https://ctb.ku.edu/en/table-of-contents/evaluate/evaluation/framework-for-evaluation/main

---

How do you evaluate a specific program?

Before your organization starts with a program evaluation, your group should be very clear about the answers to the following questions:

- What will be evaluated?
- What criteria will be used to judge program performance?
- What standards of performance on the criteria must be reached for the program to be considered successful?
- What evidence will indicate performance on the criteria relative to the standards?
- What conclusions about program performance are justified based on the available evidence?

To clarify the meaning of each, let's look at some of the answers for Drive Smart, a hypothetical program begun to stop drunk driving.

- **What will be evaluated?**
  - Drive Smart, a program focused on reducing drunk driving through public education and intervention.
- **What criteria will be used to judge program performance?**
  - The number of community residents who are familiar with the program and its goals
  - The number of people who use "Safe Rides" volunteer taxis to get home
  - The percentage of people who report drinking and driving
  - The reported number of single car night time crashes (This is a common way to try to determine if the number of people who drive drunk is changing)
- **What standards of performance on the criteria must be reached for the program to be considered successful?**
  - 80% of community residents will know about the program and its goals after the first year of the program
  - The number of people who use the "Safe Rides" taxis will increase by 20% in the first year
  - The percentage of people who report drinking and driving will decrease by 20% in the first year
  - The reported number of single car night time crashes will decrease by 10 % in the program's first two years
- **What evidence will indicate performance on the criteria relative to the standards?**
  - A random telephone survey will demonstrate community residents' knowledge of the program and changes in reported behavior
  - Logs from "Safe Rides" will tell how many people use their services
  - Information on single car night time crashes will be gathered from police records
- **What conclusions about program performance are justified based on the available evidence?**
  - Are the changes we have seen in the level of drunk driving due to our efforts, or something else? Or (if no or insufficient change in behavior or outcome,)

- Should Drive Smart change what it is doing, or have we just not waited long enough to see results?

Below is a practical non-prescriptive tool that summarizes in a logical order the important elements of program evaluation.

The six connected steps of the framework are actions that should be a part of any evaluation. Although in practice the steps may be encountered out of order, it will usually make sense to follow them in the recommended sequence. That's because earlier steps provide the foundation for subsequent progress. Thus, decisions about how to carry out a given step should not be finalized until prior steps have been thoroughly addressed.

However, these steps are meant to be adaptable, not rigid. Sensitivity to each program's unique context (for example, the program's history and organizational climate) is essential for sound evaluation. They are intended to serve as starting points around which community organizations can tailor an evaluation to best meet their needs.

- Engage stakeholders
- Describe the program
- Focus the evaluation design
- Gather credible evidence
- Justify conclusions
- Ensure use and share lessons learned

Understanding and adhering to these basic steps will improve most evaluation efforts.

Engage Stakeholders

Stakeholders are people or organizations that have something to gain or lose from what will be learned from an evaluation, and also in what will be done with that knowledge. Evaluation cannot be done in isolation. Almost everything done in community health and development work involves partnerships - alliances among different organizations, board members, those affected by the problem, and others. Therefore, any serious effort to evaluate a program must consider the different values held by the partners. Stakeholders must be part of the evaluation to ensure that their unique perspectives are understood. When stakeholders are not appropriately involved, evaluation findings are likely to be ignored, criticized, or resisted.

However, if they are part of the process, people are likely to feel a good deal of ownership for the evaluation process and results. They will probably want to develop it, defend it, and make sure that the evaluation really works.

That's why this evaluation cycle begins by engaging stakeholders. Once involved, these people will help to carry out each of the steps that follows.

Describe the Program

A program description is a summary of the intervention being evaluated. It should explain what the program is trying to accomplish and how it tries to bring about those changes. The description will also illustrate the program's core components and elements, its ability to make changes, its stage of development, and how the program fits into the larger organizational and community environment.

How a program is described sets the frame of reference for all future decisions about its evaluation. For example, if a program is described as, "attempting to strengthen enforcement of existing laws that discourage underage drinking," the evaluation might be very different than if it is described as, "a program to reduce drunk driving by teens." Also, the description allows members of the group to compare the program to other similar efforts, and it makes it easier to figure out what parts of the program brought about what effects.

Moreover, different stakeholders may have different ideas about what the program is supposed to achieve and why. For example, a program to reduce teen pregnancy may have some members who believe this means only increasing access to contraceptives, and other members who believe it means only focusing on abstinence.

Evaluations done without agreement on the program definition aren't likely to be very useful. In many cases, the process of working with stakeholders to develop a clear and logical program description will bring benefits long before data are available to measure program effectiveness.

There are several specific aspects that should be included when describing a program.

Statement of need

A statement of need describes the problem, goal, or opportunity that the program addresses; it also begins to imply what the program will do in response. Important features to note regarding a program's need are: the nature of the problem or goal, who is affected, how big it is, and whether (and how) it is changing.

Expectations

Expectations are the program's intended results. They describe what the program has to accomplish to be considered successful. For most programs, the accomplishments exist on a continuum (first, we want to accomplish X... then, we want to do Y...). Therefore, they should be organized by time ranging from specific (and immediate) to broad (and longer-term) consequences. For example, a program's vision, mission, goals, and objectives, all represent varying levels of specificity about a program's expectations.

Activities

Activities are everything the program does to bring about changes. Describing program components and elements permits specific strategies and actions to be listed in logical sequence. This also shows how different program activities, such as education and enforcement, relate to one another. Describing program activities also provides an opportunity to distinguish activities that are the direct responsibility of the program from those that are conducted by related programs or partner organizations. Things outside of the program that may affect its success, such as harsher laws punishing businesses that sell alcohol to minors, can also be noted.

Resources

Resources include the time, talent, equipment, information, money, and other assets available to conduct program activities. Reviewing the resources a program has tells a lot about the amount and intensity of its services. It may also point out situations where there is a mismatch between what the group wants to do and the resources available to carry out these activities. Understanding program costs is a necessity to assess the cost-benefit ratio as part of the evaluation.

Stage of development

A program's *stage of development* reflects its maturity. All community health and development programs mature and change over time. People who conduct evaluations, as well as those who use their findings, need to consider the dynamic nature of programs. For example, a new program that just received its first grant may differ in many respects from one that has been running for over a decade.

At least three phases of development are commonly recognized: *planning*, *implementation*, and *effects or outcomes*. In the planning stage, program activities are untested and the goal of evaluation is to refine plans as much as possible. In the implementation phase, program activities are being field tested and modified; the goal of evaluation is to see what happens in the "real world" and to improve operations. In the effects stage, enough time has passed for the program's effects to emerge; the goal of evaluation is to identify and understand the program's results, including those that were unintentional.

Context

A description of the program's *context* considers the important features of the environment in which the program operates. This includes understanding the area's history, geography, politics, and social and economic conditions, and also what other organizations have done. A realistic and responsive evaluation is sensitive to a broad range of potential influences on the program. An understanding of the context lets users interpret findings accurately and assess their generalizability. For example, a program to improve housing in an inner-city neighborhood might have been a tremendous success, but would likely not work in a small town on the other side of the country without significant adaptation.

Logic model

A *logic model* synthesizes the main program elements into a picture of how the program is supposed to work. It makes explicit the sequence of events that are presumed to bring about change. Often this logic is displayed in a flow-chart, map, or table to portray the sequence of steps leading to program results.

Creating a logic model allows stakeholders to improve and focus program direction. It reveals assumptions about conditions for program effectiveness and provides a frame of reference for one or more evaluations of the program. A detailed logic model can also be a basis for estimating the program's effect on endpoints that are not directly measured. For example, it may be possible to estimate the rate of reduction in disease from a known number of persons experiencing the intervention if there is prior knowledge about its effectiveness.

The breadth and depth of a program description will vary for each program evaluation. And so, many different activities may be part of developing that description. For instance, multiple sources of information could be pulled together to construct a well-rounded description. The accuracy of an existing program description could be confirmed through discussion with stakeholders. Descriptions of what's going on could be checked against direct observation of activities in the field. A narrow program description could be fleshed out by addressing contextual factors (such as staff turnover, inadequate resources, political pressures, or strong community participation) that may affect program performance.

Focus the Evaluation Design

By focusing the evaluation design, we mean doing advance planning about where the evaluation is headed, and what steps it will take to get there. It isn't possible or useful for an evaluation to try to answer all questions for all stakeholders; there must be a focus. A well-focused plan is a safeguard against using time and resources inefficiently.

Depending on what you want to learn, some types of evaluation will be better suited than others. However, once data collection begins, it may be difficult or impossible to change what you are doing, even if it becomes obvious that other methods would work better. A thorough plan anticipates intended uses and creates an evaluation strategy with the greatest chance to be useful, feasible, proper, and accurate.

Among the issues to consider when focusing an evaluation are:

Purpose

Purpose refers to the general intent of the evaluation. A clear purpose serves as the basis for the design, methods, and use of the evaluation. Taking time to articulate an overall purpose will stop your organization from making uninformed decisions about how the evaluation should be conducted and used.

https://ctb.ku.edu/en/table-of-contents/evaluate/evaluation/framework-for-evaluation/main

---

There are at least four general purposes for which a community group might conduct an evaluation:

- **To gain insight**.This happens, for example, when deciding whether to use a new approach (e.g., would a neighborhood watch program work for our community?) Knowledge from such an evaluation will provide information about its practicality. For a developing program, information from evaluations of similar programs can provide the insight needed to clarify how its activities should be designed.
- **To improve how things get done**.This is appropriate in the implementation stage when an established program tries to describe what it has done. This information can be used to describe program processes, to improve how the program operates, and to fine-tune the overall strategy. Evaluations done for this purpose include efforts to improve the quality, effectiveness, or efficiency of program activities.
- **To determine what the effects of the program are**. Evaluations done for this purpose examine the relationship between program activities and observed consequences. For example, are more students finishing high school as a result of the program? Programs most appropriate for this type of evaluation are mature programs that are able to state clearly what happened and who it happened to. Such evaluations should provide evidence about what the program's contribution was to reaching longer-term goals such as a decrease in child abuse or crime in the area. This type of evaluation helps establish the accountability, and thus, the credibility, of a program to funders and to the community.
- **To affect those who participate in it**. The logic and reflection required of evaluation participants can itself be a catalyst for self-directed change. And so, one of the purposes of evaluating a program is for the process and results to have a positive influence. Such influences may:
    - *Empower program participants* (for example, being part of an evaluation can increase community members' sense of control over the program);
    - *Supplement the program* (for example, using a follow-up questionnaire can reinforce the main messages of the program);
    - *Promote staff development* (for example, by teaching staff how to collect, analyze, and interpret evidence); or
    - *Contribute to organizational growth* (for example, the evaluation may clarify how the program relates to the organization's mission).

Users

Users are the specific individuals who will receive evaluation findings. They will directly experience the consequences of inevitable trade-offs in the evaluation process. For example, a trade-off might be having a relatively modest evaluation to fit the budget with the outcome that the evaluation results will be less certain than they would be for a full-scale evaluation. Because they will be affected by these tradeoffs, intended users have a right to participate in choosing a focus for the evaluation. An evaluation designed without adequate user involvement in selecting the focus can become a misguided and irrelevant exercise. By contrast, when users are encouraged to clarify intended uses, priority questions, and preferred methods, the evaluation is more likely to focus on things that will inform (and influence) future actions.

Uses

Uses describe what will be done with what is learned from the evaluation. There is a wide range of potential uses for program evaluation. Generally speaking, the uses fall in the same four categories as the purposes listed above: to gain insight, improve how things get done, determine what the effects of the program are, and affect participants. The following list gives examples of uses in each category.

Gather Credible Evidence

Credible evidence is the raw material of a good evaluation. The information learned should be seen by stakeholders as believable, trustworthy, and relevant to answer their questions. This requires thinking broadly about what counts as "evidence." Such decisions are always situational; they depend on the question being posed and the motives for asking it. For some questions, a stakeholder's standard for credibility could demand having the results of a randomized experiment. For another question, a set of well-done, systematic observations such as interactions between an outreach worker and community residents, will have high credibility. The difference depends on what kind of information the stakeholders want and the situation in which it is gathered.

Context matters! In some situations, it may be necessary to consult evaluation specialists. This may be especially true if concern for data quality is especially high. In other circumstances, local people may offer the deepest insights. Regardless of their expertise, however, those involved in an evaluation should strive to collect information that will convey a credible, well-rounded picture of the program and its efforts.

Having credible evidence strengthens the evaluation results as well as the recommendations that follow from them. Although all types of data have limitations, it is possible to improve an evaluation's overall credibility. One way to do this is by using multiple procedures for gathering, analyzing, and interpreting data. Encouraging participation by stakeholders can also enhance perceived credibility. When stakeholders help define questions and gather data, they will be more likely to accept the evaluation's conclusions and to act on its recommendations.

The following features of evidence gathering typically affect how credible it is seen as being:

Indicators

Indicators translate general concepts about the program and its expected effects into specific, measurable parts.

Examples of indicators include:

- The program's capacity to deliver services
- The participation rate
- The level of client satisfaction

- The amount of intervention exposure (how many people were exposed to the program, and for how long they were exposed)
- Changes in participant behavior
- Changes in community conditions or norms
- Changes in the environment (e.g., new programs, policies, or practices)
- Longer-term changes in population health status (e.g., estimated teen pregnancy rate in the county)

Indicators should address the criteria that will be used to judge the program. That is, they reflect the aspects of the program that are most meaningful to monitor. Several indicators are usually needed to track the implementation and effects of a complex program or intervention.

Justify Conclusions

The process of justifying conclusions recognizes that evidence in an evaluation does not necessarily speak for itself. Evidence must be carefully considered from a number of different stakeholders' perspectives to reach conclusions that are well -substantiated and justified. Conclusions become justified when they are linked to the evidence gathered and judged against agreed-upon values set by the stakeholders. Stakeholders must agree that conclusions are justified in order to use the evaluation results with confidence.

The principal elements involved in justifying conclusions based on evidence are:

Standards

Standards reflect the values held by stakeholders about the program. They provide the basis to make program judgments. The use of explicit standards for judgment is fundamental to sound evaluation. In practice, when stakeholders articulate and negotiate their values, these become the standards to judge whether a given program's performance will, for instance, be considered "successful," "adequate," or "unsuccessful."

Analysis and synthesis

Analysis and synthesis are methods to discover and summarize an evaluation's findings. They are designed to detect patterns in evidence, either by isolating important findings (analysis) or by combining different sources of information to reach a larger understanding (synthesis). Mixed method evaluations require the separate analysis of each evidence element, as well as a synthesis of all sources to examine patterns that emerge. Deciphering facts from a given body of evidence involves deciding how to organize, classify, compare, and display information. These decisions are guided by the questions being asked, the types of data available, and especially by input from stakeholders and primary intended users.

Interpretation

Interpretation is the effort to figure out what the findings mean. Uncovering facts about a program's performance isn't enough to make conclusions. The facts must be interpreted to understand their practical significance. For example, saying, "15 % of the people in our area witnessed a violent act last year," may be interpreted differently depending on the situation. For example, if 50% of community members had watched a violent act in the last year when they were surveyed five years ago, the group can suggest that, while still a problem, things are getting better in the community. However, if five years ago only 7% of those surveyed said the same thing, community organizations may see this as a sign that they might want to change what they are doing. In short, interpretations draw on information and perspectives that stakeholders bring to the evaluation. They can be strengthened through active participation or interaction with the data and preliminary explanations of what happened.

Judgements

Judgments are statements about the merit, worth, or significance of the program. They are formed by comparing the findings and their interpretations against one or more selected standards. Because multiple standards can be applied to a given program, stakeholders may reach different or even conflicting judgments. For instance, a program that increases its outreach by 10% from the previous year may be judged positively by program managers, based on standards of improved performance over time. Community members, however, may feel that despite improvements, a minimum threshold of access to services has still not been reached. Their judgment, based on standards of social equity, would therefore be negative. Conflicting claims about a program's quality, value, or importance often indicate that stakeholders are using different standards or values in making judgments. This type of disagreement can be a catalyst to clarify values and to negotiate the appropriate basis (or bases) on which the program should be judged.

Recommendations

Recommendations are actions to consider as a result of the evaluation. Forming recommendations requires information beyond just what is necessary to form judgments. For example, knowing that a program is able to increase the services available to battered women doesn't necessarily translate into a recommendation to continue the effort, particularly when there are competing priorities or other effective alternatives. Thus, recommendations about what to do with a given intervention go beyond judgments about a specific program's effectiveness.

If recommendations aren't supported by enough evidence, or if they aren't in keeping with stakeholders' values, they can really undermine an evaluation's credibility. By contrast, an evaluation can be strengthened by recommendations that anticipate and react to what users will want to know.

Three things might increase the chances that recommendations will be relevant and well-received:

- Sharing draft recommendations
- Soliciting reactions from multiple stakeholders
- Presenting options instead of directive advice

Justifying conclusions in an evaluation is a process that involves different possible steps. For instance, conclusions could be strengthened by searching for alternative explanations from the ones you have chosen, and then showing why they are unsupported by the evidence. When there are different but equally well supported conclusions, each could be presented with a summary of their strengths and weaknesses. Techniques to analyze, synthesize, and interpret findings might be agreed upon before data collection begins.

Ensure Use and Share Lessons Learned

It is naive to assume that lessons learned in an evaluation will necessarily be used in decision making and subsequent action. Deliberate effort on the part of evaluators is needed to ensure that the evaluation findings will be used appropriately. Preparing for their use involves strategic thinking and continued vigilance in looking for opportunities to communicate and influence. Both of these should begin in the earliest stages of the process and continue throughout the evaluation.

The elements of key importance to be sure that the recommendations from an evaluation are used are:

Design

Design refers to how the evaluation's questions, methods, and overall processes are constructed. As discussed in the third step of this framework (focusing the evaluation design), the evaluation should be organized from the start to achieve specific agreed-upon uses. Having a clear purpose that is focused on the use of what is learned helps those who will carry out the evaluation to know who will do what with the findings. Furthermore, the process of creating a clear design will highlight ways that stakeholders, through their many contributions, can improve the evaluation and facilitate the use of the results.

Preparation

Preparation refers to the steps taken to get ready for the future uses of the evaluation findings. The ability to translate new knowledge into appropriate action is a skill that can be strengthened through practice. In fact, building this skill can itself be a useful benefit of the evaluation. It is possible to prepare stakeholders for future use of the results by discussing how potential findings might affect decision making.

For example, primary intended users and other stakeholders could be given a set of hypothetical results and asked what decisions or actions they would make on the basis of this new knowledge. If they indicate that the evidence presented is incomplete or irrelevant and that no action would be taken, then this is an early warning sign that the planned evaluation should be modified. Preparing for use also gives stakeholders more time to explore both positive and negative implications of potential results and to identify different options for program improvement.

Feedback

Feedback is the communication that occurs among everyone involved in the evaluation. Giving and receiving feedback creates an atmosphere of trust among stakeholders; it keeps an evaluation on track by keeping everyone informed about how the evaluation is proceeding. Primary intended users and other stakeholders have a right to comment on evaluation decisions. From a standpoint of ensuring use, stakeholder feedback is a necessary part of every step in the evaluation. Obtaining valuable feedback can be encouraged by holding discussions during each step of the evaluation and routinely sharing interim findings, provisional interpretations, and draft reports.

Follow-up

Follow-up refers to the support that many users need during the evaluation and after they receive evaluation findings. Because of the amount of effort required, reaching justified conclusions in an evaluation can seem like an end in itself. *It is not*. Active follow-up may be necessary to remind users of the intended uses of what has been learned. Follow-up may also be required to stop lessons learned from becoming lost or ignored in the process of making complex or political decisions. To guard against such oversight, it may be helpful to have someone involved in the evaluation serve as an advocate for the evaluation's findings during the decision -making phase.

Facilitating the use of evaluation findings also carries with it the responsibility to prevent misuse. Evaluation results are always bounded by the context in which the evaluation was conducted. Some stakeholders, however, may be tempted to take results out of context or to use them for different purposes than what they were developed for. For instance, over-generalizing the results from a single case study to make decisions that affect all sites in a national program is an example of misuse of a case study evaluation.

Similarly, program opponents may misuse results by overemphasizing negative findings without giving proper credit for what has worked. Active follow-up can help to prevent these and other forms of misuse by ensuring that evidence is only applied to the questions that were the central focus of the evaluation.

Dissemination

Dissemination is the process of communicating the procedures or the lessons learned from an evaluation to relevant audiences in a timely, unbiased, and consistent fashion. Like other elements of the evaluation, the reporting strategy should be discussed in advance with intended users and other stakeholders. Planning effective communications also requires considering the timing, style, tone, message source, vehicle, and format of information products. Regardless of how communications are constructed, the goal for dissemination is to achieve full disclosure and impartial reporting.

Along with the uses for evaluation findings, there are also uses that flow from the very process of evaluating. These "process uses" should be encouraged. The people who take part in an evaluation can experience profound changes in beliefs and behavior. For instance, an evaluation challenges staff members to act differently in what they are doing, and to question assumptions that connect program activities with intended effects.

Evaluation also prompts staff to clarify their understanding of the goals of the program. This greater clarity, in turn, helps staff members to better function as a team focused on a common end. In short, immersion in the logic, reasoning, and values of evaluation can have very positive effects, such as basing decisions on systematic judgments instead of on unfounded assumptions.

In Summary

Evaluation is a powerful strategy for distinguishing programs and interventions that make a difference from those that don't. It is a driving force for developing and adapting sound strategies, improving existing programs, and demonstrating the results of investments in time and other resources. It also helps determine if what is being done is worth the cost.

This recommended framework for program evaluation is both a synthesis of existing best practices and a set of standards for further improvement. It supports a practical approach to evaluation based on steps and standards that can be applied in almost any setting. Because the framework is purposefully general, it provides a stable guide to design and conduct a wide range of evaluation efforts in a variety of specific program areas. The framework can be used as a template to create useful evaluation plans to contribute to understanding and improvement. The Magenta Book - Guidance for Evaluation provides additional information on requirements for good evaluation, and some straightforward steps to make a good evaluation of an intervention more feasible, read The Magenta Book - Guidance for Evaluation.

Contributor
Bobby Milstein
Scott Wetterhall
CDC Evaluation Working Group