

Outcome Evaluation



PREVENTION RESEARCH CENTER
COLORADO STATE UNIVERSITY

The guidelines below provide a step-by-step approach to conducting an outcome evaluation.

While some aspects may seem technical, they are designed to support your understanding and implementation at every stage. Certain details will be crucial to consider before launching the evaluation, while others can serve as a reference during the process. Don't hesitate to reach out to resources such as university partners (like the Prevention Research Center) who can help with data analysis, or provide experienced evaluators for support as needed.



What is an outcome evaluation and why is it important?



An outcome evaluation measures program effects by assessing participants' progress/changes in the program's targeted outcomes. For example, in a reading program designed to improve children's literacy skills, an outcome evaluation could measure increases in participants' reading comprehension scores from the beginning to the end of the program. By comparing these scores, you can determine whether the program is helping students achieve better reading outcomes.

By answering the questions below, an outcome evaluation is essential for understanding the impact of a program and ensuring it achieves positive results for as many participants as possible.

Key Questions an Outcome Evaluation Helps Answer:

- To what extent is our program influencing participant behaviors in the intended direction?
- Did the program have any unintended effects?
- Are there subgroups of individuals who benefit more or less from the program?
- Are the benefits meaningful enough to justify continuing this program?
- Can we ensure that the program is not causing harm to its participants?

Seeking Expert Guidance Before Designing Your Outcome Evaluation

Program evaluation acts as a checkpoint to assess how well programs are affecting intended outcomes. Evaluating your program is particularly important when working with unique or diverse populations, or if you are implementing a new or adapted program. Though the program being implemented may have worked well for some groups, impacts may vary for other populations/time points/locations.

Among the first questions you should ask when conducting an outcome evaluation is: **“Do we have evaluation expertise within the organization to conduct this evaluation?”** If so, then great! If not, it's a good idea to connect with someone outside your organization with evaluation experience.



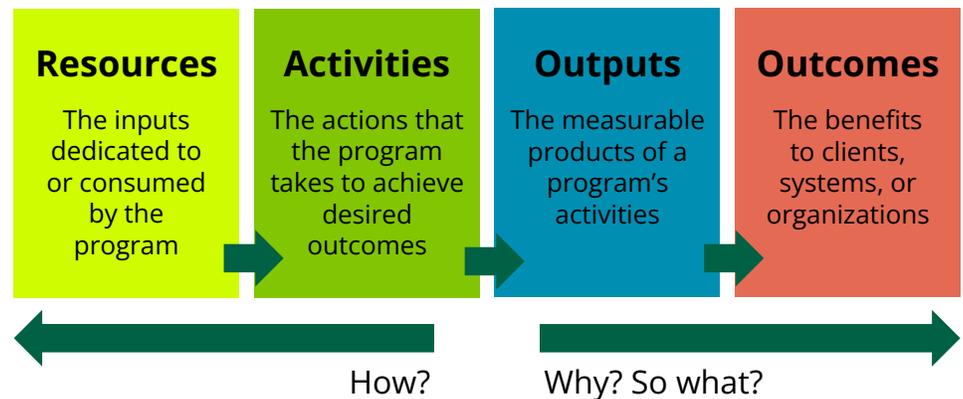
Consider reaching out to data scientists, evaluation experts such as program evaluators, or university researchers, including those at the Colorado State University Prevention Research Center (CSU PRC). They can guide you in designing evaluations, selecting measures, analyzing data, and much more.

Use a Logic Model

Another question you should ask your team is whether your prevention program has an established logic model. Logic models illustrate how your program is expected to have its effects on your desired outcomes. For more information on logic models, visit our [logic model e-course](#) hosted on our [e-learning website](#).

Logic Model

Program Goal: Overall aim or Intended Impact



Outcomes are illustrated on the far right of the above logic model. Outcome evaluations can include evaluations of both long-term outcomes related to the program's goals, as well as immediate short-term outcomes that are hypothesized to lead to change in long-term outcomes.

Design Your Evaluation



Evaluation design refers to the structured plan or framework for assessing the effectiveness, impact, and outcomes of a program or intervention. It defines how data will be collected, analyzed, compared, and interpreted to answer specific evaluation questions, ensuring that the evaluation is systematic and objective.

There are several ways to design an evaluation, each with its own strengths and weaknesses.

Below are some examples of evaluation designs and their relative pros and cons.

Design Your Evaluation

Single Group Retrospective (Posttest Only):

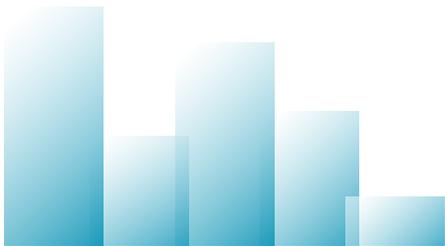
Data collected from a single group after they have experienced the program, and asked to reflect on behavior changes that may have occurred due to the program.

PROS

- **Ease of Data Collection:** Simple to administer, only requires data collection at one point in time.
- **Cost-Effective:** Fewer resources are needed.
- **Participant Recall:** Provides insights into perceived program impact when baseline data are unavailable.

CONS

- **Reliance on Memory:** Data may be affected by memory biases or inaccurate recollections.
- **No Baseline Comparison:** Without pre-data, it's harder to confirm that changes are due to the program rather than other factors.
- **Limited Causal Inference:** Not as rigorous for establishing cause-and-effect, no control or comparison group.



Single Group Pretest/Posttest:

A single group of participants is measured before (pretest) and after (posttest) they experience the program or intervention.

Note: Retrospective post tests may yield less accurate results due to challenges with memory recall and potential bias. However, they can still provide some insights if baseline data are unavailable or unable to be collected.

PROS

- **Measures Change Over Time:** This design allows evaluators to track changes within the same group.
- **Relatively Simple and Cost-Effective:** This design is straightforward and less resource-intensive.
- **Ethical:** All participants are exposed to the intervention.

CONS

- **Lacks a Control or Comparison Group:** Without a control group, difficult to rule out external factors that might have influenced the results.
- **Potential for Testing Effects:** Might perform differently at the posttest because were exposed to the test at baseline.
- **Limited Ability to Infer Causation:** Doesn't control external influences, making it less rigorous for establishing cause-and-effect relationships.

Design Your Evaluation

Two-group, intervention and control condition:

Participants divided into two groups: one that receives the intervention and one that does not. Both are measured on relevant outcomes, before (pretest) and after (posttest) the intervention.

PROS

- **Stronger Causal Inference:** Because you can compare change with participants who did not receive the intervention, you can be more confident that the change in the intervention group is due to the intervention.
- **Control for External Influences:** Intervention and control groups isolate the program's impact.

CONS

- **More Resource-Intensive:** Two groups requires more resources for recruitment and data collection.
- **Ethical Considerations:** Ethical considerations to withholding an intervention from a control group.
- **Risk of Contamination:** If members of the control group inadvertently gain access to the intervention, program effects may appear diluted.

Once you've identified your evaluation design, you can start thinking about how you want to measure change in the desired outcome(s).

Two-group, intervention and wait-list control condition:

This design has both an intervention and a control group. However, in this case, the wait-list control group receives the intervention after the program is tested with a pre-post design. Data may or may not be collected during the "wait-list" portion of the design.

PROS

- **Ethically Fairer:** The wait-list group eventually receives the intervention, addressing ethical concerns.
- **Causal Inference:** This design allows evaluators to more confidently assess program effects.

CONS

- **Potential for Delayed Benefits:** Participants in the wait-list group may experience delays in receiving program benefits causing a higher likelihood of dropout.
- **More Resource-Intensive:** Managing two groups and addressing potential participant drop-out requires significant time, budget and effort.

Key Tips for Selecting Measures

There are two basic categories of data that you can collect:

Quantitative: Focuses on numbers - This includes measures such as multiple choice questions related to the outcomes of interest.

Qualitative: Allows for open responses that provide narratives about people's experiences:

- Interviews with staff; key informant interviews
- Interviews with participants
- Focus groups
- Open ended questions on surveys

Poorly constructed surveys often tell you as little as no survey or, even worse, result in misleading conclusions. If your organization is designing an evaluation from scratch, it is recommended that you seek support from another agency or organization that knows how to find or develop accurate measures for your population. We recommend contacting a program evaluator, your local university, or the **PRC**.

Qualitative data can enrich the findings of quantitative findings. Mixed-methods studies use both methods!

Key Tips for Selecting Measures

Validity and reliability are essential for ensuring that evaluation results are accurate and trustworthy. **Validity is the extent to which a measure accurately captures what it intends to measure. Reliability refers to the consistency of a measure, ensuring that it produces stable results over time or across different evaluators.** High reliability and validity in an evaluation design increase confidence that findings genuinely reflect the program's impact and can be applied to similar contexts or populations.

If possible, use measures with previously established reliability and validity for your focal population. Keep records of where the measures were found and how valid and reliable they are. We suggest having someone on your team who is fluent in measurement validity and reliability or teaming up with university partners for this portion of the process.

- **Select measures designed to show change over appropriate periods of time.** For example, if your intervention spans 30 days, select measures that assess recent behaviors within the same timeframe. Asking about behaviors in the past 30 days, rather than the past year, ensures that responses are relevant to your intervention's specific period.
- **Multiple Items.** Asking a small number of related questions—those about the same behavior or tapping into the same concept--will improve the reliability of your measures.



Importance of Details in Data Collection

Effective outcome evaluation relies on the accuracy of the data that are collected. Here are the key considerations for ensuring high-quality data collection:

- **Timing of Surveys:** The timing of when surveys are administered can significantly impact the quality of the data. Consider factors like the timing within a program and external events that might affect participant availability and mood.
- **Anonymity of Participants:** Ensuring participant anonymity is vital for honest responses. When individuals know their answers are confidential, they are more likely to share truthful feedback. Clearly communicate how their data will be used and how their identity will remain private.
- **Unique IDs:** Using unique IDs (i.e. numeric 001, 002, date based 20241101-001, or random X4Y7Z2) for each participant helps track responses without compromising anonymity. This allows for the aggregation of data while also ensuring that individual responses remain confidential.
- **Format of Surveys:** Consider the format of your surveys to maximize response rates and data quality (Paper vs. Electronic; platform for electronic survey – Qualtrics, Google forms, etc.). Ensure questions are clear, concise, and relevant to the participants' experiences.
- **Pilot Testing:** Before rolling out surveys widely, conduct pilot tests with a small group. This will allow you to identify any issues with the survey design or instructions and make necessary adjustments to improve clarity and effectiveness. Ensure that participants understand what is expected of them and the importance of each question to reduce confusion and increase reliability.
- **Training for Data Collectors:** If your evaluation involves data collectors, ensure they receive training on best practices for collecting and handling data. This includes understanding how to maintain confidentiality and assisting participants with questions.



Data Collection and Management Plan for Implementers

Implementing a robust data collection and management plan is essential for maintaining data integrity and facilitating analysis. Consider the following components:



Data Collection Tools: Clearly define which tools (surveys, interviews, etc.) will be used and ensure they are accessible to all participants.



Data Storage: Establish secure data storage protocols to protect participant information and maintain confidentiality.



Data Management Procedures: Develop procedures for organizing, processing, and analyzing collected data. Specify who will have access to data and how it will be shared among the evaluation team.



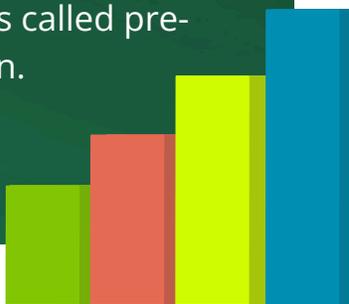
Regular Monitoring: Implement a system for regular monitoring of data collection processes to identify any issues and adjust as needed.

By addressing these considerations and implementing a comprehensive data collection and management plan, evaluators can enhance the quality and reliability of their findings.

What does it look like in practice?

Consider this...

Consider a program with the goal of reducing bullying in the classroom. To evaluate its effectiveness, it's essential to first measure how often bullying occurs before the program begins—this is known as collecting baseline data. By comparing this baseline data to measurements taken after the program is implemented, one can assess changes over time, a process called pre-post evaluation.



Plan for an evaluation early!

Planning for an outcome evaluation should start before program delivery. Program evaluation should be planned before the implementation stage to ensure your program has the tools needed to execute it. For example, if your program wants to measure the baseline issues your participants are experiencing before your program intervenes, this will include a pre-intervention survey or interview.

If baseline data were not collected before starting the program, alternative approaches can help gather meaningful insights. These include asking participants:

- To reflect on and describe their experiences before and after the intervention.
- What they feel they learned during the program.
- Key takeaways from the intervention.
- How they plan to apply the skills they've gained moving forward.

Look to your program's resources and tools:

Evidence-based programs (EBPs) often have existing valid and reliable outcome evaluation assessments aligned with their logic model. Utilizing the program's established survey and/or measurement procedures can greatly improve the quality, ease, and efficiency of the evaluation process. Established surveys often have guides for scoring these surveys too.

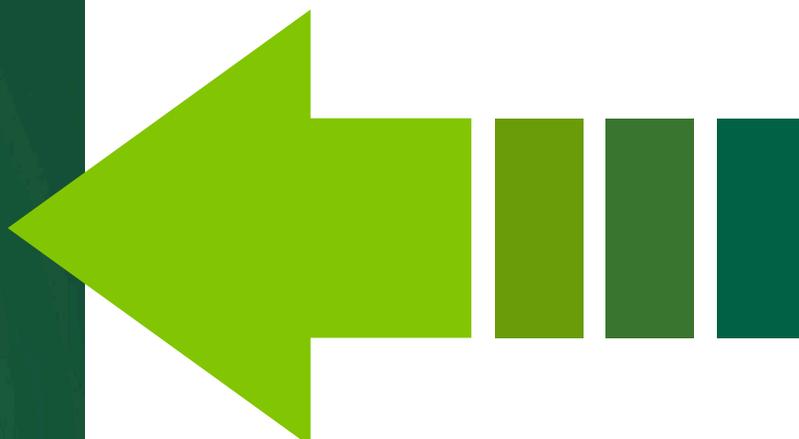
In cases where measures are not available for a given program, you can still design an outcome evaluation plan.

Evidence-based programs can be found on websites such as:

- [Blueprints for Healthy Youth Development](#)
- [Pew Clearinghouse](#).

This may also be useful when:

- You are seeking out additional outcomes the program may be affecting; this may be a call to add questions to the existing survey.
- You find the original surveys inappropriate for the participants; for example, if you've also adapted the program itself; this may be a call to also adapt your survey to better align with the group you aim to serve.
- It's always advisable to use pre-designed Evidence-Based Programs. However, if you choose to develop your own program, it's essential to conduct ongoing evaluations and collaborate with a partner to effectively measure outcomes.



In Summary:

Follow these steps to evaluate your program's impact using established resources.

- **Identify a Qualified Professional:** Engage a professional skilled in measurement and statistical analyses. This person can help with selecting reliable measures, interpreting data, and ensuring accurate coding of responses.
- **Locate Existing Evaluation Resources:** Seek out any existing evaluation tools related to your program, such as surveys, interview guides, or focus group scripts. If you can't find these resources online or in course materials, consider contacting the program developers directly for guidance.
- **Define Program-Specific Outcomes:** Review your program's defined outcomes and long-term goals to decide which specific outcomes you want to measure. Effectiveness evaluations work best when they directly assess these predefined outcomes.
- **Select Appropriate Methods:** Choose the methods that will best capture changes in these outcomes—whether through surveys, interviews, or focus groups. Whenever possible, use established measures that align with your program's objectives.
- **Collect Data:** Carefully gather data using the tools and methods you've selected, ensuring consistency and accuracy throughout the process. Stay organized and track progress to address any challenges.
- **Analyze Findings:** Examine the data to identify patterns, trends, and key insights that address your evaluation questions. Once the data are interpreted, you'll be ready to share results using the guidance in our [Dissemination: Communicating Evaluation Results tool](#).



Recommendations for Data Analysis



- **Work with a Statistical Expert:** Having a statistical expert is essential for accurate and meaningful analysis. If your team does not already have someone with this expertise, consider partnering with a local university, hiring a consultant, or reaching out to organizations that provide evaluation support.
- **Start with User-Friendly Tools:** Communities often use tools like Excel for basic analysis. Excel is accessible and allows for straightforward calculations such as averages and percentages, and visualizations like bar charts.
- **Consider Specialized Statistical Software:** These could include tools like R (a free and highly flexible option), SPSS (a user-friendly software widely used in social sciences), Stata, and others.
- **Use Online Training and Tutorials:** Platforms like YouTube, Coursera, and Khan Academy offer free or low-cost tutorials for statistical tools and methods. Take advantage of these resources to build your team's skills.
- **Match Your Methods to Your Questions:** Your data analysis should align with the types of questions your evaluation seeks to answer. For example:
 - Use basic descriptive statistics (e.g., averages, counts) for simple summaries.
 - Consider statistical tests (e.g., t-tests, chi-square tests) if comparing groups.
 - Use trend analysis if examining changes over time.
- **Seek External Validation:** Before finalizing your findings, consider sharing your results with an external evaluator or someone with advanced statistical expertise to confirm your interpretation. Once an evaluation plan is established and challenges have been addressed, it becomes a sustainable framework that can be effectively carried out year after year with minimal support.

References



1. Blueprints for Healthy Youth Development. (n.d.). Retrieved from <https://www.blueprintsprograms.org/>
2. CDC TRAIN. (n.d.). Introduction to Program Monitoring and Evaluation. Retrieved from <https://www.train.org/cdctrain/course/1079247/details>
3. Centers for Disease Control and Prevention. (2024). CDC Program Evaluation Framework. <https://www.cdc.gov/evaluation/php/evaluation-framework/index.html>
4. Centers for Disease Control and Prevention. (2020). Cost-effectiveness analysis. https://www.cdc.gov/polaris/php/economics/cost-effectiveness.html?CDC_AAref_Val=https://www.cdc.gov/policy/polaris/economics/cost-effectiveness/index.html
5. Center for the Evaluation of Value and Risk in Health. (n.d.). CEA Registry. Retrieved from <https://cevr.tuftsmedicalcenter.org/databases/cea-registry>
6. Pennsylvania State University. (n.d.). Education Policy Initiative. Retrieved from <https://epis.psu.edu/>
7. Gartlehner G, Hansen RA, Nissman D, et al. Criteria for distinguishing effectiveness from efficacy trials in systematic reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); Technical Reviews, No. 12(1), Introduction. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK44024/>
8. Gross, D., & Fogg, L. (2004). A critical analysis of the intent-to-treat principle in prevention research. *Journal of Primary Prevention*, 25, 475-489.
9. Pew Charitable Trusts. (n.d.). Results First Clearinghouse Database. Retrieved from <https://www.pewtrusts.org/en/research-and-analysis/data-visualizations/2015/results-first-clearinghouse-database>
10. Singal, A. G., Higgins, P. D., & Waljee, A. K. (2014). A primer on effectiveness and efficacy trials. *Clinical and translational gastroenterology*, 5(1), e45. <https://doi.org/10.1038/ctg.2013.13>