club") on:**
> - **OUTCOME "P" (subject area knowledge)?**          **(EQ1)**
> - **OUTCOME "D" (subject area skill/mastery)?**     **(EQ2)**
> - **OUTCOME "Q" (self-esteem)?**                              **(EQ3)**

Evaluation questions need not be outcome-oriented, but might instead ask about program implementation. For example:

> **Was ACTIVITY "A" implemented well?**                   **- or -**
>
> **Was ACTIVITY "A" implemented as planned?**      **- or -**
>
> **Were participants satisfied with ACTIVITY "A"?**

### F.      Sample

The sample section describes the specific source(s) of evaluation data for each of your evaluation questions. With that in mind, be careful to write this section focusing exclusively on who (or what) you are <u>obtaining information from </u>in order to answer your evaluation questions. Do not fall into the trap of broadly describing the population served by your Program. For instance, if your Program was for mothers and 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, <u>not</u> the moms. And if you are only going to be measuring some of the babies in the program but not all of

them, then your Sample section should explain how that subset of babies will be selected from the overall group of babies.

The Sample section of the Evaluation Plan is driven by two principal objectives: to provide guidance to whoever will be implementing the evaluation when it does take place; and to provide information that will allow readers to determine how to interpret your results. For both of these objectives, it is important to explain *what the population of interest IS, and how the sample relates to the population of interest.*

The Sample section should also describe *how the sample will be selected.* For example, if your evaluation calls for you to survey a randomly selected subset of participants who complete the program, your Sample section should explain (briefly) how this random selection will be done. If your evaluation calls for you to interview a purposively-selected subset such as the high-achievers, or the participants from distant towns, etc., then the section should describe how you define "high achievers" or "distant towns" and how you then select the subgroup you will be getting data from.

If you are going to use the evaluation results to support claims that would apply generally to a larger group of people or contexts, then the Sample section will be essential in indicating how "generalizable" your results are. The Program Description section should have included a rough estimate of the number of participants predicted for the coming year. The Sample section should indicate what % of participants will be "sampled". This is part of what will allow readers to determine to what degree your results are "generalizable". (For instance, if you expect to have 1,000 participants, yet you only plan to sample 20, you might have a difficult time making a case for the Program's likely impact on the whole population based on that relatively small sample size.) Sample size is only one part of what makes generalization possible. Other factors, such as similarity of demographic characteristics (is your sample "representative" of the population of interest?) can also be important and need to be described.

Even if you are not interested in having generalizable results – because, for example, you are evaluating in order to identify desirable program revisions rather than to demonstrate potential effectiveness of the program for other groups –it is still important for the Sample section to describe how the specific sample is determined and obtained. The difference is just that the explanation is likely to be less statistical or numerical. In the example mentioned above, about the deliberate selection of "high achievers", for example, including the definition of "high achiever" and explaining how this group will be identified within the set of participants will make it possible for readers to decide for themselves whether your results are informative in the way you wish them to be.

## G.     Measurement

The Measurement section should list and describe each measure that will be used as part of the evaluation. The descriptions should be clear and straightforward *(e.g., a short post-only questionnaire, a series of one-on-one interviews, a 10 question satisfaction survey developed by the University of Wisconsin…)*. It should be made clear which Evaluation Questions are being covered by which measure(s). If using a published or researched measure, it should be listed by name and the reference to the source(s) should be included. If your measure was assigned by an outside party (e.g., Cornell faculty, or a funder), please indicate this.

Evaluation Plans can include measures that are planned or sought but are not yet in hand; they can also include measures that are available but that will be revised in specific ways to serve the purposes of your program. As such, the Measurement section should clearly list which measures

are "existing" vs. which measures are currently "in development" (along with a description of what sorts of adaptations are planned or needed.) For measures in development (meaning they will be either located or created), the description should include the measurement type and construct or constructs to be measured. <u>The more specific you can be, in terms of which measure(s) you will use and/or how you plan to change or adapt an existing measure, the stronger your plan will be.</u> Finding or developing measures is often a difficult or time-consuming part of the evaluation work. To ensure that your evaluation does not get derailed later on, do as much advance planning in the measurement section as possible. Note: for measures in development or revision, the Timeline section should be sure to include by when these measures will be located and/or created or revised internally.

## H.      Design

The Design section explains how your Program evaluation will be conducted, focusing on the sequence of steps and on the groups that will be participating in the evaluation. The choice of Design for your evaluation question will be driven by the nature of the evaluation question and the kinds of claims you would like (and can reasonably hope) to be able to make once data collection and analysis are complete, and by the program's lifecycle phase and the actual program context.

One option for presenting your Design is to use the "tic-tac-toe method" of using subscripted Xs and Os to outline evaluation design. **O**'s stand for observations (data collection points) and **X**'s stand for programs or activities. When multiple observations or measures are used on one occasion (e.g., at the end of the program) you can use subscripts to distinguish among measures.

As long as your Design description is concise and understandable, it can be represented in whatever form you are most comfortable.

All Design sections should have text descriptions, with enough detail so that the reader can tell when, <u>relative to the program</u> (or activity), data will be obtained. Annotation (X O) is encouraged but optional. Using the (X O) notation to supplement the text description is particularly useful for complicated evaluations that include multiple measures, follow-up observations, and/or different sample groups for example. In these cases, a "picture" in the form of Xs and Os is worth a thousand words. Be sure to clearly indicate which design is being used for each Evaluation Question.

<u>**Design Example:**</u>

The evaluation design for the Expanded Food & Nutrition Education Program (EFNEP) consists of an ERS pre/post questionnaire administered once in the first lesson of the series, and then again at the last lesson of the workshop series. (See O1 below). Logs, Progress Notes and Success Stories are collected throughout and summarized at the end of the series (O2, O3 respectively).

**O1      X      O$_1$, O$_2$, O$_3$**

Where            **O$_1$** = ERS pre/post questions

O$_2$ = Logs and progress notes

O$_3$ = Success stories (during and post-program)

### I.    Data Management and Analysis

The data management and analysis sections outline <u>how you plan to handle, and obtain results from, the data your evaluation will produce</u> (keeping in mind that "data" are not always numbers and spreadsheets – they can include video, documents, recordings of interviews, and so on.)  These sections are invaluable as advance planning steps, and should cover both the data management needs and a brief description of the actual analysis that will be performed.

### Data Management:

The data management description can be concise, but it will be very useful to anticipate and think through what will be involved in obtaining results from your efforts. Note: *these considerations apply to qualitative data just as much as they do to quantitative data!*  Answers to open-ended questions on a written survey or recorded during an interview, or activities recorded on video tape, are all data. They need to be collected properly, stored in ways that preserve anonymity and confidentiality (if promised), and organized in formats that facilitate analysis.

Begin to think through data structures <u>before</u> starting to collect data, so that you can be sure that you will be able to do what you need to do with the data you have recorded. For example, for data that will be quantitatively coded (such as a multiple choice survey), you will need to record the numerical score assigned to each answer. If you are storing the data in a worksheet such as Excel you will need a column for each question and a row for each respondent. (You will also need a person who has time and enough familiarity with Excel to do the data entry and to do it reliably.) If you are interested in group averages this may be all you need to record. However if you wish to compare scores from one class against scores from another, then your data record (and the survey tool itself) will need to show which class the respondent was from (another column). If you are going further and wanted to compare the responses of different age groups, or urban vs. rural subgroups, etc., then your survey and the data record would have to include the respondent's age or urban/rural residence (more columns), and so on. If you are doing a matched pre-post, then you will have to record the unique identifier for each respondent on the survey tool and in the data record. In summary, the evaluation question and design, plus any additional sub-group differences you may want to ask about (such as how results vary by gender), will determine the variables on which to collect information and this in turn affects how you will record the data and how the analysis will be performed

For a rough introduction to data management, consider the following steps. *Keep in mind, as noted above, that many of these apply to both quantitative and qualitative evaluations!*  Some of these are really preparatory and almost amount to housekeeping, but they can be critical for ensuring the quality of your results and they take time and staff resources which need to be considered in the overall "feasibility" assessment of the plan.

### Analysis

You do not have to specify the planned analysis in enormous detail, but it is important to look ahead and think about what sort of data you will be getting, and what you will be doing with it. This is important in assuring your data are reasonably complete and in a form that lends itself well to the analysis that will support the kinds of comparisons you want to be able to make. For qualitative data, including the answers to open-ended questions on surveys, the analysis might include formal text analysis, or even just looking for patterns of language choice, body language, sequences of actions. The results may then be summarized into numbers such as frequencies or counts. That is, although the language here is all about "data", be careful to think about how this applies to ALL forms of evaluation.

As with all the other evaluation plan decisions, knowledge of the program's evaluation lifecycle phase simplifies the decision considerably. For early lifecycle evaluations (phase IA or IB, for example) involving post-only assessments, the "analysis" might be as straightforward as simply summarizing the data (average score on a satisfaction ranking question, for example, or numbers of respondents offering each of the possible answers, or descriptions of patterns observed in qualitative data.) For evaluation lifecycle phase IIA doing (for example) an unmatched pre-test and post-test of outcomes, the analysis might involve comparing group averages of pre- and post-scores and doing some basic statistical tests to see if the differences are "significant". Similar tools will apply to phase IIB programs in which there's a matched pre- and post test. For purposes of the Evaluation Plan, what's important is to describe the kind of analysis that you anticipate will be needed in order to get answers that programs can use.

When you write up the Analysis section of your Plan, include some references to or descriptions of steps that will be relevant to this evaluation. There's no fixed standard for how much detail to go into. As a guide to future work, and as a check on feasibility, the more detail you are able to provide,

84

the better. *At a minimum*, note the issues that will have to be addressed. Data management and analysis will take time, and this should be allowed for in your Timeline. If the plan includes training staff in data entry and/or data analysis methods, these should be allowed for in the Timeline as well. As with all sections in the Evaluation Plan, be sure the Analysis section is logically connected to the Evaluation Questions, Measures, and Design. There should be a description of data analysis and management for each measure being used.

### J.      Reporting and Utilization

The Reporting and Utilization section describes how results from the evaluation will be used and shared with various stakeholders – including both internal uses for program improvement, as well as internal or external reporting for accountability, impact, etc. Reflect on what will be learned about the program in <u>this</u> evaluation:  Who might be interested in hearing about those lessons, and how you can communicate with those audiences most effectively? Decide whether the results should be incorporated into a current reporting structure or if they should be presented in a separate new report or format. Think back to the stakeholder analysis produced earlier in the process: which internal and external stakeholders to your program should receive a report? Most Programs currently have basic reporting mandates both internally and externally. If the evaluation results will contribute to these existing mandated reports, note that.

Once you have decided on how to make the most of the evaluation effort and the results that will be obtained, establish an explicit reporting plan listing the timing and type of reporting you will do (such as monthly informal reports at staff meetings, quarterly formal reports to funders, etc.)

### K.      Timeline

By now, you have a good idea of the various tasks associated with the Samples, Measurement, Design, Data Management and Analysis, and Reporting and Utilization sections. The Timeline section should clearly present *when* those tasks will be accomplished. The Timeline should span the entire evaluation effort, with clear start and end dates, and should show when materials will be obtained or developed, when they will be used, when data will be entered and analyzed, and when various reports will be prepared. Think of this as a Work Plan for the period of evaluation work. It might be helpful for your own planning and internal communication efforts to also include an indication of *who* will be responsible for ensuring that individual steps are completed. This is not necessary, but might be useful.

The Timeline offers valuable guidance to staff planning their work loads. It also offers one more opportunity to assess the feasibility of the overall Evaluation Plan. As with all elements of the Evaluation Plan, be sure that the Timeline is appropriate (relative to the rest of the Plan) and feasible (given your available time and other resources.)

# *Feedback Form for Evaluation Plan*

| Program Name: | |
|---|---|
| Review completed by: | Date: |

This form is intended to be a tool for external reviewers to use for providing systematic feedback on evaluation plans. It is also meant to provide useful guidance to evaluation plan authors themselves as they develop or revise their own evaluation plans.

This document has two parts. Part I is for an overall assessment, Part II contains more detailed section-by-section comments.

**Suggested process:**

1. Read through entire evaluation plan being reviewed
2. Provide general overarching comments in Part I
3. Complete Part II as directed
4. Revisit and revise your comments in Part I

## Part I: Overall Assessment:

Evaluation plans should:

- Provide an accurate, concise and coherent description of the program
- Explain what evaluation work is being planned and how the work will be accomplished
- Be internally consistent (the planned evaluation should be appropriate for the program's content, lifecycle stage, and stakeholder needs; and the elements of the evaluation plan should be consistent with each other (evaluation purpose, scope, questions, measures, sampling strategy, design and analysis plans).)

With these things in mind, please give some overall comments on the plan, an explanation of your assessment, and any further suggestions

## Part II: Section by Section Assessment

The categories below correspond to Evaluation Plan Sections. The short description in each section is intended to guide reviewers and ensure consistent feedback. Please add helpful comments.

**Program Mission Statement**

The Mission statement should be concise and clear, specific to the program and not to the larger organization, and should convey the "big picture" motivation for the program.

**Program Description**

The Program Description should be clear, concise, and should have enough information to give outsiders a good understanding of the program. Ideally it would include information about participants (number, age, background if relevant); main program activities and overall goals; basic information about how program is implemented (setting, frequency, who leads it, and so on as appropriate); and about the history or community context of the program.

**Evaluation Purpose Statement**

The Evaluation Purpose Statement serves almost as an "Executive Summary" for the Evaluation Plan. It should describe briefly what the upcoming round (usually a year) of evaluation work will include. It should identify the specific program elements that are the evaluation's focus, and should articulate the main goals of the evaluation and how the results will be used. Reviewers please comment on whether the evaluation goals seem appropriate relative to the program lifecycle and any other specifics of the program (such as stakeholder needs) that are known from the program description.

88

**Evaluation Questions**

Evaluation Questions form the basis for the entire evaluation plan, so this section is critical. The questions should be clear, specific, and should be formatted properly. They should be appropriate for the program's lifecycle, and should be consistent with the evaluation purpose statement. Questions should be clearly related to the program's logic and/or pathway model. If the question involves a comparison, the intended basis for comparison should be made clear (pre vs post, compared to another group, etc.). Reviewers should attempt to assess feasibility – is the number and difficulty level of the questions likely to be manageable? (Subsequent sections may make this more apparent). Ideally, Evaluation Questions should be numbered so that subsequent sections can refer to them consistently and without confusing readers.

**Evaluation Sample**

This section should concisely describe the sample that will be used to answer each evaluation question. It should provide enough information to assess whether the selected sample will serve as a sound basis for the claims the evaluation is intended to address. Accordingly, the "population of interest" should be made clear, sample size and composition should be described, and the recruitment or sample selection process should be described. Reviewers should comment on whether the sample(s) appear to be sufficient for generating evidence on the desired claims.

89

**Evaluation Measures**

There should be a measure for each evaluation question, and it should be easy for readers to tell which measure covers which question. Measure description should include its type (i.e., written survey, observational checklist, structured interview, etc,) along with other relevant details (length, who will administer it, etc.), and should clearly indicate the focal construct (what will the obtained information be about?) It should be clear whether each measure is already in hand, needs to be adapted (if so, how), or needs to be located or developed. For existing measures, proper references should be cited. If reliability and validity information are available that should be included. Reviewers should comment on whether the measure will be appropriate for generating evidence on the evaluation question, and whether it appears to fit the program and stakeholder needs.

**Evaluation Design**

The Design section should lay out the sequence of observations (when measures are implemented) and activities (the whole program, or specific components) clearly and concisely and should indicate the design type (e.g. post-only, pre/post, pre/post with comparison group, etc.) For simple designs, the description will be very brief. All measure should be included. If using comparison groups, the timing of when observations will be gathered from them should be indicated clearly. If symbolic notation is used it should provide clarity, and all symbols should be defined. Reviewers should comment on whether the design is appropriate for the program's lifecycle, and whether the design is appropriate and sufficient for generating evidence for the evaluation question.

90

**Data Management**

The evaluation plan should include a description of how each measure will be administered; how the data will be collected, handled, and stored; and (if applicable) how the data will be coded or scored. (All of this is in preparation for the actual analysis.) This information may be included in or distributed across the measures and analysis sections of the plan. Reviewers please comment here on the completeness, clarity, and quality of this aspect of the plan.

**Analysis**

The analysis section should clearly and concisely describe how all the data generated from all the measures and sample groups will be analyzed in order to obtain high-quality, credible answers to each of the evaluation questions. Reviewers should comment on whether the analysis strategy is appropriate and sufficient for generating evidence to answer the related evaluation question(s).

91

**Evaluation Reporting**

The reporting section should describe how all the results of the evaluation will be shared, for both internal and external purposes. Formal and informal reporting should be included, and the format, timing, and frequency should be indicated.  It can be helpful to organize this section by evaluation question but this is not the only option. However it is organized, the section should be clear and concise and should cover all the evaluation questions. The reporting plan should fit and fulfill the purpose and scope of the current evaluation.

**Evaluation Timeline**

The timeline should be given in calendar time (not just in relative terms) so that it serves as an effective work-planning calendar. It should include program and/or activity dates (whichever is relevant). For each evaluation question the timeline should cover the detailed steps of the evaluation work, indicating start and end dates for sample identification and recruitment; measure development and testing (if needed); data collection; data entry; data analysis; reporting. Ideally, the timeline will include time for measure revision and evaluation plan updating. The timeline is a good opportunity for assessing feasibility of the plan. Reviewer should offer comments on manageability of this work plan, if possible.

**Overall Presentation Quality**

Reviewer, please provide feedback on grammar, spelling and overall readability and general appearance of the evaluation plan.

92

# Evaluation Purpose Statement

The evaluation purpose statement is an introduction to the evaluation plan document that briefly describes the scope of the evaluation being planned, how it fits in with both prior as well as intended future evaluation work, as well as a brief summary of the methodology and intended use of the evaluation. Think of the evaluation purpose statement as an executive summary of your evaluation plan. This purpose statement will be useful for readers of the plan and as a touchstone document for people implementing the plan to make sure that it stays on track.

## FAQs

### What should be included in an evaluation purpose statement?

The Evaluation Purpose Statement should be brief, but should identify the purpose of the evaluation (what are you trying to find out, and why?), provide a concise description of what will be done in the evaluation (how will you do it?), and what parts of the program will be examined (which assumptions, activities, outcomes, or relationships will be focused on?), and explain how the results will be used. It should also explain how this evaluation fits in relative to past and possible future evaluations. In some cases, it may also be appropriate to explain in your Purpose Statement why your evaluation is not going to cover some issues that might be considered a priority by some stakeholders.

## Guiding Documents

- Guidance for Evaluation Purpose Statement

# *Guidance for Evaluation Purpose Statement*

The Evaluation Purpose statement will become part of your program's Evaluation Plan. It is valuable as part of the preparation for *any* evaluation effort[1].

The Evaluation Purpose statement should provide a short description of your evaluation effort. It should describe what is and is not being evaluated and the goal/purposes of the evaluation. It sets boundaries by identifying the program elements and time frame being considered, which audiences are being addressed, and which goals or objectives are of most interest.

**Example for an earlier lifecycle program/evaluation:**

*The purpose of this evaluation is to assess the extent to which participants in the Master Forest Owners Volunteer Training Workshop feel supported and well-equipped to share their forestry knowledge with other forest owners in their local communities. Considerations within the scope of the current evaluation include program structure and processes, curricular choices, and short-term outcome assessment. Other means of supporting forest management volunteers such as our newsletter and quarterly conference calls will not be assessed. Likewise, the program's long-term impacts will not be directly addressed.*

**Example for a later lifecycle program/evaluation:**

*The purpose of this evaluation is to assess the effectiveness of the Master Forest Owners Volunteer Training Workshop in supporting and prompting MFO volunteers to extend their knowledge to other forest owners in their local communities. A secondary purpose is to provide documentation and assessment information for use by persons considering replicating the model with other forest owner groups. Considerations include assessment of contextual factors which may affect program effectiveness and medium- and long-term impacts. Other means of supporting forest management volunteers such as our newsletter and quarterly conference calls will not be assessed.*

The program modeling and the "mining the model" steps in the evaluation planning process – and the resulting priorities and considerations – form a strong foundation for determining your evaluation scope and identifying the Evaluation Questions you intend to focus on in the next evaluation cycle. All of these will contribute to a succinct, well-defined Evaluation Purpose statement.

Netway (www.evaluationnetway.com)
Document available at: http://www.evaluationnetway.com/guide/evaluation-guidance/evaluation-purpose-statement

# Evaluation Questions

Evaluation questions are the broad inquiries *about the program* that the evaluation will seek to address. The language used in evaluation questions has broad implications for both the methodology of the evaluation as well as the claims that can be made as a result of completing the evaluation. A hallmark of a good Evaluation Question is that it is answerable, meaning that it is worded and structured in a way that sets up a feasible process that will result in credible, accurate, and useful data.

## FAQs

### How are evaluation questions different from survey questions?

Evaluation questions organize and direct your whole evaluation effort. Survey questions are the questions you would ask your participants or sample group to get information that will help you answer an evaluation question. (It may take many survey questions, and perhaps more than just a survey, to properly answer an evaluation question.)

### How should evaluation questions be worded?

The more precise your Evaluation Question, the easier your evaluation planning will be. Evaluation Questions should be precise in two ways. First, they need to identify in some detail the program components that are being explored (the activity or activities you are focusing on, the specific kind of knowledge, awareness, behavior change being examined, etc.). Second, the wording of Evaluation Questions must be clear about what kind of causal relationship (if any) between components is being explored. To illustrate the importance of causality, consider the difference between asking "How effective is this program at changing [X]…?" compared to asking "To what extent is this program associated with change in [X]?" The latter means you will be looking for patterns of change in participants; the former means you will also be testing whether your program caused the changes you may observe, which means you will have to have a comparison group or some means of ruling out competing explanations for the change.

Throughout the process of developing Evaluation Questions, keep in mind that the words you choose will determine both the evaluation methods you will need in order to answer the question, as well as the claims you will ultimately be able to make once you have your results.

### What if stakeholders are pressuring me to use an evaluation strategy that is out of alignment?

Unfortunately, this is often the case. There are many realities that have to be factored into the decisions in an evaluation plan. However, your understanding of the importance of lifecycle alignment (review that section, if needed) should help you make a case to those stakeholders that their requested evaluation strategy is not appropriate (and is probably an inefficient use of resources!) Sometimes, of course, you will simply have to meet the stakeholder's needs even if it is not an ideal evaluation strategy. In this case, the benefit of understanding the lifecycle alignment issue is that you will be better able to explain to the stakeholder what the risks or consequences of a misaligned evaluation strategy might be, including the possible misinterpretation of results (whether favorable or unfavorable). You also may be able to do some appropriately aligned evaluation work on a smaller scale, concurrent with the required evaluation.

## Guiding Documents

- Developing Evaluation Questions

- Evaluation Questions and Program Lifecycle

# *Developing Evaluation Questions*

**Goal**: *This worksheet is designed to help you develop evaluation questions from the key areas (priorities) you have identified in your pathway model.*

**Program Name:**_____

**Part I:** Brainstorm what you would like to know, and what kind of claims you would like to be able to make, about the priority area(s) identified in your pathway model.

1.  Briefly identify or describe the element(s) of your pathway model (assumptions, activities, outcomes, links) that you will focus on for your evaluation.

2.  What would you like to know about this aspect of your program?

3.  What are the claims you would like to make as a result of this evaluation effort?

4.  What are the key considerations for making your final decision about evaluation questions for this evaluation? (Feasibility? Usefulness? Accuracy? Credibility?)

*Begin developing formal evaluation questions on the back of this worksheet*

**Part II:** Based on your interests in Part I, develop draft evaluation question(s) on the lines below, using wording taken directly from your program model. Pay particular attention to relationship words (causes, contributes to, etc.) and how you describe constructs. For each draft question, consider the items listed below the line.

**Example EQ:** *To what extent do program participants report that they are more engaged with their community after participating in the program than before?*
**Possible claim:** *The majority of program participants report that they are more engaged with their community after participating in the program.*
**Key construct(s) to be measured:** E*ngagement*
**Implied comparisons (if any):** *[retrospective] pre-post*
**Notes on feasibility:** *high feasibility (one time self-report measurement, post-program)*

Q1: _____

Possible claim:

Key construct(s) to be measured:

Implied comparisons (if any):

Notes on feasibility:


Q2: _____

Possible claim:

Key construct(s) to be measured:

Implied comparisons (if any):

Notes on feasibility:


Q3: _____

Possible claim:

Key construct(s) to be measured:

Implied comparisons (if any):

Notes on feasibility:


Q4: _____

Possible claim:

Key construct(s) to be measured:

Implied comparisons (if any):

Notes on feasibility:

# *Evaluation Questions and Program Lifecycle*

*This table is not intended to create rigid rules for evaluation questions relative to program lifecycle. Rather, the goal is to build awareness of how evaluation focus and evaluation question form tend to vary across programs in different lifecycle stages, and to provide examples of careful EQ phrasing.*

| Program lifecycle | Element from LM or PM being evaluated | Evaluation question forms (examples) |
|---|---|---|
| Early phase | Assumptions | What are the starting conditions (e.g., in terms of community need, demographics, participants' access or prior knowledge, etc.) for this program? |
| | Activity (Program Process) | How is [Activity A] being implemented? |
| | | What is the level of participant engagement in [Activity A]? |
| | | How well is [Activity A] being implemented? |
| | Outcome | After program participation, how do participants perform on [Outcome X]? |
| | | After program participation, how do [some subset] of participants perform on [Outcome X] compared to [some other subset]? |
| | Activity- Outcome Link | To what extent, as a group, do participants engaged in [Activity A] change on [Outcome X]? |
| | | To what extent do individual participants engaged in [Activity A] change on [Outcome X]? |
| | | To what extent do participants engaged in [Activity A] change on [Outcome X] differently than comparable non-participants? |
| Late Phase | | To what extent do participants change on [Outcome X] differently compared to a group that differs only in their exposure to [Activity A]? |

# Measurement

Measurement is the process of collecting information systematically, using appropriate methods and/or tools, to address an evaluation question. At this stage in the evaluation planning process, this involves selecting a data collection strategy that fits the evaluation question you've identified, and then finding or developing whatever tools (often referred to as "measures") are needed in order to fulfill that strategy.

## FAQs

### How is my measurement strategy influenced by my program lifecycle?

As is usually the case, there is not an absolute set of rules dictating which measurement approach should go with which lifecycle stage. But there does tend to be a pattern linking lifecycle stage and measurement strategies. Very early stage programs tend to be best served by measurement approaches that are exploratory and less intensive, providing rapid feedback about the processes of program implementation. Later stage programs tend to benefit from more intensive approaches to measurement, designed to yield appropriately structured data that will allow you to answer more complicated evaluation questions.

### What is unobtrusive measurement?

Unobtrusive measurement is, as it sounds like, a process of data collection that does not intrude on the individuals, activities or program involved. The benefits are twofold: it reduces the risk of biased response or behavior on the part of the sample group, and removes the burden of response from those who would otherwise be asked to complete a survey or be interviewed, etc.
Here are two quick examples of creative unobtrusive measurement (taken from the Research Methods Knowledge Base (RMKB), cited below): A museum wanted to measure which of its exhibits was most popular with the public. Instead of stopping visitors to survey them, the staff replaced the flooring with news tiles that wear down quickly, and then assessed which exhibits had the most worn-down tiles in front of them (of course there is room for error, such as mistaking the exhibit nearest the bathroom or gift shop as being excessively popular!) Another example is from a marketing firm that wanted to know what radio stations people were listening to. Again, instead of surveying people they asked auto mechanics to record the station to which car radios were set when cars were brought in for service. This may be an imperfect sampling strategy which contains certain biases, but does have some advantages over conducting an expensive phone or mail survey.

While it is a good idea to brainstorm possible unobtrusive measurement options, you may find that your context and constructs simply do not lend themselves to this kind of approach. For more on unobtrusive measurement, visit the RMKB page on this topic, found here: http://www.socialresearchmethods.net/kb/unobtrus.php

### Where can I learn more about possible measurement strategies?

The field of evaluation has generated an extensive array of resources and research on measurement. Here is a source with clear, concise guidance and examples, specifically designed for program implementers:
Taylor-Powell, E., & Steele, S. (1996). Collecting Evaluation Data; an Overview of Sources and Methods. Retrieved January 11, 2012, from University of Wisconsin-Extension Cooperative Extension, Program Development and Evaluation Unit Web site: http://learningstore.uwex.edu/Assets/pdfs/G3658-04.pdf

### What determines which measurement strategy is most appropriate for my evaluation?

Your well-crafted evaluation question is your starting point, because it already takes into account an array of important considerations (lifecycle, stakeholders, etc.) and has implications for the kind of data you need. The measurement strategy should serve the evaluation question, collecting data that is credible and accurate in a way that's feasible for the program and its resources.  For instance, if the evaluation question has to do with gaining a deeper understanding of how the program affected participants in potentially unexpected ways, then

a measurement strategy involving focus groups, interviews, or reflective journaling would be more appropriate than a survey with closed-ended response options. The important thing is to ensure that there is a logical and well thought out connection between the evaluation question and the chosen approach to measurement.

### What is the difference between measurement and actual measures or tools?

In an evaluation or research project, measurement refers to the overall process of gathering and recording relevant data (quantitative and/or qualitative). A measure is a specific tool designed to collect the kind of data you need. In the evaluation plan document you can think of these as companion parts. The measurement section will describe and explain your choice of measurement strategy, while the measures section will identify the specific tool(s) that will be used and how they were (or will be) obtained or developed.

## Guiding Documents

- Getting to Measures Worksheet – Explained and Illustrated

- Getting to Measures Blank Worksheet

- Key Constructs and Measurement

- Introduction to Measurement and Measures

- Obtaining a Measure – Find, Modify, or Write

- Measure Checklist

- Finding Measures

- Survey Review Form

# *Getting to Measures Worksheet*
## Explained and Illustrated

*This document follows the steps in the "Getting to Measures Blank Worksheet", using an example with explanations. A diagram at the end illustrates the thinking process. This worksheet is an exercise in fine-tuning, exploring alternatives, and considering less familiar options before making a plan. The effort is valuable for finding the best measurement strategy for what you want to know.*

## 1. Draft Evaluation Question:

Be precise about what you want to know about your program. Write this inquiry out as a formal Evaluation Question.

*The standard form for Evaluation Questions organizes the inquiry around specific elements of the program model, identifying the activity and/or outcome of interest. It also indicates clearly what strength of causality, if any, you are seeking to assess. The process of exploring options and developing a measurement strategy inevitably leads to some re-thinking of the Evaluation Question, so plan on revisiting this as you iterate through the steps.*

## 2. Clarify the "constructs" in your Evaluation Question above – what exactly do you mean by the activity and/or outcome you are focusing on?

*For example, suppose an outcome in the Evaluation Question refers to "enjoyment of cooking". Does this mean enjoying the process of cooking? Or enjoying producing a complete meal? Or enjoying preparing dishes from scratch? Does "cooking" exclude heating up prepared frozen foods to produce a complete meal? What about salads and things that don't involve a stove? Are you interested in enjoyment just during the class sessions, or beyond class? Or do you really only care about enjoyment of cooking because your program is designed on the premise that those who enjoy cooking are more likely to buy fresh and potentially locally-grown foods? (In this case the outcome might be better stated as "enjoying using fresh locally grown ingredients")…What, precisely, do you mean by this idea of enjoyment of cooking? What is your definition of this "construct"?*

*It can be helpful to get more than just one person's view of the meaning of the construct. Even among the program staff there might be different views of what is really of interest, and the views of stakeholders might be even more different. (Since stakeholders might well be the ones to whom the evaluation results will be reported, specifying a definition that would be appropriate to them and consistent with what they are thinking might be very important – and it can be surprising how much people's views about a shared item can actually differ!)*

*Another useful step for clarifying the intentions or definitions is to consult other resources – for example, if it's a classic 4-H construct (like one of the Essential Elements), look it up in published 4-H resources, find out how it has been handled in the research literature, etc. This is useful not only for clarifying the construct, but also for possibly aligning your evaluation with others that have been done before. That is, if a research-based definition "fits" what you are looking for, and if there turn out to be evaluation tools connected to that line of research, then you will have a great lead on a potential measure.*

**Note: once this thinking has been done, revisit the Evaluation Question above and see if it needs to be revised to capture this more precise specification of the construct you want to measure.**

105

3.  **Using this sharper definition of the constructs, work through the following:**

*The following are small sub-steps in an overall question of "How could you know?"  The breakdown into small steps might seem unnecessarily incremental, basically an exercise in splitting hairs, but it can help ensure that evaluation planners come up with a full range of options for answering the Evaluation Question. As the example demonstrates, there are multiple options for each of the possibilities that emerge, forming a tree that branches repeatedly (see diagram at end).  If you don't keep the options alive at each step, you might miss some.*

*It is <u>not</u> essential that you fill (a), (b) and (c) out in this order, but it is useful to cover all of them in some order or other.  Often people's thinking takes them right to types of evidence (item (b)) for example, but if that's where your mind goes then fill in (b) but don't forget to also do (a) and (c) and in both sections try to come up with multiple examples.  The point of splitting this overall "How could you know" question into pieces is to tease out a wider range of options to consider so that you can choose the best one.*

**(a) What might this "look like" or consist of in practice?** (List as many options as you can.  Be creative – think outside the "usual" options.)

*The construct identified above is an abstract statement of what you are looking for.  The next step is to "operationalize" this – identify what it would mean in the real world, or in practice.  What are the "pictures in your head" that come to mind when you think about a person or group doing this Activity, or having this Outcome?  If you know it when you see it, what would you be seeing? Brainstorm how you (or others) might recognize something when you (they) see it. As before, it is useful to get multiple contributions to this brainstorming.  Ask people to think about how they'd know it if they saw it, and then what it is that would have persuaded them.*

*For example, if the construct is defined more precisely as" enjoying producing meals from scratch", it might show up as someone taking time in the grocery store to choose fresh produce, having more meals at home and fewer at a restaurant, spending less money on pre-prepared foods, giving more invitations to friends to come over for dinner, smiling more while cooking, choosing to cook when they don't "have to", talking to friends more often about meals prepared at home, trying new recipes, expanding their collection of spices, etc. (See even more options in the diagram at the end of this document.)*

**(b) What might serve as evidence?** For the most promising candidates in (a), write down various kinds of evidence that would be informative about whether this has occurred.

I*n the above example, evidence might include the proportion of someone's grocery bill that included raw ingredients, the contents of their trash (vegetable peels vs empty prepared food containers!), the number of recipes that they actually use in a given month, the amount of time they spend on food preparation in the kitchen, the number of dinner invitations they issue, what they say about how they feel about cooking, etc.*

**(c) Review the strengths and weaknesses of the options,** taking into account several aspects that matter: "closeness" to the real thing; accuracy and reliability.  Write down a short list of the most promising candidates from (b).

*Personal statements from people about how they feel about cooking is close to the "source", so to speak.  However there might be some bias in what they say depending on who they are saying it to.  So it might be best to combine their personal statements with some outside observational evidence like the proportion of their grocery bill that goes to raw ingredients rather than pre-prepared foods*

*(you'd need some basis for deciding what constituted "more"— that is, does it change compared to pre-program levels, or does it differ from what a comparable person who doesn't like cooking might spend.)*

**4. How could you gather this evidence?** Again, think of as many options as you can and list them here. (Possibilities might include video-taping a demonstration, live observation during class sessions, directly asking participants, asking people who know the participants, testing the end-products, etc.) Identify the ways that seem best, taking into account accuracy, feasibility, fit with program context and target population. Indicate the top choices among those on your list.
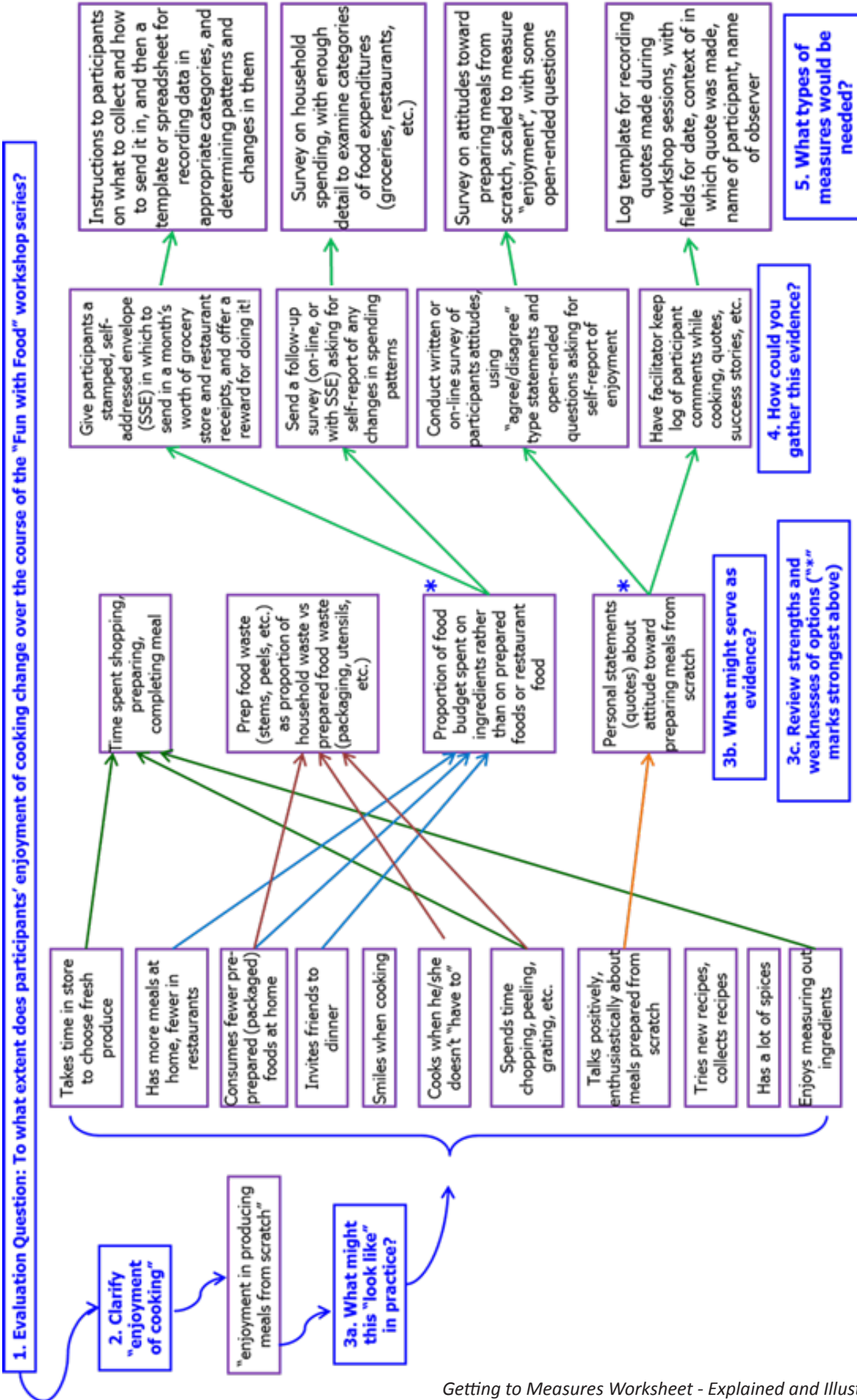
*For example, to get personal statements about how they feel about cooking you might gather it by interviewing them directly, or by interviewing their housemates to get quotes or recollections, or by asking participants to fill out a survey about attitudes toward cooking, or by keeping a log of participant comments overheard by the cooking instructors or stories the teachers noted at the end of each class. Observers might be able to capture indicators of an attitude like enjoyment of cooking through a carefully developed observational checklist.*

*If this is a large class, direct interviews might be too time-consuming. Identifying and contacting housemates is likely to be very time-consuming and since it provides second-hand information it is likely to be weak. A written survey might be feasible, but if class time is short it might not fit in well. Having instructors capture quotes at the end of each class might be a quick way of capturing basic indications of attitudes. It would not burden participants, and if your stakeholders are receptive to quotes this might be a good choice. Similarly, an observational checklist, if of sufficient quality, would be relatively unobtrusive, could be completed by a trained volunteer observer or by the facilitator, and would mean that data would be available right at the end of class time. See the diagram for additional ideas.*

*Once again, revisit your Evaluation Question phrasing in 1(b) and see if it needs updating. When finalized, it becomes a central component of the written Evaluation Plan.*

**5. What measures would be needed?** For your answer(s) to 4, what type(s) of measure would allow you to gather evidence in a way that would support analysis? Here is a list of possible measure types from which to choose: case study, interview, observation, group assessment (e.g. focus group), expert or peer review, portfolio review, testimonial, test of knowledge or skill, photograph, slide or video, diary or journal, log, document analysis, action cards, simulation, problem story, creative expression, unobtrusive measures.

*For the above, one measure would be a Quote Log with appropriate fields for recording all the data needed (date, class, comment made, etc.) Another measure would be the observational checklist of in-class "enjoying cooking" behaviors to look for. To conduct interviews, you would need an interview guide with a carefully selected list of questions in the order you want to ask them. You might also ask students to fill out a brief survey near the end of the program. Additional examples are in the last segment of the diagram.*

**1. Evaluation Question: To what extent does participants' enjoyment of cooking change over the course of the "Fun with Food" workshop series?**

**2. Clarify "enjoyment of cooking"**

"enjoyment in producing meals from scratch"

**3a. What might this "look like" in practice?**

- Takes time in store to choose fresh produce
- Has more meals at home, fewer in restaurants
- Consumes fewer pre-prepared (packaged) foods at home
- Invites friends to dinner
- Smiles when cooking
- Cooks when he/she doesn't "have to"
- Spends time chopping, peeling, grating, etc.
- Talks positively, enthusiastically about meals prepared from scratch
- Tries new recipes, collects recipes
- Has a lot of spices
- Enjoys measuring out ingredients

**3b. What might serve as evidence?**

- Time spent shopping, preparing, completing meal
- Prep food waste (stems, peels, etc.) as proportion of household waste vs prepared food waste (packaging, utensils, etc.)
- Proportion of food budget spent on ingredients rather than on prepared foods or restaurant food  *
- Personal statements (quotes) about attitude toward preparing meals from scratch  *

**3c. Review strengths and weaknesses of options ("*" marks strongest above)**

**4. How could you gather this evidence?**

- Give participants a stamped, self-addressed envelope (SSE) in which to send in a month's worth of grocery store and restaurant receipts, and offer a reward for doing it!
- Send a follow-up survey (on-line, or with SSE) asking for self-report of any changes in spending patterns
- Conduct written or on-line survey of participants attitudes, using "agree/disagree" type statements and open-ended questions asking for self-report of enjoyment
- Have facilitator keep log of participant comments while cooking, quotes, success stories, etc.

**5. What types of measures would be needed?**

- Instructions to participants on what to collect and how to send it in, and then a template or spreadsheet for recording data in appropriate categories, and determining patterns and changes in them
- Survey on household spending, with enough detail to examine categories of food expenditures (groceries, restaurants, etc.)
- Survey on attitudes toward preparing meals from scratch, scaled to measure "enjoyment", with some open-ended questions
- Log template for recording quotes made during workshop sessions, with fields for date, context of in which quote was made, name of participant, name of observer

*Getting to Measures Worksheet - Explained and Illustrated page 4 of 4*

# *Getting to Measures Worksheet*

***Complete one worksheet for each of your selected Evaluation Questions.***

**1. Draft Evaluation Question:**

**2. Clarify the "constructs" in your Evaluation Question** above – what exactly do you mean by the activity and/or outcome you are focusing on? *(When done, revise the Evaluation Question above if needed.)*

3.  Using this sharper definition of the constructs, work through the following:

   **(a) What would this "look like" or consist of in practice?** (List as many options as you can. Be creative – think outside the "usual" options.)

   **(b) What might serve as evidence?** For the most promising candidates in (a), write down various kinds of evidence that would be informative about whether this has occurred.

   **(c) Review the strengths and weaknesses of the options,** taking into account several aspects that matter: "closeness" to the real thing; accuracy and reliability. Write down a short list of the most promising candidates from (b).

Netway (www.evaluationnetway.com)
Document available at: http://www.evaluationnetway.com/guide/evaluation-guidance/measurement .

**4. How could you gather this evidence?** Again, think of as many options as you can and list them here. (Possibilities might include video-taping a demonstration, live observation during class sessions, directly asking participants, asking people who know the participants, testing the end-products, etc.) Identify the ways that seem best, taking into account accuracy, feasibility, fit with program context, and target population. Indicate the top choices among those on your list.

*Once again, take a moment to revisit your Evaluation Question phrasing and see if it needs updating. Once you are satisfied, put this Evaluation Question in the Netway.*

**5. What types of measure would be needed?** From your answer(s) to 4, now ask yourself what type(s) of measure would allow you to gather evidence in this way. Here is a list of possible measure types from which to choose: *case study, interview, observation, group assessment (e.g. focus group), expert or peer review, portfolio review, testimonial, test of knowledge or skill, photograph, slide or video, diary or journal, log, document analysis, action cards, simulation, problem story, creative expression, unobtrusive measures.*

For more explanation of measure types, see:

Taylor-Powell, E., S. Steele (1996) Collecting Evaluation Data: An Overview of Sources and Methods, Cooperative Extension Publications, University of Wisconsin, Madison, WI. (http://learningstore. uwex.edu/Collecting-Evaluation-Data-An-Overview-of-Sources-and-Methods-P1025C237.aspx )

*Getting to Measures Worksheet page 2 of 2*

# *Key Constructs and Measurement*

**Goal**: Brainstorm measurement approaches while maintaining alignment with evaluation questions.

**Instructions:** For each evaluation question, identify the construct(s). *("Construct" refers to the sometimes hard-to-define idea or variable in your EQ. For example, "knowledge", "engagement", "uptake" and "interest" are all constructs commonly found in program evaluation questions.)* For each construct, brainstorm as many indicators and possible measurement tools as you can; you can select from these options later.

| EQ: | | | |
|---|---|---|---|
| Construct to be examined | How do you define this construct? | How do you know it when you see it? (What are some possible indicators?) | How might it be measuresd? (What tool(s) would you need to capture it?) |
| | | | |
| | | | |

# *Introduction to Measurement and Measures*

In an evaluation or research project, **measurement** is the process of gathering and recording observations and information, also known as collecting data.

A **measure** is a tool used to collect this information.

Choosing and obtaining an appropriate measure(s) for each evaluation question will likely be one of the central challenges in writing and executing an evaluation plan.

Below is a brief discussion of several concepts related to measurement and measures, and how they are inter-related.

## A. Quality of Measurement and of Measures:

The quality of measurement is an important pillar of credible and useful evaluation results. Quality of measurement depends on the quality of the measure, the appropriateness of the measure in the context of the program and its evaluation, and the manner in which that measure is administered. What follows is a discussion of features of a high quality measure.

The quality of a measure is determined by three primary characteristics (1) accuracy ("validity," in evaluation terminology); (2) consistency ("reliability") and (3) fitness (to the program setting, the program lifecycle stage, target population, stakeholder needs, and staff resources.)

(1) Validity – The extent to which a measure yields information (data) that is accurate with respect to the construct being measured.

Measures are efforts to get at the "truth" of something – for example, did program participants actually change with regard to the expected outcome? The idea of the outcome may be quite clear to those conducting the evaluation, but translating the idea into the real world and measuring it can be difficult.

For example, the expected outcome might be "leadership skills". The extent to which a particular measure accurately captures the essence of "leadership skills", and allows for strong statements about the program's effect on "leadership skills" and not some other, related idea, like "confidence", is called "construct validity"[1].

A measure might have weak validity for any of a number of reasons. In a survey, for example, the questions might simply not capture what they are intended to capture. This weakness might be apparent to a reasonable outsider or it might only be evident to a specialist in the field. It could also be that the wording of the survey obscures the underlying truth. For example, a measure that is intended to assess youth self-esteem might be more revealing of participants' level of openness or capacity for self-disclosure because of the language used in the measure's questions. A measure designed for adults that is used with children may yield data that are heavily influenced by respondents' reading skills rather than by the actual construct being measured. A measure with vague or misleading language that is open to multiple interpretations will yield a range of responses, leading to biased results.

---

[1] See "Idea of Construct Validity" in Trochim, William M. The Research Methods Knowledge Base, 2nd Edition. Internet WWW page, at URL: http://www.socialresearchmethods.net/kb/considea.php.

**Assessing validity:** Unfortunately, there is no standardized objective measure of validity. This is an inherently negotiated dimension of measure quality. But there are systematic and rigorous approaches to assessing and demonstrating validity, and these are important factors to consider (or establish) when using or developing a measure

In practice, evaluation researchers approach the assessment of validity in a number of ways. "Face validity" might be viewed as the bare minimum, and refers to the idea that, "on its face", the measure appears to do a reasonable job of capturing what it is intended to capture. A more stringent criterion would be "expert validity", in which the measure has been reviewed by an appropriate set of experts, and has been deemed to be a valid measure for the construct in question (of course, "appropriate" and "experts" can be a matter of judgment as well). Additional efforts to establish validity take the form of systematically checking how well a measure performs. For example, do scores on the measure correlate well with other outcomes that it would be expected to predict? These and other aspects of validity are beyond the scope of this paper but there are excellent resources with more information[2].

For current purposes, the main points are: (1) that published research on measures should provide a description of what was done to assess the validity of the measure in other studies, and this information should be taken into account on the positive side of the ledger when selecting a measure; (2) when using a home-grown or adapted measure, it will be important to take some steps to assess its validity and to document choices made in constructing the measure[3];  and (3) in either case, reports of evaluation results should provide information on the measure's validity, as this can (rightly) be part of what makes the evaluation report credible.

---

[2] See "Measurement Validity Types" in Trochim, William M. The Research Methods Knowledge Base, 2nd Edition. Internet WWW page, at URL: http://www.socialresearchmethods.net/kb/measval.php.

[3] Reviewing the literature on a construct, even if these don't include actually measures, can be helpful in assessing and demonstrating how well the new measure meets accepted definitions of interpretations of the construct. Pilot-testing a newly-developed measure is an essential minimum level of validity assessment. For guidance on pilot-testing a measure, see "Planning and Conducting a Pilot Test" from the Corporation for National and Community Service, at http://www.nationalserviceresources.org/node/19498.

**(2) Reliability** – Reliability is the ability of a measure to yield consistent or repeatable results.

In any situation, an observation is a mixture of whatever the "true" thing is, plus some "noise" or measurement error. For example, if a test is intended to measure math skills, the score a student receives on a given day is a combination of his or her "true" math ability (as captured by the measure), plus some positive or negative variation such as whether the person slept well the night before, had a distracting student near them, got lucky when guessing, etc. Reliable measures are ones that can be expected, on average, to yield the same results over time, in different contexts, when administered by different people, and across different sample groups.

**Assessing reliability:** Statistically, the reliability of a measure can be estimated in a number of ways, using various correlation statistics[4].  Reliability coefficients range from 0 (meaning that the results of the measurement are completely due to random error) to 1 (indicating that the measure has perfect reliability.) Therefore, a reliability coefficient of .5 indicates that about half of the variance of the observed score is attributable to the true level of the construct in the participant, and the other half is attributable to random factors, such as whether they had breakfast. A reliability coefficient of .8 means the score is about 80% based in true level and about 20% random error.

If possible, choose a measure that has already been reviewed for reliability by researchers in this field (peer-reviewed literature is a good source of information.) There is no firm rule for what constitutes acceptable reliability, as this varies by field. Moreover even measures that have been found to be reliable in a range of circumstances still need to be considered carefully with respect to the particular target population and setting in the program being studied.

**Note: Validity and reliability are related but different.** It's possible to have a measure that is reliable (yielding similar scores each time) but not valid (not measuring the intended construct) or valid (questions are on topic) but not reliable (results are too sensitive to the participant's mood, the time of day, etc.)[5] . It's best to have both.

---

[4] Evaluators use a number of standardized types of reliability estimates, such as internal consistency, test-retest, inter-rater reliability, and others. (One commonly used statistic is "Cronbach's alpha", which is an estimate of the internal consistency of a measure.)  For detailed explanations and a discussion of how to compare reliability statistics, see "Types of Reliability" in Trochim, William M. The Research Methods Knowledge Base, 2nd Edition. Internet WWW page, at URL: http://www.socialresearchmethods.net/kb/reltypes.php.

[5] For a good discussion and nice graphic of the possible relationships between reliability and validity, see "Reliability & Validity" in Trochim, William M. The Research Methods Knowledge Base, 2nd Edition. Internet WWW page, at URL: http://www.socialresearchmethods.net/kb/measval.php.

**(3) Fitness –** Appropriateness of the measure given the evaluation and program context, lifecycle and stakeholders

How well a measure fits a particular program and the specific constructs intended in the evaluation is also an essential element of measure quality, separate from the validity and reliability that may have been established by researchers in other study settings. For example, even if a measure has been tested by researchers and found to be valid in the context of their study, if the construct it measures doesn't truly match the construct being measured in the current context, then it's not a valid measure <u>for the current purpose</u>. Similarly, if a high-quality measure is not well suited to the current program's context (program fidelity or lifecycle stage, participants, setting, stakeholders, etc.), then the results that would be obtained from it would not be as accurate or reliable as the published indicators of quality might suggest.

If it's not feasible to find a pre-screened measure that captures the construct at hand and is well suited to the participants and context of the current evaluation, then candidate measures may need to be modified, combined with another measure (or measures), or it may be necessary to start from scratch. In each of these cases, it is essential to pilot the new measure with the intent to assess its validity, reliability, and fit.

**Assessing fitness:** There are no standardized quantitative assessments of "fitness" in the way it is defined here. However careful attention to program context and the construct in the evaluation question can help ensure fit, and should certainly be described in a report of evaluation results, and ideally in the measures section of the evaluation plan itself.

The <u>purpose of the evaluation</u> needs to be kept in mind when choosing a measure and weighing its strengths and weaknesses. Be aware, for example, that in too closely tailoring a measure to the current program (something akin to "rigging the test"), it will be difficult or even impossible to demonstrate how the program fares relative to externally accepted definitions of the construct of interest, or to compare results to those in other programs with similar goals that are using consensus measures.

## B.    Forms of Data

In program evaluation, data collected using measures generally takes three forms: 1) demographic or descriptive data (describing participants or program); 2) process data (assessing program and its implementation); and 3) outcome data (assessing participants, communities, etc.). Each form may come in two types – either qualitative or quantitative. Qualitative data is generally text-based rather than numeric, while quantitative date generally refers to numerical representations of observations. It is important to be cautious when making this distinction, because all evidence has dimensions of both[6].

On the next page are some examples:

[6]Encyclopedia of Evaluation, Mathison, Sandra (editor), Sage Publications Inc., 2005 pp. 345-50 and p. 351.

|  | quantitative | qualitative[7] |
|---|---|---|
| descriptive | demographic characteristics of a group, rendered in percentages | information about the quality of participant skills, rendered in text form. |
| process | counts of how many participated in an activity | assessments of the rapport between program staff and program participants, rendered in text form |
| outcome | test scores for participants who completed a program | participant reflections on how their behavior has changed, rendered in text form |

Depending on the evaluation question and design, both qualitative and quantitative data may need to be coded or scored and analyzed in appropriate ways.


## C. Levels of measurement [8]

The distinctions in following categories are important. They have implications for how to interpret data from the variable in question, and for what type of analysis can be done with each kind of data. For example, it is not uncommon for a report to summarize Likert scale responses by reporting the "average" rating ... but since the rating scale only assigns meaning to the integers used in the scale (1, 2, 3, 4, for example), a 3.7 is not a defined answer. It is better, in such cases, to report results in terms of the percentage of responses in each of the four categories.

> Nominal (or "categorical", or "name") data is collected using response options that are labeled, and is frequently based on a quality or trait – boys or girls; blonde, brunette or redhead. The labels are helpful for organizing and extracting meaning from data, but the labels themselves don't necessarily imply order, ranking, or relative value.

---

[7] For references on analyzing qualitative data, see http://learningstore.uwex.edu/Analyzing-Qualitative-Data-P1023C0.aspx
For references on content analysis and thematic "coding" of qualitative data, see http://www.ischool.utexas.edu/~palmquis/courses/content.html and http://academic.csuohio.edu/kneuendorf/content/

[8] For more information, see "Levels of Measurement in Trochim, William M. The Research Methods Knowledge Base, 2nd Edition. Internet WWW page, at URL: http://www.socialresearchmethods.net/kb/measlevl.php; also "Data Levels and Measurement" by David Garson, NC State University, at http://faculty.chass.ncsu.edu/garson/PA765/datalevl.htm.

Ordinal ("order" or "rank") is used to differentiate between logical order or degrees of something such as first, second and last; or Associate's, Bachelor's, Master's and Doctorate. Likert scales are a type of ordinal data. Nominal and Ordinal data are commonly summarized using percentages.

Interval (or "continuous") data is ordered data based upon a consistent scale. Fahrenheit temperature is a common example – the difference between 65 degrees Fahrenheit and 66 degrees Fahrenheit is the same as the difference between 100 degrees Fahrenheit and 101 degrees Fahrenheit. The difference is interpretable, but there is no absolute zero (0 degrees does not mean the absence of temperature) and therefore you can't say that 100 degrees is twice as hot as 50 degrees. Dates are another type of interval data.

Ratio data is interval data that has an absolute zero, and a score of 100 is twice that of 50. Height, weight, age, etc. are ratio data. Interval and ratio data are typically summarized as averages.

In cases where audience for an evaluation requires a certain level of analysis, it may make sense to choose the level of measurement accordingly.

## D.     Types of Measurement Methods

Different methods of data collection are used in different circumstances or for different goals. For example, consider the drivers' test at the DMV. The capacity to be a good driver includes knowledge of rules of the road. The DMV issues a driving permit test for this. But they also want to measure skills, so the tool they use is an observational checklist which is completed during the driving test. If they only used a paper test, they would be missing important data crucial to the success of their Program (and the safety of the roadways). If someone wanted to look at group trends they may do secondary analysis of test scores across the country.

1) **Direct measurement** includes those methods that solicit direct feedback from participants - such as a test, survey, or interview, or that require the researcher to be physically present (participant observation). Direct measures include pencil and paper instruments (test, survey), in-person or phone interviews, or electronic assessments (Survey Monkey or other on-line instruments), and simulations. They may also include a direct observation while completing a checklist, or collecting data such as height and weight.

On the next page are some of the most common types of direct measurement[9]:

---

[9] For more information on measurement types, see

Powell, Ellen-Taylor "Collecting Evaluation Data; an Overview of Sources and Methods" http://learningstore.uwex.edu/Collecting-Evaluation-Data-An-Overview-of-Sources-and-Methods-P1025C237.aspx.

For more on the advantages and limitations of common measurement types, see
Creswell, John (2003) Research design; qualitative, quantitative and mixed methods approaches, 2cnd ed., p. 186.

| Common Types of Direct Measurement | | | | |
|---|---|---|---|---|
| | observer is: | items are: | mode is: | use to: |
| surveys | participant | standardized questions, standard or open-ended response options | researcher administered or self-administered | assess individual attitudes, knowledge, opinions, easily compare across individuals |
| interviews | participant | structured, semi-structured or unstructured questions, follow-up prompts | researcher verbal prompts, or projective techniques such as participant drawing | assess individual attitudes, knowledge, opinions in depth; develop survey questions from themes |
| observations | researcher | checklist | researcher as unobtrusive observer or researcher as participant | assess individual or group knowledge, attitudes, and especially behavior or skills. |
| focus groups | participant and researcher | semi-structured questions, facilitation techniques | researcher as facilitator in group discussion | assess individual attitudes, knowledge, opinions; especially useful for revealing group dynamics and interactions and calling up hidden information elicited by group process. |

**2) Unobtrusive measurement** refers to methods of data collection that don't require the researcher to intrude in the research context. It consists of several different methods – indirect measures, document analysis and secondary analysis.

*Indirect measures* are those measures that occur naturally within a research context – such as through video or photographs, attendance and registration records, and analysis of other outputs, such as drawing and projects (i.e. birdhouses). It is important to be aware of the ethics of collecting information without the participants' knowledge. Some researchers refer to the use of teacher and parent interviews to learn about a child as indirect measures, but since that requires the presence of the researcher, we would call those activities direct measures, even though it isn't direct interaction with the object of study.

*Document analysis* is the analysis of documents, typically to look for themes and major ideas. Examples include news articles, field notes, reports, and memos. The study may result in the text being broken down by word, phrase, sentence, or theme.

*Secondary analysis* is the analysis of already existing data, and typically refers to the re-analysis of data collected from one or more other projects or resources (databases). Some of the most common data sources for already available information:

- Historical/archival records

- Administrative records

- Outputs (attendance records, registration forms, diagrams, posters, etc)

- Notes/photos (process information)

# *Obtaining a Measure—Find, Modify, or Write*

There are three main ways to obtain a measure: (1) find a pre-validated measure, (2) modify a pre- validated measure, or (3) write a measure from scratch. To decide how to approach this task, circle the items in the table below that best correspond to the needs of the evaluation. Consider the tradeoffs before beginning to nail down measures. Note your choices and the reasons behind them for later use in reporting results.

| | Find[1] | Modify[2] | Write[3] |
|---|---|---|---|
| **Best choice when...** | Want a measure pre-screened for validity and reliability | Want as many of the benefits of a pre-screened measure as possible | Benefits of feasibility and customization outweigh need for pre-screened for validity & reliability |
| | EQ implies a strong claim about effectiveness | EQ implies moderate claims about effectiveness | EQ does not imply a strong claim about effectiveness |
| | EQ is based on a complex or contended construct | EQ construct (or target population) almost but doesn't quite match pre-screened measures available | EQ is based on a simple, non-controversial construct |
| | EQ calls for comparing results to those for similar programs using the same measure | Want or need at least loose comparison with results for similar programs | Close fit to program as delivered is more important than comparing results to similar programs |
| | Finding a measure that closely fits the program, context and target population is likely (because common, etc.) | Relatively few adjustments are needed to modify pre-screened measure | If simple measure, doesn't require much time |
| **Drawbacks** | Search costs likely to be low because measure is readily available from system | Costs of modifying likely to be less than cost of writing new measure | Costs of writing own measure likely to be less than costs of finding, modifying. |
| | Cost to get permission to use may outweigh value of added validity | Some of the validity and reliability of the pre-screened measure is lost | Difficult to write in accord with best practices, so may introduce measurement bias |
| | May not closely fit desired construct or target population | May still not adequately capture desired construct or adequately fit target population | No pre-established validity or reliability, so claims may be weaker |

**1. Find:** Look for a pre-validated measure using peers, the larger system in which the program sits, professional groups delivering similar programs, government agencies providing oversight, and, if all else fails, an online search.

**2. Modify:** Adjust wording of pre-validated measure (or assemble desired scales from multiple pre-validated measures) in order to more closely fit outcome construct of interest, target population, or context.

**3. Write:** Construct your own survey, knowledge test, interview, focus group or observation protocols in accord with best practices for writing measures.

Netway (www.evaluationnetway.com)
Document available at: http://www.evaluationnetway.com/guide/evaluation-guidance.

# *Measure Checklist*

This checklist can be used when assessing or re-assessing a potential measure, and as a guideline for things to consider when adapting, modifying, or developing a measure.

☐ **Is it the "right tool for the job"?**

Does the measure answer the evaluation question? Are you looking at an outcome about skills, knowledge, attitude, etc. – if so, then how well does the construct the measure has been developed for match the construct you want to address in your evaluation question? If the primary intention is to identify program improvements, will the measure yield helpful information? If the evaluation tool is a broad measure covering a number of constructs, does it address all outcomes you are interested in? Does it address some that you do not need? (For example, the ERS forms used by many Nutrition Programs measure multiple outcomes such as knowledge, skill, and behavior.)

☐ **Is it appropriate for the Program Lifecycle?**

Does the measure fit the appropriate stage of the program and evaluation lifecycles? Although this is not a hard and fast rule, newer programs tend to be well served by more unstructured and open-ended measurement types like observational methods, field notes, success stories, checklists, and so on. Programs that are further along in their lifecycle with some evidence already established tend to call for more quantitative testing, formal performance assessments and checklists, attitude scales, etc.

☐ **Will it "work" for the sample group?**

Consider the literacy, age level, cultural background, and other special characteristics of your target audience.

☐ **Will it "fit" into the Program setting?**

Think about the length of contact time with participants, venue (outdoors vs. indoors), and other factors that may influence how effectively and consistently the measure can be administered.

☐ **Will the results "speak" to the key stakeholder(s)?**

Consider the stakeholders who will be reviewing your results and the data you gather. Stakeholders can have different preferences or constraints, and may put more weight on some kinds of evaluations than on others. In some cases, you may want only qualitative or quantitative data, or you may want a mixture. Be sure that your measure(s) will meet the needs of your audience(s). In many cases, the strength of the research base for the measure could be very helpful in making your results credible and compelling to funding agencies, etc.

☐ **Is it feasible?**

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 the technical skills and time available to use this measure well?

☐ **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.

☐ **Quality of the Measure**

This is the "bird in the hand vs. two in the bush" decision. A program may have a choice 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. (What constitutes good values to look for may depend on other aspects of your evaluation plan.)

Note: If using established or named measures, be sure that they are properly referenced. Measures developed and field tested by others should be properly cited. If you have modified or adapted an established measure in some way, be sure to describe these changes and explain the reason(s) for them, and cite the original measure appropriately.

Finally, 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.

124

# *Finding Measures*

In order to locate existing measures that have been used by other researchers and evaluators, and have shown to be reliable and valid, it is important to know how to conduct effective literature searches for scholarly articles[1]. This type of research may lead to useful information about previous evaluations conducted on similar constructs, as well as the measures used.

In order to ensure that measurement efforts lead to results that are credible, useful, accurate, and feasible to obtain, it is important to be aware of the existing tools that may meet current measurement needs so that they can be considered alongside their "home grown" counterparts. Before beginning a search for existing measures, it is essential to do the careful thinking that goes into logic modeling and the initial steps in the evaluation planning process - namely developing high quality evaluation questions. Without prior knowledge of the program's measurement needs, it is impossible to make informed decisions about which measures will be most appropriate.

### Where to begin

Before beginning the measurement tool search process, the program description, pathway model, evaluation questions, and purpose statement should be on hand to ensure that any measure that may be considered is aligned with previous thinking and planning. In addition, all measures that are currently being used or considered should be made available for comparison. Then, begin by seeking out programs similar to the one currently being evaluated. This may provide a starting point for subsequent searches, or may lead directly to a measure that will require further consideration.

It may also be valuable to find out if there is measures repository website maintained by the state or federal funding agency responsible for overseeing the program at those levels, or by an association of professionals who deliver programs similar to the one being evaluated. These are sometimes contained within evaluation resources or evaluation "toolkit" websites specific to a program area or organization. Often, this can be a good shortcut to finding measures of interest listed by indicator or outcome. (Note that for information about the validity and reliability of these measures, you will have to check the research literature cited in them.)

### Literature Searches

Locating measures reported in scholarly articles is a two-step process. Typically, the article or resource related to the construct being measured is located first. Often the measure itself is listed as a reference in the article and must be located separately.

Extensive research has been done on many of the outcomes commonly shared across programs. However, only a few of the related scholarly articles will pertain to the specific outcomes of interest

---

[1] Scholarly articles are reports on original research or experimentation that have been written by and for scholars in a particular field of study, articles written about original research, or a new application of others' research by an author who has been credentialed in the particular field.

Netway (www.evaluationnetway.com)
Document available at: http://www.evaluationnetway.com/guide/evaluation-guidance/measurement

to a particular program, and even fewer will refer to evaluation resources with established validity and reliability. A successful search involves experimenting with different topic headings and keywords in multiple databases, and narrowing the searches down from there.

Here are some tips for getting started:

1. Before looking for articles revisit the evaluation question(s) and pull out key terms, phrases, or concepts that further specify the topic of interest. You can use these as key words in measurement searches.

2. Use the "search" or "find" function within your PDF software. With this function, you can search within the article for terms such as "measure," "instrument," and "tool;" or you can search the topic key terms. The shortcut for searching is "ctrl-f" on PCs and "apple-f" on Macs.

3. Search various databases. Databases vary widely, and different databases offer different information, even on the same resources– Some may offer only basic citation information (basic public access databases such as ERIC.gov, PubMed.gov, http://nsdl.org/, Google Scholar, AGRICOLA, etc.) and may require further searching through your University's library catalogue to retrieve the original articles. Your affiliated University most likely provides its affiliates with exclusive access to several licensed databases that offer annotated or more in-depth information than may be available at public sites. Examples of these include ISI Web of Knowledge, SocAbs, PsycINFO, and JSTOR.

Some databases are small and focused on a single or specialty topic, others cover a broader range of topics. Consider what field of academic discipline your outcomes are in, such as education, healthcare, or nutrition. Keep in mind that a University often has several libraries, and it can save time to start in the library website for that subject specialty.

| Topic | Database |
|---|---|
| Education | ERIC |
| Sociology | Sociological Abstracts |
| Science | Web of Science, SciFinder Scholar |
| Agriculture | AGRICOLA, CAB abstracts, AGRIS (international Ag) |
| Food science and nutrition | Food Science and Technology Abstracts |
| Psychology | PsycINFO |
| Business, social sciences, physical and life sciences, math, music, and other fields | JSTOR |
| Business, sciences; education; law; and others | Wiley Interscience |

1. Begin broad and narrow it down. Begin with a broad search and use that to identify other key words that are related to your searches and outcomes. There may be common jargon that is used popularly in the literature. Many databases will provide options for similar or related topics or other keywords to use, others may provide a thesaurus for exploring

126

optional keywords – a thesaurus search for "attitudes" at the Educational Resources Information Center (ERIC) is shown below. You can see how the related terms and the narrower terms may lead to more focused searches.



"Interest in science" is a common construct, as well as its relationship to career choice. A Google Scholar search shows almost 5.5 million hits. But if you put quotations around the phrase it is reduced to 20,500.



At the National Science Digital Laboratory (NSDL) there were far fewer results (133,134), and a focus on K-12 dropped that down to less than 13,000. But a search at the ERIC had less than 5,000, and can be sorted alphabetically or according to relevance or date.
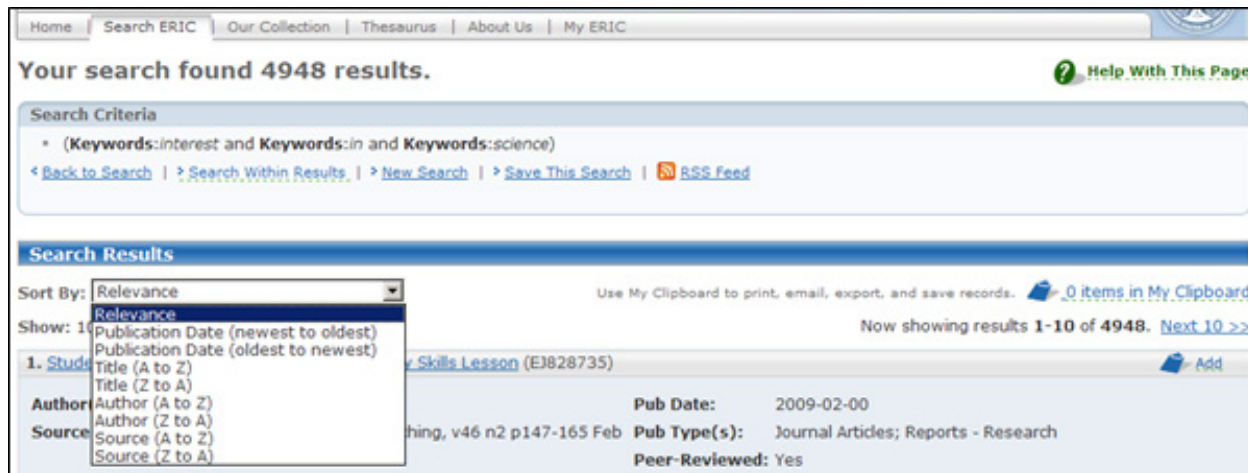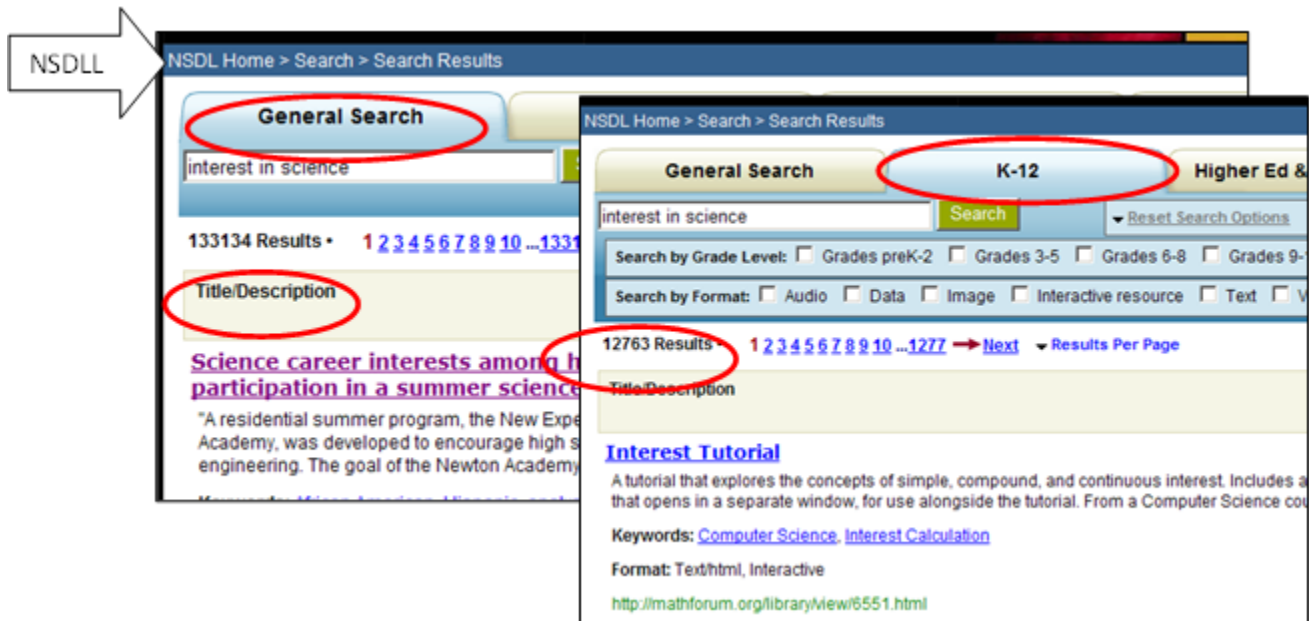
At the NSDL, a search for "interest in science" brings up an article "Science career interests among high school girls one year after participation in a summer science program." This record suggests other "keywords" including: "Career influences," "career interest," and "gender differences". At ERIC the other "descriptors" or key words for the same article included "career choice," "science education," "student attitudes," and "sex differences".

Each search phrase leads to more relevant studies. Experience will show that different terms may be more or less effective depending upon the database you are searching. Skim an article summary or abstract while keeping in mind your research questions to help you determine whether a particular article is of interest.

With these three searches you can begin to see the potential of literature searches for finding measurement resources for your programs and useful background information from scholarly work done in the area of your construct of interest.

1. Keep a record of your searches. Select a method of tracking the date of your search, where you have searched, what you have found, and what your key terms were. You may also want to record the number of results you received, and where you went from there. Keep track of your search so that you or the next researcher to work on the project does not spend energy duplicating a search that was already done.

These records can be kept electronically in a text document or spreadsheet; others may wish to track their searches in a notebook. Another tip provided by the Cornell Library is to use Zotero (to download this tool, please visit http://www.zotero.org/), which is integrated into your Firefox web browser and assists in keeping track of the records of interest.

2. Explore both basic and advanced searches. This requires learning how to search each database. Basic search functions usually provide for simple key word, title, author, or article searches within a particular database or journal, whereas advanced searches allow for multiple pieces of information to simultaneously direct a search. Different databases may have different rules for searching, and they usually offer some guidance on how to search. The use of quotes around phrases, Boolean searches (which use true/false logical operators such as "and", "or", "not", "if" and "then") and the use of wildcards (*), question marks and the like are database specific. Look for directions on how to conduct an advanced search at each database.

129

Try one of these methods for linking to more resources after you have identified a good article:

> • Search directly for authors whose research has come up in several previous searches, or for additional articles by authors of your article

> • Search for keywords within a journal that has provided several valuable articles

> • Return to an article you have already used and skim its references for relevant titles to search, or follow links to references (see Web of Science, right)

> • Follow links that may be provided to similar articles (see related items from ERIC and Google Scholar, below.)

> • Follow links that tell you who has cited particular articles (see citations from Web of Science, right, Google Scholar, below)
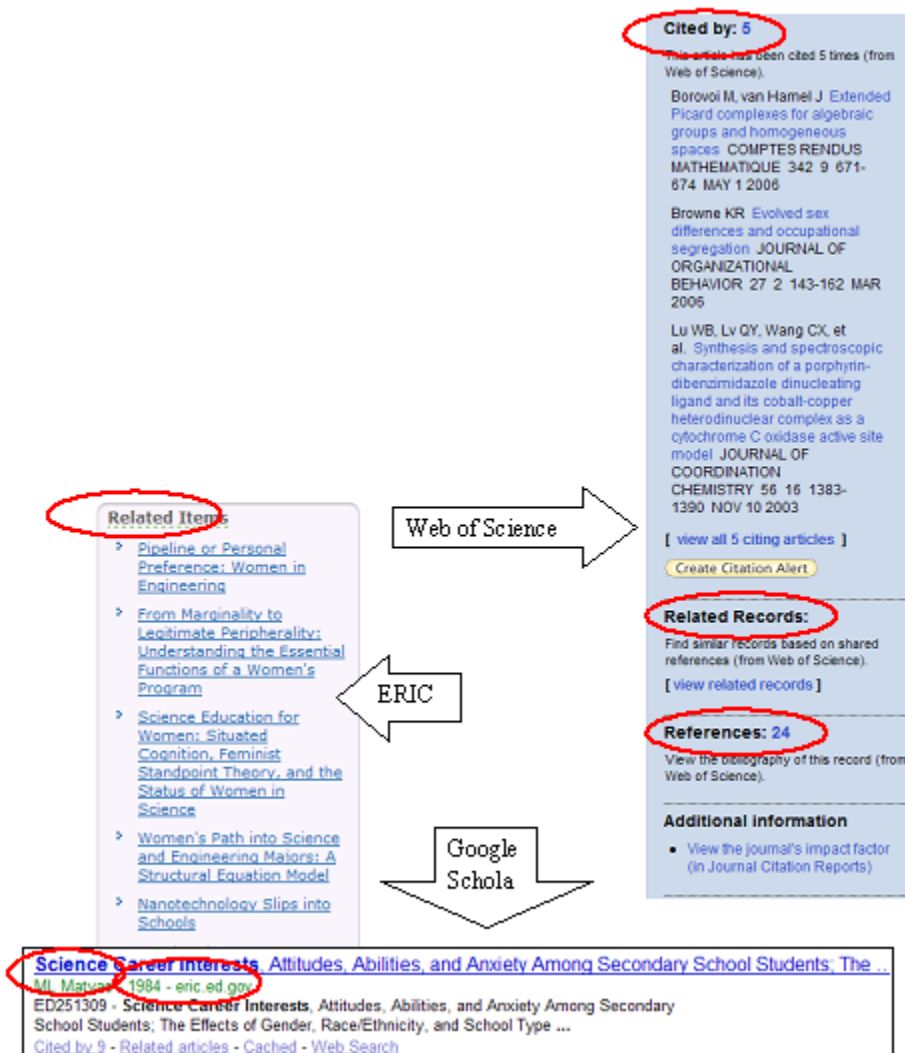
**Locating Measures that Work**

Having located the article, the next task will be to locate the actual measure. Different situations call for different approaches, but here are some basic steps you can take to find the measures referred to in the article:

- Start out by looking for the measures within the article itself. Sometimes, the measures are included in the research results section, or may be included in an appendix.

- Look for references to the measure. Find and read the section where the author explains the "method," "procedure" or "measures" used in conducting the research. Usually, there is a sentence about where or how they obtained these measures. From there you can determine if the measure was, for example, purchased or developed and how to ultimately obtain it. At times the author cites the article where the measure was obtained. In that case, search for the referenced article to obtain the measure. At other times, any reference to the measure may only be a reference to an article where you can get more information on the measure.

- If there is no appropriate citation to the measure, conduct a search on the title of the measure. Any well-documented measure should be revealed by a direct search by title.

**Other Resources**

Below are some potential web resources that can assist you in your literature review. At the time of this publication, they were all active links, but it's possible that they will become outdated. Additionally, the quality of the sites may change over time, and this list doesn't even begin to exhaust all of the good websites out there!

**Measurement Tool Databases:**

Field-tested Learning Assessment Guide
http://flaguide.org/tools/newsearch.php.

Out-Of-School time database and bibliography
http://hfrp.org/out-of-school-time/ost-database-bibliography

Compendium of Assessment and Research Tools
http://cart.rmcdenver.com/index.cgi?autoid=7258

Assessment Tools in Informal Sciences
http://atis.pearweb.org/

Online Evaluation Research Library
http://oerl.sri.com/

**Instrument Toolboxes**

Place-Based Education Evaluative Collaborative: Research and Evaluation Instruments
http://www.peecworks.org/PEEC/PEEC_Inst/

Authentic Assessment Toolbox
http://jonathan.mueller.faculty.noctrl.edu/toolbox/index.htm

Learning in a Technology-Rich Environment (LITRE) Student Learning Toolkit
http://litre.ncsu.edu/sltoolkit/ToolKitEntry.html

131

**Boolean Searches**

http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Boolean.pdf

**Literature Reviews and searching**

http://library.usm.maine.edu/tutorials/esp/module2/module2.htm

**Data management resources**

http://www.zotero.org/ ( Free)

http://www.endnote.com/

http://www.refworks.com/

**Evaluating Resources**

http://www.library.cornell.edu/resrch/intro#2Findingbooks,articles,andothermater

http://library.usm.maine.edu/research/researchguides/webeval.php

# *Survey Review Form*

Measure Name: _____

Program Name: _____ Date: _____

This is a review form for polishing a newly developed survey or similar measure, such as a set of interview questions, using four main criteria. Feedback comments in the four categories should be filled into the boxes below, following the guidance in each section. Revised measures should be pilot-tested before they are used.

**1) ALIGNMENT:** Survey items should be aligned with evaluation questions, and all intended evaluation questions should be covered.

Are there extraneous items – ones that impose a burden on respondents without pulling their weight by helping answer EQs? Do all intended EQs (those that are supposed to be addressed by this measure) have corresponding items in the survey? If causal claims are implied in the EQs, does the measure include items that help rule out competing possible causes? Do questions match the desired unit(s) of observation (individual, group?) Do measures include items asking for demographic and other information needed for subgroup comparisons (program site or date, participant demographic data, duration of participation, etc.?)

**2) RESPONDENT FRIENDLY:** Surveys should respect respondents' time and effort, need for anonymity or confidentiality, sensitivities.

Is there an introduction explaining the purpose of the survey and why their contributions are valued? Are there transitional explanations between different sections where needed? Are there expressions of appreciation, and assurances of confidentiality if applicable? Do reading level and cultural frame of reference match intended audience? Do the questions ask for information that is feasible for respondents to know, or easily look up? Is there a chance that the questions could make someone uncomfortable or even distressed?

## 3) TECHNICAL QUALITY

Does question wording follow best practices (avoiding double-barreled questions and matching question format to analytical purpose)? Does question order follow best practices (start with an easy question and get harder, start with most important questions and end with demographics)? Is the layout uncluttered and easy to read? Is the wording bias-free, jargon-free and easy to understand? Are there enough but not too many questions? Are instructions clear and easy to follow? Are questions phrased in a way to ensure consistent interpretation? Do response categories make sense?

## 4) ANALYSIS-READY

Does the structure of questions accord with the analytical purposes of the evaluation? Does the response format allow for the type of analysis that can support claims implied in EQs? (If assessing change, for example, will the questions yield the kinds of answers that would make it possible to detect change?)

**References:**

"Questionnaire Design: Asking questions with a purpose", Ellen Taylor-Powell, University of Wisconsin Extension (1998), http://learningstore.uwex.edu/assets/pdfs/G3658-2.PDF .

See also other University of Wisconsin Extension materials on surveys and questionnaires at http://www.uwex.edu/ces/pdande/evaluation/evaldocs.html.

134

# Sample

In the context of program evaluation, the sample is the group (of people, objects, etc.) that you will collect data about in order to address the evaluation question(s). For some evaluations, it will be appropriate to collect data about the entire population of interest (for example, all program participants); for other evaluations, the sample may be a sub-group of the population of interest, selected based on the needs of the evaluation and practical considerations (feasibility, resources, etc.).

## FAQs

### Where can I learn more about different sampling strategies?

There are many good sources of information on sampling.  Here are a few that we recommend:

Research Methods Knowledge Base: Trochim, William M. The Research Methods Knowledge Base, 2nd Edition. Internet WWW page, at URL: http://www.socialresearchmethods.net/kb/sampling.php

Taylor-Powell, E., (1998). Sampling. Retrieved May 5, 2015, from University of Wisconsin-Extension Cooperative Extension, Program Development and Evaluation Unit Web site:  http://learningstore.uwex. edu/Assets/pdfs/G3658-03.pdf

### How should I describe my sample in my evaluation plan?

In describing your sample, it is important to explain how your sample group relates to the population of interest. The population of interest depends on your evaluation question. If your evaluation question is about your program participants (for example about their demographics, their satisfaction with the program, or about their level of knowledge on a topic covered by the program) then we would say that your "population of interest" is the set of program participants. If you intend to collect data from all of your program participants (assuming that's feasible) then the sample is the same as the population of interest. This kind of sample is referred to as a "census." If you only intend to collect data from some of the program participants, then you need to describe how that subset (your sample) is defined and how it relates to the population of interest.

However, if your evaluation question is about a larger group (for example if you are interested in assessing how well your program is likely to work for members of a community, or youth in a certain age range, and so on) then the description of your sampling plan should explain the extent to which this a representative sample (in terms of demographics, experience levels, place of residence, or other critical characteristics). Note that if you are interested in generalizing to a larger group, but the only individuals you have access to are the ones who took your program, then you really have a "convenience sample." In order to make your claims accurate, you should take care to describe the critical characteristics of your actual sample group and note how similar or dissimilar they are from the population of interest.

### What determines which sampling strategy is most appropriate for my evaluation?

First, be clear about the purpose of your evaluation – who are you interested in learning about and being able to make claims about? This will be your "population of interest." If it's a small and easily accessible group, such as current program participants, then you would probably choose to sample that entire group because that's feasible and provides exactly what you want to know.

However if you are interested in generalizing from the results in your sample group to a larger population, then you will need to be careful about how your sample relates to the population of interest. At a minimum, you might want to gather data about the characteristics of your sample group in order to compare and contrast them to the population of interest. Or you might be engaged in a more challenging evaluation and might need to ensure that your sample is drawn appropriately from the population of interest, so that your claims based on the sample results can legitimately be extended to the population of interest. Random sampling strategies would be appropriate in this case.

In contrast, sometimes you might have reason to deliberately not choose a representative sample. If, for example, your purpose is to examine characteristics of participants who succeeded particularly well in your program, then your sampling strategy would be to identify the high-achievers and obtain data from all or some of them. This would be one example of a non-probability sampling strategy.

## Guiding Documents

- Developing a Sampling Plan

- Sampling Plan Worksheet

# *Developing a Sampling Plan*

## Preface

There are some key threads that provide structure to the intertwined decisions in an Evaluation Plan. These are:

- **What do you want to know about your program?** This is a key determinant of the Evaluation Questions, and is itself shaped by the program's structure and logic, by stakeholder interests, and by program development needs.

- **Closely related: What kinds of claims do you want to be able to make?[1]** These considerations help sharpen the Evaluation Questions, and influence the choice of Sample (who will you get information from?), Design (what will be compared to what?), Measurement strategy (what will you be looking for?), etc.

Key questions in determining your Measurement Strategy:

- **What would provide "evidence" of what you are trying to assess?**

- **How might you gather that evidence?**

So it's all about <u>gathering evidence</u> that will be <u>useful for YOUR purposes</u>. These purposes might be purely internal, if you are focusing on program development and improvement; or external if you need to communicate to stakeholders about program activity and/or effectiveness; or both. "**Sampling**", or <u>selecting the group from which to gather information,</u> extends this decision-making process, to make sure that the evidence you gather will be able to support the kinds of claims you want to be able to make, or conclusions you want to be able to draw.[2]

## Key terms in Sampling:

A. "POPULATION" (or "population of interest"): *The "population" is the entire group that you are interested in learning about or making claims about.*

*If the data are being gathered from non-participants or from objects (databases, or physical records for example), then the "population" is the entire set of sources related to or produced by the group of interest.\**

---

[1] Obviously, regardless of the "claims you might want to make," the evaluation might give the opposite answers. This is still valuable information, as it will promote program improvements or a better allocation of scarce time and resources, etc.

[2] There are a number of good references on sampling that supplement what is discussed here. Taylor-Powell, E. (1998) Sampling, Cooperative Extension Publications, University of Wisconsin, Madison, WI. (http://learningstore.uwex.edu/Sampling-P1029C237.aspx ) covers most of the concepts here and offers careful explanations and how-to instructions for obtaining different types of samples (random samples, purposeful samples, etc.)  Another excellent reference is http://www.socialresearchmethods.net/kb/sampling.php

*A subtle distinction to clarify:

In a lot of evaluation writings and discussion, the terms "population" and "sample" are described only in terms of sets of people rather than (as included in the definitions above) sets of data. If your measurement strategy involves getting data about people from the people themselves (participants, for example) then it is appropriate to describe your sample as consisting of the subset of people.

However, it is important to keep in mind that your sources of data might be different from the people you are actually interested in. For this reason the definitions above include the language about sets of data. For example, it might be a good measurement strategy to obtain data about the people you are interested in by obtaining it from *other people* (observers, teachers, etc.) or from *objects* (such as journals, record books, food prepared in a cooking demonstration, etc.)

> **Illustration**: Suppose you are interested in evaluating the post-program equipment handling skills of your participants. Suppose your measurement strategy is to test the pieces they produce from using the equipment (in order to assess consistency, quality, amount of material wasted, or whatever is of interest). The *population* in this example would consist of all the products produced by the participants in the program you are studying. The *sample* would consist of the selected subset of the products you actually intend to test.

*Note*: This distinction is an example of how the decision about measurement strategy interacts with the definition of sample and the decision about sampling.

**Examples (these will be carried forward in subsequent sections):**

(1) For an early lifecycle evaluation focusing on program implementation (such as participant reactions to the program that was offered) or program outcome (such as post-only knowledge of the participants in the spring of 2010), the "population" could be the full set of participants in the program in the spring of 2010.

(2) For a later lifecycle evaluation that is intended to yield conclusions about the program's likely benefits for future teen participants in the community, for example, the population could be all eligible teens in the community.

(3) For an evaluation that is intended to assess which program elements were most effective for the high achievers in past sessions of the program, the population could be all the high achievers in all past (presumably you'd have a date range of past sessions you wanted, and you'd have a definition of what you mean by "high achievers").

B. "SAMPLE": *The "sample" is the set of individuals you intend to get data from, or the set of data you intend to collect in order to conduct the evaluation*. The sample is either equal to, or is a portion of, or is systematically related to, the population of interest (see Figure 1). The process of figuring out what the sample will be, relative to the population, is the sampling decision.

**Figure 1: Sample relative to population**



Sample selected from population
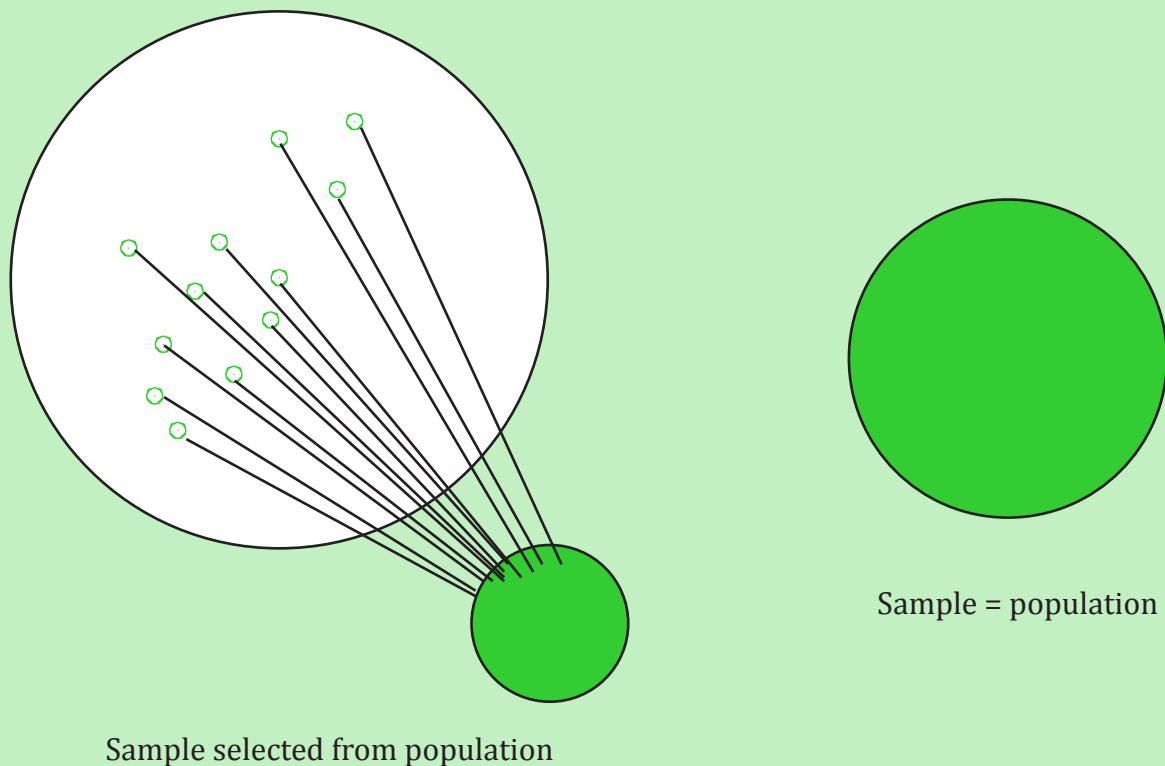
Sample = population

Examples (cont'd from above):

(1) For this evaluation, the sample could be the whole set of participants (sample = population). However, if the group is large, or if time is just too limited to include everyone, the sample could be a selected subset of the participants. How you select that subset, and how it relates to the overall population, is what has to be decided in sampling.

(2) For this evaluation, where there is a need to "generalize" the evaluation results in order to make claims about the program's likely effect on individuals who have not yet participated in the program, the sample might be all or some of the recent program participants, selected so that they will be "representative" of all eligible teens in the community. So, there has to be a systematic relationship or correspondence between the characteristics of the sample and the characteristics of the population[3].

(3) For this evaluation, where you are not necessarily attempting to generalize, you still need to do "purposeful" sampling. This example would involve "outlier" or "extreme case" sampling – deliberately selecting the "high achievers" from among past participants, and

---

[3] In order to draw really strong conclusions you'd have to address the fact that, by definition, the program participants are "different" from other teens in the County because for reasons you might or might not know, these youth chose to do the program and the others did not. So you might need to control for that unavoidable bias by incorporating a matched comparison group of non-participants. But it will often be sufficient for your purposes to make less aggressive claims and be able to say, for example, that future participants are "likely" to experience the same benefits.

selecting the sample according to a systematic process. If the measurement strategy is to do intensive case studies then your sample might be quite small[4].

*Note*:  The term "sample" refers to the set of sources from which you intend to get information. Because of limited response, the actual sample you end up with may be smaller than your intended sample. This potential shortfall should be taken into account when you decide how large your (intended) sample should be, so that in the end you can expect to have "enough" data to make the assessments you wish to make. There are also things you can do to try to increase the "response rate" so that you don't have too many incompletes or non-respondents[5].

### The Sampling Decision – what to consider:

There are options, and choices to be made when sampling. As noted in the examples, the sampling decision governs how the Sample relates to the Population, and from there, what claims can legitimately be made about your evaluation results. It also governs how you will go about selecting the actual sample.

If you wish to *generalize* your results, then you will have to ensure that your sample is representative of the larger group to which you wish to generalize. Depending on how strong a generalization claim you want to make, you will have to engage in varying degrees of rigor in ensuring "representativeness", or the extent to which conclusions drawn about the sample can be said to represent or speak for the population.  It may be sufficient to ensure that your sample roughly matches the population on a simple list of characteristics (demographics, etc.) in order to claim a "likely" result or to claim that the fact that your program is "associated" with certain outcomes suggests that these benefits would occur for others who are similar.

However if you want to make a stronger claim that your program <u>caused</u> the effects you observe and to begin to prove that it would be likely to cause them in other members of the population of interest, then you would need to obtain more detailed information about your sample and select them randomly from the population of interest. You would also need to select a comparison group from the population of interest. To make a full claim of causality and generalizability you would need to randomly select participants and assign some to the program and some to a "control" group. Situations like this are beyond the scope of this document.

*Sample size* is a particular element in the sampling decision that matters if you want to draw conclusions you want claim also hold for the larger population of interest. In general, to assure that the sample fairly represents the population, the sample should be above a certain percentage of the population. How large that percentage needs to be depends on the size of the population. Generally, the smaller the population, the larger the sample percentage needs to be, because smaller populations tend to have more variability. For guidance on deciding sample size, see Appendix 2 in Taylor-Powell, "Sampling", which includes a table of recommended sample sizes for different precision levels.

---

[4] For a list of purposeful sampling strategies, see Table 1 in Taylor-Powell, p. 7, which adapts a table from Patton, Michael Q. Qualitative Evaluation and Research Methods second ed., Newbury Park, CA; Sage Publications.

[5] For more on response rates and how to adjust sample sizes accordingly see Ellen Taylor Powell's "Sampling" referred to earlier.  Also see "Random Samples: How and Why?" in the Pennsylvania State Extension's Evaluation Tip Sheets, at http://extension.psu.edu/evaluation/pdf/TS57.pdf .

If you do not need to generalize your results, but are instead interested in more in-depth information about a program, about participants or about a delivery method, you may want to do *purposeful sampling,* in which you select cases likely to shed light on a question or issue of interest. In this type of sampling, the above considerations on sample size do not apply. It will simply be important to explain and justify the sampling decisions you make.

For both kinds of sampling, random and purposeful, it is important to consider *feasibility*. If you find you have chosen an approach or sample size that is not feasible given the time frame and/or staff resources available, it may be necessary to scale back. So it is best to think this through **before** finalizing sampling plans.

*How-to:* See Taylor-Powell for how-to steps on basic sample selection methods including simple random sampling, systematic samples, etc. See also the Penn State Extension and Outreach "Evaluation" web page, which also has a number of resources on sampling, including practical, step-by-step directions, at: http://extension.psu.edu/evaluation/sampling.html.


## What to put in the Sample Section of the Evaluation Plan

In the Sample section of your Evaluation Plan for each evaluation question, briefly describe the population of interest, and say whether you intend to include in the sample the entire population of interest, or generalize from a smaller sample to a larger population, or do purposive sampling. Give a rough idea of your thinking about sample size.  If selecting a subset from the population, indicate how it will be selected. (If you only have time to evaluate half of the products, will you select every other one? Or draw numbers from a hat in order to see whose products get tested and whose don't? Will half be enough to give you a reliable answer? And so on.) If you will use a comparison group, give the above kinds of information for that group as well.

# *Sampling Plan Worksheet*

***Complete one worksheet for each of your Evaluation Questions.***

Evaluation Question:


For this evaluation question, what is the *population of interest?* (The group you are interested in learning about or making claims about.)


What portion of that group do you intend to try to obtain information or data from? Why them?


If your sample is a subset of the population, how will you select that subset?


How large is this intended sample likely to be?


What kind of response or participation rate do you expect to have from that intended sample? (Will the resulting actual sample be large enough for you to obtain the data you need?)


Draft the Sample Section for this Evaluation Question:
It should provide enough information to assess whether the selected sample will serve as a sound basis for the claims the evaluation is intended to address. Accordingly, the "population of interest" should be made clear, sample size and composition should be described, and the recruitment or sample selection process should be described.

# Design

The evaluation design lays out the sequence of program implementation and data collection. In its most basic form, the evaluation design tells when information is collected and from whom. For example, an evaluation design might specify a "pre-post" data collection strategy, a "post-only", or might include plans for collecting data from a comparison group, etc.

## FAQs

### Where can I learn more about possible design strategies?

To learn more, go to the Introduction to Design section of the Research Methods Knowledge Base: http://www.socialresearchmethods.net/kb/desintro.htm

Elsewhere on the same website there is much more in-depth information about experimental and quasi-experimental designs.

### What role do lifecycles play in determining appropriate evaluation design?

Evaluation design is driven by the evaluation question. Evaluation questions, in turn, are strongly influenced by program lifecycle. The wording of the evaluation questions should provide clear guidance on what kind of design will be needed. In particular, pay close attention to what kind of relationships are referenced in the evaluation question. For example, if the evaluation question refers to the extent to which program participation is associated with some kind of change in participants (in their knowledge, awareness, behavior, etc.) then you would need some way to contrast the before and after condition. This could be achieved with a retrospective pre-post, or an actual pre-post design. However, if the evaluation question refers to the extent to which the program has caused a change in participants then a stronger design is needed, in order to control for the impact of non-program factors.

In general, earlier stage programs might choose designs focused on implementation processes or post-only designs, while later stage programs might choose designs involving pre and post measurement with a comparison group. Also, it's important to consider which option is most likely to give you credible, accurate, and useful data (and, eventually, findings) in a way that is feasible, keeps stakeholders in mind, builds on prior evaluations, and prepares you for subsequent evaluations.

## Guiding Documents

- Overview of Evaluation Design

- Alignment with Design

# *Overview of Evaluation Design*

**<u>Design</u>**: The evaluation design provides an outline of how the evaluation will be conducted, laying out the sequence of program implementation and data collection. In its most basic form, the evaluation design tells when information is collected and from whom. It also answers the question (about evaluation findings): "compared to what?" Choice of design is keyed to the evaluation question, and should also take into consideration lifecycle stage and feasibility. In choosing an evaluation design, keep in mind that there is a tradeoff between the strength of evidence and the feasibility of collecting and analyzing data. Try to document the greatest causal effect possible without going beyond what's appropriate to your program's lifecycle stage. In reporting results, always keep claims in line with the evaluation design used to generate results. For example, when claiming effectiveness, make sure the design supports such a claim by including measures of change, not just post measures, and a strategy that can rule out other possible causes, such as a comparison group.

One way of communicating your chosen design is the "tic-tac-toe method" where X's stand for programs or activities and O's stand for observations (data collection points). When multiple observations or measures are used on one occasion (e.g., at the end of the program) you can use subscripts to distinguish among measures.

For example:

| Type of Design | "Tic-tac-toe" representation |
|---|---|
| Post-only | X O |
| Pre-Post | O X O |
| Longitudinal | O X O O O O |
| Mixed Methods | $O_1\ O_2$ X $O_1\ O_2$ |

As long as your Design description is concise and understandable, it can be represented in whatever form you are most comfortable. The next page provides a grid showing common evaluation designs (in the "tic-tac-toe" format) and their implications.

**Implications of common evaluation designs§**

| Design | Description | Advantages | Disadvantages |
|---|---|---|---|
| A. Post-only<br>X O | Collect data after a program occurs | Provides a snapshot or cross-sectional view at a particular point in time | Doesn't show change over time. |
| B. Post-only with comparison group<br>X O<br>O | Collect data after a program occurs from both participants and a comparison group | Allows for comparison with another group | Doesn't distinguish program effect from possible effects of initial differences in groups. |
| C. Post-only with control group and random assignment<br>R X O<br>R O | Randomly assign eligible individuals to either program or comparison group; measure after program participation for both groups. | Random assignment is most effective way to eliminate effects of any possible systematic bias in selection into program. Post-only design eliminates any bias which may be introduced by pre-testing. | Random assignment often not possible in program evaluation. Doesn't distinguish program effect from effects of outside factors or from effects of testing. |
| D. Retrospective post-then-pre<br>X OO | Collect data after program participation; ask participants about both post and pre states. | Eliminates "shift bias" which occurs when participants shift in their understanding and interpretation of the same questions. | Risks introducing recall bias. |
| E. Simple pre-post*<br>O X O | Collect data both before and after program participation | Shows change over time | Doesn't distinguish program effect from effects of other factors that could be causing change. (However, other potential competing theories may be ruled out in other ways) |
| F. Pre-post with comparison group<br>O X O<br>O O | Collect data before and after program participation from both participants and comparison group | Combines strengths of B. (pre-post) and C. (comparison group) to show change and help distinguish program effect from effect of some other possible influences on outcomes | Doesn't eliminate possibility of bias in selection into program and/or comparison group. |
| G. Pre-post with control group and random assignment<br>R O X O<br>R O O | Randomly assign eligible individuals to either program or comparison group; collect data before and after program participation for both groups. | Combines strength of E. (random assignment) with strength of F. (pre-post with comparison) for best possible isolation of program impact from other factors that might influence outcomes. | Random assignment often not possible in program evaluation. Doesn't distinguish program effect from effects of testing. |

*Pre-post designs may be: (1) matched at the individual level (2) matched at the group level or (3) unmatched.

*Overview of Evaluation Design page 2 of 2*

# *Alignment with Design*
## For Evaluations Involving Program Outcomes

Several aspects of an evaluation plan help determine how credible the results will be. One important pillar of credible results is the alignment between purpose (and implied claims) and evaluation design.

Key determinants of appropriate design are (1) whether the evaluation is intended to assess *change*; and (2) whether the evaluation is intended to assess *effectiveness* of the program in causing change. If the former, then the design will need to incorporate measurement both before and after program participation. If the latter, then the design will need to capture change <u>and</u> rule out other possible causes by comparing with a non-program group.

To check alignment of an evaluation plan: use the grid below, and note which Purpose, Claim, and Design best match those in the evaluation plan. If all three are in the same row, the design supports the evaluation goals. If they are not, Purpose, Claims, and/or Design must be adjusted as needed, keeping in mind stakeholder demands, the lifecycle stage of the program, and feasibility.

*In making adjustments, keep in mind that there is a tradeoff between the strength of evidence and the feasibility of collecting and analyzing data. In reporting results, keep claims appropriate for the evaluation design used.*

| Evaluation Purpose | Associated Claim | Design where X=program O=observation R indicates random assignment | Does design show change? | Does design rule out other possible causes?* |
|---|---|---|---|---|
| ...to assess the post-program state of participants. | After program, these participants show a certain level of outcome Z. | X O (post only) | No | No |
| ...to compare the post-program state of participants with that of non-participants. | After program, these participants show a different level of outcome Z than non-participants. | X O    O (post only with comparison group) | No | Somewhat |
| " | " | R X O R    O (post-only with comparison group, random assignment) | No | Mostly |
| ...to assess participant perceptions of program effectiveness. | According to these participants, in this setting and context, the program caused a change on outcome Z. | X Opost/Opre (retrospective "post then pre") | Yes | No |
| ...to assess the extent to which participation in the program is associated with change in outcome Z. | Participation in the program is associated with a change in outcome Z in this setting and context, with these participants. | O X O (simple pre/post) | Yes | No |
| ...to assess the effectiveness of the program in increasing or improving outcome Z. | The program is effective in this setting and context, with these participants. | O X O O    O (pre/post with comparison group) | Yes | Somewhat |
| " | The program is effective in this setting and context, with these participants. It may also be effective in other settings and contexts, with other participants. | R O X O R O    O (pre/post with random assignment) | Yes | Mostly |

 * Some possible outside causes of participant change include: a) bias in how participants are selected into the program; b) bias in how participants were assigned to program versus comparison groups, c) maturation of program participants, and  d) events not related to the program that caused change

149

# Data Management

Data management is the ongoing process of recording, documenting, tracking, securing, and organizing data that is collected during the course of an evaluation project. It's important to plan this out in advance, in order to ensure that nothing gets lost as data come in, that there is a record of when it came in and through what means, and that data are stored in a way that is retrievable and organized to facilitate the analysis.

## FAQs

### What are the most common considerations when planning for data management?

When planning for data management, consider (1) what your data will be like, and (2) what you will want to do with your data to analyze them. For the first consideration, ask yourself: Will I have qualitative data, quantitative data, or both? How much data will I have? Will there be a need for data entry (e.g., entering the responses to a paper-based survey into a spreadsheet or database). If so, who will do so, following what guidelines? What will I need to do to ensure that data are well organized and safe (especially if anonymity or confidentiality must be maintained)? For the second consideration, ask: Will I want to aggregate data or keep it all separate? What variables or factors will I summarize, and how? What, if any, comparisons among variables will I want to make? Because there are so many different possible answers to the set of questions above, it is difficult to prescribe any particular way of managing data. However, considering these questions (and testing out whatever approach you decide on, to be sure it will work for your data entry and analysis) is really important.

Generally, data are managed with computer software, such as an Excel spreadsheet or a specialized data management and analysis program. If using a spreadsheet program, the spreadsheet should be set up where rows contain responses (e.g., individual respondents' survey answers) and the columns are variables (e.g., gender, individual survey questions, etc.) However, remember that hard copies of data (paper surveys, audio recordings of interviews) should be retained as a backup or for future review, and protected for confidentiality, even after the data have been entered into a computer program.

# Analysis

Analysis is the process of describing and making meaning from data. The appropriate analysis strategy for a given evaluation depends on the evaluation question(s), the type, quantity and quality of the data gathered, as well as the insights and potential claims that emerge from initial review of the data.

## FAQs

### Where can I learn more about how to do analysis?

**Quantitative Analysis:** One good, succinct source on analyzing quantitative data is: Taylor-Powell, E. (1996). Analyzing Quantitative Data. Retrieved May 5, 2015, from University of Wisconsin-Extension Cooperative Extension, Program Development and Evaluation Unit Web site: http://learningstore.uwex.edu/Assets/pdfs/G3658-06.pdf . Another useful source is: Research Methods Knowledge Base: Trochim, William M. The Research Methods Knowledge Base, 2nd Edition. Internet WWW page, at URL: http://www.socialresearchmethods.net/kb/analysis.php

**Qualitative Analysis:** Although it is rather lengthy and in-depth, a very good, highly readable source on qualitative data analysis (and qualitative evaluation and research in general) is: Patton, M. Q. (2015). Qualitative Research & Evaluation Methods. Thousand Oaks, CA: Sage.  For a more succinct source (12 pages) specifically about qualitative analysis, see: Taylor-Powell, E., & Renner, M. (2003). Analyzing Qualitative Data. Retrieved May 5, 2015, from University of Wisconsin-Extension Cooperative Extension, at http://learningstore.uwex.edu/Assets/pdfs/G3658-12.pdf

### How is evaluation data analysis influenced by program lifecycle?

Evaluation data analysis is not directly influenced by program lifecycle, but it is heavily influenced indirectly, by way of the choices that go into the evaluation purpose, evaluation questions, and, especially, measurement and design sections of your evaluation plan (which are very directly related to your program's lifecycle stage). In other words, your program lifecycle strongly influences your measurement approach, and your measurement approach, in turn, strongly influences your data analysis plan. As such, analysis for earlier lifecycle programs often involves summarizing, averaging and comparing data in relatively simple ways and can be done internally; analysis for later lifecycle stage programs will tend to be more difficult and complicated, due to the statistical tests required--here, it may be advisable to seek external help from a statistician or research methodologist.

### What determines which analysis strategy I should use?

Analysis is essentially about making sense of the data in such a way that you answer or otherwise gain insights in response to your evaluation questions. So it is important that your plan for analysis is closely aligned to your evaluation questions and especially to your chosen measurement strategy. For example, if your evaluation question is about how participants experienced the program, then your measurement would likely include interviews. Then, your analysis strategy can be selected from among the options which exist for analyzing interview data (many of which involve categorizing the data, or text, into codes). If your evaluation question is about whether or not your program caused a desired outcome to occur among participants, your measurement and design would likely have involved a pre-post quantitative survey administered to two randomly assigned, representative samples of participants and non-participants. In this case, your analysis will require specific statistical tests, usually performed using a statistical software package such as SPSS. In such cases, it is common to seek outside assistance from a statistician.

# Reporting and Utilization

"Reporting" refers to the process of communicating results and recommendations to stakeholders. This includes internal (program staff, participants) and external (funders and administrators) stakeholders. "Utilization" refers, more generally, to the process of using evaluation results to make decisions, if any, about program design, support, staffing, and so on.

## FAQs

### What should I cover in the Reporting and Utilization section of my evaluation plan?

A first step for crafting this section of the plan is to refer back to the Evaluation Purpose Statement, which should have included a description of how the results of the evaluation will be used. This should help identify the key stakeholders (internal and/or external) who will be most interested in the evaluation results and putting them to use. Think through who those stakeholders are, which part(s) of the evaluation they each care about, and how best to communicate the results to them (format, and so on). This might entail a presentation, a meeting and roundtable discussion, participatory review and analysis, or other ways of engaging stakeholders. Then describe this communication plan in this section. There may also be other stakeholders who would benefit from seeing the evaluation results. Perhaps, for example, the Marketing Team would value seeing results about program outcomes, in addition to the funders who were your primary intended audience? It is also a very good practice to share results back with the individuals or groups who contributed to your evaluation by completing surveys, participating in interviews or focus groups, and so on.

Because each stakeholder or stakeholder group may be interested in different parts and/or may need a different type of report, be sure to reflect these specific needs in this section.

### Does "reporting" just mean writing a written report of evaluation results?

We hope not! The reporting stage is potentially an exciting time, when there is important and useful new information to be shared. Think of it as an opportunity to inform and engage a range of people associated with the program – including potential future participants or funders, who may not even know about the program yet. Sometimes a formal written report, including descriptions of the evaluation process and data used, is just what the stakeholder needs. For some stakeholders, other formats can also be very effective or possibly more effective. Consider video clips, colorful posters, a dynamic website, or other creative communication – as long as it will be seen as credible and useful to whichever stakeholder you are trying to reach. Keep in mind, even in these less formal formats, it can still be important to include some information about process and sources, as appropriate.

A great resource for smart and creative reporting strategies is Stephanie Evergreen's blog on intentional data visualization, at http://stephanieevergreen.com/. Additional resources may be found on the American Evaluation Association website, http://www.eval.org/.