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 system[1]. 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:

[1] Cornell Office for Research on Evaluation. (2012). The Netway [Software for Evaluation Planning]. Available from www.evaluationnetway.com

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.

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 <u>not</u> the kind of questions you'd include on a questionnaire for participants,. The Evaluation Questions are questions you are asking <u>about your program</u> 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 obtaining information from 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, not 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.

Design Example:

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 O_1 below). Logs, Progress Notes and Success Stories are collected throughout and summarized at the end of the series (O_2, O_3 respectively).

O_1 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 how you plan to handle, and obtain results from, the data your evaluation will produce (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 before 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.

<u>Typical "steps" in data management and analysis process:</u>

Prior to collecting data:

- **Organize**
  - *Decide what variables you need to collect data on, which spreadsheet software, statistical software, and/or qualitative software you will use*
  - *Set up spreadsheets to receive coded data and set up any pre-determined qualitative coding structures that may be needed*
  - *For quantitatively coded data, create a codebook telling what any codes or labels you set up for data entry will mean (e.g., 1=female, 2=male, etc.*)

During or shortly after data collection:

- **Convert to analyzable format**
  - *Score tests, numerically code survey responses, qualitatively code interviews, etc.*
- **Enter into electronic form**
- **Clean**
  - *Examine data and determine what's valid and usable, what has to be thrown out*
  - *Ensure uniform formatting, etc.*
- **Analyze**
  - *Compute a change score for pre/post measures*
  - *If an effect is found, perform statistical analysis to see if it is significant*
  - *For qualitative data, summarize patterns with textual description, word-count analysis or graphic depiction*
- **Synthesize and interpret** (involving stakeholders, as appropriate)
- **Draw conclusions, make inferences** to the extent the nature and structure of the data supports them

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, 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 this 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.)