
Evaluation Learning Community: Summary & Key Takeaways

Key Takeaways

1. Program evaluation is a systematic process used to determine the value, impact, and quality of a program.
2. While evaluation and research use similar methods that are designed to be rigorous, evaluation helps to make judgements about a program's value, impact and quality whereas research adds to a field's understanding about a topic.
3. A process evaluation assesses whether a program was delivered as planned. This provides useful information to help programs run more effectively and efficiently (e.g., program records).
4. An outcome evaluation measures the immediate, or short-term effects, of a program that occur immediately following the program's conclusion (e.g., behaviors, attitudes, and perceptions related to substance use).
5. An impact evaluation measures the longer-term effects of a program after it has ended, typically measuring the program's desired goal (e.g., preventing substance use).
6. A program's desired goal is also referred to as the criteria of merit. In a substance use prevention program, the criteria of merit would be preventing or delaying the initiation of use, or reducing current use. At least one evaluation question should be about the criteria of merit and a standard, or expected performance target, for that criteria of merit.
7. Program objectives represent the specific, measurable steps that collectively lead toward the achievement of the program's goal. Objectives are derived from the program's logic model.
8. A logic model describes how a program is designed to work, from the inputs that go into to the program to the intervention components and how they operate to achieve the program's desired goal.
9. A Behavior-Determinant-Intervention (BDI) logic model is created by working backwards in a stepwise fashion to demonstrate the evidence-based relationships among interventions, determinants of behaviors, the behaviors themselves, and the ultimate health goal(s) as follows:
 - a. Operationalize the goal (e.g., decreasing substance use in teens)
 - b. Identify the specific behaviors research has shown lead to that goal (e.g., increase number of times that teens use refusal skills; increase use of decision-making skills)
 - c. Identify the determinants, or attitudes, beliefs, and knowledge, that research has shown lead to each of the identified behaviors (e.g., self-efficacy to use refusal skills is proven to predict teens using refusal skills)
 - d. Identify the intervention components research has shown to lead to the identified determinants (e.g., teaching and practicing refusal skills predicts self-efficacy to use refusal skills).
10. Identifying change as a result of a program intervention requires 1) Assessing outcome and impact variables at multiple time points (e.g., measuring knowledge, attitudes, beliefs, and behaviors pre and post an intervention); and 2) Measuring these variables in another group who did not receive the program.

- a. A pre/post evaluation is used to measure progress or change in a given area (e.g., knowledge, attitudes, and perceived abilities) before and after a program, training, or intervention. This approach is frequently used to measure acquired learning in a specific area. In substance use prevention, pre/post evaluations can tell if change occurred, but is limited in answering whether the change occurred as a result of the program.
 - b. To attribute progress or change to a program (as opposed to another factor that may simultaneously impact these outcomes) requires replicating the same pre-post assessment with a similar group who did not receive the program intervention. If there is a difference in the program (or intervention group) and not in the control group, you have greater confidence in attributing change to the effects of the program.
11. Statistics are a set of procedures used to answer questions and draw conclusions from large amounts of information.
- a. Descriptive statistics are used to *describe* data to enhance understanding of that data. Comparisons and inferences cannot be drawn from descriptive statistics. In other words, we cannot make a judgement about an increase or decrease in some outcome from looking at percentages, average scores, or counts.
 - b. Inferential statistics are used to draw *inferences* about group differences or associations between variables.
12. Evaluation reports tell the story of the evaluation and answer the questions: What does the data tell us? What does this mean about our program? What do we do with this information? Focus on how you (or others) will use the evaluation information and tailor reports to meet their needs. Do this before you collect data.
13. Reports can be presented in a detailed technical fashion or as an infographic highlighting the most relevant data. The time period covered in a report can also vary, ranging from ongoing continuous quality improvement reports to annual reports to program completion reports. It is advisable to engage in ongoing and cumulative reporting of evaluation data.
14. Graphs and charts help to visualize data and should allow the reader to quickly and easily pull out the most important information. However the types of visuals one uses matter. For instance, research shows that pie charts are NOT an effective way of visualizing data.

What is Evaluation?

Evaluation is **the systematic determination of the value, effectiveness, significance, or quality of something using credible evidence and standards**. Evaluation collects information to make judgements about programs, initiatives, and interventions. Program evaluation answers the questions of: How well is a program doing? What is the program's value, and is it worth continuing?

Evaluation is important because it shows the work being done, identifies success, identifies ways to improve, demonstrates results, builds public support, and helps to make informed decisions.

Evaluation requires a **systematic collection of information**, meaning that information is collected in a way that is organized, planned, and methodical. Evaluation planning should start when a program is first being planned. Evaluation may be **formative** (conducted to give program staff information on how to improve a program while it is being implemented) or **summative** (conducted after a program is implemented to make decisions on whether a program should continue, be replicated, be expanded, etc.).

Types of evaluation activities include **process evaluation** (Was the intervention delivered as planned?), **outcome evaluation** (What changes occurred *right after* the program?), and **impact evaluation** (Did the program achieve what it set out to, such as reducing substance use?).

There are several core phases in the program evaluation life cycle. These include 1) Engaging stakeholders, 2) Describing the program, 3) Developing evaluation questions, 4) Gathering credible evidence, 5) Analyzing and interpreting data to justify conclusions, and 6) Ensuring use of report results.

Describing a Program

To describe a program, it is necessary to identify the **criteria of merit**, or what the program should accomplish. For example, a substance use prevention program aims to prevent substance use.

A logic model is a graphical representation of how the program is supposed to achieve its goal. A logic model should focus on outcomes, should explicitly connect program activities to outcomes, and should be grounded in evidence. **The Behavior-Determinant-Intervention (BDI) model** connects intervention components to determinants, behaviors, and a health goal. To create a BDI logic model, work backwards by identifying a goal, identifying the specific behaviors tied to that goal, identifying factors that impact these behaviors, and matching these factors to intervention components. Note that the program cannot influence the health goal directly but can influence determinants that impact the behaviors tied to it.

If the goal is to reduce opioid overdose deaths, a contributing behavior is increasing the number of law enforcement officers that administer naloxone. To increase this behavior, law enforcement officers need to feel confident in their ability to administer naloxone. Therefore, a matching intervention would be to train law enforcement officers to administer naloxone. The program cannot directly control overdose deaths and the behaviors associated with it (substance use, naloxone administration, etc.); however, the program can influence related factors, such as the confidence of law enforcement officers in administration of naloxone.

Writing Goals, Objectives, and Evaluation Questions

A mission statement, goals, and objectives of a program provide direction, aid with planning, and provide a way to monitor and evaluate. **A mission statement** states the overall direction of an organization and is very broad, providing a foundation for what the organization wants to do. **Goals** are the expectation of what the program should do, which should be linked to the criteria of merit. **Objectives** are the most specific, and are small, measureable steps towards goals. Objectives can focus on specific activities (process objectives), whether knowledge, attitudes, beliefs, behavior, or the environment has changed (impact objectives), and whether the outcome, such as the number of overdoses, has changed (outcome objectives). Objectives should be SMART, or specific, measureable, achievable, realistic, and time-bound.

When generating evaluation questions, at least one evaluation question should be about the criteria of merit and a standard for that criteria of merit. A criteria of merit is what the program should do. For example, if you have a law enforcement officer training to increase use of a computer system, you must determine if “use of system” has changed. Then, did that change meet the standard. For example, staff and stakeholders would expect X% increase to say that the program was effective.

Gathering Credible Evidence

Process evaluation examines how well a program was implemented, or executed. A program has high implementation fidelity when it closely matches the implementation plan, i.e. the program is conducted in the way it was planned. This can include factors such as dosage, attendance, adherence, quality, adaptation, and participant responsiveness. If an evidence-based program is being implemented, it is important to match these factors as closely as possible, because this will mirror the environments in which the program was found to be effective. The impact may decrease if significant changes are made, such as changing the dosage from 5 sessions to 1 session.

To determine if a program works, or causes a change in the outcome such as substance use, information needs to be carefully collected. Many variables impact substance use that are unrelated to a program and out of your control, such as new city ordinances, a change in employment levels, or outside training programs. How and when program participants are surveyed influences what conclusions can be made. For example, only surveying someone after completion of a program can provide immediate reactionary data but is limited in that it cannot tell you if any change occurred as a result of your program. Surveying a group before and after a program will show changes in the outcome you are measuring but will not tell you if those changes are related to your program or other factors. **To identify change and to attribute that change to your program** requires a survey (or other data collection methods) before *and* after a program for a group that received the program, *and* for a similar group, *that did not receive the program*. This allows for comparisons to be made between the two groups because the only difference is whether the program was experienced. For example, a school-based prevention program may survey middle school students before a program is implemented and afterwards, with one group (Middle School A) receiving the program and another similar group (Middle School B) not receiving the program.

Many pre-made data collection tools (surveys, questionnaires, etc.) are available. It is recommended that you do not create your own survey and use an existing survey if one exists.

Analyzing Evaluation Data

Statistics are a set of procedures used to answer questions and draw conclusions from large amounts of information. **Descriptive statistics** describe a set of data and most commonly include summary statistics such as mean, median, range, standard deviation, etc. Certain types of graphs (e.g., histograms) and tables (e.g., frequency table) are also considered descriptive statistics. One example of a descriptive statistic would be the average number of drinks consumed per week in a sample of college students. Let's imagine among this sample of college students, it was found that males, on average, consumed 6.2 drinks per week and women consumed 3.6 drinks per week. It is important to note that while we may see these numbers and attempt to draw a conclusion that males drink more than females, comparisons and inferences cannot be made from descriptive statistics alone. In other words, we cannot make a judgement about an increase or decrease in some outcome from looking at percentages, average scores, or counts.

To make judgements about group differences or associations between variables, **inferential statistics** must be used. This can include t-tests, correlations, ANOVAs, etc. An example of an inferential statistic would be using a t-test, which is designed to compare the means of two groups, to determine if there is a statistically

significant difference between the number of drinks consumed by college males and females. Another example of an inferential statistic is using a paired samples t-test to compare pre- and post-test scores to determine if any changes between these time points are real or due to random chance.

When the appropriate statistics are used to answer an evaluation question, more accurate conclusions can be drawn. These conclusions can then be used to meaningfully inform programs and policies. A consultant or academic partner with experience in statistics can be extremely helpful in supporting analysis of evaluation data.

Reporting and Using Evaluation Information

Evaluation reports are used to tell the story of the evaluation process and present evaluation data. They answer the questions of *what, so what, and now what?* In other words, what does the data tell us? What does this mean about our program? What do we do with this information? Focus on how you (or others) will use the evaluation information and tailor reports to meet their needs. Do this before you collect data.

Evaluation reports can be technical and comprehensive or more informal, drawing attention to the most relevant information. It is often helpful to consider making both versions of a report. The time period covered in a report can also vary, ranging from ongoing continuous quality improvement reports at meetings to annual reports to program completion reports. It is advisable to engage in ongoing and cumulative reporting of evaluation data.

One of the first steps in preparing to report evaluation data is to determine *what* information to report and *how often* to report this information. These decisions are typically informed by the information needs of your stakeholder(s). Once this is determined, collaborative decisions can be made as to the most appropriate ways to communicate this information (e.g., report, infographic) and the communication channels through which it will be shared.

When preparing a technical evaluation report, the report will typically include the following:

- Title page
- TOC
- Summary
- Context of the Evaluation
- Description of the program
- Logic model
- Evaluation Questions, COMs, and Standards
- Research Design
- Data Collection
- Results
- Recommendations
- Rejoinder
- Appendices

When presenting evaluation data, it is important to consider how to most effectively use graphs, charts, and figures. Visuals help the audience to make sense of the information you are sharing. An effective visual will allow the reader to quickly pull out the most pertinent information. **Of note, research shows that pie charts are NOT an effective way of visualizing data** as the intended message is not accurately received or recalled at a later time.

Steps to Developing an Evaluation Plan

- (1) What is the purpose of the evaluation?
- (2) What are your evaluation questions?
- (3) What are your indicators, questions, COMs?
- (4) What is your research design?
- (5) What data will you collect?
- (6) How will you analyze the data?
- (7) How will the evaluation information be disseminated and used? How will you ensure that the info will be used?