



# The Program Manager's Guide to Evaluation

Third Edition

OPRE Report No. 2022-208



# Preface to the Third Edition

The Office of Planning, Research, and Evaluation (OPRE), a unit within the Administration for Children and Families (ACF), advises the Assistant Secretary for Children and Families on increasing the effectiveness and efficiency of programs to improve the lives of children and families.

In collaboration with ACF program offices and others, OPRE studies ACF programs and the populations they serve through rigorous projects, including evaluations, research syntheses, and exploratory studies. OPRE also supports ACF programs in the responsible management and use of data to improve the effectiveness and efficiency of human services programs.

OPRE's research portfolio spans a wide array of ACF program areas, including welfare and family self-sufficiency, employment and training, early care and education, child and youth development, child welfare, family strengthening, and more. The office's broad-reaching studies explore program effectiveness and strategies to improve efficiency, test innovative service delivery models or strategies, and identify areas for further research.

Toward these ends, OPRE is proud to present an updated edition of *The Program Manager's Guide to Evaluation*. The previous editions of the Guide have been popular and well-received resources. To assess potential users' needs for the third edition, we solicited input from ACF management, program staff, intermediary organizations that work with ACF programs, and practitioners and researchers who have experience developing and using evaluation and technical assistance resources. To validate and expand on important learnings from these discussions, we recruited four evaluation experts to provide input on strategic content priorities and identify timely and relevant resources.

The new edition has been updated to reflect currently accepted practices, up-to-date terminology, and issues to consider. For example, we added a section titled "Practice Culturally Responsive and Equitable Evaluation" to each chapter to provide program managers with clear, actionable guidance on implementing diversity, inclusion, and equity principles in each phase of evaluation. Based on feedback within ACF, we have incorporated many stylistic changes to help program managers navigate more quickly to the information they seek. Each chapter now begins with a "roadmap" to summarize the main takeaways and provide hyperlinks to relevant subsections. Finally, to strengthen the real-world application of the guide's content, we have thoroughly updated and reorganized the appendices to include complementary resources aligned with different evaluation topics covered in the Guide.

# Preface to the Third Edition (continued)

As with the first two editions of *The Program Manager's Guide to Evaluation*, this updated edition explains what program evaluation is, why evaluation is important, how to conduct an evaluation and understand the results, how to report evaluation findings, and how to use evaluation results to improve programs that benefit children and families.

Emily Schmitt  
*Deputy Director*  
Office of Planning, Research, and Evaluation

Connect with OPRE





# Acknowledgements

Numerous individuals contributed to the development of this third edition. At the Administration for Children and Families, contributors include Harmanpreet Bhatti, Kim Clum, Amanda Coleman, Nicole Denmark, Kathleen Dwyer, Calonie Gray, Kriti Jain, Emily Schmitt, and Maria Woolverton. The Insight Policy Research (Insight) team received invaluable guidance from four expert consultants: Matthew Courser at the Pacific Institute for Research and Evaluation, Karin Coyle at ETR, Myra Park at the University of Washington, and Teresa Eckrich Sommer at Northwestern University's School of Education and Social Policy. At Insight, Scott Cody and Debra Wright provided quality assurance guidance.

Suggested citation:

El Mallah, S., Gutuskey, L., Hyra, A., Hare, A., Holzwart, R., & Steigelman, C. (2022). *The program manager's guide to evaluation* (OPRE Report 2022-208). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.



# Contents

Chapter 1. An Introduction to Program Evaluation and This Guide.....	1.1
Introduction .....	1.2
Defining Program Evaluation.....	1.3
Understanding the Benefits of Program Evaluation.....	1.6
Considering Federal Program Evaluation Standards.....	1.8
Guidelines for Conducting a Successful Evaluation.....	1.9
References .....	1.11
Chapter 2. An Overview of the Program Evaluation Process.....	2.1
Introduction .....	2.2
Basic Steps in Program Evaluation.....	2.2
Practice Culturally Responsive and Equitable Evaluation Throughout the Evaluation Process.....	2.4
References .....	2.8
Chapter 3. Engage an Evaluation Team.....	3.1
Introduction .....	3.2
Decide Whom to Engage in Your Evaluation .....	3.2
Hire and Manage Your External Evaluator .....	3.6
Practice Culturally Responsive and Equitable Evaluation When Engaging an Evaluation Team.....	3.14
References .....	3.16
Chapter 4. Prepare for the Evaluation .....	4.1
Introduction .....	4.2
Decide What to Evaluate .....	4.2
Develop Evaluation Questions.....	4.4
Build and Use a Logic Model.....	4.5
Develop Measurable Objectives.....	4.7
Prepare an Evaluation Budget .....	4.9

# Contents (continued)

Practice Culturally Responsive and Equitable Evaluation When Preparing for an Evaluation .....	4.11
References .....	4.13
Chapter 5. Design Your Evaluation.....	5.1
Introduction .....	5.2
Refine Your Implementation Evaluation.....	5.3
Choose Your Evaluation Designs .....	5.4
Choose Evaluation Data Collection Methods .....	5.9
Protect Evaluation Participants .....	5.11
Choose Evaluation Samples.....	5.12
Manage and Monitor the Evaluation.....	5.13
Navigate Unexpected Changes.....	5.14
Practice Culturally Responsive and Equitable Evaluation When Designing an Evaluation .....	5.16
References .....	5.19
Chapter 6. Gather Credible Evidence.....	6.1
Introduction .....	6.2
Identify the Best Data Sources.....	6.2
Select or Develop Data Measures.....	6.5
Develop Instruments .....	6.8
Create Data Collection Procedures .....	6.9
Protect Study Participants.....	6.12
Monitor Data Collection Activities.....	6.13
Practice Culturally Responsive and Equitable Evaluation When Gathering Data .....	6.15
References .....	6.18
Chapter 7. Analyze Data .....	7.1
Introduction .....	7.2
Preparing Data for Analysis.....	7.2



# Contents (continued)

Assessing Data Quality .....	7.5
Analyze Implementation Data .....	7.7
Analyze Outcome Data .....	7.11
Interpreting Your Findings .....	7.14
Practice Culturally Responsive and Equitable Evaluation When Analyzing and Interpreting Data .....	7.15
References .....	7.17
Chapter 8. Share Lessons Learned.....	8.1
Introduction .....	8.2
Plan for Communication and Dissemination.....	8.2
Develop Reports and Communication Products.....	8.4
Communicate Results .....	8.10
Practice Culturally Competent and Equitable Evaluation When Sharing Lessons Learned .....	8.12
References .....	8.14
Appendix A. Additional Resources.....	A.1
Evaluation Design and Planning .....	A.1
Evaluation Reports.....	A.9
Performance Measurements and Indicators .....	A.11
Program Design and Implementation.....	A.17
Professional Associations .....	A.19
Appendix B. Templates and Examples.....	B.1
Logic Model.....	B.1
Implementation Objectives Stated in Measurable Terms .....	B.3
Participant Outcome Objectives Stated in Measurable Terms.....	B.4
Analyzing Information on Implementation Objectives.....	B.5
Final Report Outline.....	B.6

# Contents (continued)

## Figures and Tables

Figure 1.1. Federal Components of Evidence Framework.....	1.3
Table 1.1. Common Types of Evidence-Building Activities.....	1.4
Figure 3.1. Continuum of Engagement in Evaluation.....	3.4
Figure 3.2. Potential Evaluation Team Composition.....	3.5
Table 3.1. Types of Evaluators.....	3.6
Table 3.1. Types of Evaluators (continued).....	3.7
Figure 4.1. Potential Evaluation Target Options.....	4.3
Figure 4.2. Basic Elements of a Logic Model.....	4.6
Table 4.1. Example of Implementation Objectives Stated in Measurable Terms.....	4.8
Table 4.2. Example of Outcome Objectives Stated in Measurable Terms.....	4.9
Table 4.3. Common Cost Drivers and Cost Savers in Program Evaluation.....	4.10
Table 5.1. Possible Designs for Outcome Evaluations.....	5.7
Table 5.1. Possible Designs for Outcome Evaluations (continued).....	5.8
Table 5.2. Possible Data Collection Methodologies.....	5.9
Table 5.2. Possible Data Collection Methodologies (continued).....	5.10
Table 5.2. Possible Data Collection Methodologies (continued).....	5.11
Table 6.1. A Comparison of Existing Versus New Data Collection Measures.....	6.7
Table 7.1. Terms Used to Establish Trustworthiness of Qualitative Data.....	7.6
Table 7.2. Sample Table for Analyzing Information on Implementation Objectives.....	7.10
Table 8.1. Communication Product Formats, Messages, and Audiences.....	8.5
Table 8.1. Communication Product Formats, Messages, and Audiences (continued).....	8.6
Table 8.2. Communication Channels.....	8.11
Table A.1. Professional Associations and Descriptions.....	A.19
Table B.1. Sample Logic Model: Child Abuse Prevention Program.....	B.1
Table B.2. Worksheet: Logic Model.....	B.2
Table B.3. Worksheet: Describing Program Implementation Objectives in Measurable Terms.....	B.3
Table B.4. Worksheet: Participant Outcome Objectives Stated in Measurable Terms.....	B.4
Table B.5. Worksheet: Analyzing Information on Implementation Objectives.....	B.5



# Chapter 1. An Introduction to Program Evaluation and This Guide

## What's Inside?



### ▶ What this chapter contains

The importance of program evaluation and this Guide

### ▶ Who can use this chapter

Program managers new to evaluation or seeking an overview of this Guide

Click the links below to view the relevant section



# Introduction

As a program manager, you already know human services programs aim to improve individuals' families' and communities' health and well-being. However, you may be wondering why it is important to evaluate programs, how to conduct an evaluation, or how to ensure an evaluation is equitable for your community. This Guide will address these questions and connect you to other resources that can help.

Each chapter of this Guide addresses specific steps in the evaluation process and provides guidance on how to tailor an evaluation to your program's needs. The following features will help you navigate the Guide.

## Stand-alone chapters

Reading the full Guide from start to finish is an excellent way to build an understanding of the evaluation process. However, each chapter also functions as a stand-alone document, enabling navigation to specific topics of interest.

- **An initial overview of the chapter.** At the start of each chapter, you will find a "What's Inside?" box describing what the chapter addresses, who might be interested in reading it, and what subsections are included.
- **Links for further reading.** This Guide will explain important concepts at a high level. However, if you are looking for more information on a topic, you will find links to other chapters or appendices for further reading.
- **Illustrative examples.** This Guide has examples and explanations embedded throughout in callout boxes to illustrate important concepts.
- **Appendices with additional resources and tools.** If you have reviewed the Guide and want more resources, check out the appendices. [Appendix A](#) provides additional resources such as relevant evaluation materials and lists of professional organizations. [Appendix B](#) provides helpful templates, tools, and worksheets to support your evaluation efforts.

### Practice culturally responsive and equitable evaluation

At the end of every chapter, the Guide provides approaches and examples of ways to incorporate principles of equitable evaluation in each stage of evaluation. These principles will help ensure your evaluation benefits from contributions from program participants and members of the communities your program serves.

## Defining Program Evaluation

Program evaluations can determine how a program is operating, reveal if it is working as intended, determine whether it has achieved its objectives, and identify potential areas for improvement. What distinguishes program evaluation from the more informal feedback program managers and staff obtain from program users is the systematic approach.<sup>1</sup> This approach ensures the information is gathered and analyzed objectively.

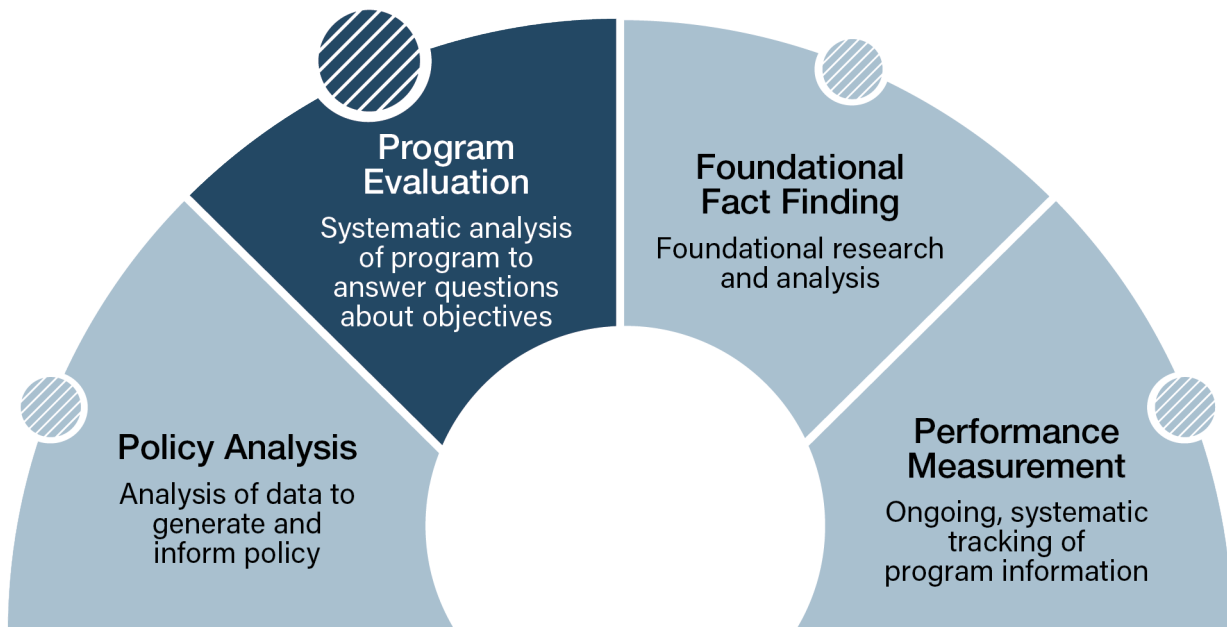
By following well-established steps to collect, analyze, and use data to answer questions about a program, you can achieve results that clearly assess the quality of program activities. Your evaluation findings can then be used to inform local decision-making, build organizational capacity, and/or facilitate the use of rigorous evidence among federal, state, and local agencies.

### Federal guidance

For more details on federal evidence building, as identified by the Office of Management and Budget, see Young (2021).

The recently enacted Foundations for Evidence-Based Policymaking Act of 2018 (also referred to as the Evidence Act) reinforces the importance of using data and evidence to solve complex issues and challenges. Program evaluation is just one approach to generating evidence-based knowledge. The Office of Management and Budget identifies four ways evidence can be collected (evidence-building activities)—foundational fact finding, policy analysis, program evaluation, and performance measurement (Young, 2021) (see figure 1.1). While each of these evidence-building activities can provide useful information about

Figure 1.1. Federal Components of Evidence Framework



Source: Young, 2021

<sup>1</sup>A systematic approach is methodological and repeatable and can be learned via a step-by-step procedure. The Law Dictionary. (n.d.). *What is systematic approach*. <https://thelawdictionary.org/systematic-approach/#:~:text=The%20approach%20that%20is%20methodical,step%2Dby%2Dstep%20procedure>

your program, this Guide focuses on one specific component of evidence-based policymaking—program evaluation.

Program evaluation is a broad term that can include many types of research activities, all involving a systematic process of collecting, analyzing, and using data to answer questions about a program’s objectives. This Guide focuses specifically on evaluation that assesses the implementation of a program or its outcome objectives, defined as follows:

- **Program implementation objectives.** What you plan to do in your program, how you plan to do it, and your intended target population (e.g., the services or training you plan to provide, the number of people you plan to reach, the staff training you plan to conduct)
- **Participant outcome objectives.** Your expectations about how your program will change participants’ knowledge, attitudes, behaviors, or awareness

You may be familiar with other types of evaluation and research activities as highlighted in table 1.1. These activities may also yield critical information, complement implementation or outcome evaluation activities, and offer relevance to one or more evidence components. However, they are not the primary focus of this Guide, which describes program evaluation activities most relevant to a broad range of program managers. You can find more detail about other types of evidence-building activities in the appendices.

**Table 1.1. Common Types of Evidence-Building Activities**

Type of activity	When is it typically used?	What is the focus?	Why is it useful?	
Primary focus of Guide	Implementation (also called Process) Evaluation	<ul style="list-style-type: none"> <li>■ Early stages of program implementation</li> <li>■ Ongoing</li> </ul>	<p>Assesses whether program activities are being implemented as planned, whether the expected program services are being delivered, and how the program is operating in practice</p> <p>Might collect information on processes, content, quantity, quality, and structure of program activities</p>	Provides periodic feedback that compares actual performance with target objectives or standards the program seeks to achieve
	Outcome (including Effectiveness or Impact) Evaluation	<ul style="list-style-type: none"> <li>■ At defined intervals</li> <li>■ At end of program delivery</li> </ul>	<p>Offers understanding as to whether change has occurred as intended</p> <p>For certain types of outcome evaluations (i.e., impact evaluations), can focus on long-term program effects or results and whether change can be attributed to program activities</p>	<p>Can provide evidence of program effectiveness</p> <p>Can explain the population that benefits most from the program and the conditions needed for success</p>

Table 1.1. Common Types of Evaluation and Research Activities (continued)

Type of activity	When is it typically used?	What is the focus?	Why is it useful?
Foundational Fact Finding	<ul style="list-style-type: none"> <li>Anytime</li> </ul>	Systematically describes a program without inferring causality or measuring effectiveness	Helps describe what is happening in the program or among the target population. Can provide insights into the demographic characteristics of the target population or what characteristics are related to a particular outcome
Formative Evaluation	<ul style="list-style-type: none"> <li>Before program is fully implemented</li> </ul>	Assesses whether program—or some aspect of the program—is feasible, appropriate, and acceptable before it is fully implemented. Can include process and outcome measures	Determine what aspects of the program are working as intended and whether the program can be implemented as designed
Developmental Evaluation	<ul style="list-style-type: none"> <li>Early stages of developing new program</li> <li>Ongoing</li> </ul>	Provides rapid and real-time feedback to promote innovation and support program adaptation	Offers methodological flexibility and can be better suited in less predictable situations or complex contexts
Continuous Quality Improvement, Progress Monitoring, or Performance Measurement	<ul style="list-style-type: none"> <li>Throughout program delivery</li> </ul>	Offers understanding of how the program is working and not working, who it is reaching, how it operates differently across contexts, and how it is progressing toward established goals	Supports improvement in program design and delivery
Economic or Efficiency Evaluation	<ul style="list-style-type: none"> <li>After program delivery</li> </ul>	Assesses the cost-effectiveness (i.e., cost per desired outcome) and cost-benefit (i.e., cost per overall benefit)	Facilitates comparison with other programs designed to achieve the same outcomes
Implementation Science Research	<ul style="list-style-type: none"> <li>Late stages of scale-up and replication</li> </ul>	Identifies what is needed to bring effective strategies to scale by examining factors that promote the uptake of evidence-based practices into routine settings	Maximizes evaluation investment and promotes the translation of evaluation findings into practice (i.e., closes the gap between “what we know” and “what we do”)

\* Note: While this Guide does not discuss needs or evaluability assessments in detail, you will find information throughout about how to design an appropriate and feasible evaluation.

Source: Table content compiled from multiple sources including Types of Evaluation (Centers for Disease Control and Prevention [CDC], 2020), Components of Evidence (Office of Management and Budget, 2019), and Cost Analysis Standards Project (American Institutes for Research, 2021).

This Guide provides general advice for conducting or participating in implementation and outcome evaluations, with links to helpful resources, templates, and tools throughout. It is not intended to make you an evaluation expert but rather to provide you with an understanding of the overall process and considerations to keep in mind along the way.

# Understanding the Benefits of Program Evaluation

Familiarity with the program evaluation process and its many benefits can help as you begin to engage your team and community.

## Benefit 1

### Program evaluation helps programs use limited resources more effectively

Evaluation is a worthwhile investment in your program. While it's true that implementing an evaluation will require additional staff time and financial resources, evaluations can provide actionable information and evidence that, over the long term, can help you improve and revise services to make them even more effective and efficient (e.g., eliminating program components not necessary to achieve intended outcomes). For more information about the steps to conduct program evaluation, see [chapter 2](#). For more on evaluation costs and budgeting, see [chapter 4](#).

## Benefit 2

### Program evaluation supports the work of program staff

In addition to answering basic questions about a program's effectiveness (e.g., how the program is being implemented, how participants experience the program, ways participants might benefit from program participation), evaluation activities can also identify strategies program staff should employ to improve their work. Ideally, evaluation should become part of your program, integrated into the way you do business. In many cases, evaluation processes can be incorporated into ongoing program activities (e.g., collecting evaluation information from clients as part of normal intake processes) and program management tasks (e.g., reviewing data trends during management meetings). For more about how your team may support evaluation, see [chapter 3](#). For more about ways to obtain the information you need for an evaluation, see [chapter 6](#).

## Benefit 3

### Program evaluation methods can be adjusted to your needs

Program evaluation can build your personal and organizational capacity to collect, use, and understand data and research. As a practitioner, you have built a specialization in serving children, youth, and families. But you also routinely examine information about your program and make decisions.

### Community members

This Guide uses the term "community members" throughout the chapters to refer to people who are eligible for or receive your program services or who share geographical or cultural identities with your program participants.

Participating in a program evaluation is a great way to build your knowledge around data collection and analysis. It can refine your ability to collect accurate information about your program participants and your ability to use that information to make decisions. Participating in evaluations will also enhance your ability to interpret and incorporate new research findings in the future. For more information about hiring an outside evaluator, see [chapter 3](#). For more information about preparing for an evaluation and developing an evaluation plan, see [chapters 4](#) and [chapter 5](#), respectively.

## Benefit 4

### Program evaluation supports program improvement

An evaluation will likely document facilitators and barriers to program effectiveness and may reveal areas where a program is exceeding goals. All this information helps you make evidence-based decisions and provide opportunities to improve your program. As evaluation becomes part of your organization's culture, you will find you regularly look toward evaluation results to manage your program, remain accountable to community members and funders, and secure funding and support for your organization's future. For more information about how to use the results of your evaluation, see [chapter 7](#). For more details about reporting your findings, see [chapter 8](#).

## Benefit 5

### Program evaluation offers unique benefits compared with other types of research and evidence building

This Guide walks you through the steps of conducting implementation and outcome evaluations. An evaluation will help you see how a program is carrying out its work (implementation evaluation) or the effectiveness of a program in producing change (outcome evaluation). While other evidence-building activities (such as an analysis of Census data to understand local community demographics) can help inform your work generally, only program evaluation efforts are specific to your program, your staff, your participants, your community, your context, and your world today. Program evaluation asks different questions and provides unique benefits (see [figure 1.1](#) and [table 1.1](#)).

#### Evaluation principles

To learn more about the guiding principles to evaluation design in the Administration for Children and Families' Evaluation Policy, see their website (OPRE, n.d.).

## Benefit 6

### Good evaluation can engage your community

As a program manager, you know programs work best when they meet the needs of a specific community. When program activities—including evaluation—are respectful and inclusive of the communities they serve and reflect the communities' aspirations, they can be more successful. You can structure your evaluation to provide frequent opportunities to engage your community throughout the research cycle (e.g., during the design phase, through data collection, sharing results when the evaluation is complete). Evaluation can also

empower your community to be partners (e.g., engaging them to collectively affirm your program goals) and co-creators of knowledge and information generation. Throughout this Guide, you will find examples of ways to ensure your evaluation is appropriate for and inclusive of the populations you serve. You can find more information about community engagement and culturally appropriate research practices embedded throughout this Guide.

## Benefit 7

### Evaluation can inform policy

Policymakers can make better decisions when they have credible evidence—evidence collected through high-quality, carefully constructed and executed evaluations—about how to design and implement programs, and what does and does not work. By conducting an evaluation and sharing your findings, you contribute to collective knowledge about human services programs and potentially influence decisions that can inform oversight and funding for programs like yours.

## Considering Federal Program Evaluation Standards

Careful planning is needed to design an effective, high-quality evaluation, but methods and approaches may vary, depending on your program's unique circumstances and needs. The Administration for Children and Families' evaluation policy outlines the following broad evaluation principles to consider when designing an evaluation:

- **Rigor.** Does the evaluation use the most technical and credible methods appropriate to the evaluation questions and feasible within budget and other constraints?
- **Relevance.** Does the evaluation address questions of importance and serve the information needs of program staff, community members, funders, and the research field?
- **Transparency.** Is information about the planning, implementation, and reporting phases of the evaluation available and accessible to enable accountability, and does it enable users of the evaluation to understand and critique the design and methods?
- **Independence.** Does the evaluator operate with an appropriate level of independence from programmatic, regulatory, policymaking, and funding organizations activities?
- **Ethics.** Does the evaluation safeguard the dignity, rights, safety, and privacy of participants, community members, and affected entities?

These principles can also be found in the evaluation policies of several other federal agencies and in guidance related to the Foundations for Evidence-Based Policymaking Act of 2018.

Balancing these principles will help protect you against launching an evaluation that is, for example, rigorous and independent but not useful to your organization's decision-makers or sensitive to the cultural background of your program participants.



# Guidelines for Conducting a Successful Evaluation

As you prepare for and conduct your evaluation, the following guidelines can help you enhance its benefits. We address many of the guidelines in more depth throughout this Guide.

## Invest in planning

Invest time and effort in deciding what you want to learn from your evaluation. This is the single most important step you will take in this process. Consider what you would like to discover about your program and how it affects participants, and use this information to guide your evaluation planning.

### What about evidence standards?

You may have heard of “evidence standards,” which are standardized criteria that must be met for programs to be judged as tested and effective. Sometimes program evaluations are expected to generate specific types of evidence. For more details on evidence standards, see [chapter 5](#).

## Tailor your evaluation to your program’s needs

There is no one-size-fits-all approach to program evaluation. An evaluation’s design will depend on several factors, including the types of research questions you need to address; your program’s structure, objectives, and resources; your community and program participants’ information needs, past experience with evaluation, and trust/comfort with the program; and how evaluation results will be used. Take the time to thoroughly assess these factors, and tailor your evaluation accordingly, rather than assuming you must use a specific type of evaluation design or data collection methodology.

## Integrate the evaluation in ongoing program activities

Some program staff may see evaluation as something an outsider does after a program is over or as an activity “tacked on” to please funders. Unfortunately, many programs are evaluated in this way. You can increase the benefits of an evaluation by planning it and the program simultaneously so you can use evaluation feedback to inform and improve program operations.

## Participate in the evaluation and show program staff you think it is important

An evaluation needs the program manager’s participation to succeed. Even if you hire an outside evaluator, that person or team cannot do the needed work without your input, and they will need you to teach them about your program, your participants, and your objectives. Staff will value the evaluation if you, the program manager, value it. Talk about it with staff individually and in meetings. If you hire an outside evaluator to conduct the evaluation, have them attend staff meetings and give evaluation updates and receive input and feedback from program staff. Your involvement will encourage a sense of ownership and responsibility for the evaluation among all program staff.

Identify and engage people and organizations who are interested in the program as much as possible and as early as possible. Program staff, partners, participants, community members, and others have considerable interest in the evaluation's success, and their ongoing engagement can make the evaluation more meaningful, relevant, and potentially more likely to inform future practice. These individuals and their representing organizations will have questions and issues the evaluation can address. Because of their experiences and expertise, program staff, community members, and policymakers can ensure the evaluation questions, design, and methodology are appropriate. Community members and program participants may also be more willing to participate in data collection and other evaluation-related tasks if they have been invested in the process along the way.

## Be realistic about timing and burden

Evaluations take work. Even if your evaluation uses an outside evaluator to help design, collect, and analyze the data, time is needed to arrange for the evaluator to have access to records, administer questionnaires, or conduct interviews. Agencies and evaluators often underestimate how much additional effort these activities involve. When program managers and staff brainstorm about all the questions they want answered, they often produce a long list. Creating buy-in through brainstorming is often a good start to program evaluation, but it is important to narrow goals and information requested so your final evaluation plan is not too complicated. Focus on your top priority questions to ensure your evaluation is feasible within your budget, timeframe, and scope.

## Address cultural and ethical issues

Good evaluation aligns with the social and cultural context of program participants and their communities. For example, it respects the cultural backgrounds and individuality of program participants and staff, makes use of their knowledge and strengths, and incorporates culturally sensitive data collection methods and instruments. It may also engage community members in various steps in the evaluation. Participants must be informed that they are taking part in an evaluation, and they have the right to refuse to participate in this activity without jeopardizing their participation in the program. You must also ensure confidentiality of participant information will be maintained.

## Manage expectations

Make sure you clearly communicate with program staff, your participants, and others for whom the program's success is important, such as community members and funders, about the evaluation's goals, processes, scope, and the intended use of the findings. This will help ensure everyone is on the same page and knows what to expect regarding products the evaluation will produce, questions it will answer, and decisions its findings can inform.

## References

- American Institutes for Research. (2021). *Standards for the economic evaluation of educational and social programs: Cost standards project*. <https://www.air.org/sites/default/files/Standards-for-the-Economic-Evaluation-of-Educational-and-Social-Programs-CASP-May-2021.pdf>
- CDC (Centers for Disease Control and Prevention). (n.d.). *Types of evaluation*. <https://www.cdc.gov/std/program/pupestd/types%20of%20evaluation.pdf>
- OPRE (Office of Planning, Research, and Evaluation). (n.d.). *ACF evaluation policy*. U.S. Department of Health and Human Services, Administration for Children & Families. <https://www.acf.hhs.gov/opre/report/acf-evaluation-policy>
- Peters, D., Adam, T., Alonge, O., Agyepong, I. A., & Tran, N. (2013). Implementation research: What it is and how to do it. *British Journal of Sports Medicine* 48(8). <https://doi.org/10.1136/bmj.f6753>
- Vought, R. T. (2019). *Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning agendas, personnel, and planning guidance* [Executive Office of the President, Office of Management and Budget, Memorandum for Heads of Executive Departments and Agencies]. <https://www.whitehouse.gov/wp-content/uploads/2019/07/M-19-23.pdf>
- Young, S. D. (2021). *Evidence-based policymaking: Learning agendas and annual evaluation plans*. [Executive Office of the President, Office of Management and Budget, Memorandum for Heads of Executive Departments and Agencies]. <https://www.whitehouse.gov/wp-content/uploads/2021/06/M-21-27.pdf>

# Chapter 2. An Overview of the Program Evaluation Process

## What's Inside?



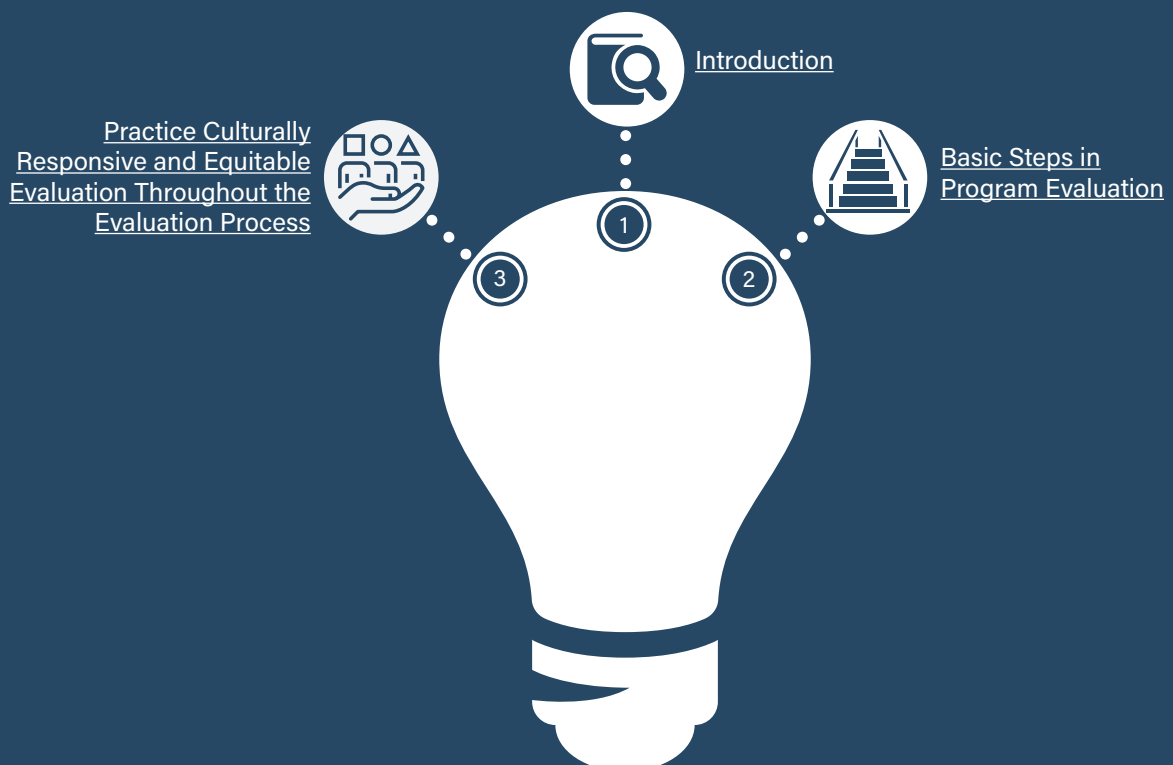
### What this chapter contains

- An overview of six common steps to conducting a program evaluation
- An introduction to culturally responsive and equitable evaluation

### Who can use this chapter

- Program managers who are new to program evaluation and seeking a better understanding of how to plan and conduct evaluations equitably

Click the links below to view the relevant section



# Introduction

To achieve a well-designed and well-executed program evaluation, planning is critical. This chapter lays out a six-step process, or framework, to help support your evaluation planning. These steps are adapted from CDC's Framework for Program Evaluation (1999), a practical, nonprescriptive tool that will help you summarize and organize the essential components of your evaluation. This framework is just one of several tools available to guide you through the evaluation process and support planning. Some frameworks are tailored to programs targeting specific behaviors, conditions, or populations. Others—including the CDC framework—apply across a range of settings.

This chapter introduces a central theme of the Guide: applying a culturally responsive and equity-focused approach to the design, implementation, and management of evaluation will improve its quality and utility (Inouye et al., 2005). A culturally responsive and equitable approach is one that is multiculturally valid<sup>1</sup>; values the voices, knowledge, and expertise of systemically minoritized and underrepresented groups; and aligns evaluation objectives to address equity (Dean-Coffey, 2018). The remaining chapters of the Guide are organized around the six steps, offering you an in-depth look at important decisions you will encounter at each stage and practical recommendations for adapting your evaluation to your program's unique context.

## Basic Steps in Program Evaluation

This section summarizes common activities you may conduct during each stage of an evaluation. Although presented in sequential order, all six steps are interrelated and overlap, and they may be iterative. Often the first three occur at the same time and provide a foundation for the last three. At each step, you should tailor the evaluation to your program's unique needs and continuously seek alignment with shifting priorities:

### Step 1

#### Engage an evaluation team

- Define the roles and responsibilities of individuals on your evaluation team.
- Identify individuals, organizations, and/or communities interested in or affected by your evaluation (e.g., representatives from your service community).
- Decide whether an internal or external evaluator better suits your program's needs.
- When applicable, select and hire an external evaluator.

### Step 2

#### Prepare for the evaluation

- Cultivate opportunities to engage community representatives in meaningful participation.
- Carefully consider all relevant factors when determining the size and scope of your evaluation.

<sup>1</sup>Multiculturally valid refers to a measure or technique that is accurate or authentic across cultural differences (Fey, 2018).

- Bring together an evaluation team that includes evaluation subject matter experts, program staff, program managers, and other important perspectives such as those of community members.
- Build and use a logic model.
- State implementation and/or outcome objectives in measurable terms.
- Prepare an evaluation budget.
- Plan to communicate lessons learned and your evaluation findings.

### Step 3

#### Focus the evaluation designs

- Choose appropriate designs and methodological approaches for your evaluation.
- Seek community input and collaboration on your evaluation design.
- Develop a method to select participants and collect data from members of your target population who will represent the whole group.
- Establish clear procedures for managing and monitoring the evaluation.
- Ensure the safety, respect, and privacy of all evaluation participants.
- Safeguard the confidentiality of data and data sources.

### Step 4

#### Gather credible evidence

- Identify data sources that provide accurate information.
- Select or construct measures to capture all the information you need for the evaluation.
- Build and test data collection instruments that systematically and thoroughly capture the information you need to answer your evaluation questions.
- Ensure measures and instruments are culturally appropriate and reflect community member perspectives.
- Develop data collection procedures that promote consistency.
- Monitor the quality of data collected periodically.

### Step 5

#### Analyze data

- Prepare your data for analysis and assess its initial quality.
- Use appropriate data analysis methods.
- Share provisional interpretations of results with program staff and community members and solicit feedback.

- Interpret your findings to develop an understanding of your results.
- Limit conclusions to the situations, time periods, persons, contexts, and purposes where they are applicable.

## Step 6

### Share lessons learned

- Communicate lessons learned, following your communication plan, to relevant audiences including community members.
- Ensure results are communicated in ways that are transparent and accessible and that facilitate the use of evidence to make programmatic decisions.
- Adapt content and language of evaluation products for different audiences.
- Identify actions or decisions consistent with your evaluation's conclusions.

## Practice Culturally Responsive and Equitable Evaluation Throughout the Evaluation Process

In an environment of equity,<sup>2</sup> all people, regardless of factors such as income, identity, or skin color, would live in thriving communities with access to the resources and opportunities to live healthy, happy lives. Equity is in place when everyone, regardless of who they are or where they come from, has the opportunity to thrive (Expanding the Bench, 2022). Taking a culturally responsive and equitable approach to program evaluation means rethinking how evaluation design, implementation, and the sharing of findings are related to equity. Organizations such as the Equitable Evaluation Initiative, the Center for Culturally Responsive Evaluation and Assessment, and We All Count have developed approaches to help researchers develop culturally responsive and equitable evaluation (CREE) frameworks.

Using a CREE approach in your evaluation can help incorporate cultural, structural, and contextual factors (e.g., historical, social, economic, racial, ethnic, gender) through a participatory process.<sup>3</sup> Such an approach shifts power towards the individuals who are most strongly affected by the evaluation (Expanding the Bench, 2022). CREE is not just one method of evaluation; it is an approach that can be infused into all evaluation methodologies. CREE advances equity by informing strategy, program improvement, decision-making, policy formation, and change.

To date, much of the information about programs and their effectiveness has been generated by educated, higher income, predominately White evaluators using conventional evaluation approaches. Evaluators

<sup>2</sup> Equity is apparent when everyone, regardless of who they are or where they come from, has the opportunity to thrive. Equity requires acknowledging root causes of inequities, eliminating barriers, elevating community strengths, and relentlessly pursuing justice (Expanding the Bench, n.d.).

<sup>3</sup> Participatory processes are specific methods employed to achieve active participation by all members of a group in a decision-making process. The approach shifts power to individuals most impacted by evaluation (Expanding the Bench, 2022).

should practice reflexivity to understand how they bring their experiences, values, and assumptions to their evaluation work. Doing so can help mitigate implied or explicit assumptions that White is the normative, standard, or default position, assumptions which can reinforce stereotypes and perpetuate disparities.

Reflexivity is one way to approach evaluation with cultural humility<sup>4</sup> to address power imbalances and develop mutually beneficial and nonpaternalistic partnerships with communities. Cultural humility requires an understanding that people are experts of their own culture and experiences. Evaluators who analyze data, measure outcomes, and make recommendations for programs and systems have the responsibility to examine the role of power, privilege, and oppression in their work and actively avoid contributing to systemic inequality or sociodemographic disparities.

Beyond the internal reflection of evaluators, evaluation practices should routinely explore the context of programs and how to incorporate the voices of individuals potentially affected by the program. As CREE approaches become more common, new resources have emerged to support evaluators (see additional resources at the end of the chapter). One resource is CDC's guide, *Practical Strategies for Culturally Competent Evaluation* (2014). This resource highlights opportunities for program staff and evaluators to integrate equitable practices throughout the evaluation process. This Guide applies examples of CREE-based solutions to address conventional evaluation approaches from CDC as follows.

## Step 1

### Engage an evaluation team

- Conventional approach: The input and participation of community representatives may be undervalued and overlooked throughout the evaluation process.
- CREE solution: Engage community representatives in meaningful roles throughout the evaluation, including determining evaluation questions, testing data collection instruments, interpreting findings, and developing communication plans.

## Step 2

### Prepare for the evaluation

- Conventional approach: Program descriptions (e.g., logic models) can draw on deficit-based perspectives, which focus on individual and cultural factors viewed as “deficiencies” while disregarding the larger historical and sociopolitical contexts that perpetuate challenges for historically oppressed populations.
- CREE solution: During the development of an evaluation plan, adopt a strengths-based, community-driven approach to clarify community members’ perspectives and affirm what is known about the historical and social context of the program.

<sup>4</sup> Cultural humility is a lifelong commitment to self-evaluation and critique, to redressing power imbalances, and to developing mutually beneficial and nonpaternalistic partnerships with communities on behalf of individuals and defined populations (Greene-Moton & Minkler, 2020).



## Step 3

### Focus the evaluation design

- Conventional approach: Evaluation questions can overlook what potential users of the evaluation findings seek to learn about a program. The choice of design and methods may not align with the needs of those engaged in the evaluation or those with strategic interests in the evaluation.
- CREE solution: Partner with community members during all stages of the research process to ensure the evaluation addresses the needs of the community and potential users of the program.

### What is reflexivity?

Developing cultural humility as an evaluator is an individual, personal, and lifelong journey requiring reflexivity. Reflexivity involves questioning and exploring your underlying values, assumptions, and beliefs that influence the evaluation process:

- Reflection on your own cultural position
- Consideration for the wider and political context
- Intentional efforts to gain perspective of those whose backgrounds differ from one's own

For details on how to strengthen an evaluation team's capacity to operate reflexively, see Attia and Edge (2017).

## Step 4

### Gather credible evidence

- Conventional approach: Evaluation instruments do not always undergo the necessary testing to ensure they accurately and reliably measure what they are intended to measure when used in culturally specific contexts.
- CREE solution: When selecting measures, assess available options for cultural bias in language and content. Be sure you collect numerous perspectives on what and how the evaluation should measure so you collect data that addresses different groups' understanding of credible evidence (e.g., funders, the evaluation field, the communities that participate in the program under evaluation).

## Step 5

### Analyze data

- Conventional approach: Cultural humility is not always demonstrated when interpreting findings. Failing to recognize how your own beliefs, values, biases, and social position can influence how information is seen, heard, and interpreted increases the likelihood of holding a self-focused rather than other-oriented interpersonal stance (Hook et al., 2013).
- CREE solution: Collaborate with community members to uncover your assumptions and engage in reflexive practices that support examination of evaluator backgrounds, beliefs, or biases.

## Step 6

### Share lessons learned

- Conventional approach: Evaluation teams do not effectively communicate knowledge gained from an evaluation with people outside the team.
- CREE solution: Work closely with community members to develop a communication plan that aligns with their needs and emphasizes community benefit, positive change, and social justice.

Each subsequent chapter in this Guide concludes with a section titled “Practice Culturally Responsive and Equitable Evaluation.” Refer to these sections for clear, actionable guidance on applying CREE practices during that stage of your evaluation.

### To learn more ...

- [A Guide to Incorporating a Racial and Ethnic Equity Perspective Throughout the Research Process](#) (Andrews et al., 2019a)
- [Center for Culturally Responsive Evaluation and Assessment](#) (University of Illinois-Urbana-Champaign, n.d.)
- [Considerations for Conducting Evaluation Using a Culturally Responsive and Racial Equity Lens](#) (Public Policy Associates, 2015a)
- [Equitable Evaluation Framework](#) (Equitable Evaluation Initiative, n.d.)
- [Equity as a Perspective for Implementation Research in Early Childhood](#) (Nores, 2020)
- [Evaluating Health Promotion Programs: Introductory Workbook](#) (Snelling & Meserve, 2016)
- [Guiding Principles for Evaluators](#) (American Evaluation Association, n.d.)
- [How to Design and Manage Equity-Focused Evaluations](#) (Bamberger & Segone, 2011)
- [How to Embed a Racial and Ethnic Equity Perspective in Research](#) (Andrews et al., 2019b)
- [Is My Evaluation Culturally Responsive?](#) (Public Policy Associates, 2015b)
- [Key Considerations for Managing Evaluations](#) (Sonko et al., 2011)
- [Materials and Resources Based on CDC’s Program Evaluation Framework](#) (CDC, 1999)
- [RE-AIM Framework](#) (RE-AIM, n.d.)
- [Reflections on Applying Principles of Equitable Evaluation](#) (Stern et al., 2019)
- [Using a Culturally Responsive and Equitable Evaluation Approach to Guide Research and Evaluation](#) (Woodson, 2021)
- [Utilization-Focused Evaluation Checklist](#) (Patton, 2013)
- [We All Count](#) (We All Count, n.d.)
- [WHO Evaluation Practice Handbook](#) (World Health Organization, 2013)

## References

- American Evaluation Association. (n.d.). *Guiding principles for evaluators*. <https://www.eval.org/About/Guiding-Principles>
- Andrews, J., Parekh, J., & Peckoo, S. (2019a). A guide to incorporating a racial and ethnic equity perspective throughout the research process. *Child Trends*. <https://www.childtrends.org/publications/a-guide-to-incorporating-a-racial-and-ethnic-equity-perspective-throughout-the-research-process>
- Andrews, J., Parekh, J., & Peckoo, S. (2019b). How to embed a racial and ethnic equity perspective in research. *Child Trends*. [https://www.childtrends.org/wp-content/uploads/2019/09/RacialEthnicEquityPerspective\\_ChildTrends\\_October2019.pdf](https://www.childtrends.org/wp-content/uploads/2019/09/RacialEthnicEquityPerspective_ChildTrends_October2019.pdf)
- Attia, M., & Edge, J. (2017). Be(com)ing a reflexive researcher: A developmental approach to research methodology. *Open Review of Educational Research*, 4(1), 33–45. <http://dx.doi.org/10.1080/23265507.2017.1300068>
- Bamberger, M., & Segone, M. (2011). *How to design and manage equity-focused evaluations*. UNICEF. [https://evalpartners.org/sites/default/files/EWP5\\_Equity\\_focused\\_evaluations.pdf](https://evalpartners.org/sites/default/files/EWP5_Equity_focused_evaluations.pdf)
- CDC (Centers for Disease Control and Prevention). (1999). Framework for program evaluation in public health. *Morbidity and Mortality Weekly Report*, 48, No. RR-11. <https://www.cdc.gov/mmwr/PDF/rr/rr4811.pdf>
- CDC. (2014). *Practical strategies for culturally competent evaluation*. U.S. Department of Health and Human Services. [https://www.cdc.gov/dhdsp/docs/cultural\\_competence\\_guide.pdf](https://www.cdc.gov/dhdsp/docs/cultural_competence_guide.pdf)
- Dean-Coffey, J. (2018). What's race got to do with it? Equity and philanthropic evaluation practice. *American Journal of Evaluation*, 39(4), 527–542. <https://doi.org/10.1177/1098214018778533>
- Equitable Evaluation Initiative. (n.d.). *Equitable evaluation framework*. <https://www.equitableeval.org/framework>
- Expanding the Bench. (2022). *Spreading knowledge of CREE*. <https://expandingthebench.org/about/terms/>
- Fey, B. B. (2018). Multicultural validity. *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. <https://methods.sagepub.com/reference/the-sage-encyclopedia-of-educational-research-measurement-and-evaluation/i14024.xml>
- Greene-Moton, E., & Minkler, M. (2020). Cultural competence or cultural humility? Moving beyond the debate. *Health Promotion Practice*, 21(1). <https://journals.sagepub.com/doi/full/10.1177/1524839919884912>
- Hook, J., Davis, D., Owen, J., Worthington, E. L., & Utsey, S. O. (2013). Cultural humility: Measuring openness to culturally diverse clients. *Journal of Counseling Psychology*, 60(3). [https://www.researchgate.net/publication/236641214\\_Cultural\\_Humility\\_Measuring\\_Openness\\_to\\_Culturally\\_Diverse\\_Clients](https://www.researchgate.net/publication/236641214_Cultural_Humility_Measuring_Openness_to_Culturally_Diverse_Clients)

- Inouye, T., Yu, H., & Adefuin, J. (2005). *Commissioning multicultural evaluation: A foundation research guide*. The California Endowment in partnership with Social Policy Research Associates. [http://leadershiplearning.org/system/files/multicult\\_eval\\_rpt.pdf](http://leadershiplearning.org/system/files/multicult_eval_rpt.pdf)
- Nores, M. (2020). *Equity as a perspective for implementation research in the early childhood field*. National Institute for Early Childhood Research. Section 3, Chapter 12. [https://www.fcd-us.org/assets/2020/06/GettingitRight\\_UsingImplementationResearchtoImproveOutcomesinECE\\_Chapter12\\_2020.pdf](https://www.fcd-us.org/assets/2020/06/GettingitRight_UsingImplementationResearchtoImproveOutcomesinECE_Chapter12_2020.pdf)
- Patton, M. Q. (2013). *Utilization-focused evaluation (U-FE) checklist*. The Evaluation Center, Western Michigan University. [https://wmich.edu/sites/default/files/attachments/u350/2014/UFE\\_checklist\\_2013.pdf](https://wmich.edu/sites/default/files/attachments/u350/2014/UFE_checklist_2013.pdf)
- Public Policy Associates. (2015a). *Considerations for conducting evaluation using a culturally responsive and racial equity lens*. <https://publicpolicy.com/wp-content/uploads/2017/04/PPA-Culturally-Responsive-Lens.pdf>
- Public Policy Associates. (2015b). *Is my evaluation practice culturally responsive?* [http://jordaninstituteforfamilies.org/wp-content/uploads/2018/06/Self-Assessment\\_6-pages.pdf](http://jordaninstituteforfamilies.org/wp-content/uploads/2018/06/Self-Assessment_6-pages.pdf)
- RE-AIM. (n.d.). *Checklist for inclusion of RE-AIM issues by RE-AIM dimension*. <https://re-aim.org/learn/checklist-for-inclusion-of-re-aim-issues-by-re-aim-dimension/>
- Snelling, S., & Meserve, A. (2016). *Evaluating health promotion programs: Introductory workbook*. Public Health Ontario. <https://www.publichealthontario.ca/-/media/documents/e/2016/evaluating-hp-programs-workbook.pdf?la=en>
- Sonko, R., Berhanu, A., & Shamu, R. (2011). *Key considerations for managing evaluations (brief reference guide)*. <https://www.betterevaluation.org/sites/default/files/Key%20Considerations%20for%20Managing%20Evaluations.pdf>
- Stern, A., Guckenburger, S., Persson, H., & Petrosino, A. (2019). *Reflections on applying principles of equitable evaluation*. WestEd Justice & Prevention Research Center. <https://www.wested.org/wp-content/uploads/2019/07/resource-reflections-on-applying-principles-of-equitable-evaluation.pdf>
- University of Illinois Urbana-Champaign. (n.d.). *CREA, Center for Culturally Responsive Evaluation and Assessment*. <https://crea.education.illinois.edu/>
- We All Count. (n.d.). *The project pieces. We All Count project for equity in data science*. <https://weallcount.com/about-us/>
- Woodson, T. T. (2021). *Using a culturally responsive and equitable evaluation approach to guide research and evaluation*. Mathematica. <https://www.mathematica.org/publications/using-a-culturally-responsive-and-equitable-evaluation-approach-to-guide-research-and-evaluation>
- World Health Organization. (2013). *WHO evaluation practice handbook*. [https://apps.who.int/iris/bitstream/handle/10665/96311/9789241548687\\_eng.pdf;jsessionid=20F23B46EFD93A44A7EFECDA4217405B?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/96311/9789241548687_eng.pdf;jsessionid=20F23B46EFD93A44A7EFECDA4217405B?sequence=1)

# Chapter 3. Engage an Evaluation Team

## What's Inside?



### What this chapter contains

- A discussion about identifying and engaging people with an interest in your program in the evaluation
- A description of several approaches to developing an evaluation team
- A plan for securing and working with an external evaluator
- Examples of ways to apply culturally responsive and equitable principles to evaluation team engagement
- A discussion of potential hurdles related to evaluation and how to overcome them

### Who can use this chapter

Program managers working to bring together a group of individuals who will guide and conduct a program evaluation

Click the links below to view the relevant section

[Practice Culturally Responsive and Equitable Evaluation When Engaging an Evaluation Team](#)



[Introduction](#)



[Hire and Manage Your External Evaluator](#)



[Decide Whom to Engage in Your Evaluation](#)



## Introduction

Like most work in human services, evaluation requires a team effort with many individuals contributing expertise and knowledge to generate credible, relevant, and actionable findings. Typically, this team includes individuals with professional evaluation skills and experience, individuals with strong program knowledge, and individuals with experience collecting data on program implementation and outcomes.

Participatory evaluation<sup>1</sup> focuses on creating meaningful roles for people affected by an evaluation in the design, execution, and application of research. It acknowledges and addresses power differentials and considers the diversity of experiences and perspectives within an evaluation team. For example, past program participants, community leaders, and advocates may help shape the evaluation questions, identify important measures and outcomes, and provide their perspectives to interpretations of evaluation findings. When done well, collaborative evaluations build team capacity and better reflect the lived experiences and expertise of individuals served by the program being evaluated.

This chapter focuses on how to identify interested parties and engage them in program evaluation, factors to consider when selecting internal or external evaluators, and the roles of various evaluation team members. This chapter also provides more information and advice about securing and working with an external evaluator and how to address potential challenges.

## Decide Whom to Engage in Your Evaluation

Build your evaluation team to meet your information and knowledge goals. The size and membership of your team will vary based on expertise, budget, timeline, and/or outside requirements (e.g., program, evaluation funder requests). Identifying individuals who are interested in the program and the evaluation will help you effectively design the evaluation; collect the best evidence; analyze it appropriately; and share progress updates, products, and reports.

### Identify People With an Interest in Your Program

All program evaluations benefit from the engagement and participation of individuals who have an interest in your program, how it operates, and the evaluation findings. These individuals are often grouped under the umbrella term “stakeholders.” Because this term can refer to a wide range of individuals with different goals, perspectives, relationships to the evaluation, and levels of power or authority, this Guide uses more specific terms to describe individuals’ relationships to a program. Broadly, this chapter refers to three groups, though individuals may belong to more than one:

- Those engaged in program operations (staff, staff at referral organizations, organizational leadership, or curriculum or program developers)

<sup>1</sup> Community Tool Box (2022) explains that participatory evaluation involves the individuals affected by the program or policy in the evaluation process. It acknowledges and addresses power differentials and considers the diversity of experiences and perspectives within an evaluation team.

- Those served or affected by the program (current participants, past participants, individuals who qualify for but have not taken up services); see [Practice Culturally Responsive and Equitable Evaluation When Engaging an Evaluation Team](#) at the end of this chapter for more information about including community members in your evaluation team
- Other intended users of the evaluation findings (current and future funders, policymakers and legislators, researchers, and evaluators)

Your evaluation team may engage representatives from the above groups in different ways:

- **Provide critical context and background to an evaluation plan.** For example, long-time community residents might share the community's history with evaluation and research and reveal whether previous partnerships with evaluators left a negative impression.
- **Provide input on an evaluation's design.** For example, current program participants can help identify outcomes of interest relevant to those served or affected by the program, such as increased connections to other parents in the community.
- **Participate in the implementation of an evaluation.** For example, you may hire and train community members to co-facilitate focus groups to create opportunities for program participants to engage with individuals who reflect their lived experiences.
- **Help contribute to the interpretation of evaluation findings and ensure findings are contextualized.** For example, program staff can share insights into local policy changes, such as waitlists at a local childcare center that may influence a program's ability to improve employment rates. Without this knowledge, the evaluation may overlook an important contributing factor affecting changes in participant employment.
- **Ensure evaluation findings are meaningfully shared.** For example, policymakers who are engaged in the evaluation may be more likely to share findings with their colleagues and draw from them in future decision-making efforts.

Having people who are interested in your program participate in your evaluation is not an all-or-nothing effort. Rather, engagement activities can be viewed as options along a continuum, ranging from providing information, to co-creation, to execution of an evaluation (see figure 3.1). The level of participation may vary throughout the life of the evaluation, across different groups of people, and even across representatives of the same group.

### Who should work on your program evaluation?

This chapter uses several terms to describe the people who should be part of an evaluation. The term evaluation team refers here to all the people who will be at the table during decision-making, including focusing the evaluation, troubleshooting, and monitoring evaluation progress, interpreting findings, and communicating and using findings. That group should include project managers, project staff, people with evaluation expertise (see below) and other individuals such as community members.

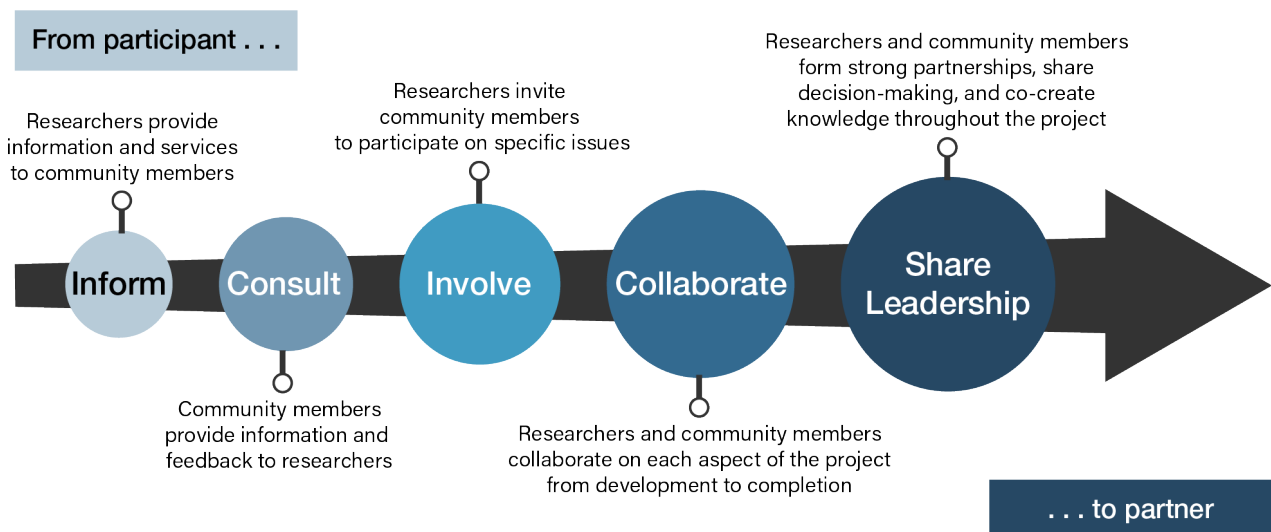
The term evaluator describes the person, set of persons, or organizations with the specific technical and subject matter expertise and knowledge to execute an evaluation (see section [Choose an Evaluator](#)). They are members of your larger evaluation team and should work with the other members of the evaluation team to successfully complete your program evaluation.

Many benefits can accrue from including community members and others in your evaluation. Evaluation findings are more likely to document program participant experiences accurately and align with participant goals when community members participate in the planning stages. Program staff may be more likely to integrate evaluation findings into their routine operations when they have been meaningfully engaged in the process. These benefits will ultimately help you conduct stronger and more useful evaluations.

However, you should also consider the potential challenges. Community engagement can affect your evaluation budget because participatory evaluation activities take significant time and effort. Ensure your timeline and budget align with your expectations for community engagement. It is also important to maintain independence and objectivity to ensure rigorous standards (OPRE, 2021). You can prepare for and respond to concerns about independence by being transparent. Clearly communicate to community representatives and the evaluation team about the level of influence, guidance, and input you expect (see figure 3.1 for one way to conceptualize different levels of engagement).

Document where community members have input in the evaluation (such as helping to revise measures) and where they do not (such as requesting certain findings not be made public). Share that information in final public reports. This approach will help ensure each team member understands their role and creates transparency about how you use input and feedback.

**Figure 3.1. Continuum of Engagement in Evaluation**



Source: Home Visiting Applied Research Collaborative (2018)



## Choose an Evaluator

Your evaluation team will likely include people served or affected by the program, program staff, you as the program manager, and evaluation subject matter experts as evaluators. A single individual, a group of individuals, or an entire organization or firm with technical and subject matter expertise can execute an evaluation. See figure 3.2 for various groups that should have representation on your evaluation team.

An evaluation can be conducted by internal or external evaluators. Internal evaluators are members of the program being evaluated. They may be staff at your organization or otherwise affiliated with your organization or the program, such as staff who develop your curriculum. External evaluators do not have a role within your organization or the program. They may be researchers at universities, evaluation firms, or independent consultants.

Figure 3.2. Potential Evaluation Team Composition



### Federal expectations for grantee evaluations

The Administration for Children and Families (ACF) provides funding to many programs. Often, funding recipients are required to conduct or participate in evaluation efforts.

Common requirements follow:

- Using specific constructs (e.g., elements such as depression) or measures (e.g., the specific item used to measure depression) to collect data common across grant recipients
- Engaging an independent, third-party evaluator
- Engaging in evaluation technical assistance as provided by ACF or a contractor
- Developing evaluation plans for ACF to review and approve
- Participating in a national evaluation of multiple grant recipients

Your selection of an evaluator (including teams or groups of evaluation staff) will depend on factors such as internal evaluation capacity, budget, and availability of external evaluation options. You should weigh the relative importance of having an evaluator with insider knowledge of the program (an internal evaluator) against the perceived objectivity, neutrality, and credibility associated with an externally conducted evaluation. Your evaluator selection will need to align with funder requirements, including the evaluator's ability to work with an institutional review board (IRB)<sup>2</sup> (see [chapters 5 and 6](#) for more information). Double-check all funder

<sup>2</sup> Institutional review boards are administrative bodies established to protect the rights and welfare of human research subjects recruited to participate in research activities conducted under the auspices of the institution it is affiliated with (Oregon State University, n.d.).

requirements related to evaluators' capacities and affiliations to ensure your plan complies with your funding agreement.

Table 3.1 describes three potential options for subject matter expert evaluators. In addition to all the other members of your evaluation team, you may choose to select an external evaluator or an internal evaluator, or have both external and internal experts work together as evaluators. Whichever choice you select, the evaluator(s) will join your larger evaluation team, which includes you as the program manager providing management of the evaluators, offering program expertise, and maintaining representation of program staff.

**Table 3.1. Types of Evaluators**

Considerations	Option 1: External Evaluator	Option 2: Internal Evaluator Supported by External Evaluator (Consultant)	Option 3: Internal Evaluator
Possible advantages	<ul style="list-style-type: none"> <li>Can be considered more objective or credible than other options</li> <li>Can bring more technical expertise than internal evaluator</li> <li>Can be more efficient than internal evaluator because of experience with evaluation</li> <li>Can offer new perspective and unique insights better than internal evaluator</li> </ul>	<ul style="list-style-type: none"> <li>Can (sometimes) be less expensive than an external evaluator alone</li> <li>Can increase likelihood evaluation is consistent with program objectives and needs of target population compared with external evaluator alone</li> </ul>	<ul style="list-style-type: none"> <li>Can (sometimes) be least expensive option of three</li> <li>Can increase engagement and participation of program staff compared with external evaluator</li> <li>Can build organizational capacity for future evaluation efforts compared with external evaluator alone</li> </ul>
Possible disadvantages	<ul style="list-style-type: none"> <li>Can be more expensive than internal</li> <li>Can increase risk, compared with internal evaluators, because evaluator may not have adequate understanding of program components or target population</li> </ul>	<ul style="list-style-type: none"> <li>Can increase burden for program staff</li> <li>Can be considered less objective than external evaluators alone</li> </ul>	<ul style="list-style-type: none"> <li>Might lack sufficient knowledge or experience to design and implement the evaluation</li> <li>Can be considered less independent, objective than an external evaluator</li> </ul>

The remainder of this chapter focuses on engaging and working with an external evaluator. If all your evaluation subject matter expertise on your evaluation team will come from internal staff, the rest of this chapter may be less relevant to your evaluation. Feel free to skim or skip to [chapter 4](#).

# Hire and Manage Your External Evaluator

If you decide to work with an external evaluator, you will need to identify and hire that support. This section will guide you through the process.

## Find an External Evaluator

Four basic steps are useful for finding an evaluator. These steps are similar to those you would use to recruit and hire new program staff. Public agencies may need to use a somewhat different process and work with other divisions of the agency. If you are managing a program in a public agency, check with your procurement department about regulations for hiring outside evaluators or consultants.

### Step 1

#### Develop a job description

The first step in the hiring process is to develop a job description showing the materials, services, and products you expect an evaluator to provide, including the activities the evaluator will be expected to perform, the general budget, and an estimated timeline. Evaluator responsibilities may include the following:

- Develop an evaluation plan.
- Secure evaluation permissions (such as IRB approvals).
- Manage the evaluation team.
- Lead community engagement in evaluation efforts.
- Provide progress reports.
- Develop data collection instruments, forms, and procedures.
- Collect and analyze data.
- Write reports.
- Participate in communication efforts to share information about the evaluation.

In some cases, a job description is sufficient. In other cases, you may need to work with agency procurement offices to develop a request for proposals (RFP). If you need assistance in developing a job description, consider asking another organization with experience in hiring outside evaluators. Advisory board members could also assist with this task.

## Step 2

### Locate sources for evaluators

After you have created a job posting for an evaluator, the next step is to develop a strategic recruitment strategy. You can find evaluators through numerous channels, such as the following:

- **Professional associations.** Some examples include the [American Evaluation Association](#) (AEA), the [American Sociological Association](#), the [Association for Public Policy Analysis & Management](#) (APPAM), and the [Society for Research on Child Development](#). Several of these professional associations can support recruitment efforts. For example, AEA provides a [Find an Evaluator](#) search tool to connect interested agencies with member evaluators. APPAM provides a [Job Board](#) where you can post your needs (for a fee). Some of these organizations can provide a list of members located in your area (for a fee) and/or offer tips on how to tailor advertisements to attract an evaluator who best meets your needs. Additional information on these organizations appears in [appendix A](#).
- **Other agencies that have used outside evaluators.** Agencies in your community may be able to recommend an external evaluator, suggest methods of advertising, and/or provide other useful information. These agencies can represent one of the best ways to find an evaluator who understands your program and is sensitive to the community you serve.
- **Evaluation divisions of state or local agencies.** Most state or local government agencies have planning and evaluation departments. You may be able work with individuals from these agencies on your evaluation. Some evaluation divisions offer their services at no cost as an “in-kind” service. If they are unable to respond to an RFP or provide you with in-kind services, staff members may be able to direct you toward other organizations interested in conducting outside evaluations.
- **Local colleges and universities.** Departments of sociology, psychology, social work/social welfare, education, public health, public administration, and university-based research centers are all possible sources for locating an external evaluator. Well-known academic researchers affiliated with these institutions may be readily identifiable. Even if they cannot personally assist you, they may be able to refer you to a colleague with interest in performing local program evaluations.
- **Technical assistance providers.** Some federal grant programs include a national or local technical assistance provider. If your agency is participating in this kind of grant program, assistance in identifying and selecting an evaluator is an appropriate technical assistance request.
- **Research institutes and consulting firms.** Many experienced evaluators are part of research institutes and consulting firms. Try entering “program evaluation firms” or “human service evaluators” in a search engine. ACF also provides information about grant recipient evaluations on their various program office web pages. You can browse these reports to identify firms that partnered with other grant recipients. Federal evaluation clearinghouses such as the [What Works Clearinghouse](#), the [CLEAR Clearinghouse](#), and the [Title IV-E Prevention Services Clearinghouse](#) maintain lists of studies that they review. Each site provides a way to see the citations of the studies that contribute to the evidence of the programs or approaches under their purview. By reviewing those lists, you can generate names of evaluators who might be a good fit to evaluate your program. Finally, your state human services departments may have a list of firms that have bid on recent contracts for evaluations of state programs they can provide a list of approved vendors on their procurement website.

- **National advocacy groups and local foundations.** Some examples include [United Way](#), [American Public Human Services Association](#), [Child Welfare League of America](#), and the [National Urban League](#). The staff and board members of these organizations may be able to provide you with names of local evaluators. They may also offer insight into evaluations that were done well or evaluators particularly suited to your needs.

## Step 3

### Advertise and solicit applications

After you have developed a job description or RFP, identified possible sources for evaluators, and found ways to advertise the position, you are ready to post an advertisement and solicit applications. You can distribute your request for an evaluator through many channels. For example, use social networks (Facebook, Twitter, LinkedIn), professional networks, or collaborations with your local government's human resource department (if you are a public agency). Depending on your agency's procurement rules, you may also be able to email the job description or RFP directly to evaluators and/or evaluation firms you identified in step 2 asking them to respond to your request.

Advertise as widely as possible, particularly if you are in a small community or you are undertaking an evaluation for the first time. Using several advertising sources increases the likelihood of receiving many responses. You should build in as much time as possible between posting the position and when you plan to review applications.

If you have sufficient time, consider a two-step process for applications. For example, you can release a complete solicitation and ask potential respondents to submit a letter of intent several weeks before their applications are due. This helps you gauge the number of potential respondents and either revise your plans to accommodate a higher-than-expected number or increase advertising efforts if you receive fewer than expected. Alternatively, you can ask first for a one- to three-page initial proposal from applicants. Then, you can review these short proposals and identify the top candidates to receive the full solicitation or job description. This will reduce the number of lengthy applications you need to review.

## Step 4

### Review applications and interview potential candidates

In reviewing applications, consider the candidate's writing style, type of evaluation plan proposed, language (jargon free), experience working with your type of program and staff, familiarity with the subject area of your program, experience conducting similar evaluations, and proposed costs.

After you have narrowed your selection to two or three candidates, you are ready to schedule an interview with each finalist. This interview will give you an opportunity to determine whether you, your staff, and the evaluator are compatible. Like other job applicants, you will need to check references from other programs that have worked with your candidate. If you are hiring a firm, ask for past client referrals and contact those references.

Despite best efforts, you may encounter difficulties in hiring an outside evaluator, including the following:

- **You receive few or no responses to your advertisement.** Many programs, particularly those in isolated areas, can struggle to obtain even a few responses to their advertisements. Be sure to circulate your advertisement widely, using organizational and professional memberships, networks, contacts, and social media accounts. If you sent your job description or RFP to specific evaluators or evaluation firms, consider following up and asking why they decided not to respond to the solicitation. You might make adjustments to address these concerns, such as extending the due date, revising contractual requirements, or reducing the scope or number of evaluation questions. If you have access to evaluation technical assistance or have a funding agency, you can ask for feedback or suggestions from those organizations.
- **The outside evaluator's proposed costs are higher than your budgeted amount.** Evaluations can be tricky to budget. You may find the evaluators responding to your solicitation estimate the work will be significantly more expensive than you had originally budgeted. In this case, consider scheduling a call with the finalists to discuss each party's budget assumptions and the tradeoffs associated with bringing their costs and your expectations in agreement. Several approaches can help resolve the issue:
  - ▶ Reduce the scope of the evaluation (e.g., fewer participants, fewer site visits, reduced communication efforts).
  - ▶ Ask the evaluator to delegate additional work to more cost-effective staff (including graduate students) with strong senior supervision.
  - ▶ Find additional funds for your evaluation or ask the evaluator to donate some of their services (in-kind services).

#### A good evaluator ...

- ... is willing to work collaboratively to develop an evaluation plan that meets your needs
- ... can communicate in simple, practical terms
- ... has experience evaluating similar human service delivery programs
- ... has experience with statistical methods
- ... has the time available to do the evaluation
- ... has experience developing data collection forms or using standardized instruments
- ... will work with a national evaluation team (if one exists)
- ... will treat data confidentially

See [chapter 5](#) for more information on developing a budget for your evaluation.

## Create a Contract

A major step in managing an evaluation is developing a contract with your outside evaluator. Your contract should consider the following elements.

### How you will pay for and monitor evaluation services

Work with experts such as your procurement office or lawyer to establish payment agreements. Many contract types and payment/reimbursement models are options for hiring evaluation work. You may also

need to specify a timeline for payments, invoicing instructions, and what supporting materials are needed to document the hours/budget the evaluator is invoicing. In some contracts, payments are tied to assessments of the quality of the work provided by the evaluator. Your contract should spell out expectations for quality management and control, standards of minimum quality, and repercussions for work of lower-than-expected quality.

### Who “owns” the evaluation information

You may want to specify that your organization owns data collected under the evaluation and that the evaluator must request your permission to share the data or findings from the data. Your contract should require the evaluator to receive clearance for any plans to publish evaluation results.

#### Maintaining an evaluation firewall

In some evaluations, it will be important to develop and maintain a firewall, or clear separation of activities and knowledge, between the evaluator and program staff. For example, some program developers are involved in evaluations of their own programs. It is important to determine when and how the program developer has input or influence on the evaluation. For example, they may be involved in outcome measurement selection, but be firewalled from analysis activities to not influence analysis and interpretation. You should determine where your evaluation firewalls are, and make sure all individuals engaged understand the firewall, why those firewalls are in place, and what happens if they are not respected.

The contract should also address the need to align with any publication restrictions from the funding agency. In some instances, the funding agency may have requirements about the use of data and the release of reports.

You should also confirm who owns data when your evaluation uses any software or program to collect or analyze data, such as an online survey platform.

### Who will perform each evaluation task

The contract should identify those who will perform each evaluation task and the level of contact between the evaluator and the program. In most cases, program staff will need to support the evaluator in some tasks; for example, obtaining consent for program participants to conduct data collection or follow-up phone calls.

If a problem occurs even after specification of tasks, you may want to speak with your evaluator to offer the option of renegotiating their level of effort or tasks. The resolution should be mutually agreeable to both program staff and the evaluator to avoid compromising the integrity of the evaluation. Again, the responsibilities of program staff and the evaluator may vary, depending on the structure of your evaluation and the amount of money available.

### Expectations about evaluator and program staff contact and communications

At a minimum, an outside evaluator needs to keep program staff informed about the status of the evaluation and/or uphold agreements related to engagement, including engagement of program staff. Ideally, your evaluator will work with your larger evaluation team as a unit to ensure staff buy-in and representation of different perspectives on the program and evaluation (including community members).

Depending on the structure of the evaluation, the evaluator may also be able to provide information to the program in real time to inform ongoing program improvement efforts. Accordingly, the contract can specify the frequency of evaluation team meetings, the purpose of the larger evaluation team, and evaluator reporting requirements.

## Collaborate With an External Evaluator

After you have selected an evaluator, build rapport, relationships, and a partnership. You and your staff should participate as full partners with external evaluators throughout the evaluation process as members of the evaluation team. A strong collaborative relationship will reap benefits for both parties. For example, you and your staff will learn more about evaluation and better ensure the evaluation addresses your program improvement information needs. Your engagement with the evaluation also means the evaluator will have a clearer sense of how the program functions and will be better positioned to provide more relevant and useful feedback.

As with any partnership or relationship, working with an evaluator is a learning process for both parties. Even with a solid contract in place, problems can arise during an evaluation. Mutual respect and clear communication can go a long way in identifying and resolving challenges before they become problems. You may want to discuss common problems as a team preemptively. Examples of problems you may encounter and potential remedial steps follow.

### **Evaluation approaches differ (the program staff and evaluator do not see eye-to-eye)**

Expect some communication challenges when two different sets of expertise collaborate. Developing a shared vocabulary is a good first step. This helps ensure all parties understand what the others mean when they use terms such as participant, data collection instrument, or comparison condition. Next, try to clarify each party's main concerns, goals, and constraints. Often, issues arise when idealistic evaluation goals meet practical challenges in real-world implementation, such as what evaluation questions can be answered without a comparison group, how many people can be interviewed with a certain budget, or why a certain measure is not appropriate to ask of program participants. The goal is to reach common ground where both programmatic and evaluation constraints and needs are met.

If many reasonable attempts are made to resolve differences and significant conflicts persist that jeopardize the program or evaluation, program staff should consider terminating the contract. This decision should be weighed carefully because a new evaluator will need to be recruited and brought up to speed midstream. In some situations, finding a new evaluator may indeed be the best option. Prior to making this decision, however, talk to your program funders, particularly if they are providing financial support for the evaluation.

### **Make sure you follow proper procurement procedures**

Check with your human resources or legal offices to ensure you are adhering to organizational regulations and legal requirements when hiring or contracting with an external evaluator. Engage these offices at step 1, when developing a job description. For example, Native American communities may need the Tribal council to approve the evaluator selection. Finally, if you are a federal grant recipient, check your grant requirements or consult with your funding agency to determine whether you need federal approval of your evaluator.



## **Evaluation of the program requires technical skills outside your original plan**

Despite best intentions, you may find the evaluation requires additional technical skills your current evaluation team does not have. For example, you may have an evaluation design that requires a complicated statistical approach for data analysis. If this is the case, your evaluator will likely agree with your assessment or may even be the one who identifies the issue. Many federal grant programs provide evaluation technical assistance and may be able to augment that skill set. Other funders may be willing to connect your evaluation team to other experts. Alternatively, your evaluator may be able to hire an expert as a consultant or staff member to provide the additional support. Programmers, statisticians, and others can augment the evaluation team without fundamentally changing the evaluation team's structure.

## **The evaluator leaves, terminates the contract, or does not meet contractual requirements**

Rarely, an evaluator may need to exit an evaluation. This can happen because of unexpected personal circumstances or unanticipated organizational changes. You can reduce the chances your evaluation will face serious disruption by contracting with a team of evaluators (e.g., a university research center, two evaluators working together) or a firm rather than a single individual. You can also be more prepared for a transition or disruption by maintaining close management of your evaluation. Maintain copies of all study materials; contact information for the IRB and data collection web portal; and if applicable, contact information for your evaluator's supervisor, such as the university department chair.

In other cases, you may determine the evaluator is not meeting contractual obligations. In that case, you may get support from your funder (to help mediate the discussion) or from another staff member at the evaluator's organization (e.g., their supervisor). If you ultimately decide the relationship cannot continue and you choose to terminate the contract, it is important to determine who has the rights to any materials developed and request copies of datasets, documents, and guides as a condition of the termination. When your evaluator does not meet contractual requirements and efforts to resolve the dispute have failed, public agencies should turn the case over to their procurement office, and private agencies should seek legal counsel.

## **The evaluator is not culturally competent or does not have experience working with your community and the participants**

It is not always possible to locate an evaluator with experience in the type of evaluation you need plus experience working with specific groups and subgroups in the community. That lack of familiarity may negatively affect the quality and relevance of your evaluation. Depending on your program and service population, your evaluator may need to better understand the racial and ethnic backgrounds of your participants and their cultures, religions, languages, gender identities, sexual orientations, disability status, and other lived experiences. Evaluators can help mitigate their lack of lived or technical experience through education, participant observation in community events, interviews with community members, or hiring community representatives to become part of their evaluation staff. You can also deepen community representatives' participation in the evaluation.

## You are not happy with the evaluator's findings

Sometimes program managers and staff discover the evaluator's findings are not consistent with their impressions of the program's effectiveness. Program staff may perceive participants are demonstrating the expected changes in behavior, knowledge, or attitudes, but the evaluation results do not match this perspective. In this situation, you may want to work with your evaluator to ensure the instruments being used are designed to measure the changes you previously observed in program participants.

Your evaluator will continue to need input from program staff in interpreting evaluation findings. You may also want your evaluator to assess whether some of the participants are changing and whether participants share any common characteristics that are or are not showing change over time. However, be prepared to accept findings that do not support your initial perceptions. Not every program will work the way it was intended to, and you may need to make some program changes based on your findings. Remember, findings that indicate your program is not operating as intended or not having the impact intended can be positive information you can use to refine your program. Your ultimate goal is to help participants, and if you identify barriers or challenges that impede your goal, you can develop a plan to address them and better serve program participants.

### Virtual and hybrid evaluations

In response to the COVID-19 pandemic, many organizations have begun conducting evaluations on virtual platforms. This has led to new benefits, such as working with evaluators located far from a program site and the additional flexibility to host and attend virtual meetings. At the same time, virtual evaluations present logistical challenges, such as obtaining consent from participants, managing confidential data collection, and building rapport in an online environment. Future evaluation efforts will likely employ a hybrid model, using virtual activities where they make logistical and financial sense and meeting in person when necessary.

## Practice Culturally Responsive and Equitable Evaluation When Engaging an Evaluation Team

As discussed in the section [Decide Whom to Engage in Your Evaluation](#), involving community members using a participatory evaluation approach can strengthen your evaluation. Community members bring their lived experience and participant understanding of your program to ensure an evaluation is more relevant, accurate, and credible. At the same time, it is important to acknowledge the complex power dynamics within evaluations, where evaluators and funders are often seen as experts and decision-makers, while program staff and community members might feel less power, authority, and agency. CREE practices can help engage community members respectfully and authentically in evaluation efforts. Consider the following recommendations:

- Identify individuals and groups who reflect your program's audience and invite them to participate in the evaluation process as advisors.

- Emphasize the importance of having evaluators who have worked with your program's service population in the job description or RFP.
- Understand the evaluation team's social identities and lived experiences. Seek consultants or hire additional staff as needed to ensure a variety of voices and experiences contribute to your work.
- Develop evaluators' awareness of how and why to engage community members in authentic ways. This could include training on CREE, participatory evaluation, and collaborative evaluation methods and benefits.
- Account for time to facilitate rapport and trust building with community representatives and ensure their meaningful contribution to the evaluation.

### Assembling an evaluation team to promote inclusivity, cultural sensitivity, and empowerment

The Children Services Council of Broward County (Florida) recently presented their approach to engaging community members in data projects at the annual OPRE Methods Meeting (DuCille et al., 2021). To build a team of researchers, child welfare system professionals, and system-involved parents and youth, Broward County made use of relationships professionals already had with families. By inviting families with connections to the professionals on the research team, the team started with some foundation of trust and familiarity. The researchers then built rapport with youth and caregivers through shared meals and validating families' experiences and stories. The whole team began their work with a two-day training on antiracism, implicit bias, and local, race-related history. Finally, the researchers were prepared to provide healing responses to traumas as they surfaced, such as breathing, body movement, and one-on-one discussions.

### To learn more ...

- [An Introduction to Collaborative, Participatory, and Empowerment Evaluation Approaches](#) (Fetterman et al., 2018)
- [Checklist for Building Organizational Capacity](#) (Volkov et al., 2007)
- [Empowerment Evaluation Principles in Practice](#) (Fetterman, 2005)
- [Five Steps for Selecting an Evaluator](#) (Bronte-Tinkew et al., 2007)
- [Guidelines for Working With Third-Party Evaluators](#) (Heinemeier et al., 2014)
- [Identifying and Determining Involvement of Stakeholders](#) (CDC, n.d.)
- [When and How to Use External Evaluators](#) (Rutnik et al., 2002)

## References

- Bronte-Tinkew, J., Joyner, K., & Allens, T. (2007). *Five steps for selecting an evaluator: A guide for out-of-school time practitioners* (Research-To-Results Brief, Pub No. 2007-32). Child Trends. [https://www.childtrends.org/wp-content/uploads/2013/04/Child\\_Trends-2007\\_10\\_01\\_RB\\_SelectingEvaluator.pdf](https://www.childtrends.org/wp-content/uploads/2013/04/Child_Trends-2007_10_01_RB_SelectingEvaluator.pdf)
- CDC (Centers for Disease Control and Prevention). (n.d.). *Identifying and determining involvement of stakeholders*. <https://www.cdc.gov/std/Program/pupestd/Identifying%20and%20Determining%20Stakeholders.pdf>
- Community Tool Box. (2022). *Section 6, participatory evaluation*. <https://ctb.ku.edu/en/table-of-contents/evaluate/evaluation/participatory-evaluation/main>
- DuCille, A., Gallagher, S., & Csonka, T. (2021). *Community participatory action research: Case study #1 Broward County, Florida* [Presentation at OPRE's Methods Meeting]. <https://opremethodsmeeting.org/wp-content/uploads/2021/09/Community-Participatory-Action-Research-Case-Study-1-Broward-County-Florida.pdf>
- Fetterman, D. M. (2005). *Empowerment evaluation principles in practice: Assessing levels of commitment*. <https://evalparticipativa.net/wp-content/uploads/2021/08/45.-Empowerment-Evaluation-Principles.pdf>
- Fetterman, D. M., Rodriguez-Campos, L., Wandersman, A., O'Sullivan, R. G., & Zukoski, A. P. (2018). *An introduction to collaborative, participatory, and empowerment evaluation approaches*. <https://evalparticipativa.net/wp-content/uploads/2021/08/48.-Introduction-to-collaborative-participatory-and-empowerment-evaluation.pdf>
- Heinemeier, S., D'Agostino, A., Lammert, J., & Fiore, T. A. (2014). *Guidelines for working with third-party evaluators*. Westat. [https://osepideasthatwork.org/sites/default/files/Guidelines\\_3rdEvaluators\\_508\\_0.pdf](https://osepideasthatwork.org/sites/default/files/Guidelines_3rdEvaluators_508_0.pdf)
- Home Visiting Applied Research Collaborative. (2018). *The importance of participatory approaches in precision home visiting research*. <http://www.hvresearch.org/wp-content/uploads/2019/09/HARC-2018-The-Importance-of-Participatory-Approaches.pdf>
- OPRE (Office of Planning, Research, and Evaluation). (2021). *ACF Evaluation Policy*. U.S. Department of Health and Human Services. <https://www.acf.hhs.gov/opre/report/acf-evaluation-policy>
- Oregon State University. (n.d.). *What is the institutional review board (IRB)?* <https://research.oregonstate.edu/irb/frequently-asked-questions/what-institutional-review-board-irb>
- Rutnik, T. A., & Campbell, M. (2002). *When and how to use external evaluators*. The Association of Baltimore Area Grantmakers. <https://efc.issuelab.org/resources/11630/11630.pdf>
- Volkov, B. B., & King, J. A. (2007). *Checklist for building organizational evaluation capacity*. Western Michigan University, Evaluation Checklist Project. <https://wmich.edu/sites/default/files/attachments/u350/2018/eval-cap-bldg-volkov%26king.pdf>

# Chapter 4. Prepare for the Evaluation

## What's Inside?



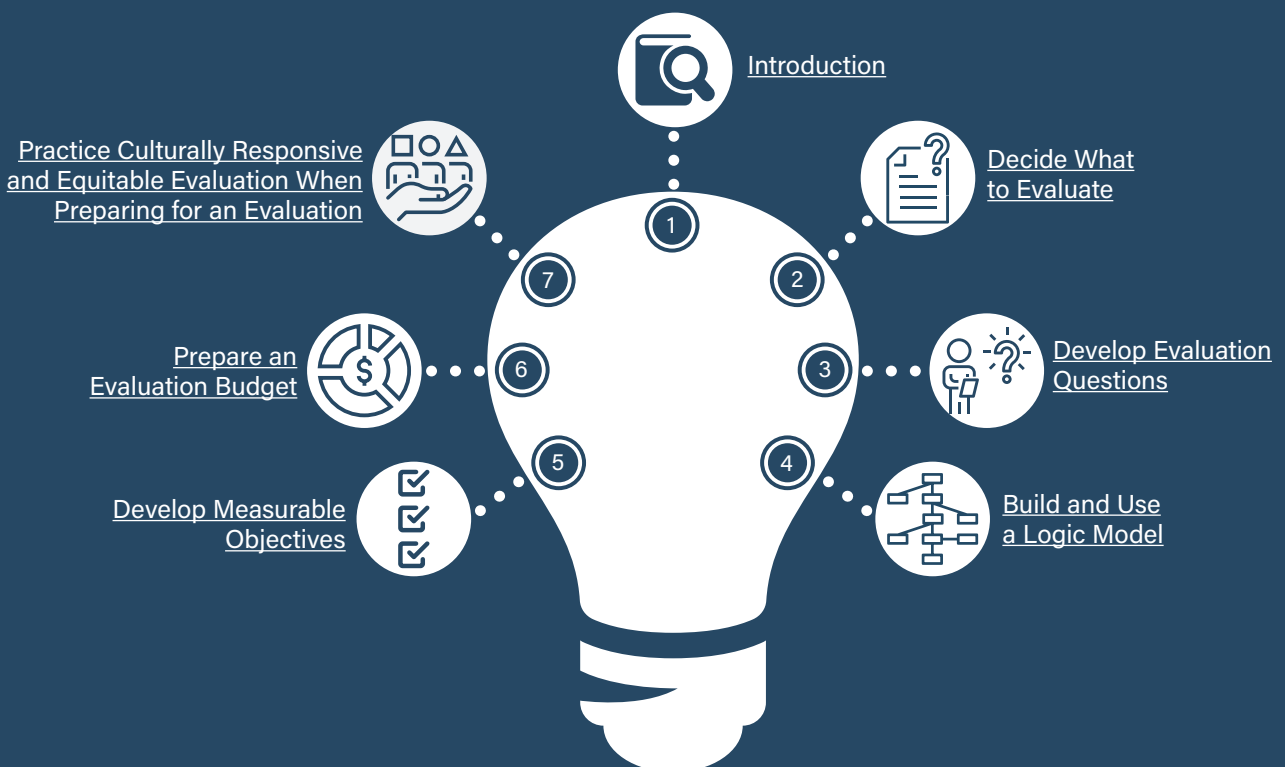
### What this chapter contains

- An introduction to the importance of planning for an evaluation
- A discussion about deciding what program, component, service, or activity to evaluate
- A description of the basic questions an evaluation can answer
- A guide for developing a logic model that will provide a structural framework for your evaluation
- A plan for stating program objectives in measurable terms
- A discussion of the common cost drivers and cost savers in an evaluation
- Examples of ways to apply culturally responsive and equitable principles when preparing for an evaluation

### Who can use this chapter

- Program managers preparing to conduct a program evaluation

Click the links below to view the relevant section



## Introduction

Once you have assembled your evaluation team, the next step is to look closely at the purpose of your evaluation to determine what evaluation questions can be asked and answered and how to get the best return on your evaluation investment. A shared understanding of the purpose, use, and users of the evaluation findings should drive the development of evaluation questions. This understanding should in turn drive the evaluation design, data collection, analysis, and reporting. Beyond facilitating good evaluation practice, the planning phase can—

- Foster transparency for the evaluation.
- Increase program staff buy-in for evaluation activities.
- Connect and align various evaluation activities (especially for programs employing different contractors or contracts).
- Improve transitions during staff turnover.
- Establish whether sufficient program resources and time are available to accomplish the intended evaluation activities.

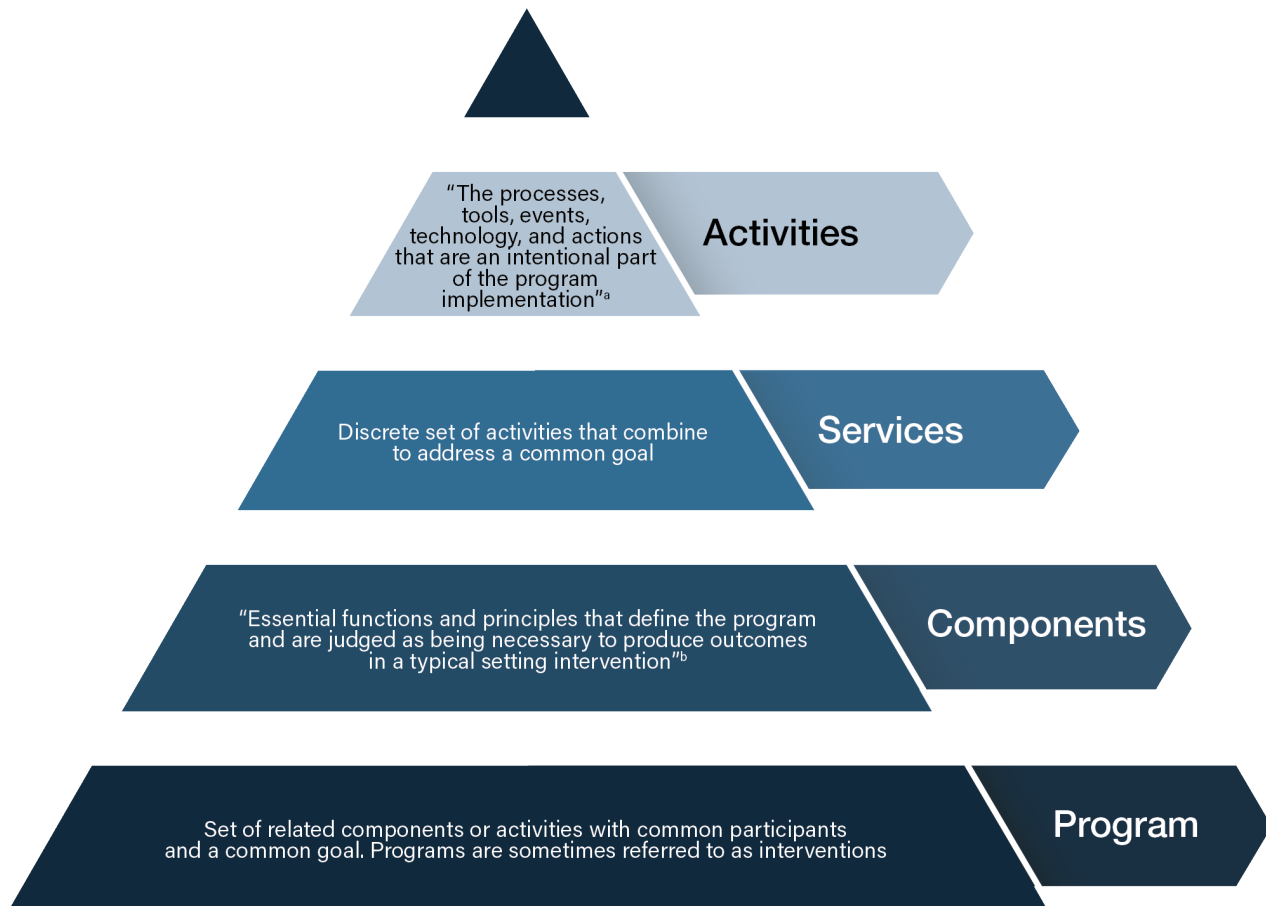
The important decisions of what to evaluate and how should involve the outside evaluator or consultant (if you decide to hire one), all program staff who are part of the evaluation team, and anyone else in the agency who will be engaged. As noted in [chapter 3](#), evaluation teams should engage potential users of the evaluation and community members early and often. Their engagement during the initial decision-making processes will improve the ultimate usefulness of the evaluation and help balance the power between evaluators and evaluation participants. Ideally, the planning process should begin before implementing the program, component, or service you wish to evaluate. When that is not possible (i.e., the program is already operational), take time to understand and articulate program goals and strategies.

This chapter offers guidance on preparing for the evaluation, including defining its size and scope, identifying the evaluation questions, building a logic model to provide a structural framework, stating program objectives in measurable terms, and budgeting for the evaluation. It concludes with strategies to support conducting a culturally responsive and equitable evaluation.

## Decide What to Evaluate

Some programs have many components, while others have only one or two. You can evaluate your entire program, one or two program components, or even one or two services or activities within a component (see figure 4.1). Consider, for example, a Head Start grantee providing seasonal Head Start to migrant farmworker families. A successful evaluation will distinguish whether it is evaluating the early learning and child development services, health and nutrition services, family well-being services, or all three components.

Figure 4.1. Potential Evaluation Target Options



<sup>a</sup> W.K. Kellogg Foundation, 2004, p. 2

<sup>b</sup> Blase & Fixsen, 2013, p. 3

To a large extent, your decision about what to evaluate will depend on program staff and leadership, the funder, and potentially the local community's priorities. The decision will also be subject to available financial resources, staff and contractor availability, and the amount of time committed to the evaluation.

Several options are available to work within limited evaluation resources. For example, you might simplify the design or narrow the scope of your evaluation. It is better to conduct an effective evaluation of a single program component than attempt to evaluate several components or an entire program without sufficient resources. Sometimes, the decision about what to evaluate is made for you, as when funders require specific evaluation elements as a condition of a grant award. At other times, you or your agency administrators will decide what to evaluate.

#### Program as shorthand for an evaluand

You can evaluate almost anything. In addition to the examples of a program, program component, service, or activity, you can study policies, laws, websites, a training, etc. In the interest of readability, the Guide uses the term "program" as a placeholder for any evaluand (a generic term for the object or thing that is the subject of an evaluation).

If your program is already operational, you may decide to evaluate a particular service or component because you are unsure about its effectiveness for some participants. The introduction of a new service or component may be another reason to focus your evaluation on that specific service or component. Alternatively, you may choose to evaluate your entire program because you believe it is effective and you want evidence of effectiveness to help you obtain additional funding to continue or expand it. Defining what you will evaluate helps you determine at the outset whether your new efforts are being implemented successfully and are effective at attaining expected participant outcomes.

## Develop Evaluation Questions

Once you have decided what programs, components, services, or activities to evaluate, you should decide which questions you want the evaluation to answer. These questions will play a central role in guiding the evaluation, so plan them carefully. Strong evaluation questions should be clear, relevant, and rigorous. They must stem from a program's objectives.

As described in chapter 1, the two types of objectives are program implementation objectives and participant outcome objectives. While implementation evaluations help you determine whether program activities have been implemented as intended, outcome evaluations measure program effects (CDC [Centers for Disease Control and Prevention], n.d.-b). Sometimes, evaluating program implementation objectives is referred to as a process evaluation (OPRE [Office of Planning, Research, and Evaluation], 2010). However, because many types of process evaluations are possible, this guide uses the term implementation evaluation.

Implementation and outcome evaluations can be used to determine whether you have been successful in attaining both types of objectives by answering the following questions:

- **Has the program been successful in attaining the anticipated implementation objectives?** For example, are you implementing the services or training you initially planned to implement? Are you reaching the intended target population? Are you reaching the intended number of participants? Are you developing the planned collaborative relationships?
- **Has the program been successful in attaining the anticipated participant outcome objectives?** For example, are participants exhibiting the expected changes in knowledge, attitudes, behaviors, or awareness? Can these changes be attributed to the program?

A comprehensive evaluation must answer both questions. You may be successful in attaining your implementation objectives, but if you do not have information about participant outcomes, you will not know whether your program is having the intended outcome or effect. Similarly, you may be successful in changing participants' knowledge, attitudes, or behaviors, but you will need information on implementation to guide program adoption, replication, and scale-up.



One common framework for formulating concise but rigorous outcome evaluation questions is known as *Population, Intervention, Comparison, and Outcome (PICO)*. This framework encourages evaluators to consider the target population that will participate in the intervention and evaluation, the intervention to be evaluated, the comparison that will be used to assess whether the intervention makes a difference, and the outcomes you expect the intervention to achieve (Tribal Evaluation Institute, n.d.). Strong evaluation questions should specify all four of these elements. An example of an evaluation question that specifies the four elements of PICO might be, “Do student parents (P) of children who attend Head Start (I) miss fewer classes (O) than student parents whose children do not attend Head Start (C)?”

### PICO framework

The PICO framework is a widely used strategy for breaking down evaluation questions into four elements that facilitate the identification of relevant information: population, intervention, comparison, outcome. To learn more about how PICO can clarify evaluation questions, see the Tribal Evaluation Institute (2016) or the evaluation plan template in Blocklin et al. (2019).

Although this section focuses on implementation and outcome evaluations, other categories of questions may be relevant to your program: questions regarding the need for services (needs assessment) and questions regarding the program’s economic benefits (economic evaluation). These topics are beyond the scope of this Guide, but a basic understanding of them may be helpful.

A needs assessment is a study of the problem a program intends to address and the need for the program, such as determining the number of children who are chronically absent from school and the likely reasons why they miss school (GAO [Government Accountability Office], 2021). An economic evaluation<sup>1</sup> is a study that measures program costs and compares them with either a monetary value of the program’s benefit (cost-benefit analysis) or a measure of the program’s effectiveness in achieving its outcome objectives (cost-effectiveness analysis). For more information on these types of assessments, see the resources in the [“To learn more”](#) section at the end of the chapter.

## Build and Use a Logic Model

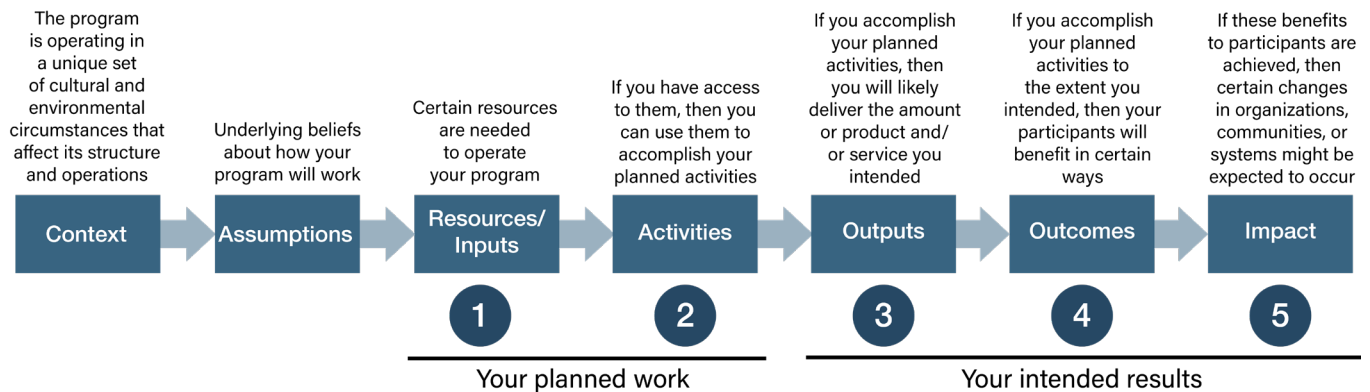
Whether you decide to evaluate an entire program, a single component, or a single service, you will need to build a logic model. A logic model<sup>2</sup> is typically represented as a flow chart that tracks how inputs drive activities to produce outputs, outcomes, and ultimate impact (OPRE, 2010). A variety of formats can be used to create a logic model; the key is to develop a clear understanding of the program and its context for operation. A logic model may also be referred to as a program model, program theory, and theory of change.

<sup>1</sup> Economic evaluation is an effort to use analytic methods to identify, measure, value, or compare the costs and consequences of one or more alternative programs or interventions (CDC n.d.-a).

<sup>2</sup> A logic model is a picture of how your organization does its work—the theory and assumptions underlying the program. A program logic model links outcomes (both short and long term) with program activities and processes and the theoretical assumptions and principles of the program (W.K. Kellogg Foundation, 2004).

In general, all logic models represent a series of logically related assumptions about the program's participant population and the changes you hope to bring about in that population as a result of your program. Evaluators and program staff should work together to jointly build the logic model to ensure it reflects how the program will work and how it will influence the target population. Figure 4.2 presents the basic elements of a logic model.

**Figure 4.2. Basic Elements of a Logic Model**



Source: Adapted from W.K. Kellogg Foundation Logic Model Development Guide (2004)

Logic models can inform program improvement and program evaluation. Regarding program improvement, logic models can help advance strategic planning and program management by identifying the target population (those the program is designed to serve), clarifying the program goals and any conceptual gaps, tracking progress and changing needs, and describing the program to internal and external audiences.

Regarding program evaluation, logic models can provide a structural framework for your evaluation by informing the development of a data collection plan and helping your evaluation team understand why desired outcomes are or are not attained. For example, tracking program outputs can help evaluators determine whether ineffectiveness is the result of (1) insufficient resources or inputs or other implementation challenges or (2) other issues (i.e., the intervention is implemented with fidelity but did not have the intended effects).

Logic models are not difficult to construct, and they lay the foundation for your evaluation by clearly identifying your program implementation and participant outcome objectives. These models can then be stated in measurable terms for evaluation purposes. See "To learn more" and the appendices for resources and templates for building a logic model.

### Falsifiable logic model

A logic model is a helpful tool for thinking through causal pathways by linking outcomes with program inputs and activities. Taking this idea one step further, falsifiable logic models expand the role of the logic model by including detailed—and falsifiable—goals for components of a conventional logic model. Falsifiable logic models can help evaluation teams determine whether a program is satisfying its own stated goals.

To learn more about how falsifiable logic models can help a program strengthen its implementation and increase the likelihood of success in a rigorous impact evaluation, see Epstein and Klerman (2013).

## Develop Measurable Objectives

The logic model serves as a foundation for identifying your program's implementation and participant outcome objectives. Initially, focus your evaluation on assessing whether implementation objectives and immediate participant outcome objectives were attained. This will help you assess whether it is worthwhile to commit additional resources to evaluating attainment of intermediate and long-term outcome objectives.

Program managers often believe that stating objectives in measurable terms means establishing performance standards or some arbitrary "measure" the program must attain. This is not true. Stating objectives in measurable terms simply means you describe what you plan to do in your program and how you expect the participants to change in a way you can measure. From this perspective, measurement can involve anything from counting the number of services (or determining the duration of services) to using a standardized test that will result in a quantifiable score. Some examples of stating objectives in measurable terms appear below.

Stating implementation objectives in measurable terms. Examples of implementation objectives follow:

- **How will you know the planned activities occurred?** For example, the number, duration, and frequency of services or activities implemented
- **Who will do it?** What the staffing arrangements will be; the characteristics and qualifications of the program staff who will deliver the services, conduct the training, or develop the products; and how these individuals will be recruited and hired
- **What population do you plan to reach? How many individuals?** A description of the participant population for the program; the number of participants to be reached during a specific timeframe; and how you plan to recruit or reach the participants

To state these objectives in measurable terms, be specific about your program's operations. The example in table 4.1 demonstrates how general implementation objectives can be transformed into measurable objectives. A blank worksheet for stating your implementation objectives in measurable terms is provided in [appendix B](#).

Table 4.1. Example of Implementation Objectives Stated in Measurable Terms

<b>How will you know the planned activities occurred?</b>	<b>General objective:</b> Provide drug abuse education services.
	<b>Measurable objective:</b> Provide 2-hour drug abuse education classes 5 days a week, eight sessions per year.
<b>Who will do it?</b>	<b>General objective:</b> Program staff will be experienced, certified addictions counselors.
	<b>Measurable objective:</b> One hundred percent of program staff will have an addictions counseling certification; program staff will have a minimum of 2 years' experience.
<b>What population do you plan to reach? How many individuals?</b>	<b>General objective:</b> Recruit and serve runaway and homeless youth.
	<b>Measurable objective:</b> Participants will include youth aged 8–14 residing in a shelter during time of classes. Reach six participants per session; recruit the participants to the classes by intake counselors and clinical director.

From your description of the specific characteristics for each objective, the evaluation will be able to assess in an ongoing way whether the objectives were attained, the types of problems encountered during program implementation, and the areas where changes may be needed. Using the example above, you may discover the first class session included only two youth from the crisis intervention services. Based on the findings from the evaluation, you might examine your data to gain more insights into the recruitment process:

- How many youth resided in the shelter during that timeframe?
- How many youth agreed to participate?
- What barriers to participation did youth encounter (such as youth or parent reluctance to give permission, lack of transportation, or lack of interest among youth)?

Based on your answers, you may decide to revise your recruitment strategies, train crisis intervention counselors to be more effective in recruiting youth, visit a youth's family to encourage the youth's participation, or offer transportation to youth to make it easier for them to attend the classes.

**Stating participant outcome objectives in measurable terms.** Be specific about the changes in knowledge, attitudes, awareness, or behavior you expect to occur as a result of participation in your program. One way to be specific about these changes is to ask yourself the following questions:

- What change is expected to occur?
- How much change is expected to occur?
- For whom will the expected change occur?
- How will you know the expected change occurred?

To answer these questions, identify the evidence needed to demonstrate your participants have changed. The example in table 4.2 demonstrates how participant outcome objectives may be stated in measurable terms. A worksheet for defining measurable participant outcome objectives appears in [appendix B](#).

**Table 4.2. Example of Outcome Objectives Stated in Measurable Terms**

<b>How will you know expected change occurred?</b>	<b>General objective:</b> Expect to reduce the use of alcohol by youth.
	<b>Measurable objective:</b> Youth who complete the program will demonstrate a 10 percent decrease in alcohol use compared with preprogram, as measured by the Alcohol Timeline Followback instrument.

## Prepare an Evaluation Budget

Program managers are often concerned about the cost of an evaluation. This is a valid concern. Evaluations do require time, money, and expertise. Many program managers and staff believe it is unethical to use program or agency financial resources for an evaluation because available funds should be spent on serving participants. However, evaluation is essential if you want to know whether your program is benefiting participants. It is more accurate to view money spent on evaluation as an investment in your program and in your participants rather than as a diversion of funds away from helping participants.

Unfortunately, calculating evaluation costs is not strictly defined. The amount of money needed depends on many factors:

- Aspects of your program you decide to evaluate
- Number of people who will contribute to the evaluation (e.g., how many evaluators; how many community members and their level of engagement)
- Size of the program (i.e., the number of staff members, participants, components, and services)
- Number of outcomes you want to assess
- Who is conducting the evaluation
- Your agency's available evaluation-related resources

Costs also vary according to economic differences in communities and geographic locations. Table 4.3 describes other common factors that influence the costs and resources needed to conduct a program evaluation, such as the source and condition of data, how the data will be collected, the statistical complexity of data analyses, and the program staff's evaluation capacity.

Table 4.3. Common Cost Drivers and Cost Savers in Program Evaluation

Factor	Considerations	Lower Cost	Higher Cost
Data source	<ul style="list-style-type: none"> <li>What data are already available (secondary data), and what will need to be collected (primary data)?</li> <li>How many sources of data are needed?</li> </ul>	<ul style="list-style-type: none"> <li>Using previously collected data (e.g., administrative data) that are readily available and inexpensive to obtain</li> <li>Collecting data at a single site and/or from a single informant group</li> </ul>	<ul style="list-style-type: none"> <li>Using previously collected data will require an agreement (e.g., data use agreement, memorandum of understanding) to obtain</li> <li>Collecting data at many geographically spread sites and/or from multiple sources (particularly from comparison groups)</li> </ul>
Data condition	<ul style="list-style-type: none"> <li>Will the data require extensive cleaning or manipulation?</li> <li>Is the file easy to interpret and use (e.g., is a data dictionary provided)?</li> </ul>	<ul style="list-style-type: none"> <li>Available datafile(s) cleaned and ready for use</li> </ul>	<ul style="list-style-type: none"> <li>Datafiles will require entry, cleaning, and coding</li> </ul>
Data collection methods	<ul style="list-style-type: none"> <li>How will data be collected?</li> <li>Does the evaluation require computer-assisted data collection methods?</li> <li>Will data collection require travel?</li> <li>Will participants be compensated?</li> </ul>	<ul style="list-style-type: none"> <li>Using current measures</li> <li>Using simple data collection techniques (e.g., existing portals, easily programmed survey software)</li> </ul>	<ul style="list-style-type: none"> <li>Developing new measures</li> <li>Using complex data collection techniques (e.g., building a new data collection portal, programming complicated web surveys, using computer-assisted telephone interviewing)</li> </ul>
Statistical complexity	<ul style="list-style-type: none"> <li>What amount of time and level of technical expertise are required to conduct data analysis and interpretation?</li> </ul>	<ul style="list-style-type: none"> <li>Data analysis requires descriptive methods to summarize data (e.g., average participant age, proportion of participants who are employed)</li> <li>Evaluation intends to establish evidence of a relationship between intervention and outcomes that is suggestive, not causal</li> </ul>	<ul style="list-style-type: none"> <li>Data analysis requires advanced inferential statistical methods to establish evidence of a causal relationship</li> </ul>
Evaluation capacity	<ul style="list-style-type: none"> <li>Can the evaluation be conducted by staff from the program or organization being evaluated (internal evaluators), or does it require outside support (external evaluators)?</li> <li>Do any supplies or equipment need to be purchased or rented for the evaluation?</li> </ul>	<ul style="list-style-type: none"> <li>Program staff have sufficient knowledge or experience to design and implement the evaluation (e.g., relevant training, data-driven culture, experience engaging community representatives)</li> <li>Program staff can build in time needed to conduct the evaluation and/or have access to evaluation resources</li> </ul>	<ul style="list-style-type: none"> <li>Need to hire staff with sufficient knowledge, experience, time, and resources to design and implement the evaluation</li> <li>Independent external evaluator needed to enhance credibility of findings</li> </ul>

In general, as you increase the budget for your evaluation, you gain a corresponding increase in knowledge about your success in attaining your program objectives. In many situations, the lowest cost evaluations may not be worth the expense, and realistically, the highest cost evaluations may be beyond the scope of most agencies' financial resources. When possible, consider dropping evaluation components rather than reducing the quality of the evidence collected. For example, lowering your study recruitment budget may reduce your survey response rates because your team does not have time to follow up with nonrespondents. This would diminish the quality of your data and the conclusions you can draw about your program's effectiveness. Instead, maintain data quality and reduce the scope of your evaluation (e.g., focus on one component rather than an entire program).

Depending on budgeting and planning processes in your organization, you may be asked to roughly estimate evaluation costs before evaluation planning starts and develop a more detailed budget later.

## Practice Culturally Responsive and Equitable Evaluation When Preparing for an Evaluation

Evaluation teams often fail to include community members as co-creators or consider cultural assumptions and norms, the community's history and context, and the structural inequities. Use a culturally responsive and equitable evaluation (CREE) approach to gain a better understanding of your program's setting. While it is important to engage community members, especially those eligible to receive the program's services, they are not responsible for educating evaluators. Evaluators must do the work to understand the factors that can influence an evaluation.

Ideally, systems to collaborate with local organizations and community members will be in place before the planning process begins. Engaging community members in the logic model development can help you identify perspectives previously not explored. This approach can also help program staff understand how community members' expectations may differ from their own. If the evaluation design is already underway (i.e., the logic model and/or objectives are set), it is still worthwhile to include community members and other collaborators to the extent possible.

When learning about factors such as historical and current systemic sources of racism, communities cannot be considered as having identical experiences. Collect information from many sources offering a variety of

### Sources for understanding factors that could influence a program and its evaluation

- Review written materials, such as literature or evaluation reports of similar programs in comparable communities, local news stories, or even blog posts by local influencers
- Local public officials or records
- Business and nonprofit leaders
- Neighborhood associations
- Program partner organizations
- Community members
- Current and past participants
- Other evaluators working in the community

perspectives. Potential, current, or past participants all have valuable perspectives about why they would or would not participate in the program and what they would expect from program participation.

When thinking through *what to learn*, focus on factors that could influence the program based on the emerging or final logic model design. New information can help shape overarching objectives and ways to measure specific implementation and outcome objectives. For example, if an implementation objective relates to the number of program participants, understanding barriers to participation is important.

When thinking through *how to apply* what you learned, consider how development of evaluation questions can reflect a focus on equity based on community members' experiences of underlying systems of inequity (e.g., examine how institutional practices or policies affect individuals differently based on race, gender, income). In addition to shaping the logic model and development of objectives, your understanding will likely influence the data you seek (e.g., anticipated and actual program access barriers, determination of whether the program is culturally appropriate and meeting the expectations of participants, and participant outcomes and feedback).

Following are general considerations when incorporating a CREE approach to program evaluations:

- Allow time in your evaluation development process to learn about factors that could influence the program's implementation or outcomes. Time is needed to develop rapport with community members and include many perspectives.
- Include budget needs for evaluation team time and effort, including community members and other local partners and any other necessary resources for the planning process.
- Form an inclusive evaluation team as early as possible to gather more diverse perspectives on planning aspects, such as the logic model and program objectives.
- Develop a common understanding of how decisions will be made to ensure all members of the evaluation team, including community members and study participants, can contribute in meaningful and authentic ways.

### To learn more ...

- [A Guide to Assessing Needs](#) (Watkins et al., 2012)
- [Budget Preparation Guidelines Procurement and Grants Office](#) (CDC, n.d.-c)
- [Checklist for Developing and Evaluating Evaluation Budgets](#) (Horn, 2001)
- [Evaluability Assessment: Examining the Readiness of a Program for Evaluation](#) (JRSA, 2003)
- [Evaluation Questions Checklist](#) (Wingate et al., 2016)
- [Logic Model Tip Sheet](#) (FYSB, n.d.)
- [Needs Assessment Guide](#) (WHO, n.d.)
- [Refining Your Question](#) (DeCarlo, 2018)
- [Logic Model Development Guide](#) (W.K. Kellogg Foundation, 2004)
- [Tools and Methods for Evaluating the Efficiency of Development Interventions](#) (Palenberg, 2011)



## References

- Blase, K., & Fixsen, D. (2013). *Core intervention components: Identifying and operationalizing what makes programs work*. <https://aspe.hhs.gov/report/core-intervention-components-identifying-and-operationalizing-what-makes-programs-work>
- Blocklin, M., Hyra, A., Kean, E., & Porowski, A. (2019). *Community collaborations evaluation plan template and quality indicators*. Abt Associates. <https://omb.report/icr/201906-0970-001/doc/98252801.pdf>
- CDC (Centers for Disease Control and Prevention). (n.d.-a). *Types of evaluations*. <https://www.cdc.gov/std/program/pupestd/types%20of%20evaluation.pdf>
- CDC. (n.d.-b). *Program evaluation tip sheet: Economic evaluation*. Evaluation and Program Effectiveness Team, Division for Heart Disease and Stroke Prevention. [https://www.cdc.gov/dhdsp/docs/program\\_evaluation\\_tip\\_sheet\\_economic\\_evaluation.pdf](https://www.cdc.gov/dhdsp/docs/program_evaluation_tip_sheet_economic_evaluation.pdf)
- CDC. (n.d.-c). *Budget preparation guidelines: Procurement and Grants Office (PGO)*. <https://www.cdc.gov/hiv/pdf/funding/announcements/ps17-1704/cdc-hiv-ps17-1704-budget-preparation-guidelines.pdf>
- DeCarlo, M. (2018). *Refining your question*. Scientific Inquiry in Social Work. <https://scientificinquiryinsocialwork.pressbooks.com/chapter/3-3-refining-your-question/>
- Epstein, D., & Klerman, J. A. (2012). When is a program ready for rigorous impact evaluation? The role of a falsifiable logic model. *Evaluation Review*, 36(5), 375–401. <https://doi.org/10.1177/0193841X12474275>
- FYSB (Family and Youth Services Bureau). (n.d.). *Logic model tip sheet*. U.S. Department of Health and Human Services, Administration for Children and Families. [https://www.acf.hhs.gov/sites/default/files/documents/prep-logic-model-ts\\_0.pdf](https://www.acf.hhs.gov/sites/default/files/documents/prep-logic-model-ts_0.pdf)
- GAO (Government Accountability Office). (2021). *Program evaluation: Key terms and concepts*. <https://www.gao.gov/assets/gao-21-404sp.pdf>
- Horn, J. (2001). *Checklist for developing and evaluating evaluation budgets*. Western Michigan University, Evaluation Checklist Project. <https://wmich.edu/sites/default/files/attachments/u350/2018/budgets-horn.pdf>
- JRSA (Justice Research and Statistics Association). (2003). *Evaluability assessment: Examining the readiness of a program for evaluation*. Office of Juvenile Justice and Delinquency Prevention. <https://www.jrsa.org/pubs/juv-justice/evaluability-assessment.pdf>
- OPRE (Office of Planning, Research, and Evaluation). (2010). *The program manager's guide to evaluation*. Second Edition. U.S. Department of Health and Human Services. <https://www.acf.hhs.gov/opre/report/program-managers-guide-evaluation-second-edition>
- Palenberg, M. (2011). *Tools and methods for evaluating the efficiency of development interventions*. Global Public Policy Institute. <https://www.gppi.net/2011/05/31/tools-and-methods-for-evaluating-the-efficiency-of-development-interventions>

- Tribal Evaluation Institute. (2016). *Using PICO to build an evaluation question*. <https://www.tribaleval.org/evaluation/using-pico/>
- Watkins, R., Meiers, M., & Visser, Y. (2012). *A guide to assessing needs essential tools for collecting information, making decisions, and achieving development results*. World Bank. <https://openknowledge.worldbank.org/bitstream/handle/10986/2231/663920PUB0EPI00essing09780821388686.pdf?sequence=1&isAllowed=y>
- WHO (World Health Organization). (n.d.). *Needs assessment*. [http://apps.who.int/iris/bitstream/handle/10665/66584/WHO\\_MSD\\_MSB\\_00.2d.pdf;jsessionid=942102247F724CABD2490AC87B924C34?sequence=4](http://apps.who.int/iris/bitstream/handle/10665/66584/WHO_MSD_MSB_00.2d.pdf;jsessionid=942102247F724CABD2490AC87B924C34?sequence=4)
- Wingate, L. & Schroeter, D. (2016). *Evaluation questions checklist for program evaluation*. Western Michigan University, Evaluation Checklist Project. <https://wmich.edu/sites/default/files/attachments/u350/2018/eval-questions-wingate%26schroeter.pdf>
- W.K. Kellogg Foundation. (2004). *Logic model development guide*. <https://wkkf.issuelab.org/resource/logic-model-development-guide.html>

# Chapter 5. Design Your Evaluation

## What's Inside?



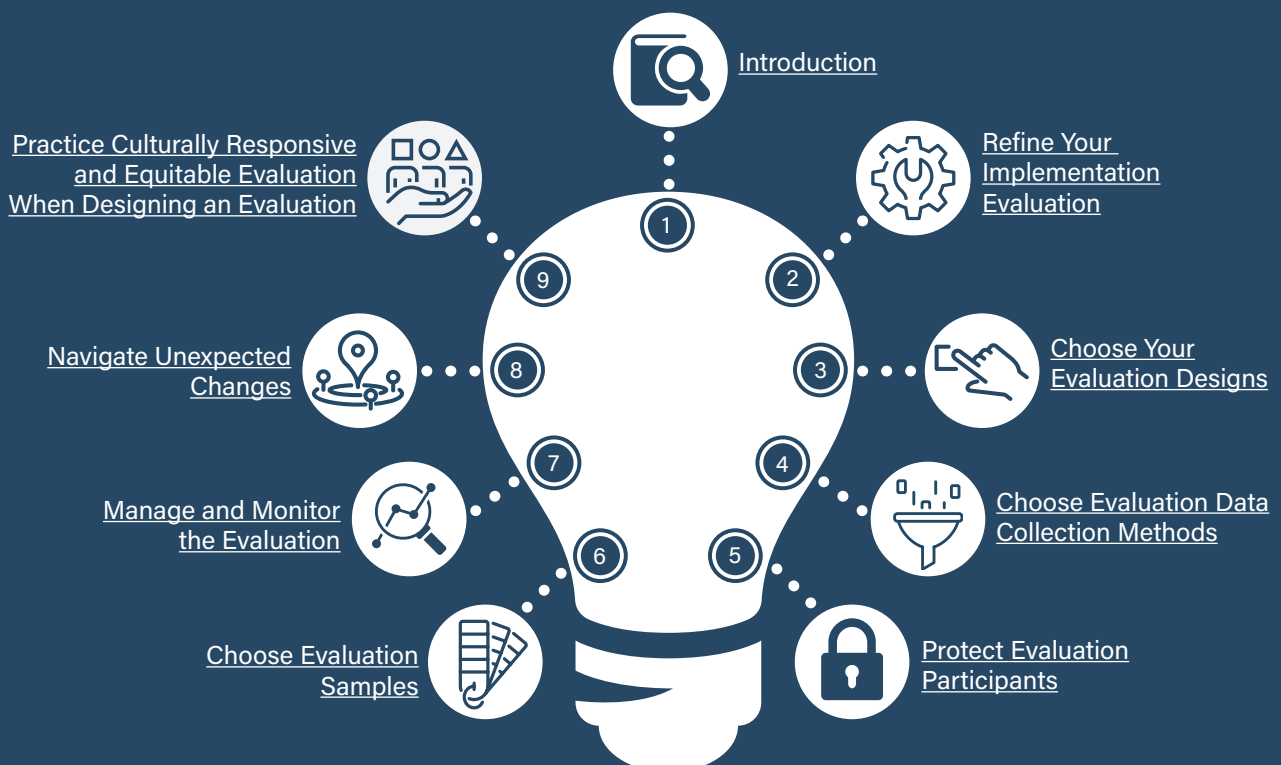
### What this chapter contains

- A brief introduction to implementation and outcome evaluation designs and common data collection methods
- A discussion of evaluation management, including ways to protect evaluation participants

### Who can use this chapter

- Evaluation team members preparing to develop an evaluation plan

Click the links below to view the relevant section



# Introduction

Now that you have developed your evaluation questions, you need to determine how you will answer each one. That means some big decisions to develop an approach for each question:

- Determine the appropriate evaluation designs.
- Select the appropriate data collection methodologies.
- Establish each data collection's source of data.
- Identify the appropriate measures for each evaluation question's concepts.

This chapter covers the first two steps, and [chapter 6](#) addresses the last two steps. We recommend reviewing [chapter 6](#) (to learn about identifying and collecting data) in conjunction with this chapter on design and methods. Selecting the appropriate designs and methods is an important part of evaluation planning because it helps you collect high-quality, relevant data to best answer your evaluation questions.

## Evaluation terminology

In an evaluation, information is often referred to as data.

## Evaluation plan templates

Many federal funding programs require grantees to develop and submit evaluation plans for agency approvals. Funders often have specific guidance for how grantees should develop their evaluation plans, but the general purpose is to make as many evaluation decisions as possible in advance of implementation to better ensure the timely and smooth execution of a program evaluation.

This Guide supports and aligns with common evaluation plan components, such as explaining evaluation questions, determining designs, identifying measures, and developing data collection procedures. For examples of evaluation plan advice or templates, see Children's Bureau, Administration for Children and Families (ACF), 2019, and Blocklin et al., 2019.

You may want to pull all your plans together into a formal evaluation plan. An evaluation plan is a "written document that describes how you will monitor and evaluate your program, as well as how you intend to use evaluation results for program improvement and decision-making" (CDC, 2011a). These plans are particularly helpful if membership in your evaluation team changes over time. The evaluation plan provides much information a new member will need to get oriented to the evaluation.

If made public, evaluation plans can also support transparency by sharing information with interested parties about how you plan to conduct your evaluation.

If you are conducting a high-profile evaluation, you may also use your evaluation plans to register your study (i.e., add your evaluation plan to a third-party study registry). Study registrations offer a more formal way to prespecify your evaluation approach and sometimes even your analysis plans. Registration brings transparency and credibility to your evaluation because you have registered a plan for how your study will unfold and what outcomes you will report. This disclosure helps prevent selective sharing of only positive findings in later reports.

## Evaluation Design Versus Data Collection Methods

The distinction between evaluation design and data collection method is important. The two are different but closely related and easily confused. Evaluation design is the approach you use to answer each evaluation question. Evaluation data collection methods refer to the data collection strategies you use to execute that plan. No specific research design requires a specific data collection method. Therefore, you may select data collection methods after you select your design.

Because each program is unique, choose designs and data collection methods that fit your evaluation's goals, objectives, and expected ability to attribute outcomes as an effect of the program.

This chapter briefly introduces three types of evaluation designs (nonexperimental, quasi-experimental, and experimental) and describes the most frequently used data collection methods (surveys, secondary data analysis and archives, interviews, focus groups, and observations). These designs and data collection methods do not represent an exhaustive list but rather a starting point when considering the most suitable option for your program evaluation.

### Treatment groups, comparison groups, and control groups

- **Treatment groups** are sets of individuals, classrooms, schools, departments, families, or other groups who are offered or receive the program being tested in the evaluation. In this Guide, we use the term treatment or program group to stay consistent with program as the focus of the evaluation.
- **Comparison groups** are sets of the same units as treatment or program who do NOT receive or are not offered the program. Comparison groups can be selected using numerous methods (see quasi-experimental designs below).
- **Control groups** are a specific type of comparison group created through random assignment. All control groups are comparison groups, but not all comparison groups are control groups. For simplicity, we use only the term comparison groups.

## Refine Your Implementation Evaluation

Implementation evaluations can answer different types of questions, and each of those questions may call for different methods. Below are common implementation evaluation questions and their associated data needs (adapted from Blocklin et al., 2019):

- **Questions about reach.** Reach measures the scale of program activities. You could calculate reach at the participant level by counting the number of participants served by the program over a time period (such as yearly, or life of the program or evaluation). Community-level reach could be measured as the number of communities (or neighborhoods, schools, cities, or housing developments) served by the project.
- **Questions about saturation.** Saturation measures how widespread the program was. This concept is related to reach in that the count of people (or families, houses, schools, etc.) becomes a numerator. The denominator is a measure of the size of the overall population. You could calculate saturation as the number of children served over the number of children eligible for services or the number of children in the entire county, for example. Saturation is particularly important for programs that take a place-based or community-based approach to services.

- **Questions about service receipt.** Service receipt measures how participants engage in your program. This measure is more detailed than reach. It could measure the number and type of services participants (or different types of participants) received (e.g., 90 percent of men participated in job search activities, while 15 percent participated in substance prevention education). It can also measure dosage such as the percentage of people who completed the number of service hours required to graduate or the average number of mentor sessions youth attended.
- **Questions about fidelity.** Fidelity measures the extent to which the program was implemented as planned or as designed. Fidelity is measured separately for each program activity (e.g., outreach, recruitment, curriculum-driven sessions, case management meetings, coaching sessions). It is assessed by establishing a threshold for each activity (e.g., conduct 10 outreach sessions a month, recruit 200 families a year, 80 percent of enrolled participants receive at least 7 of 10 curriculum-driven sessions). Ideally, fidelity calculations occur at least annually as a way to inform program management decisions.
- **Questions about implementation drivers, barriers, and solutions.** These questions typically collect qualitative data from program staff, leadership, and partnering organizations to document how program implementation went. They focus on drivers (what helped the program function), barriers (challenges program staff encountered when running the program), and solutions (what program staff did to overcome or mitigate the barriers). Drivers and barriers can occur at many levels, such as federal, societal, state, local, system, agency, community, or individual.

## Choose Your Evaluation Designs

Evaluation questions are often grouped in two categories: implementation (sometimes referred to as “process”) and outcome. Implementation evaluation questions are descriptive: They help you collect systematic information about how the program was delivered, who staffed the program, who participated in the program, how well program activities were delivered, and how external factors influenced program delivery.

Outcome evaluation questions document changes associated with the program, such as improvements in participant income, reductions in staff turnover, or changes in data interoperability. Typically, outcome evaluations are strengthened by accompanying implementation evaluations that provide context to outcome findings.

In designing evaluations to address outcome questions, evaluation teams must determine how to isolate the impact of a program from other factors that could influence the same outcome. Accordingly, outcome evaluation questions require a comparison condition; that is, a way to compare observed program results with those you would expect if the program had not been implemented

### What are comparison conditions?

Outcome evaluations need to answer the question, compared to what? How will you know if the value of an outcome (e.g., average income after training) is “good?” You need to compare the outcome finding to something else. That something else is a comparison condition. Common comparison conditions follow:

- Pretreatment measures of the outcome from the treatment/program group
- A benchmark measurement such as state-level target for standardized test results
- An evidence-based target such as gains demonstrated by other similar program evaluations (e.g., a 10-percentage-point drop)
- Outcomes from a nontreated group, such as a comparison or control group, measured at the same time as the treatment/program group outcome

(i.e., the counterfactual<sup>1</sup>). Evaluations establish a comparison condition for outcome evaluation questions via three designs: nonexperimental, quasi-experimental, and experimental.

These designs appear below in order from least to most able to attribute changes in outcomes to the program and not other factors:

**Nonexperimental designs** provide a hypothetical prediction of what would have happened in the absence of the program. The most common of these designs is a single group pre-posttest: Participants provide data on outcomes of interest, the program is implemented, and participants once again provide data on the same outcomes. For example, an evaluation captures participants' knowledge of child development before and after a parenting education program to examine whether participants demonstrate change (improvement) over time.

Other nonexperimental designs compare participant outcomes with benchmarks or national statistics (e.g., 85 percent of program children demonstrated grade-level reading skills compared with 70 percent of children of a similar age nationwide). Nonexperimental designs are often used when quasi-experimental or experimental designs are not feasible or practical.

**Quasi-experimental designs** identify a comparison group—individuals who are as similar as possible to the evaluation participants, but they did not participate in the program. Evaluators use many approaches to develop a quasi-experimental comparison group; for example, including—

- Individuals eligible for but uninterested in participating in the program
- Individuals not able to participate in the program yet because of program space constraints (wait list control group)
- Similar individuals in another community or school (matched comparisons; see textbox in [Choose Evaluation Samples](#) below)
- Artificial comparison groups created using advanced statistical techniques (such as synthetic comparison groups)

The evaluator can also use a design called comparative interrupted time series<sup>2</sup> to conduct quasi-experimental cluster evaluations, such as tests of community-, city-, or state-level interventions. Comparative interrupted time series use multiple data collection waves to establish patterns for both treatment/program and comparison clusters. Other advanced quasi-experimental designs make use of statistical techniques to develop a comparison condition, such as regression discontinuity or propensity score matching. To provide rigorous results, treatment groups and comparison groups must be statistically identical or similar to each other on pre-intervention measures of important outcomes.

<sup>1</sup>Counterfactuals allow evaluations to make comparisons between the observed results to those expected if the intervention had not been implemented (Better Evaluation, n.d.-a).

<sup>2</sup> Some evidence suggests comparative interrupted time series designs may be as internally valid as experimental designs (St. Clair et al., 2014)

**Experimental designs** also highlight comparison groups of people who were not offered the program. The difference is in how the comparison groups are constructed. In experimental designs, individuals are randomly assigned to be offered or not offered the program. Random assignment seeks to ensure the two groups are nearly identical in factors that may influence the outcome being examined. As a result, any difference in outcomes between the program participants and comparison group after the program has been implemented can be attributed to effects of the program.

Experimental designs offer the strongest evidence that changes in outcomes are caused by the program. Experimental designs are considered the gold standard in generating causal evidence and are important because they provide strong conclusions about whether a program should be replicated or expanded to more people, or whether the program is ineffective and should be discontinued or significantly revised.

Sometimes, randomly assigning some people to not be offered a program can seem unpalatable or objectionable to program staff or community members. Program staff may feel it is unfair to deny services to interested community members. It is important to remember the evaluation is providing information about whether a program works or even if it might have negative impacts. It is not unethical to offer a program to only a portion of interested people to learn whether it is working. This is similar to clinical trials in medicine (e.g., we know the polio vaccine is effective and so it is universally available, but we have not yet identified an effective vaccine against HIV, so clinical trials are ongoing).

If program staff or community members still have reservations about random assignment, evaluations may be able to overcome those reservations by addressing the services or programs offered to the comparison group. For example, not every evaluation needs to have a no-treatment comparison group, in which comparison group members receive no services. In some cases, you may be able to provide an alternate service or a small intervention such as a book or gift card. Alternatively, you could offer the program to comparison group members after final data collection ends.

Each of these designs offers different strengths and limitations (see table 5.1). Your evaluation team will need to select the most appropriate design for each outcome evaluation question based on factors such as the following:

- **Required level of rigor in causal attribution.** Your design choice will be influenced by the extent to which you must be able to document a change in an outcome has been caused by the program. The level of rigor you need may be dictated by your funder, your advisory board, or the goals of your evaluation (e.g., meet evidence requirements established by an evidence clearinghouse). In general,

### Focusing on evaluation rigor

For each type of evaluation, several strategies can increase rigor or improve the quality of the information gathered. Examples follow:

- Nonexperimental designs can establish a priori goals for the magnitude of change expected to be seen if the program meets the level of improvement goals.
- Experimental and quasi-experimental evaluations can apply a “difference in differences” approach to calculate change over time, which helps to account for change that can be attributed to factors other than the program.

Work with your evaluator to identify feasible strategies to strengthen your evaluation designs.



experimental designs provide evaluators the highest confidence that any differences in outcomes are caused by the intervention. Quasi-experimental designs also enable evaluators to make inferences about whether interventions cause impacts, but there is some uncertainty about whether factors not observed by the study are causing changes in outcomes. Nonexperimental designs do not enable evaluators to demonstrate that changes in outcomes are caused by the intervention; these designs cannot rule out other factors as causing the changes. If you must have an evaluation at the highest level of rigor, experimental designs are most likely to meet that need.

- **Data availability.** Quasi-experimental and experimental designs need units of observation (e.g., families, centers, communities, children, classrooms) that did not participate in the program. That means you will need to identify comparison units and be able to collect identical data from both the comparison units and the units that were offered the program.
- **Resource availability.** Typically, more complex evaluation designs require additional resources. These resources include financial, technical skills/evaluation capacity, data capacities, and time to collect data on long-term outcomes (outcomes that happen months or years after the program ends). Some random assignment designs need a surplus of people interested in participating in the program to build large enough treatment/program and comparison groups to meet statistical power requirements.
- **Social, cultural, and political context.** Not all outcome evaluation designs are feasible in the “real” world. You will not be able to ask some families to not participate in a universal or mandated program (e.g., not attend public school). In other situations, community representatives may not approve of certain designs, particularly those that withhold an accessible program.

**Table 5.1. Possible Designs for Outcome Evaluations**

Design	Description	Example(s)	Use	Strengths	Limitations
Nonexperimental	<ul style="list-style-type: none"> <li>■ Designs without comparison groups or randomized assignment</li> </ul>	<ul style="list-style-type: none"> <li>■ Single group pre-post test</li> </ul>	<ul style="list-style-type: none"> <li>■ Describe individuals, settings, or events within the context of their occurrence</li> </ul>	<ul style="list-style-type: none"> <li>■ Can be used when baseline data and/or comparison groups are not available</li> <li>■ Requires fewer resources</li> </ul>	<ul style="list-style-type: none"> <li>■ Minimal ability to infer causality</li> </ul>
Quasi-experimental	<ul style="list-style-type: none"> <li>■ Designs with comparison groups but no randomized assignment</li> </ul>	<ul style="list-style-type: none"> <li>■ Matching and propensity score designs</li> <li>■ Comparative interrupted time series designs</li> <li>■ Regression discontinuity designs</li> <li>■ Instrumental variables estimations</li> </ul>	<ul style="list-style-type: none"> <li>■ Conduct evaluations in field settings or in situations when comparable groups are created by differences that already occur in real world</li> <li>■ More appropriate for complex community and systems change initiatives</li> </ul>	<ul style="list-style-type: none"> <li>■ Can infer moderate level of causality when not logistically feasible or not ethical to conduct randomized controlled trial</li> </ul>	<ul style="list-style-type: none"> <li>■ Offers moderate confidence in inferring causality</li> <li>■ Differences between groups may generate a confound<sup>a</sup></li> </ul>

Table 5.1. Possible Designs for Outcome Evaluations (continued)

Design	Description	Example(s)	Use	Strengths	Limitations
Experimental	<ul style="list-style-type: none"> <li>▪ Designs with randomized assignment (inclusion of a control group) to definitively answer cause-effect questions</li> </ul>	<ul style="list-style-type: none"> <li>▪ Randomized controlled trial</li> </ul>	<ul style="list-style-type: none"> <li>▪ Establish cause-effect relationship</li> <li>▪ More appropriate for programs seeking highest level of rigor (considered gold standard for studying causal relationships)</li> </ul>	<ul style="list-style-type: none"> <li>▪ Most robust design for testing causal hypotheses</li> </ul>	<ul style="list-style-type: none"> <li>▪ Most resource intensive</li> <li>▪ Can be difficult to generalize to “real world”</li> </ul>

Sources: CDC (2011b); Moore (2008)

<sup>a</sup> Confounds are “any factor, other than the intervention, that is both plausibly related to the outcome measures and also completely or largely aligned with either the intervention group or the comparison group” (Wilson et al., 2019, p. 35). For example, if all treatment group members receive the treatment from a single individual, a confound is present because you cannot parse out whether the treatment (like a reading intervention) or the provider is responsible for the increase in reading scores in comparison to the comparison group.

### Wait! Shouldn't we always conduct a randomized control trial?

Sometimes evaluators choose a design requiring random assignment because of its high prestige in the research community. Prestige, however, is not a relevant criterion for design selection. The best designs for your evaluation are those you can implement well and with design fidelity. Use the advice in this text box to determine whether a high-quality random assignment design is feasible in your situation. The evaluation literature describes common challenges to different evaluation designs; your evaluation team should critically evaluate their ability to overcome or ameliorate those challenges.

Many random assignment evaluations are underpowered: They are unable to recruit enough people in the treatment/program and comparison groups to actually detect differences in outcomes. What evidence do you have that you'll be able to recruit a sufficient number of evaluation participants?

Some programs undergo rigorous evaluation too soon—looking for program impacts before the program model has been refined and without any evidence the model is implemented with fidelity. This can make an otherwise effective program appear ineffective. Programs may want to conduct implementation evaluations first and test program improvements before turning to a rigorous outcome evaluation.

Regardless of which outcome evaluation designs you select, an implementation evaluation should ideally accompany an outcome evaluation. Pairing these two designs lets you provide context and strengthen interpretations of any outcome findings (e.g., an implementation evaluation can indicate if the program was poorly implemented to explain the lack of improvement in a related outcome).

## Choose Evaluation Data Collection Methods

The next step in evaluation planning is to decide on the methods of data collection you will use to collect the information needed to answer your evaluation question. Many people think the term “data” refers to only numerical information, but data can be facts, statistics, images, quotes, or any other information collected about your program or participants.

Most evaluation questions can be answered using several data collection methods. Select those that best meet your needs, accessibility to various data sources, budget, and timeline. Common data collection methods include surveys, administrative data, interviews, focus groups, observations, and document reviews. The method you choose should be based on the type of data you want to collect (i.e., qualitative<sup>3</sup> versus quantitative<sup>4</sup>). For example, focus groups, interviews, and observations are best for collecting qualitative data, while quantitative data are typically collected from survey and administrative data sources. However, most methods can produce both qualitative and quantitative data. For example, you can pose open-ended questions in a survey or provide percentages of the number of interviewees who identified the same implementation challenge in a focus group. See table 5.2 for a discussion of common data collection methods and their strengths and limitations.

### Evaluation consent

For all methods that involve human subjects, it is imperative evaluations receive consent from people to collect data from or about them. See [Protect Study Participants](#) for more information.

**Table 5.2. Possible Data Collection Methodologies**

Design	Description	Example(s)	Use	Strengths	Limitations
Surveys	<ul style="list-style-type: none"> <li>Data collection efforts that use a formal, prespecified, instrument to collect data; can be large-scale</li> <li>Can be paper and pen, online, phone fielded, or use combination of collection strategies</li> <li>Typically collect closed-ended questions (e.g., Do you like this Guide? Answer yes or no) but can also use open-ended questions (e.g., Describe how you will use this Guide)</li> </ul>	<ul style="list-style-type: none"> <li>Online data collection form that evaluation participants complete</li> </ul>	<ul style="list-style-type: none"> <li>Collect identical data across all evaluation participants; when you know all the questions you want to ask and what response options are</li> </ul>	<ul style="list-style-type: none"> <li>Collect large amount of data from many people</li> <li>Produce quantitative data to inform statistical analyses</li> </ul>	<ul style="list-style-type: none"> <li>Unlikely to collect new perspectives</li> <li>Can be costly to develop and field</li> <li>Need to invest in measurement selection and order of questions</li> </ul>

<sup>3</sup> Qualitative data are information that are difficult to measure, count, or express in numerical terms. For example, a participant's impression about the fairness of a program rule/requirement is qualitative data (OPRE, 2010).

<sup>4</sup> Quantitative data are information that can be expressed in numerical terms, counted, or compared on a scale. For example, using a score developed from a reading test to document a child's reading level (OPRE, 2010).

Table 5.2. Possible Data Collection Methodologies (continued)

Design	Description	Example(s)	Use	Strengths	Limitations
Administrative data	<ul style="list-style-type: none"> <li>Data that programs collect as part of providing services</li> <li>Can be data from your specific program, such as attendance records, but can also be data collected from your participants but by another organization</li> </ul>	<ul style="list-style-type: none"> <li>Temporary Assistance for Needy Families recipient database</li> </ul>	<ul style="list-style-type: none"> <li>Access data that have already been collected; reduce burden on evaluation participants; can be cost-effective</li> </ul>	<ul style="list-style-type: none"> <li>Can be cost-effective; often very little missing data</li> </ul>	<ul style="list-style-type: none"> <li>Measures of interest may not be available in existing datasets</li> <li>May be difficult to access based on data ownership</li> <li>May be difficult to link individuals in administrative data to other data sources</li> <li>May have challenges with data quality, accuracy, and/or thoroughness</li> </ul>
Interviews	<ul style="list-style-type: none"> <li>In-depth conversations with one or more individuals</li> <li>Interviewers typically use standardized protocols or lists of questions to guide the conversations</li> <li>Most questions are open ended</li> </ul>	<ul style="list-style-type: none"> <li>In-depth, structured discussion with program manager</li> </ul>	<ul style="list-style-type: none"> <li>Collect information about experiences, perceptions, or activities not easily captured with closed-ended questions</li> </ul>	<ul style="list-style-type: none"> <li>Generate data when response options are unknown or too complex</li> <li>Able to shift and change as interviewee surfaces additional topics</li> </ul>	<ul style="list-style-type: none"> <li>Resource intensive</li> <li>Requires significant time commitment on part of interviewee</li> <li>Data may be seen as less rigorous than other methods</li> </ul>
Focus groups	<ul style="list-style-type: none"> <li>Conversation held with multiple individuals at once</li> <li>Typically use written set of questions but also provide significant amount of space for focus group participants to react to, build on, and engage with comments from their fellow participants</li> </ul>	<ul style="list-style-type: none"> <li>Focus group of program participants</li> </ul>	<ul style="list-style-type: none"> <li>Collect data from multiple, similar evaluation participants at the same time</li> <li>Generate additional data through evaluation participant interaction and discussion</li> </ul>	<ul style="list-style-type: none"> <li>Collect much data in short time</li> <li>May be more comfortable for evaluation participants than one-on-one interviews</li> <li>Can benefit from group synergy</li> </ul>	<ul style="list-style-type: none"> <li>Can be difficult to schedule</li> <li>Group format may affect honesty of participant responses</li> <li>Need to be conducted by skilled facilitator</li> <li>Need to address issues of confidentiality with focus group participants</li> </ul>
Observations	<ul style="list-style-type: none"> <li>Members of data collection team "sit in on" event or process and document what they see</li> <li>Observations typically use tool to ensure consistent information is documented for each event</li> </ul>	<ul style="list-style-type: none"> <li>Evaluator observes participant workshop session</li> </ul>	<ul style="list-style-type: none"> <li>Document the "feel" of an event or process</li> </ul>	<ul style="list-style-type: none"> <li>Give context and depth to an evaluation</li> </ul>	<ul style="list-style-type: none"> <li>Typically need to be combined with other data source to answer evaluation questions</li> </ul>

Table 5.2. Possible Data Collection Methodologies (continued)

Design	Description	Example(s)	Use	Strengths	Limitations
Document reviews	<ul style="list-style-type: none"> <li>Make use of existing written materials as data sources</li> <li>Typically guided by extraction tools to help capture relevant information</li> </ul>	<ul style="list-style-type: none"> <li>Program management decisions as documented in meeting notes</li> </ul>	<ul style="list-style-type: none"> <li>Collect data already available in written sources</li> </ul>	<ul style="list-style-type: none"> <li>Cost-efficient</li> <li>Less burdensome to evaluation participants</li> </ul>	<ul style="list-style-type: none"> <li>Rely on quality, accuracy, and thoroughness of source documents</li> </ul>

Typically, strong evaluations do not rely on just one type of data; instead, they employ a mixed-methods approach. Mixed-methods evaluations use more than one type of data to tell the program's story (i.e., they collect both qualitative and quantitative data). A specific kind of mixed-methods approach—triangulation—uses multiple methods to collect data on the same outcome. For example, an evaluation could triangulate customer satisfaction by surveying participants (quantitative), interviewing participants (qualitative), and conducting observations of participant and program staff interactions (qualitative).

## Protect Evaluation Participants

### Rapid cycle evaluation (RCE)

RCE is an approach to evaluation that relies on innovative design and methods to quickly test program components and provide actionable results to integrate improvements into further testing. With RCE, program or process changes can be tested in a shorter time and decision-makers can have increased confidence in results. For a more detailed look at RCE, see Atukpawu-Tipton and Poes (2020).

All evaluation efforts, including data collection, must respect and protect the privacy of the individuals who contribute information to the evaluation. Evaluators and social science researchers have developed procedures designed to ensure individuals who provide data do so voluntarily, have their information safeguarded, and have their privacy respected.

Institutional review boards (IRBs) are oversight agencies that review study procedures to ensure study participants' rights and welfare are protected. Many large evaluation

firms, almost all universities, and numerous state-level agencies have IRBs that approve evaluations conducted by their staff or with their funding. Independent evaluators can hire private IRBs to approve their evaluations. Any human subjects research or evaluation conducted with federal dollars is required to receive IRB approval.

Broadly speaking, protecting study participants (i.e., any human who provides information to a study, including evaluation) includes the following:

- Informed consent.** People who contribute information about themselves for an evaluation should understand what information you are asking for, what you are doing with the information, how you will protect their information and identity, what risks there may be if they participate in the evaluation,

and what would happen if they choose to not participate in the evaluation (e.g., they would still receive services but wouldn't receive financial incentives for data collection). Consent procedures should take into account participants' preferred languages, levels of literacy, comfort with research, and power dynamics between the evaluation team and potential participants.

- **Voluntary nature.** Individuals must retain their rights to autonomy. Potential evaluation participants should not be coerced into participation in the evaluation, nor should they face significant consequences for not participating in the evaluation. Evaluation participants should be able to refuse to answer questions or provide information they are not comfortable providing. Evaluation participants should also be able to revoke their consent at any time. Evaluations should provide contact information if an evaluation participant wants to have their data removed from the evaluation and destroyed at a later point.
- **Data security procedures.** Your evaluation team will need to develop procedures and safeguards to ensure only the people who need to see the data (i.e., evaluation team members) can see the data. This includes safeguarding paper copies of surveys, online databases, participant contact information, and datafiles.
- **Privacy procedures.** Your evaluation team will also need to determine if, and if so how, you will ensure

information provided by individuals won't be linked to them. Such procedures include maintaining participant contact information separate from other datafiles, using unique identifier codes rather than people's names or Social Security numbers, and having reporting standards around cell sizes.

### Matching comparison and treatment/program groups

When choosing a sample for a quasi-experimental evaluation, you need to consider how to develop a comparison group that is as similar as possible to your treatment/program group. These similarities reduce the strength and number of alternate explanations for differences in outcomes between the two groups.

This means you will want to select demographic characteristics (e.g., age, gender, race/ethnicity), characteristics associated with your outcomes of interest (e.g., programs aiming to improve educational attainment should choose a comparison group with a similar mixture of educational credentials to the treatment/program group), and temporality (e.g., avoid measuring treatment/program group wages in 2018 and 2022 and using a comparison group with wage data from 2004 and 2008).

### Confidentiality and privacy

What's the difference? Confidentiality and privacy are both important concepts in protecting evaluation participants' data, but they have different meanings. Privacy is about people, while confidentiality refers to data. Evaluations protect participants' privacy by collecting only information the evaluation needs and using discreet data collection procedures (e.g., interviews in a private room, not in a public cafeteria). Confidentiality extends privacy by protecting the information participants provide. It includes procedures used to ensure only authorized people have access to data, and they won't purposely or inadvertently identify evaluation participants and share their information (UCI Office of Research, 2021).

## Choose Evaluation Samples

For each data collection method described above, you will need to determine the data sources for the evaluation. Two terms evaluators use when talking about how units of observation can be selected follow:

- **Census.** A census means you collect data from each unit of observation eligible to provide data. For example, you may survey every single individual who participated in a program (and consents to data collection).

- **Sample.** A sample means you select some number of the eligible units of observation to provide data. For example, your evaluator will not be able to observe every single interaction between case managers and program participants. In that case, you and your evaluator will need to determine how you will develop your sample.

Each different technique for sampling has different implications for your evaluation budget, evaluation timeline, and the extent to which your data can be generalized or seen as representative of the whole group of units of observation. The higher the generalizability, the more likely the data collected from the sample are similar to data you would have gotten if you had collected data from every member of that group. Below are a few common approaches to sampling:

- **Random sampling** means everyone has the same probability of being chosen to be a part of the sample. For example, if you need to collect neighborhood data through a survey, you could knock on every 10th door in the community. Random sampling has strong generalizability and can be cost-effective by helping save time and resources.
- **Convenience sampling** collects data from units of observation easiest to reach. If you recruited for a staff focus group by emailing a request for volunteers, you would have a convenience sample. Convenience samples are easier to generate because you know participants are interested in engaging in the evaluation. However, they provide poor generalizability because it's hard to know if eager individuals differ in important ways from people who didn't see or answer a call for engagement.
- **Purposeful sampling** takes into consideration the purpose of the evaluation, along with the understanding of the target audience. For example, for a purposeful interview sampling frame, your evaluator may call five people who never completed their intake, five people who attended only the first session of the program, and five people who completed the whole program. This strategy would enable the evaluator to answer questions about the program from numerous perspectives and level of engagement.

Sampling approaches can be challenging to develop, implement, and document. If you think your evaluation will need to collect data from a sample of units of observation, it is important a member of your evaluation team has sampling experience.

## Manage and Monitor the Evaluation

As part of designing your evaluation, you will need to build in systems and time to track and manage the evaluation, including data collection and program staff engagement in the evaluation and procedures for updating and revising your evaluation as needed.

Common strategies and tools to manage an evaluation follow:

- **A written evaluation plan.** Drawing from the work you do to develop the evaluation, consider formalizing the final decisions in a written evaluation plan. See the end of the chapter ([To learn more...](#)) for resources with guidance on developing evaluation plans. Most include evaluation questions, designs, data collection methodologies, measures, analysis plans, roles and responsibilities, and an evaluation timeline.
- **Staff trainings and manuals.** Often program staff are engaged in aspects of evaluation data collection, such as fielding intake forms or documenting attendance at events. Everyone engaged in the evaluation must understand how to support the evaluation correctly. This will ensure consistency in information collection and be useful for staff who are hired after the evaluation begins. Training materials help explain the purposes of the evaluation and data collection, how to collect and input data, common challenges to accurate data collection, and advice for solving common challenges.
- **Data dictionaries and coding manuals.** Evaluation team staff engaged in assessing and analyzing data should follow standardized practices to code, clean, and analyze data. Manuals and guides, typically developed by the lead evaluator if delegating work to others, can help support rigorous data analysis.
- **Data quality monitoring.** Evaluation team members should regularly check all data collected to ensure forms are completed accurately, identification numbers are used correctly, and no more than an acceptable amount of data is missing (the amount of acceptable missing data will depend on your study type, expected level of rigor, and funder guidance).
- **Continuous quality improvement procedures.** Programs often engage in continuous quality improvement efforts to identify ways to strengthen program implementation or management. Evaluations may want to apply those same concepts. For example, conversations with data collection staff can help identify if any procedures are hard to follow or burdensome.
- **Regular evaluation team meetings.** Check timelines, review current work, identify, and navigate challenges.

Finally, evaluations often have yearlong timelines or timelines across many years. It is unrealistic to expect all program operations to stay static over that time. For example, you may add or discontinue a particular service or program component. Your evaluation should keep track of changes in program operations through procedures for documenting the time this change occurred, the reasons for the change, and whether particular participants were engaged in the program prior to or after the change. This will help you determine whether the change had any impact on attainment of expected outcomes.

## Navigate Unexpected Changes

Over the course of your evaluation, you likely will not implement the evaluation exactly the way you intended. Changes are expected; you might even make changes to improve the quality or rigor of your evaluation. Examples of possible evaluation changes include program changes, changes in funding, staff turnover, losing access to a data source, slipping timeline, or changes in your expected sample size.



## Conducting evaluations during a pandemic

COVID-19 upended much of daily life, program evaluations included. As programs cancelled in-person activities and pivoted to virtual services, program evaluations also needed to adapt to new public health restrictions. Evaluators revisited their evaluations and worked with programs to determine which evaluation questions were still applicable, which data collection efforts could be adapted to virtual efforts, and how to consent participants and protect their data in a virtual environment.

Evaluators who successfully mitigated the setbacks and challenges posed by COVID-19 worked as partners with their program staff. They brainstormed innovative ideas and pressure-tested them for feasibility. They looked to colleagues and other fields for advice and support. Many evaluators also introduced new evaluation questions to understand and document how programs adapted and responded to the massive changes brought about by COVID-19.

When changes occur, you should take the following steps:

### Step 1

#### Determine how the changes will affect the evaluation

For example, if you will end up with a smaller sample size, your study may have less statistical power to detect changes in outcomes. If you lose access to a data source, you may no longer use those data to measure program outcomes.

### Step 2

#### Determine whether those effects are acceptable, and if not, develop an alternative plan

Continuing with our examples, your statistical expert may indicate your expected magnitude of change can still be detected with a smaller sample size. Conversely, you may determine the outcome measure collected by the lost data source is key to your evaluation and develop another method for collecting similar information.

#### Internal and external validity: What are they and why do they matter?

Evaluators and researchers use these terms to assess certain elements of credibility of studies. Internal validity refers to the extent to which an evaluation identifies the true impact for the individuals included in the evaluation. Many evaluation decisions or quality of execution can affect internal validity. For example, random assignment to treatment/program and comparison groups and high response rates for data collection efforts increase internal validity. Differential attrition (where individuals in the treatment/program and comparison groups drop out of the study at different rates), crossover (where comparison group members receive the treatment/program), and having program staff collect data (as opposed to an independent, “unbiased” data collector) can threaten internal validity.

External validity refers to the extent to which the results of an evaluation might be applicable in other situations (also known as generalizability). Both program and study elements can affect external validity. For example, large incentives or program supports that aren’t feasible in the “real world” reduce external validity, while implementing the program in a standard setting, as opposed to a laboratory or clinical setting, increases external validity.

### Step 3

#### Document the change, and if needed, the solution

Good evaluators develop sufficient documentation of their evaluation to enable them to answer questions about the process and potentially for another evaluator to replicate the evaluation. You should document when the changes occurred (potentially even coding your dataset like a variable for whether participants received the initial or revised program) and how you handled them. This information is also good “institutional knowledge.” Evaluation teams sometimes experience turnover, and strong documentation can help a new team member quickly get up to speed.

### Step 4

#### Prepare and plan for changes in the future

While this Guide may support you through one evaluation, you will likely conduct others in the future. Be sure to draw from your experience with changes and challenges during your current evaluation and proactively plan for them. For example, if you had sample size issues, in future evaluations you may want to plan for a much larger than needed sample (by recruiting more people to participate in a program, or by enrolling people in the evaluation over a longer time period).

## Practice Culturally Responsive and Equitable Evaluation When Designing an Evaluation

A CREE approach to selecting your evaluation designs, data collection methodologies, and sampling frames requires a critical dialogue with community members. Engaging with community representatives will help ensure the evaluation is co-created and community members have buy-in and view the evaluation as credible and meaningful.

**Evaluation designs.** Reflect on whether, or to what extent, certain evaluation designs may be a poor match for the program participants and community members. For quasi-experimental and randomized control groups, can you match treatment and comparison groups on similar characteristics determined as noteworthy, such as race? Also consider the implications or appropriateness for assigning potential participants to a comparison group. For example, communities with a history of purposeful exclusion from beneficial programs and policies, such as difficulties accessing the GI bill for Black veterans (Smithsonian American Art Museum, n.d.), may have particular difficulty accepting a random assignment model where the comparison group is offered no or few services. It may be possible to balance evaluation information needs with community needs. For example, the comparison group of an evaluation of a job training program could be offered weekly social support groups rather than no services at all, or the comparison group could be offered services after all waves of data collection.

**Data collection methods.** Have conversations with evaluation users and community members about what “counts” as credible evidence. Often evaluations tend to prefer quantitative data and use qualitative data to supplement or add context to quantitative results. Data collection methods that capture more individualized experiences (e.g., interviews, focus groups, photovoice, appreciative inquiry, ripple effects mapping) could resonate better with evaluation users and help the evaluation team develop a more complete understanding of possible implementation and outcome findings.

**Survey tools and sample sizes.** Conventional evaluation encourages the use of standardized measures and instruments. These measures may not have been validated with people who would respond or interpret questions similarly to your evaluation respondents. They might also call for larger sample sizes for generalizability than you could reach if also disaggregating data based on demographic characteristics, such as race and religion. That doesn't mean you should or shouldn't use them. Just be aware of the considerations for each choice. Ideally, you can use a mixed-methods approach where some data sources could lead to generalizable findings, while others might need to be considered within the context of the group where they were collected.

**Protecting evaluation participants.** If a CREE approach has been incorporated into your evaluation design, consider how it could influence the steps needed to protect evaluation participants. For example, if using data collection methods that have participants share their opinions in front of one another, that approach changes what confidentiality and privacy entail. Each participant must commit to confidentiality and trust in one another other to carry through. Although not a data collection topic, this could also apply to community members who are on the evaluation team or an advisory board. Consider if the level of vulnerability or trust you are requesting is necessary and appropriate.

**Data collection protocol.** It is important to establish rapport before expecting participants to share their experience, story, or personal or private information. Build in time before data collection for data collectors to connect with evaluation participants on a personal yet professional level. Discuss aspects of privacy and confidentiality; if there are multiple participants in the room, also discuss considerations for creating a safe space for sharing.

### What is credible evidence?

Consider the following example of determining what “counts” as credible evidence. A program helping parents ensure their children develop in an enriching, stable home environment is conducting an outcome evaluation of their efforts. They know they need to measure financial outcomes for families. The evaluator initially recommends they measure whether parents were able to buy a vehicle for child transportation.

During discussions with former program participants, the evaluators hear that many parents don't want a car, per se. Instead, they felt a more accurate marker of stability was whether they were able to transport their family to places they needed to be, regardless of transportation modality (e.g., rideshare, bus, bicycling, car rental). If you measure only car ownership, your evaluation will miss the evidence of impacts the community wants to see through program participation. In that sense, the evaluation will not generate evidence that is of interest to, or credible to, the community.

### To learn more ...

- [Building Strong Evidence in Challenging Contexts: Alternatives to Traditional Randomized Controlled Trials](#) (Malin & Deterding, 2017)
- [Evaluation Plan Template](#) (CDC, n.d.)
- [Examining the Internal Validity and Statistical Precision of the Comparative Interrupted Time Series Design by Comparison With a Randomized Experiment](#) (St. Clair et al., 2014)
- [Manager's Guide to Evaluation](#) (Better Evaluation, n.d.-b)
- [Quantitative Research Designs: Experimental, Quasi-Experimental and Descriptive](#) (Drummond & Murphy-Reyes, 2018)
- [Quick Guide to Sampling, Sample Sizes, and Representation](#) (Washington State University, 2020)
- [Sampling and Evaluation: A Guide to Sampling for Program Impact Evaluation](#) (Lance & Hattori, 2016)

## References

- Atukpawu-Tipton, G., & Poes, M. (2020). *Rapid cycle evaluation at a glance* (OPRE Report 2020-152). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/rapid-cycle-evaluation-glance>
- Better Evaluation. (n.d.-a). *Compare results to the counterfactual*. <https://www.betterevaluation.org/en/rainbow-framework/understand-causes/compare-results-to-counterfactual>
- Better Evaluation. (n.d.-b). *Manager's guide to evaluation*. <https://www.betterevaluation.org/managers-guide>
- Blocklin, M., Hyra, A., Kean, E., & Porowski, A. (2019). *Building capacity to evaluate child welfare community collaborations to strengthen and preserve families (CWCC) grantee local evaluation plan and implementation plan templates*. Abt Associates. <https://omb.report/icr/201906-0970-001/doc/98252801.pdf>
- CDC (Centers for Disease Control and Prevention). (n.d.). *Evaluation plan template*. [https://www.cdc.gov/tb/programs/Evaluation/Guide/PDF/Evaluation\\_plan\\_template.pdf](https://www.cdc.gov/tb/programs/Evaluation/Guide/PDF/Evaluation_plan_template.pdf)
- CDC. (2011a). *Developing an effective evaluation plan: Setting the course for effective program evaluation*. <https://www.cdc.gov/obesity/downloads/cdc-evaluation-workbook-508.pdf>
- CDC. (2011b). *Introduction to program evaluation for public health programs: A self-study guide*. U.S. Department of Health and Human Services. <https://www.cdc.gov/evaluation/guide/CDC Eval Manual.pdf>
- Children's Bureau, ACF (Administration for Children and Families). (2019). *Evaluation plan development tip sheet*. U.S. Department of Health and Human Services. <https://www.acf.hhs.gov/cb/policy-guidance/im-19-04>
- Drummond, K. E., & Murphy-Reyes, A. (2018). *Quantitative research designs: Experimental, quasi-experimental, and descriptive*. In *Nutrition research: Concepts and applications*. [http://samples.jbpub.com/9781284101539/9781284101539\\_CH06\\_Drummond.pdf](http://samples.jbpub.com/9781284101539/9781284101539_CH06_Drummond.pdf)
- Lance, P., & Hattori, A. (2016). *Sampling and evaluation: A guide to sampling for program impact evaluation*. [https://www.researchgate.net/publication/311805268\\_Sampling\\_and\\_Evaluation\\_A\\_Guide\\_to\\_Sampling\\_for\\_Program\\_Impact\\_Evaluation/citations](https://www.researchgate.net/publication/311805268_Sampling_and_Evaluation_A_Guide_to_Sampling_for_Program_Impact_Evaluation/citations)
- Malin, J., & Deterding, N. (2017). *Building strong evidence in challenging contexts: Alternatives to traditional randomized controlled trials*. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. [https://www.acf.hhs.gov/sites/default/files/documents/opre/methodsmeetingsummary2016\\_final\\_112017\\_508.pdf](https://www.acf.hhs.gov/sites/default/files/documents/opre/methodsmeetingsummary2016_final_112017_508.pdf)

- Moore, K. A. (2008). *Quasi-experimental evaluation: Part 6 in a series on practical evaluation methods* (Publication 2008-040). Child Trends. [https://www.childtrends.org/wp-content/uploads/2008/01/Child\\_Trends-2008\\_01\\_16\\_Evaluation6.pdf](https://www.childtrends.org/wp-content/uploads/2008/01/Child_Trends-2008_01_16_Evaluation6.pdf)
- OPRE (Office of Planning, Research, and Evaluation). (2010). *The program manager's guide to evaluation, Second Edition*. U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/report/program-managers-guide-evaluation-second-edition>
- Smithsonian American Art Museum. (n.d.). *After the war: Blacks and the G.I. Bill*. <https://americanexperience.si.edu/wp-content/uploads/2015/02/After-the-War-Blacks-and-the-GI-Bill.pdf>
- St. Clair, T., Cook, T. D., & Hallberg, K. (2014). Examining the internal validity and statistical precision of the comparative interrupted time series design by comparison with a randomized experiment. *American Journal of Evaluation*, 35(3), 311–327. <https://doi.org/10.1177/1098214014527337>
- UCI Office of Research. (2022). *Privacy and confidentiality*. <https://research.uci.edu/human-research-protections/research-subjects/privacy-and-confidentiality/>
- Washington State University. (2020). *Quick guide to sampling, sample sizes, and representation*. <https://ace.wsu.edu/documents/2015/03/sample-size-and-represen>
- Wilson, S. J., Price, C. S., Kerns, S. E. U., Dastrup, S. D., & Brown, S. R. (2019). *Title IV-E Prevention Services Clearinghouse Handbook of Standards and Procedures, version 1.0* (OPRE Report 2019-56). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. [https://www.acf.hhs.gov/sites/default/files/documents/opre/psc\\_handbook\\_v1\\_final\\_508\\_compliant.pdf](https://www.acf.hhs.gov/sites/default/files/documents/opre/psc_handbook_v1_final_508_compliant.pdf)

# Chapter 6. Gather Credible Evidence

## What's Inside?



### What this chapter contains

- A discussion of identifying data sources and selecting measures
- Recommendations for developing data collection instruments and data collection procedures

### Who can use this chapter

- Evaluation team members completing an evaluation plan
- All staff engaged in or responsible for data collection

Click the links below to view the relevant section

[Practice Culturally Responsive and Equitable Evaluation Methods when Gathering Data](#)



[Introduction](#)



[Monitor Data Collection Activities](#)



[Identify the Best Data Sources](#)



[Protect Study Participants](#)



[Select or Develop Data Measures](#)



[Create Data Collection Procedures](#)



[Develop Instruments](#)



## Introduction

This chapter serves as a companion to [chapter 5](#), delving deeper into data collection, beginning with planning activities related to evidence gathering. The following are data collection decisions you will need to make:

- Which sources you will use to obtain data
- What data elements you need and what measures you will use to collect your data
- How you will structure your data collection instruments
- What procedures you will use to collect data
- How you will continue to monitor data collection and protect study participants

While these decisions are presented in sequential order, you will likely go back and forth developing answers to these questions according to data and data source availability, budget, and effort until you have a complete and feasible plan.

## Identify the Best Data Sources

While methodologies presented in chapter 5 describe techniques for how to collect data, you also need to determine where you will collect data and from whom or what. Two types of data sources are available: primary<sup>1</sup> and secondary.<sup>2</sup> An example of a primary data source is a survey of individuals participating in the program being evaluated for the specific purpose of collecting information to be used in an evaluation. An example of a secondary data source is program administrative data. Program administrative data, such as household composition, income, and program attendance, may be collected to inform program management, service eligibility, service tracking, and reporting needs, but the data can also provide important information on variables of interest for the evaluation.

### Making use of administrative data

Administrative data are often underused when building knowledge to inform social service policy and program design. Using such data can be a cost-effective way to answer policy-relevant evaluation questions by eliminating the need for gathering primary data. However, consider the logistics and time required to access administrative data and assess quality or usability. Potential considerations related to administrative data follow:

- Data quality, including amount of missing data and consistency in coding and timing of data collection
- Time lag between waves of data collection and data availability
- Approvals for using administrative data for evaluation purposes
- Developing data use scope and rules for a data sharing agreement
- Data transfer from data storage entity to evaluator
- Creating or clarifying information for a data dictionary or codebook (OPRE, 2019)

For additional guidance on the use of administrative data, see OPRE (n.d.) for sample reports that demonstrate and discuss the use of administrative data for evaluation supporting social services.

<sup>1</sup> A primary data source is an original data source; that is, the data are collected firsthand by the researcher for a specific research purpose or project (Salkind, 2010).

<sup>2</sup> Secondary data refer to data that have already been collected for some other purpose (Allen, 2017).



One strategy to identify data sources is to brainstorm all the possible data sources that could inform your evaluation questions and identify the benefits and limitations of each.

*Example:* You are interested in knowing whether children in foster care feel safe and comfortable during visits from birth parents at your family visitation center. Potential primary data sources include the participating children themselves, family visitation center staff, children's case managers, children's foster parents, and children's birth parents. Potential secondary data sources include visitors' logs tracking the frequency and setting of visitations from birth parents, court reports written by social workers, and documentation of the birth parents' adherence to court-ordered case plans (e.g., parenting classes, drug treatment services).

Each data source provides unique information about the children's experiences. While the evaluation team might first consider collecting data directly from the children, this might not be the best choice. Some children may be too young to engage in the data collection effort, or your IRB may determine such questions would be too emotionally difficult for children to answer. You may determine birth and foster parents are too busy to respond to a data collection request, and collecting data from children's case managers is best given your constraints.

Assess how well each potential source will provide accurate and high-quality data. Consider how complete and thorough the data are likely to be. For example, individuals generally provide more accurate information about themselves than can be obtained from a secondary source. However, program participants may not have access to or be able to recall the specific data needed. Another consideration is survey response. If you plan to conduct a survey of program participants, will they be particularly difficult to reach or track (such as people with unstable housing)? If so, you may decide an administrative data source that captures information about all or most evaluation participants is a better data source than a survey. A survey could suffer from low response (and therefore would not be representative based on your ability to contact evaluation participants).

### Metadata: data about your data

Did you know you can improve your evaluation skills and future evaluation efforts by learning about your current data collection efforts? Many online survey platforms offer access to survey metadata. Metadata is information about how your survey data collection effort unfolded. Depending on your survey platform, you may be able determine the average time it took people to complete your survey, the items they spent the most and least time on, which types of equipment survey respondents used, and information such as the time of day and location where they completed the survey. This information can help you identify items that may need to be revised, improve communications around expectations of time needed to complete survey, and help time reminders for survey completion.

While you need at least one data source for each data element you collect, you do not need to limit yourself to one. Using more than one data source to measure the same data element is called triangulation.<sup>3</sup> Triangulation can increase the quality and rigor of your study. For example, you could measure substance use through both a standardized self-report measure on a study and through an evaluation participant blood test. You could measure child behavior by interviewing both a child's primary parent and primary

<sup>3</sup> Triangulation refers to the use of multiple methods or data sources in qualitative research to develop a comprehensive understanding of a phenomenon. Triangulation also has been viewed as a qualitative research strategy to test validity through a convergence of information from different sources (Carter et al., 2014).

teacher. This would let you balance the relative advantages and weaknesses of each source and check for consistency across sources.

In deciding the best sources for information, the goal is to obtain the most accurate and complete data available within the cost, time, and burden constraints of your evaluation. Consider the following questions when evaluating data sources:

- Are useful secondary data sources available to address the evaluation question?
- Is the data source accurate (i.e., does it provide relevant and correct information about the concepts you intend to measure)?
- Is the data source reliable (i.e., does it yield consistent results)?
- Is the data source timely (i.e., is it available for analysis and interpretation when needed; does it cover the time period needed) and within the budget of an evaluation?
- Is the data source comprehensive and complete (i.e., does it provide sufficient detail or contextual information to be meaningfully interpreted; do you anticipate high levels of missing data)?
- Will collecting information from a particular data source pose an excessive burden (i.e., will it take much evaluation team time or budget to secure, or much participant time)?

Identifying the appropriate data source often involves considering tradeoffs. For example, you may obtain more in-depth information about services from interviews with program staff, but because of time and budget constraints, you choose to rely on case records or program logs. In this case, you must document the limitations of using a secondary data source when obtaining this information and discuss the implications when sharing results.

The rest of this chapter focuses on primary data collection. Many factors should be considered when using secondary data. For example, you must understand the limitations of a specific dataset, such as data from the American Community Survey or the need to develop an extraction tool for systematic data extraction from documents. Such elements are outside the scope of this Guide. See the “To learn more” textbox at the end of this chapter for secondary data resources.

### Identifying and prioritizing information needs

When making decisions about your information needs, you and your evaluation team may consider collecting certain types of interesting data. However, if the data are not directly related to at least one of your evaluation questions, you should resist the urge. Limiting data collection to “must know” is more time-efficient, cost-effective, and respectful of data providers’ time. The greater the time commitment required, the harder it will be to recruit participants. It is preferable to achieve a high response to your survey using a concise data collection instrument.

# Select or Develop Data Measures

## Data Elements

You will need a measure for each of your data elements. Data elements are ideas you want to capture information about. They are more than just outcome<sup>4</sup> domains. They include every concept you need to answer all your evaluation questions (implementation and outcome) and any that can provide context for your findings (such as gender, age, geographic location, or educational attainment). For each evaluation question, develop a list of all the data elements you need to measure, then develop measures to capture or operationalize the data. Below are two examples.

You are evaluating your mentoring program targeting adolescents who would be first-generation college students, and you want to know who is participating in the program. Your implementation evaluation question is, “What are the characteristics of adolescents who enrolled in the mentoring program?” As a first step, you will identify all the characteristics you want data about. Those characteristics might include age, gender, race/ethnicity, family/household structure, career/education aspirations, and relationships with peers. You will also define enrollment (e.g., completed all the registration paperwork). For each data element, you will identify exactly how you measure it. Some constructs, such as gender, might seem simple on the surface, but you will need to think through how you will define them. For example, will you ask for gender assigned at birth or current gender identity? Will you offer noncisgender options, such as nonbinary, transwoman, or transman? Concepts such as career aspirations are even more complicated to convert to measures. Luckily, many other evaluators and researchers have grappled with developing strong, valid measures for many concepts, as discussed below under data measures.

### Measurement terminology

Measurement can take numerous forms. A measure is all the information you will use to operationalize a data element. Sometimes you will use a single item (e.g., how old are you) as a measure. Other times you may use multiple items to collectively measure a data element. This typically happens when you are measuring a more complex data element. For example, you may develop several interview questions (items) to capture staff experience with program recruitment or use a scale to measure child development that has 20 or more items, such as the Ages and Stages Questionnaires (Squires & Bricker, 2009).

In another example, you are evaluating your local Head Start program and want to know the impact of the program on socioemotional learning. Your outcome evaluation question is, “To what extent do children at Alexandria’s Head Start demonstrate improvements in socioemotional learning after we institute a mindfulness program?” You need to determine what elements of socioemotional learning you expect will change as a result of the mindfulness program. The CASEL framework (CASEL, 2020) identifies five elements of socioemotional learning: self-awareness, self-management, responsible decision-making, relationship skills, and social awareness. You will need to visit your logic model and program materials to select specific types of socioemotional learning you will measure. You will then need to determine how you will measure self-awareness, for example, among preschool children.

<sup>4</sup> Outcomes are variables that are monitored during a study to document the impact of a given intervention or exposure (Ferreira & Patino, 2017).

## Data Measures

Measures are the tools you will use to assess each of your data elements precisely. All data elements must be operationalized. Operationalization means turning your more abstract concepts into measurable observations. If you want to measure earnings, for example, you could operationalize this element as participant-reported wages per hour, weekly take-home pay as documented on a pay stub, average salary for participants' work titles as documented by local labor market information, or the amount indicated on a participant's W-2. Much like data sources, you will need to assess the options for measuring each of your data elements based on quality, accessibility, feasibility, and precision.

Quality measures typically have the following characteristics (Blocklin et al., 2019):

- **They are reliable.** Good measures have demonstrated capabilities to collect information consistently. This means the same result can be achieved repeatedly using the same methods under the same circumstances. Typical reliability measures include internal consistency, test-retest reliability, or interrater reliability.
- **They are valid.** Good measures are those that truly capture the concept they intend to measure. Measurement validity has several different aspects, such as face, content, and criterion (Price et al., 2017).

Outcome measures should also exhibit the following qualities:

- **They are sensitive to change.** Good outcome measures can be expected to capture change over time and within the timeframe of data collection and expected sample size. For example, a measure such as Adverse Childhood Experiences (Felitti et al., 1998) is static; you cannot experience a reduction in the number of childhood experiences. Some measures may not be sensitive enough to capture change during your data collection timeframe. For example, a drug or alcohol use measure that asks about use over the previous 6 months would not be appropriate for a short-term program. If the time between pre- and posttest measurement is less than 6 months, the posttest measurement is still capturing behavior before the program started. Finally, some measures may not be sensitive enough to capture your expected level of change. If you expect your program to make small improvements in quality, a single item (e.g., that asks "is your marriage good" with yes or no as response options) may not capture enough nuance or detail. A more complex measure that captures multiple aspects of a marriage would be better suited to your needs.
- **They are not overaligned.** Good outcome measures are not too closely aligned with or tailored to the intervention being tested. Overalignment can occur when a measure is developed for a study of a specific program or when the intervention's creator develops the measure. Overaligned measures are a particular problem in impact analyses because they give the treatment group an advantage in appearing to have improved in that data element. For example, if you are providing a parenting course based on the book "1-2-3 Magic," you shouldn't measure parenting techniques by asking a question based directly on the book such as, "Do you follow the no talking, no emotion rule?" This kind of measure will skew your posttest comparisons. Only people who have completed the program will understand this measure; people who have enrolled but not yet started the program and people in your comparison group will be unable to answer this question accurately.

Selecting measures for each data element. Continuing with the three measurement examples above (gender, career aspirations, and self-awareness), you may find your program already captures some elements in an administrative database (such as gender). As discussed in section A, secondary data sources can provide a low-cost source of measures. However, ensure the data contained in those secondary sources truly reflect the concept you wish to measure. Review how the data are collected and coded to ensure you are meeting your information needs.

If you do not already have the data, you could develop your own measure to capture the data, or you could use a measure developed by another researcher. Available measures already in use typically offer many advantages.

Researchers often invest heavily in developing measures to ensure they will produce high-quality data. They are typically composed of scales, meaning they have many components or questions, which help to measure a complicated concept such as self-awareness. They have often undergone testing to confirm they are reliable and valid. Some measures also have corresponding instructions for use, coding, scoring, and even interpreting scores (such as values that indicate thresholds for high, medium, and low values).

However, no measure is suitable for every concept for every target population and type of data collection effort. For example, you may find self-awareness measures, but they are designed for elementary aged-children to self-report, and you need a measure a teacher can complete about a 3-year-old. Or, you may find available measures of career aspirations that have been validated with White, urban males, but they will likely not resonate with the Vietnamese American girls in rural Louisiana participating in your program. Table 6.1 shows considerations to help you compare the advantages and disadvantages of available versus new measures. You may end up with a combination of available and new measures.

### Developing qualitative measures

While some evaluations use existing measures for some or all their quantitative data, almost all evaluations develop at least some new measures to support their qualitative data collection. For example, you may write your own questions or items for interviews and focus groups. You should consult question development resources to provide advice and principles related to good question development, such as Jacob and Furgerson (2012) and Krueger (2002). Sources such as the Office of Management and Budget provide copies of interview and focus group protocols used in federal research and evaluation. You may be able to copy or adapt questions used in similar data collection efforts.

**Table 6.1. A Comparison of Existing Versus New Data Collection Measures**

Type of Measure Used	Advantages	Disadvantages
Using an available measure	<ul style="list-style-type: none"> <li>Often standardized</li> <li>Usually established as valid and reliable</li> <li>Can offer comparisons or benchmarks from other studies or surveys</li> </ul>	<ul style="list-style-type: none"> <li>Not always appropriate for all cultural or ethnic populations</li> <li>May not be useful for specific program</li> <li>May have use implications (e.g., cost, administration restrictions)</li> </ul>
Developing a new measure	<ul style="list-style-type: none"> <li>Can align more closely with program objectives</li> <li>Offers more flexibility to increase cultural sensitivity and relevance of measurement content and language</li> </ul>	<ul style="list-style-type: none"> <li>Might not be seen as valid and reliable</li> <li>Can be less psychometrically sound (i.e., less likely to measure what it is meant to with reliability and validity)</li> <li>Good measurement development is difficult (time and resource investment)</li> </ul>

### Criteria for selecting data measures

Use the following questions to assess your measurement options:

- Does the measure address a program domain (e.g., for a parenting training, choosing a measure that captures change in knowledge or skills)?
- Does it have acceptable values of validity and reliability (e.g., consistency measures, test-retest statistics, face validity)?
- Is the measure appropriate for participants with regard to age or developmental level, language, and ease of use?
- Does the measure respect and reflect participants' cultural backgrounds (e.g., are definitions, concepts, and items in the measure relevant to the participants' community and experience)?
- Have you pilot tested the measure to uncover any difficulties and ensure it can be completed in a reasonable timeframe?

## Develop Instruments

After you've identified all your measures, data sources, and methodologies, you will likely need to develop one or more instruments to collect primary data. The most common instruments are interview protocols, focus group protocols, and surveys (which can be self-administered electronically, on paper, or through a data collector asking questions and recording answers). Important elements of instruments follow:

- **Introductions and consent.** Respondents should understand the purpose of the evaluation they are participating in, why their participation is being requested, what their participation entails, and how their data will be used (see [Protect Study Participants](#) below for more information).
- **Ground rules.** You will likely develop instructions or expectations for each type of data collection. Focus groups need to operate using common understanding of participation (developed by either the evaluator or by group consensus), such as confidentiality, respect, and ensuring all participants have space to speak. Interviewees should receive information about how their names will be recorded (if not in consent form), how you will handle recordings, and how interviewees can indicate they do not want to answer a question. Survey respondents need to know how to mark their responses, where to pose questions if they have problems, and how to save and submit their responses.
- **Item flow.** You will need to order all your measures (single or multiple item) in a logical fashion. You might start with easier, more factual questions, and save more personal questions for later in the survey or interview. You may not want to leave vital measures until the end of any instrument in case you run out of interview time or the survey respondent quits the survey before it is over. You will need to carefully plan any skip logic patterns (where, depending on the answer to one question, your respondent is sent to a different subsequent question). Skip logic is often used to make a survey easier for respondents because they skip questions that do not apply to them (e.g., there is no need to ask a respondent the ages of their children if they indicate they have no children). Additional attention and time might be needed to ensure surveys are coded correctly and to clean and prepare a dataset with skipped items.

- **Thank-you, closing, next steps.** Evaluation participants give their precious time to the evaluation, and you should thank them for that. Before you end data collection, you may also consider giving evaluation team contact information in case participants have questions later. This offers a way for participants to follow the evaluation if they want to read public reports, and if appropriate, for the evaluation team to provide referrals or hotlines to address potential traumas or issues that sensitive evaluation questions may have revealed.

After your instruments are compiled, test them before data collection. Your testing should ensure your instrument is visually correct (on screens or on paper), automatic skip patterns work or paper instructions are clear, response options are coded correctly (e.g., an item that directs respondents to “choose all” doesn’t limit them to one response), paper instructions are clear, and electronic data are saved in your system after a survey is submitted. For all instruments, you should practice fielding—or pilot testing—them to see how long it takes to collect the data, make sure the questions flow, and if appropriate, the data collector is comfortable with the instrument.

To the extent possible, data collection staff should conduct the pilot testing. Ask them to take notes and make comments on the process of administering or using each instrument. Then, review these notes and comments to determine whether changes to the instruments or data collection procedures are needed. As part of pilot testing, instruments should be reviewed to assess the number of incomplete answers, unlikely answers, comments on items that may be included in the margins, or other indicators that suggest revisions are necessary.

## Create Data Collection Procedures

Important components of evaluation rigor are consistency and accuracy in data collection. Ideally, you will develop written data collection procedures and use those procedures to train data collection staff, monitor data collection efforts, and describe your methodology in final reports and articles. Note that some evaluation funders may have specific requirements for data security which programs will need to follow.

### Be mindful

Understand you may be asking personal, uncomfortable, and potentially traumatizing questions of your evaluation participants. Data collectors should be trained to understand how evaluation participants may experience data collection and understand how to respond to participants’ concerns or reactions. Determine whether you should develop a resource list to share with evaluation participants at the end of data collection. A resource list should provide local or national organizations and hotlines that can respond to issues that surfaced during data collection. For example, a federally funded evaluation on romantic relationships probed on intimate partner violence, and all instruments closed with information about the National Domestic Violence Hotline.

Data collection procedures should consider the following:

- **Who is in charge of and participates in each data collection effort.** You may name staff roles (e.g., case manager) or staff names and their role on each data collection effort (e.g., Dr. Hernandez will lead 10 focus groups). You may also identify the skills, knowledge, training, and experience individuals engaged with data collection should have. For example, you may require all data collectors to receive training about trauma and retraumatization and all evaluation team members to agree to evaluation participant protections such as signing nondisclosure agreements or holding human subjects research training certificates.
- **When each data collection effort occurs.** This includes dates (e.g., staff interviews occur annually in May) and time in relation to evaluation participation (e.g., baseline data are collected within 2 weeks of study enrollment, and posttests take place within 2 weeks after the 10-week program ends).
- **Where and how each data collection effort occurs.** Data collectors need checklists to ensure data collection occurs as intended and consistently across data collectors. Procedures can include instructions, for example: Intake forms are completed on a tablet by evaluation participants alone in a private room. Evaluation staff wait outside to answer questions if needed. You should also address the use of any incentives such as gift cards and when you provide them (e.g., after consent, after completion of all forms) and any other requirements, such as signed acknowledgement forms.
- **How online data collection sites are accessed.** You need to keep your online data collection and storage sites secure. Research the security features of any online survey or data collection system you might use. Ask about protections they have in place, certifications of their systems, and their other clients. Employ strong access controls, such as two-factor authentications, strong and frequently changed passwords, and access limited to only necessary evaluation team members. Consider downloading the data at the end of data collection to analyze on a physical computer and deleting the copies online.
- **How data are handled and protected.** A clear chain of control for transfer of data is important. Strong data security procedures are also needed to reduce the likelihood of breaches or identification of evaluation participants.
- **How data are stored.** Both paper instruments and electronic files need to be kept as secure as possible. Security related to paper forms could include storage in a locked hotel room safe, secure shipping to the office, and ultimately storage in a locked file folder. Electronic files should be stored on restricted access folders that only necessary evaluation team members can access. Develop procedures to keep original and working copies of electronic files in case of accidental deletions or corruptions.
- **How evaluation participant privacy is maintained.** Many evaluations ask sensitive or personal questions related to income, involvement with child welfare, parent-child relationships, or experience with violence. Evaluation designs may require two or more data collection waves, with information collected from the same participants. Evaluations need to connect data across waves to the same person but keep the participant's identity separate from their sensitive data.

Most evaluations address this problem by using unique identifiers. Each evaluation participant is assigned an identification (ID) number. Only the ID number is included on files containing responses and other data. Never include information that can be used to identify a person in the same file that contains



the data. A separate crosswalk file should be maintained with contact and personal identification information so you can link a participant to an ID number as needed. The datafile and crosswalk should be kept in separate locations on a cloud or computer drive with different access restrictions. Individuals involved in the data collection must respect participant privacy by using this information only to track participation (not to review an individual's data). Data collectors must not discuss anything they have learned about an individual during the data collection.

- **What to do if things go wrong.** It is almost inevitable your evaluation will encounter a problem at some point. Good data collection procedures plan for common challenges (e.g., completed intake form goes missing, a data collector quits) and provide guidance to the evaluation team on how to handle problems and whom to contact (e.g., evaluation team leadership). You may also want to discuss how to handle rarer but challenging problems; for example, a situation where a data collector feels unsafe or is threatened.

### Data security principles

Several frameworks define data security, including the CIA Triad (Samonas & Coss, 2014), the General Data Protection Regulation (Tamburri, 2020), and the Precision Medicine Initiative (PMI) Data Security Policy Framework (All of Us Research Program, n.d.). The PMI framework is based on work conducted by the National Institute for Standards and Technology, offering five principles:

- **Identify:** Develop a data security plan, use risk management approaches to develop protection decisions, have your plans reviewed by independent parties, be transparent in your approaches and plans.
- **Protect:** Employ access control measures, conduct awareness and training efforts, execute data security plans, maintain data infrastructures.
- **Detect:** Audit events and logs, employ a detection and alert system, share information threats with similar organizations, report anomalies to organizational leadership for future prevention planning.
- **Respond:** Develop and employ an incident response plan, test the incidence response system regularly, notify individuals when their data were part of a breach, assign a staff member to be the accountability point of contact for breaches.
- **Recover:** Establish and implement incident and breach recovery plans, communicate when the data platform is resecured, document lessons learned.

Everyone engaged in collecting evaluation data must be trained in data collection procedures. Training should include the following:

- An item-by-item review of each of the instruments to be used in data collection, including a discussion of the meaning of each item, why it was included in the instrument, and how it is to be completed
- A review of all instructions on administering or using the instruments, including instructions to the respondents
- A discussion of potential problems that may arise in administering the instrument, including procedures for resolving the problems

- A practice session when data collection staff administer the instrument to one another, use it to extract information from existing case records or program logs, or complete it themselves if it is a written questionnaire
- A discussion of respondent confidentiality, including administering an informed consent form, answering respondents' questions about confidentiality, keeping completed instruments in a safe place, and procedures for submitting instruments to the appropriate person
- A discussion of the need for frequent reviews and checks of the data and for meetings of data collectors to ensure consistent data collection

Ideally, you will develop a training guide or manual for data collection staff to read. You would then conduct a training or a series of trainings with data collection staff with ample time for questions and discussions. To anticipate staff turnover, record the trainings so new staff can be onboarded quickly. If your data collection period is lengthy (e.g., over 6 months), you should schedule regular refresher trainings. You should also update the manual and hold additional trainings if data collection procedures change, or if you move into a new phase of data collection (e.g., begin a follow-up data collection phase that involves a new mode of data collection, such as a phone survey).

It is useful to develop a manual that describes precisely what is expected in the information collection process. This will be a handy reference for data collection staff and useful for new staff hired after the initial evaluation training occurred.

## Protect Study Participants

[Chapter 5](#) provides an overview of the IRB process and discusses the importance of obtaining IRB approval prior to starting any data collection. All research and evaluation efforts must take adequate steps to protect the privacy and confidentiality of all human data. This includes working with available or secondary data and any primary data collection efforts.

### IRB protections for “vulnerable populations”

The Code of Federal Regulations dictates the way IRBs work. The code requires that the IRB pay special attention to evaluations conducted with certain kinds of people: pregnant women, fetuses, and newborns; children; and incarcerated people. While you should not avoid conducting an evaluation with these populations, you will need to spend additional time thinking through how to ensure their safety in your study. For example, a study of a certain medication or therapy may have different (negative) effects on fetuses, children may not be able to understand the full ramifications of participating in an evaluation, and people who are incarcerated may be vulnerable to coercion based on their imprisonment. Your IRB may extend the spirit of protections for vulnerable populations to other groups such as non-Native English speakers or individuals with cognitive impairments. If your evaluation will collect data from any of these populations, you must communicate closely with your IRB to ensure your evaluation adequately protects all participants. For more information, see Shivayogi (2013).

Ensure all your data collection efforts align with the procedures your IRB approved. Make sure any deviations, including adverse effects or data breaches, are reported to your IRB within the approved timeframe. Adverse effects include a negative reaction an evaluation participant has to the evaluation, such as crying during an interview. A data breach occurs whenever an unauthorized individual gets or could have gotten access to data. Common data breaches occur when someone uses a nonsecure method of data transfer or leaves data unattended, or a computer with data stored on it is lost.

## Monitor Data Collection Activities

Throughout the data collection timeframe, monitor and examine data collection efforts. Data collection needs to proceed according to your plans and be consistent and accurate over time; across data collectors; and across people, sites, and treatment and comparison groups. Failure to administer data collection instruments correctly or consistently can significantly degrade the quality of your data and in some cases make it unusable.

Data collection monitoring activities can include the following:

**Establish a routine and timeframe for submitting completed instruments.** This information may be included in your data collection manual. Data collectors should submit their completed instruments to the appropriate receiver as soon as possible (e.g., upload completed observation tools to the required folder each night of a site visit). A process should be in place to quickly check completed instruments for accuracy and completeness. Ideally, errors should be identified in time to resolve them (e.g., call the evaluation participant back and complete missing questions). Many data collection efforts use software to create automatic checks, such as pausing a survey if a required item is not answered or flagging responses that seem likely to be incorrect (e.g., a birthday in the 1800s). You may need to retrain some or all data collectors if some mistakes occur frequently.

**Conduct random observations of the data collection process.** A member of the evaluation team may be assigned to observe the data collection process at various times during the evaluation. This person, for example, may sit in on an interview session to ensure all procedures are being conducted correctly.

**Conduct checks of data coding and privacy procedures.** If you use paper-and-pen data collection forms, someone will need to enter the data into a datafile. Develop a data coding quality check plan where another individual spot checks the datafiles against the original forms to ensure all data are entered and entered correctly. Monitor procedures to assign evaluation participants unique identifiers and the construction of separate contact and evaluation datafiles.

**Conduct quality checks with respondents.** As an additional quality control measure, it can be helpful to assign someone on the evaluation team to routinely check with a sample of respondents to determine whether the instruments were administered in the expected manner. This individual may ask respondents if they were given the informed consent form to sign and if it was explained to them, where they were interviewed, whether their questions about the interview were answered, and whether they felt the attitude or demeanor of the interviewer was appropriate.

### Sampling error: A threat to the evaluation

Chapter 5 discusses internal and external validity and how they affect the credibility of an evaluation. Sampling errors are one way an evaluation can have reduced internal validity.

### Encourage staff to view the evaluation as an important part of the

**program.** If program staff are given responsibility for data collection, they will need your support for this activity. Their priority is providing services or training to participants, and collecting evaluation information may be a secondary goal. You will need to emphasize to your staff that the evaluation is part of the program, and evaluation information can help them improve their services or training to participants. The best way to demonstrate the value of evaluation data is to provide concrete examples, such as a mock data report (at project beginning) along with regular, complete reports and periodic findings presentations.

**Monitor completion rates.** Aim to collect data from as many members of your sample as possible. Missing data can affect the accuracy of your findings across the whole population, or eligible units of observation, you are studying (see textbox). Monitor overall completion rates compared with your goal and assess completion rates by subgroups. Relevant analyses could compare completed intake forms by intake coordinators, survey completion rates by community, and interview rates by treatment and comparison groups. If you find differences in completion rates, work to address them. For example, retrain intake coordinators or spend more time and effort to obtain responses from underresponding community members.

### Monitor for missing data

Track your response rates and missing data rates. Overall, the higher the rate of missing information or nonresponse, the less confident you can be that the data you collect represents responses you would have gotten from all the eligible entities. For example, if people who didn't find employment after completing your training program were less likely to complete the postprogram survey, your evaluation may find artificially high estimates of program outcomes on employment. In this way, missing data affects the validity and generalizability of your evaluation. See [chapter 7's section](#) on assessing data quality for more information.

After evaluation information is collected, you can begin to analyze it. To increase the benefits of the evaluation to you, program staff, and program participants, this process should be ongoing or occur at specified intervals during the evaluation. Information on the procedures for analyzing and interpreting evaluation information are discussed in the following [chapter](#).

## Practice Culturally Responsive and Equitable Evaluation When Gathering Data

Sources for data, either primary or secondary, can have inherent bias. Whether you are selecting or developing instruments to use with program participants or compiling already available administrative data, use a CREE approach to think through how bias might be built into the data. One type of bias is the selection of those who will have the opportunity to provide data. Think about those you might reach with a mailed survey versus by telephone or online. How does the time of day and day of the week of a focus group exclude some and give advantage to others? For secondary data, it might be difficult to identify biases because the instruments and data collection methods were decided in the past by others. Depending on the data source, a description of evaluation methods might be available for you to review.

Community members, members of your program's target population, and other advocates for human service program recipients can help improve your data and data collection efforts. Community member input can benefit data instrument decisions and identify concerns respondents may have about participating in data collection. Consider the following recommendations:

- Select measures that are culturally appropriate and credible. [Chapter 5](#) discussed ensuring you select appropriate outcomes. Measures selected should also appropriately operationalize those outcomes. Community members should vet measurement options and provide advice related to their selection. For example, you may learn that one positive parenting measurement uses more culturally relevant language than another.

### Community engagement in Tribal evaluation initiatives

In addition to following CREE best practices, some evaluations may be required to work alongside community members throughout the data collection process. Evaluations subject to multijurisdictional oversight, such as Tribal governments, need to abide by all Tribal regulations and recommendations. Tribal members are protected by their own sovereign nation's oversight. Evaluators should recognize that Native American and Indigenous populations have a history of being harmed by research and evaluation. The National Congress of American Indians recommends that evaluations that include tribal members adhere to the following principles:

1. Indigenous knowledge is valid and should be valued.
2. Research is not culturally neutral.
3. Responsible stewardship includes the task of learning how to interpret and understand data and research.
4. Tribes must exercise sovereignty when conducting research and managing data.
5. Research must benefit Native people.

For more information, see National Congress of American Indians Policy Research Center (2009).

- Select measures that have been “normed” and tested with members of your program’s target population. Measures developed for and tested with people like your evaluation participants are much more likely to accurately capture the concept they intend to measure.
- Work closely with many members of your community and target population if you decide to develop your own new measures or adapt current measures. Consider consulting with a subject matter expert who is also a member of the target population to help develop and test the new measure.
- Ask members of your target population to review and help pilot test your data collection instruments. Ensure the measures and response options are relevant, the order of items is sensible, the instructions are clear, and the language is accessible and at the appropriate reading level. If the instrument is online, make sure it works on mobile devices, tablets, and computers. Talk with community members to get a sense of accessibility issues such as Wi-Fi, public computers, and how much data a respondent might need to complete the instrument. Test sending the instrument to various email platforms to determine if the transmission is flagged as spam and ask community members to help develop email language more likely to be seen as authentic and not a scam.

### Culturally and linguistically responsive data collection

To understand the effectiveness of federally funded nutrition assistance programs in Puerto Rico, a study team developed a household survey on food security, economic well-being, and coping strategies following natural disasters such as hurricanes. The study team used a standard measure of food security, validated in Spanish, that a Puerto Rican member of the study’s technical expert panel vetted. The study team needed to develop new items or revise others that had been used with other populations to capture data on other topics such as shopping habits. To ensure these items were clear and response choices were appropriate, the team pretested the survey instrument in Puerto Rico. A local, trusted business helped recruit pretest participants and advised the study team on the incentive amount. Two Puerto Rican bilingual evaluators translated the survey into Spanish and back into English (back translation is a method to assess the quality of the translation). During the pretest, participants read each question aloud and recited their thought process so the interviewer could discern any confusion or hesitation. Interviewers also asked pretest participants about likely reasons a sample member might not respond. Based on the responses, the survey materials were revised to emphasize that all responses would be kept private and would not affect the respondent’s benefits in any way (Wilson, 2021).

- Consider including community members in developing your data collection training. Community members can help elevate concerns evaluation participants may have, highlight important cultural practices to build trust and show respect, and give other relevant advice.
- Hire members of your evaluation’s target community to be data collectors. Evaluation participants may be more likely to consent to evaluation data collection and provide more thorough or accurate information if they feel they are speaking with someone who understands and values them. Community data collectors are building important evaluation skills and can use your employment as job and education references.

- Consider compensating community members for their valuable time spent participating in evaluation planning and the evaluation participants' effort in providing data. If you can't provide financial remuneration, think about covering transportation costs, providing food and meals, and asking community members what other contributions they would find valuable.

### To learn more ...

- [Best Practices in Creating and Adapting Quality Rating and Improvement System \(QRIS\) Rating Scales](#) (Burchanai, Tarullo, & Zaslow, 2016)
- [Enhancing Rigor, Relevance, and Equity in Research and Evaluation Through Community Engagement](#) (OPRE, 2021)
- [Supporting the Use of Administrative Data in Early Care and Education Research: Resource Series](#) (OPRE, 2019)
- [Types of Data Used for Impact Evaluation](#) (Courtney, 2021)

## References

- All of Us Research Program. (n.d.). *Achieving the principles through a precision medicine initiative data security policy framework*. U.S. Department of Health and Human Services, National Institutes of Health. <https://allofus.nih.gov/protecting-data-and-privacy/precision-medicine-initiative-data-security-policy-principles-and-framework-overview/achieving-principles-through-precision-medicine-initiative-data-security-policy-framework>
- Allen, M. (2017). *Secondary data*. SAGE Encyclopedia of Communication Research Methods. <https://dx.doi.org/10.4135/9781483381411.n557>
- Blocklin, M., Hyra, A., Kean, E., & Porowski, A. (2019). *Building capacity to evaluate child welfare community collaborations to strengthen and preserve families (CWCC) grantee local evaluation plan and implementation plan templates*. <https://omb.report/icr/201906-0970-001/doc/98252801.pdf>
- Burchinal, M., Tarullo, L. & Zaslow, M. (2016). *Best practices in creating and adapting Quality Rating and Improvement System (QRIS) rating scales* (OPRE Research Brief 2016-25). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. [https://www.acf.hhs.gov/sites/default/files/documents/opre/ccepra\\_qris\\_531\\_508compliant\\_66\\_b508.pdf](https://www.acf.hhs.gov/sites/default/files/documents/opre/ccepra_qris_531_508compliant_66_b508.pdf)
- Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum*, 41(5), 545–547. <https://pubmed.ncbi.nlm.nih.gov/25158659/#:~:text=Triangulation%20refers%20to%20the%20use,of%20information%20from%20different%20sources>
- CASEL. (2020). *CASEL's SEL framework*. <https://casel.org/casel-sel-framework-11-2020/>
- Courtney, M. (2021). *Types of data used for impact evaluation*. U.S. Department of Health and Human Services, Children's Bureau. <https://www.acf.hhs.gov/sites/default/files/documents/opre/Types%20of%20Data%20Used%20for%20Impact%20Evaluation-oct-2021.pdf>
- Felitti, V. J., Anda, R. F., Nordenberg, D., Edwards, V., Koss, M., & Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. *American Journal of Preventive Medicine*, 14(4), P245–258. [https://www.ajpmonline.org/article/S0749-3797\(98\)00017-8/fulltext](https://www.ajpmonline.org/article/S0749-3797(98)00017-8/fulltext)
- Ferreira, J. C., & Patino, C. M. (2017). *Types of outcomes in clinical research*. National Library of Medicine. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5790671/>
- Jacob, S. A., & Furgerson, S. P. (2012). Writing interview protocols and conducting interviews: Tips for students new to the field of qualitative research. *The Qualitative Report*, 17, 1–10. <https://files.eric.ed.gov/fulltext/EJ990034.pdf>
- Krueger, R. (2002). *Designing and conducting focus group interviews*. <https://www.eiu.edu/ihec/Krueger-FocusGroupInterviews.pdf>



- National Congress of American Indians Policy Research Center. (2009). *Research that benefits Native People: A guide for Tribal Leaders*. <https://www.ncai.org/policy-research-center/research-data/NCAIModule1.pdf>
- OPRE (Office of Planning, Research, and Evaluation). (n.d.). *OPRE publications, administrative data, and research*. U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/project/opre-publications-administrative-data-and-research>
- OPRE. (2019). *Supporting the use of administrative data in early care and education research: Resource series*. U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/report/supporting-use-administrative-data-early-care-and-education-research-resource-series>
- OPRE. (2021). *Enhancing rigor, relevance, and equity in research and evaluation through community engagement*. U.S. Department of Health and Human Services, Administration for Children and Families. <https://opremethodsmeeting.org/meetings/2021/>
- Price, P. C., Jhangiani, R. S., Chiang, I-C. A., Leighton, D. C., & Cuttler, C. (2017). *Research methods in psychology* (3rd ed). <https://opentext.wsu.edu/carriecuttler/>
- Ryan, K., Gannon-Slater, N., & Culbertson, M. J. (2012). Improving survey methods with cognitive interviews in small- and medium-scale evaluations. *American Journal of Evaluation*, 33(3) 414–430. <https://journals.sagepub.com/doi/full/10.1177/1098214012441499>
- Salkind, N. (2010). *Primary data source*. Encyclopedia of Research Design. <https://dx.doi.org/10.4135/9781412961288.n333>
- Samonas, S., & Coss, D. (2014). The CIA strikes back: Redefining confidentiality, integrity and availability in security. *Journal of Information System Security*, 10(3). <http://www.proso.com/dl/Samonas.pdf>
- Shivayogi, P. (2013). Vulnerable population and methods for their safeguard. *National Library of Medicine*, 4(1), 53–57. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3601707/>
- Squires, J., & Bricker, D. (2009). *Ages & stages questionnaires* (3rd ed.) <https://products.brookespublishing.com/Ages-Stages-Questionnaires-Third-Edition-ASQ-3-P569.aspx>
- Tamburri, D. (2020). Design principles for the General Data Protection Regulation (GDPR): A formal concept analysis and its evaluation. *Information Systems*, 91. <https://doi.org/10.1016/j.is.2019.101469>
- Wilson, C. (2021). *Food security status and well-being of Nutrition Assistance Program (NAP) participants in Puerto Rico: Revised pretest findings and recommendations* [Insight Policy Research internal memo]. September 21.

# Chapter 7. Analyze Data

## What's Inside?



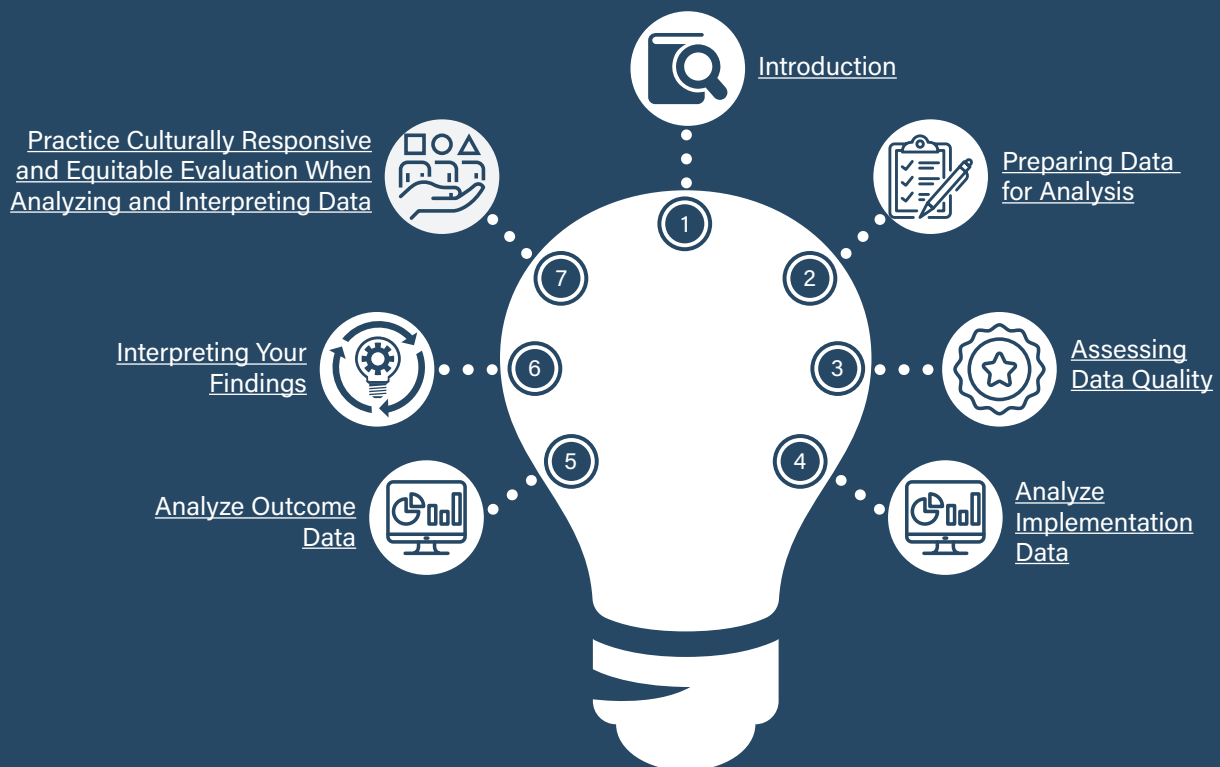
### What this chapter contains

- An introduction to analyzing collected data and interpreting what the data (findings) mean
- A description of common procedures for preparing data for analysis
- Recommendations for assessing data quality
- A discussion of procedures for analyzing implementation and outcome data
- Examples of ways to apply culturally responsive and equitable principles when analyzing and interpreting evaluation data

### Who can use this chapter

- Program managers preparing to analyze and interpret evaluation data

Click the links below to view the relevant section



## Introduction

After you have gathered evaluation data, the next stage is to analyze those data and interpret what the findings mean. Although someone experienced in working with quantitative and/or qualitative data should lead data analysis, all members of the evaluation team should be engaged in making decisions throughout this stage. You will need to make decisions related to how your evaluation team will—

- Prepare the data for analysis (data cleaning and transformation).
- Conduct an initial assessment of data quality.
- Conduct analyses needed to answer evaluation questions.
- Potentially, conduct additional exploratory analyses.
- Discuss initial findings with community representatives to ensure their perspectives inform your interpretations.
- Interpret findings and make meaning.

While this chapter is not a manual for conducting statistical tests to analyze evaluation data, it provides basic information about approaches to analyzing evaluation data to help you understand and participate more fully in this process. Many ways to analyze and interpret evaluation data are available, and the methods discussed in this chapter are not the only possibilities. Whatever methods your evaluation team decides on, be sure your evaluation questions guide your analysis. The following evaluation questions are discussed throughout this manual:

- **Has the program been successful in attaining the anticipated implementation objectives?** If not, why not? What types of barriers impeded implementation objectives, or what factors facilitated their attainment?
- **Has the program been successful in attaining the anticipated outcomes?** If not, why not? What types of barriers impeded outcome objectives, or what factors facilitated their attainment?

The following sections discuss various approaches to analyzing data to answer both types of questions. The chapter concludes with examples of how to apply culturally responsive and equitable principles when analyzing and interpreting evaluation data.

## Preparing Data for Analysis

Data are not ready to be “crunched” or analyzed right after collection. Both qualitative and quantitative data need to be prepared.

**Clean the data.** Cleaning data involves examining your data to ensure accuracy and completeness. Accurate data do not have any remaining incorrect or erroneous values for any element. For example, if you

have seven racial/ethnic codes with values of 1–7, you should not have any individuals with an out-of-range value of 12 or -6. You might explore whether some variables have values that seem improbable, such as birth dates in the 19th century or a parent indicating they have 26 children.

**Ensure the type of data is correct.** For example, numeric data elements should not have any character responses, such as a response of “&” when you expect a value between 0 and 3. Invalid or unlikely values may indicate an error was made in entering the data. You and your team will need to decide which values seem invalid and how you will handle those data, such as recoding erroneous values as missing.

**Assess data accuracy.** Develop an approach to assess accuracy of the data. For qualitative data, you could share interview notes back with the interviewee to see whether you captured the conversation correctly; compare notes with transcripts; or if you had more than one interviewer, have them compare and reconcile notes. For quantitative data, you might verify agreement on responses to related items. For example, if a survey respondent indicates in one response that they live with their spouse but in another says their household size is one, one response is likely incorrect.

**Transform the data.** Another common preparation step is to transform data. This task is particularly common with quantitative data, and it will depend on your evaluation questions and approach to the work. One transformation is the calculation of a scale score. Standardized instruments often come with guidance about how to convert responses into a numerical score. You may also want to collapse categories; for

example, recoding the number of children to 0, 1, 2–3, and 4 or more (if you have indications those differences matter).

**Link the data.** Often, evaluations collect many waves of data that must be combined for analyses. Studies need to match respondents across datafiles using identifying information. Ideally, each individual has a unique identifier value to combine datasets. If you don't have an identification number, or the number isn't applied accurately across all cases, you might be able to use techniques such as probabilistic matching to make likely merges of same-respondent data across different files (see Asher et al., 2020, for an introduction). If you merge data to create an analytic file, examine your new datafile to ensure accurate merging.

### Analysis is an ongoing activity

For many reasons, you should not wait until 6 months before your final report is due to start conducting analysis. Interim analysis conducted regularly throughout the course of your program evaluation (1) helps identify any major challenges to the data and facilitates midcourse corrections, (2) informs continuous quality improvement efforts by identifying challenges with program implementation, and (3) gives a preview when changes in outcomes are generally proceeding the way you expect (even if your interim samples are too small to detect statistical significance). Consider conducting annual analysis efforts.

### Research and evaluation clearinghouses

Clearinghouses are repositories of high-quality program evaluations that try to answer questions of effectiveness. The federal government has funded numerous evidence clearinghouses such as the What Works Clearinghouse (Department of Education, n.d.), CLEAR (U.S. Department of Labor, n.d.), the Prevention Services Clearinghouse (ACF [Administration for Children and Families], n.d.-a) and the Home Visiting Evidence of Effectiveness project (ACF, n.d.-b). Most of these clearinghouses examine data quality issues such as study attrition in their review of impact evaluations. It is beneficial to use information about clearinghouse standards to guide the review of your impact evaluation (quasi-experimental or experimental) data.

**Transcribe the data.** In the case of qualitative approaches, you may need to transcribe and clean interview or focus group data. Transcription is the process of converting speech (either live or recorded) into a written or electronic text document to facilitate coding of qualitative data. While transcribing may appear to be a straightforward technical task, the process of transcription may differ according to its end use (see Bailey, 2008, for more guidance).

### Memoing

Memoing refers to informal record-keeping by qualitative researchers that references ideas, hypotheses, research literature, or observations about evaluation questions, design, methods, and theory as they arise throughout the evaluation process (see Satterlund, n.d.). Memoing helps you keep track of your thoughts and support evaluation team communication.

Decisions about the level of detail (e.g., whether to transcribe or omit nonverbal communication) should be discussed in advance. For example, if you plan to use transcripts to identify quotes and sound bites for supporting evidence, you may not need the same level of detail as for those that will be systematically reviewed, grouped into themes, and analyzed for content.

When transcripts are completed, the evaluation team should engage in data familiarization,<sup>1</sup> a common practice in all forms of qualitative data analysis. Researchers may begin identifying and notating features of the data that are potentially relevant to the evaluation questions, a helpful step in preparing for the next phase, the coding process.

### Types of transcripts

Several kinds of transcripts can be used in qualitative research, depending on the methodology used and purpose of an evaluation:

- *Verbatim transcripts* are the most common type of transcript used for thematic analysis; they aim to capture every word and nonverbal auditory communication (e.g., sighs, laughing, stutters, pauses).
- *Jeffersonian transcripts* are designed to capture what was said and the way it was said, using symbols to represent sound, pace, intonation, and interaction in the conversation.
- *Gisted transcripts* are less detailed than verbatim or Jeffersonian transcripts; they aim to capture the essence (i.e., "the gist") of an audiofile or videofile's content.
- *Multimodal transcripts* are commonly used when analyzing video recordings of interviews, focus groups, or other forms of social interaction. All nonverbal forms of communication (e.g., gaze, head shake, gestures, eye rolls, posture) and verbatim communication are transcribed. Other variables that may influence participant responses are noted (e.g., cell phone ringing) to produce a highly comprehensive set of analyzable data.

<sup>1</sup> Data familiarization is the process of repeatedly reading or listening to each item of data to develop a deeper understanding of participants' perspectives (Statistics Solution, n.d.).

## Assessing Data Quality

Before you invest great effort analyzing data, understand the quality of data that will generate your findings. You cannot create better data through analysis.

One of the most important markers of data quality is the extent of missing data. Calculate the response rate for data collection efforts, such as the percentage of evaluation participants who completed a customer satisfaction form over the total number of evaluation participants asked to complete the form. Calculate the rate of missing data for each item also. For example, your overall survey may have had a high response rate, but one item about the amount of meat eaten per week had much missing data.

**Study attrition.** You must also calculate study attrition.<sup>2</sup> For example, if 100 people completed the pretest, how many of them completed the posttest? If your evaluation used a comparison group, calculate two types of attrition for each data collection wave: overall attrition, for the whole sample, and differential attrition, the difference in attrition rates from the treatment versus comparison group. High levels of differential attrition can indicate your two groups are too different to produce reliable comparisons that could be attributed to the program being evaluated. Attrition is calculated on both the sample and the measure level. In other words, report the number and percentage of your sample that provided any data at each wave of data collection. Also report the number and percentage of your sample that provided data for each outcome measure at each wave of data collection. This means that measures with more missing data (as described above) have higher attrition rates than measures from the same instrument and data collection wave with fewer missing responses.

**Measurement reliability.** You may want to assess the reliability of some of your measures (see [chapter 6](#) for more information). If you have developed your own scales or are using a scale not yet validated with your population, check whether it seems to have captured the construct. One approach to this task is to calculate a Cronbach's alpha statistic (UCLA, n.d.), which is a measure of how well all the items in your scale relate to one another (i.e., together capture the same construct).

Similarly, if you have a measure captured by observation (e.g., data collectors rated child behavior in the classroom), consider calculating its interrater reliability.<sup>3</sup> To do this, you will need more than one data collector to collect the same observation data on the same evaluation participants. These analyses will demonstrate the level of agreement between data collectors. High agreement means the observation tool is reliable regardless of which data collector is completing it.

**Qualitative data quality.** When conducting qualitative research, data collection often runs concurrently with data analysis, and a high level of rigor in qualitative data is often discussed as the level of trustworthiness. Trustworthiness is established when findings as closely as possible reflect the meanings as described by

<sup>2</sup> Study attrition refers to any loss in responses from the study sample (Deke et al., 2015).

<sup>3</sup> Interrater reliability is the degree to which different raters or judges make consistent estimates of the same phenomenon; also known as interobserver reliability (Multon & Coleman, 2018).

the evaluation study respondents. The researcher is often the primary instrument for data collection in qualitative approaches, so researcher biases not adequately addressed or errors in judgement can affect the quality of data and subsequent interpretation of findings. Unlike quantitative methods' strong emphasis on reliability<sup>4</sup> and validity,<sup>5</sup> it is not possible to use the same metrics when judging the quality of conclusions in qualitative studies. Instead, more viable alternatives have been proposed to serve as evaluative criteria (see table 7.1).

**Table 7.1. Terms Used to Establish Trustworthiness of Qualitative Data**

Conventional Terms in Quantitative Research	Alternative Evaluative Criteria in Qualitative Research	Description of Criteria for Demonstrating Rigor in Qualitative Research
Objectivity	Confirmability	Requiring researchers to be reflexive or self-critical about their biases
Internal validity	Credibility, authenticity	Presenting an accurate description or interpretation of a human experience
External validity	Transferability, fittingness	Transferring findings or methods from one group (or setting) to another
Reliability	Dependability, audibility	Following the decision chain, so other researchers can determine the credibility of the findings

Several strategies are available to evaluation teams to establish trustworthiness of qualitative data. Throughout the research process, the evaluation team should practice reflexivity.<sup>6</sup> Similarly, the use of an audit trail<sup>7</sup> offers flexibility to make decisions not previously prescribed while still requiring justification of those decisions to be recorded.

Triangulation, peer debriefing, and member checking can avoid or minimize error or bias and boost the accuracy in data collection and analysis processes:

- Triangulation involves identifying convergence of data obtained through multiple data sources and methods (e.g., observation field notes and interview transcripts).
- Peer debriefing, sometimes referred to as analytic triangulation, involves consulting with researchers outside an evaluation project who have experience with the topic, the population, or methods being used to better explain how the evaluation team's own values and interests are influencing the conduct, interpretation, and analysis of the research project. Peer debriefing is often compared with internal validity.

<sup>4</sup> Reliability refers to the extent to which a measurement (such as an instrument or a data collection procedure) produces consistent results over repeated observations or administrations of the instrument under the same conditions. It is important that reliability be maintained across data collectors; this is called interrater reliability (OPRE, 2010).

<sup>5</sup> Validity refers to the extent to which a measurement instrument or test accurately measures what it is supposed to measure. For example, a reading test is a valid measure of reading skills but is not a valid measure of total language competency (OPRE, 2010).

<sup>6</sup> Reflexivity refers to actively acknowledging how one's own identity, beliefs, and values are inevitably assisting or hindering the process of co-constructing the meaning of the experience under investigation (University of Melbourne, n.d.).

<sup>7</sup> An audit trail refers to clear documentation of research procedures throughout the data analysis process (Robert Wood Johnson Foundation, 2006).

- Member checking refers to a set of processes evaluation teams can use to “check in” on how participants in qualitative data collection respond to comments in the data or to researchers’ interpretations of the data. Ideally, member checks are used in combination with other methods to establish a study’s credibility (i.e., ensure the research findings are believable to participants).

Document all data quality issues in your reports, and discuss the implications and limitations associated with your findings based on data quality. In some cases, your data quality issues may be so severe you cannot use some elements or even an entire dataset.

## Analyze Implementation Data

As a reminder (see [chapter 1](#)), implementation evaluations use data on program implementation to assess whether and to what extent program activities are being implemented as planned, expected program services are being delivered as planned, and how the program is operating in practice. Examples of basic program implementation evaluation questions follow:

- **How will we know the planned activities occurred?** For example, the number, duration, and frequency of services or activities implemented
- **Who will do it?** What the staffing arrangements will be; the characteristics and qualifications of the program staff who will deliver the services, conduct the training, or develop the products; and how these individuals will be recruited and hired
- **What population do you plan to reach, and how many individuals?** A description of the participant population for the program, the number of participants to be reached during a specific timeframe, and how you plan to recruit or reach the participants

Implementation evaluations typically collect data about implementation barriers and facilitators and how staff and program participants experienced the program. Because implementation evaluations do not try to ascribe changes to the program, they rely on descriptive analyses. Descriptive analyses paint a picture of the setting and provide details but do not attempt to measure the association or relationship between measures. Descriptive analytical techniques can be applied to both qualitative and quantitative data.

## Quantitative Data

Quantitative implementation evaluation data can be analyzed using the following:

- Counts (e.g., 1,000 families were served over the program period)
- Averages (e.g., on average, 32 workshop sessions were provided per month)
- Frequencies (e.g., 40 percent of caseworkers had 5 or more years’ experience)



How you calculate each descriptive quantitative statistic depends on your evaluation questions. For example, you may need to calculate weekly attendance rates or monthly rates. You may need to report the mean number of workshops attended or the percentage of participants who attended at least 8 of 10 workshops.

## Qualitative Data

Numerous approaches to analyzing qualitative data are available, each with different levels of rigor and requiring different levels of expertise and effort. Coding<sup>8</sup> is a ubiquitous part of qualitative analysis. In general, coding processes fall into one of two categories, deductive<sup>9</sup> or inductive:<sup>10</sup>

- Deductive or “theory-driven” coding is a top-down approach that applies predetermined codes.<sup>11</sup> The codes can be drawn from the literature or represent issues an evaluation team is seeking to better understand. For example, you may decide to apply the codes “transportation,” “child care,” and “work schedule” to interview transcripts with program participants based on previously reported barriers to participation. In this case, deductive coding may save time and ensure key areas of interest are coded. However, starting with predefined codes also increases the risk of researcher bias and/or could overlook other important themes.
- Inductive or “data-driven” coding is a bottom-up approach that generates codes based on the data. These codes are iteratively developed throughout a coding process that typically involves reading through the data to establish a general understanding of the issue (e.g., experience, behavior, decision, relationships), identifying meaning units,<sup>12</sup> assigning codes to those meaning units, and grouping codes according to themes. For example, the use of inductive coding may lead to assigning the code “cultural incongruence” to capture any participant discussion about how the program content and/or delivery may be lacking cultural sensitivity.

Both deductive and inductive strategies can be combined to facilitate a foundational understanding of the topic (from a previous evaluation of the same program, for example), while also facilitating the addition of new, unanticipated information to emerge from the data as a codebook is being developed.

Finally, qualitative analysis is not a linear process, and coding is rarely a one-time event. First-level coding mainly uses descriptive, low inference codes that are useful for summarizing segments of data (i.e., to answer questions such as who, what, when, where) and provide the basis for higher level order coding. For example, when coding an interview transcript with a program participant, any mention of gas cards or ride share payments might be coded as “supportive service payment.” Second-level codes tend to focus on patterns across multiple informants or sources of data and often require some degree of inference beyond the data.

<sup>8</sup> Coding is the process of systematically categorizing excerpts in qualitative data to find themes and patterns (Delve, n.d.-a).

<sup>9</sup> Deductive coding is also called a top-down approach: you start with a set of predetermined codes and then find excerpts that fit those codes (Delve, n.d.-b).

<sup>10</sup> Inductive coding is also called a bottom-up approach: you start with no codes and develop codes as you analyze the dataset (Delve, n.d.-b).

<sup>11</sup> Codes are descriptive labels assigned to data (CESSDA Training Team, 2020).

<sup>12</sup> Meaning units are segments of text that describe some information about the evaluation question (Elo et al., 2014).

The previous code for “supportive service payment” might be grouped under the broader code of “participation barriers” or broken down further into subcodes such as “insufficient compensation” or “gift card challenges.”

In summary, qualitative analysis is a flexible, reflective, and continuous process of coding, recoding, and categorizing, with subsequent return to the raw data to tell a story about how program implementation occurred. Qualitative data analysis can provide insights into how planned activities occurred and why, who implemented the activities, program reach, and participant characteristics. You can then compare this information with your initial objectives and determine whether there is a difference between objectives and actual implementation. Qualitative data analysis is also used to contextualize outcome analysis findings (mixed-methods approach) as described in [chapter 5](#). This process will answer the question: Has the program been successful in attaining the anticipated implementation objectives?

If your objectives and your actual implementation differ, you can analyze your evaluation information to determine the reasons for the differences. This step answers the question: If not, why not?

You can also use your evaluation information to identify barriers that impeded implementation and facilitating factors that contributed to implementation. This information can be used to “tell the story” of your program's implementation. Recall the measurable objectives introduced as examples in [chapter 4](#) for the planning of a substance use prevention program:

- The program will provide eight substance use education class sessions per year.
- Each session will involve 2 hours of classes per day.
- Classes will be held for 5 days.

An example of how this information might be organized is provided in table 7.2. The table represents an analysis of the program's measurable implementation objective concerning what the program plans to do. The first column lists the measurable objectives. The actual program implementation information is provided in the second column. For this program, differences between objectives and actual implementation were apparent for three of the four measurable objectives. Column 3 notes the presence or absence of differences, and column 4 provides the reasons for those changes. Columns 5 and 6 identify the barriers encountered and the facilitating factors. These factors are important to identify regardless of whether implementation objectives were attained. They provide the context for understanding the program and will help you interpret the results of your analyses.

Table 7.2. Sample Table for Analyzing Information on Implementation Objectives

Implementation Objective	Actual Implementation	Differences? (Yes/No)	If Yes, Reasons for Change	Barriers Encountered	Facilitating Factors
Eight substance use prevention class sessions per year	Six substance use prevention class sessions the first year	Yes	Delay in startup time during the first year	Difficulty finding qualified staff Delay in curriculum development	Agency experience in implementing similar types of programs Assistance of volunteers with sessions
Each session will last 2 weeks	First two sessions lasted 2 weeks; last four sessions lasted 1 week	Yes	Participants could not consistently attend for 2 weeks	Youth lost interest during second week	Available participants in shelter
Each class will last 2 hours	First two sessions, classes were 2 hours each day; last four sessions were 3 hours each day	Yes	Because the time was shortened, had to extend intensity of classes to cover curriculum material	None	Experienced staff able to cover curriculum during shortened time span
Classes will be given 5 days of each week	5 days a week	No		Problems with crisis intervention youth attending all 5 days	Staff availability

By reviewing the information in this table, you could say the following about your program:

- The program implemented only six substance use prevention sessions instead of the intended eight sessions.
  - ▶ A delay in starting the first set of sessions caused the program to complete fewer sessions during the program evaluation timeline than expected.
  - ▶ The delay was caused by the difficulty of recruiting and hiring qualified staff, which took longer than expected.
  - ▶ With staff now on board, we expect to be able to implement the full eight sessions in the second year.
  - ▶ After staff were hired, the sessions were implemented smoothly because there were several volunteers who helped organize special events and transport participants to the events.
- For the first two sessions, the class time was 2 hours per day, as originally intended. After the number of sessions was decreased, the class time increased to 3 hours per day.
  - ▶ The increase was caused by the need to cover the curriculum material during the session.
  - ▶ The extensive experience of the staff and the assistance of volunteers facilitated covering the material during the 1-week period.
  - ▶ The youth's interest was high during the 1-week period.

- The classes were provided for 5 days, as intended.
  - ▶ This schedule was facilitated by staff availability and the access to youth residing in the shelter.
  - ▶ It was more difficult to get youth from crisis intervention services to attend for all 5 days.

You would then apply this approach to data relevant to all your other implementation objectives, such as staffing (who will do it) and the population (reach and characteristics of participants). To begin organizing the implementation information from your own program, see the blank template of table 7.2 provided in [appendix B](#).

## Analyze Outcome Data

[Chapter 1](#) defines outcome evaluations as studies that intend to understand the extent to which change has occurred as intended. An impact evaluation can attribute outcomes (typically those that occur sometime after program completion) to the program. The analysis of participant outcome information typically answers the following questions:

- Did the expected changes occur in participants' knowledge, attitudes, behavior, or awareness?
- And for impact designs:
  - ▶ Did the expected changes occur in other outcomes such as participants' incomes, parenting, educational attainment, or relationship?
- If changes occurred, were they the result of your program's interventions?

If you employed a quasi-experimental or experimental impact design (and executed it well), you likely will answer questions such as the following:

- Did the program improve participant outcomes (such as increases in wages, education, or family stability, or decreases in smoking, number of missed school days, or financial hardships)?

Another question that could be included in your analysis of participant outcome information follows:

- Did some participants change more than others, and if so, what explains this difference? (For example, characteristics of the participants, types of interventions, duration of interventions, intensity of interventions, or characteristics of staff)

Your evaluation planning must include a detailed description of how you will analyze information to answer these questions. Know exactly what you want to do before you begin collecting data, particularly the types of statistical procedures you will use to analyze participant outcome information.

All outcome evaluations assess changes, so they all have a “compared with what” component. At a basic level, your analysis will calculate the value of the outcome at each postprogram time point and compare it, using a statistical procedure, with your selected comparison condition. As described in chapter 5, common comparison conditions follow:

- For nonexperimental designs: data from the same individuals before program start; benchmarks, such as national or state-level averages; or targets, such as funder expectations, or those your evaluation team develops based on evidence from other similar evaluations
- For impact (quasi-experimental and experimental) designs: outcome data on the same measures from a randomly or nonrandomly selected but similar population. To strengthen the rigor of impact studies, you can conduct difference-in-difference analyses. In difference-in-difference, you compare the change over time within your treatment group on an outcome to the change over time on the same outcome for your comparison group

### Baseline equivalence

If you have an impact evaluation model, you should conduct baseline equivalence tests for each outcome at pretest data collection period. These tests are particularly important for quasi-experimental designs or random assignment designs that had high attrition. It helps to ensure that treatment and comparison groups were equal on outcomes of concern before the treatment group received the program.

## Understanding statistical procedures

For outcome analyses, you will conduct inferential analyses. Unlike descriptive analyses (mentioned above) that aim to describe, inferential analyses aim to test relationships among data elements.

### Inferential analyses

- **Measure the degree to which the outcome variable and other variables are associated.** At a basic level, you will test the relationship between the outcome of interest and one or more independent variables<sup>13</sup> (such as characteristics of the program). This type of analysis can indicate whether the levels of individuals' outcomes are correlated with the independent variables. For example, you may determine that outcomes are positively correlated with hours of program services received. Remember that correlation does not mean causation. Researchers can include additional independent variables to “control” the analysis for factors that are associated with both the outcome and the program characteristic. Factors such as socioeconomic and demographic characteristics and “pretest” measures of the outcome often serve as strong control variables. Impact analyses will contrast individuals offered the program with individuals not offered the program by including a program indicator as an independent variable.

<sup>13</sup> An independent variable is one that stands alone and isn't changed by the other variables being measured (National Center for Education Statistics, n.d.)

- **Can be conducted using multiple types of models.** It is important the evaluation expert on your team knows which types of tests are appropriate for which types of data. For example, ordinary least squares regression models work with continuous (e.g., weight) or ordinal (e.g., 5-point attitude scale) dependent variables, while logistic regressions are typically used to test binary dependent variables (e.g., did or did not drop out of school). In contrast, an ANOVA test is applicable for categorical dependent variables (e.g., when the outcome of interest is a set of different categories that can't be ranked, such as marital status). Your expert should confirm your data meets the assumptions of proposed models, such as normal distribution for a multiple regression model.
- Produce two important pieces of information:
  - ▶ **Statistical significance:** The  $p$ -value is the probability that an impact greater than your estimate would occur by chance when the true impact is zero. See U.S. Department of Education (2021) for a more technical explanation.
  - ▶ **Measure of magnitude:** an indication of how much change in your outcome (dependent variable) occurred. Impact analyses often use a calculation called an effect size. You can also demonstrate magnitude by calculating difference in outcomes through the model (e.g., on average, program participants scored 10 points higher on a parenting measure at program completion compared with pretest). You can also report on clinical or meaningful changes; for example, of the children in the program with below grade-level reading skills before enrollment, half scored at or above grade level at posttest.

After you have answered your primary evaluation questions, you may want to conduct some exploratory analyses, such as the following:

- **Subgroup analyses.** In subgroup analyses, you test to see if certain participants experienced larger or smaller differences in an outcome. For example, you may have a hunch, or a hypothesis, that women who participate in your parenting program were more likely to report increased confidence in their parenting than men. A subgroup analysis would separate the outcome change for those two groups and help verify or refute your hunch.
- **Dosage analyses.** In dosage analyses, you can assess the extent to which individuals who received more of your program had better outcomes than those who received less. Be careful with the interpretation of these findings. People who take more programming may differ in important ways from those who take less programming, and those differences might drive changes in outcomes as opposed to the program. For example, these two groups could differ on motivation to change, access to transportation, health, or level of stress.
- **Sensitivity analyses.** Sensitivity analyses are approaches that test the robustness of your model (and its findings). In sensitivity analyses, you change the assumption in your model to see the extent to which slightly different assumptions lead to different findings.

# Interpreting Your Findings

The interpretation step of program evaluation might be the most meaningful. When you make sense of the findings, you can tell the story of your program through the evaluation results. Without bringing meaning to these numbers and tables, you can't make use of the evaluation. With evaluation, you can explain how your program will perform to future funders, make program adaptations and improvements, provide potential program participants with evidence about what their experience might be like, and potentially scale and replicate your program in other locations with fidelity.

Interpretation involves setting down program information and evaluation results and asking questions such as the following:

- Was this finding what we expected? Why or why not?
- Why did one outcome show statistical improvement? Why didn't others?
- Do we think the program is responsible for this outcome change? What else could be affecting change at the same time?
- What did we do well? What do we need to change or do better?

## Information to have at hand during interpretation efforts

During interpretation conversations and efforts, you should be able to reference, compare, and revisit much relevant program information. To be fully prepared for that conversation, you should gather important materials such as your program logic model, staff training materials, funder requirements, program objectives and goals, and any other relevant information such as evaluations of similar programs.

Findings from both your outcome and implementation evaluation should come together during this process. For example, if you did not see expected improvements in participant knowledge of community resources for social support, you should look toward your implementation evaluation findings. Your implementation evaluation found that most program facilitators were new to the community and didn't have a strong understanding of the different community-based organizations in the area.

Pay attention to the magnitude of your findings. While general evaluation practice suggests evaluation teams should focus on findings that demonstrate statistical significance, discussing the magnitude of those statistically significant findings brings important nuance to your evaluation. Large, statistically significant findings are more meaningful with respect to real-life improvement than small ones. At the same time, there's an adage, "The absence of evidence is not necessarily the evidence of absence." In other words, spend time exploring why you might have received unexpected, nonsignificant findings. If you can't tie that finding to implementation problems, it could be a function of your evaluation design or evaluation quality. Reassess whether you had a large enough sample size (it's easier to find statistical significance with larger samples), used an appropriate measure, measured the outcome at the right time to see change, or had a low response rate (which may have skewed which types of people provided outcome data).

Throughout the interpretation process, keep methodological and situational limitations in mind. For example, a large magnitude of change measured by a survey with low response rates should be viewed with great

caution because the underlying quality of the data could be problematic. Your evaluation design affects the extent to which outcome changes can be seen as caused by the program. In addition, situate your findings within the current evaluation context. Many of your evaluation audience members will want to know if your findings are likely to happen in other situations. For example, if your program showed significant improvements in reading scores, they may want to know if those improvements would occur if they implemented the same program. Be sure to think about and document the context in which your program operated, and the extent to which you think certain elements were key to your findings.

## Practice Culturally Responsive and Equitable Evaluation When Analyzing and Interpreting Data

A CREE approach does not stop with the evaluation design or data collection, but it is also helpful when analyzing and interpreting the data. Numerous ways are available to meaningfully engage community members during the analysis stage to ensure a more robust understanding of the data. For example, data walks, a strategy for visually sharing data with community members, create an opportunity for program participants, community residents, and service providers to jointly review data presentations in small groups, interpret what the data mean, and collaborate to use their individual expertise to improve policies, programs and other factors of community change (Murray et al., 2015).

Understanding when and how to disaggregate data is also a valuable way to practice CREE during data analysis. This practice enables evaluators to identify and address findings for groups of participants, instead of just participants as a whole. Data should be disaggregated to the level where it can still be understood in a meaningful way for the group it is exploring. For example, can you look at data by race, age, and gender to understand if there is a commonality in experience for Asian American females over 60? Data disaggregation can be informative when reviewing or analyzing data, conducting root cause analyses, reporting findings, or presenting information. Before analyzing data at a subgroup level, think about what types of bias could reside in the measurement tool or data collection process that could influence differences between groups. Carefully frame your findings so you don't inadvertently reinforce racial stereotypes.

A CREE approach is vital when interpreting the results of your analyses. Evaluation teams should consider engaging program participants and community members when interpreting analyses to ensure conclusions drawn are informed by the community's cultural values and perspectives of the program's quality and effectiveness. Evaluation teams should also be thoughtful about contextualizing research evidence to the local settings and systems (e.g., examining historical or structural factors that might shape findings). This can be particularly important when interpreting negative or unexpected findings. More specifically, evaluators must be careful to draw conclusions that account for both individual and system-level factors that could contribute to negative findings.



In summary, consider strategies to promote CREE practices both when analyzing and interpreting the data they've collected and the following recommendations:

- Disaggregate data and findings to illuminate disproportionality.
- Contextualize data using information about lived experiences.
- Include findings that communicate assets and strengths within a community.
- Provide context for findings, such as relevant institutional and environmental factors that influence individual behaviors.
- Ensure community voices are heard when making judgments about the data.
- Demonstrate cultural humility when interpreting findings.

### Analyzing and interpreting data through a culturally responsive and equitable lens

Consider the first-time parent-child development education program discussed earlier in this chapter. Disaggregating the data may reveal your child development program is effective for White parents but not for parents of color. In such a case, you could examine your analysis of program implementation information to understand why this may have happened and provide recommendations for program improvement. By contrast, if you had not disaggregated the data in this manner, this disproportionality likely would not have been unearthed and addressed.

If you find the program is effective for White parents but not for parents of color, have you considered the potential role of structural racism? For instance, structural racism may have a negative effect on the number of instructors of color, which in turn, may reduce the program's effectiveness for parents of color.

### To learn more ...

- [Analyzing and Interpreting Data](#) (Wilder Research, 2009)
- [Analyzing Quantitative Data for Evaluation](#) (CDC, 2018)
- [Disaggregating Data by Race Allows for More Accurate Research](#) (Sharpe, 2019)
- [Ethics and Empathy in Using Imputation to Disaggregate Data for Racial Equity: Recommendations and Standards Guide](#) (Brown et al., 2021)
- [Forum Guide to Collecting and Using Disaggregated Data on Racial/Ethnic Subgroups](#) (National Forum on Education Statistics, n.d.)
- [Methods, Challenges, and Best Practices for Conducting Subgroup Analysis](#) (Breck & Wakar, 2021)
- [Practical Strategies for Culturally Competent Evaluation](#) (CDC, 2014)
- [Qualitative Methods in Monitoring and Evaluation: Analyzing Qualitative Data](#) (Peters, 2022)
- [The Essentials of Disaggregated Data for Advancing Racial Equity](#) (Race Matters Institute, 2019)

## References

- ACF (Administration for Children and Families). (n.d.-a). *Title IV-E prevention services clearinghouse*. U.S. Department of Health and Human Services. <https://preventionservices.acf.hhs.gov/>
- ACF. (n.d.-b). *Home visiting evidence of effectiveness*. U.S. Department of Health and Human Services. <https://homvee.acf.hhs.gov/>
- Asher, J., Resnick, D., Brite, J., Brackbill, R., & Cone, J. (2020). An introduction to probabilistic record linkage with a focus on linkage processing for WTC registries. *International Journal of Environmental Research and Public Health*. <https://doi.org/10.3390/ijerph17186937>
- Bailey, J. (2008). First steps in qualitative data analysis: Transcribing. *Family Practice*, 25(2), 127–131. <https://academic.oup.com/fampra/article/25/2/127/497632>
- Breck, A., & Wakar, B. (2021). *Methods, challenges, and best practices for conducting subgroup analysis* (OPRE Report 2021-17). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/methods-challenges-and-best-practices-conducting-subgroup-analysis>
- Brown, S., Ford, L. D., & Ashley, S. (2021). *Ethics and empathy in using imputation to disaggregate data for racial equity: recommendations and standards guide*. Urban Institute. <https://www.urban.org/research/publication/ethics-and-empathy-using-imputation-disaggregate-data-racial-equity-recommendations-and-standards-guide>
- CDC (Centers for Disease Control and Prevention). (2014). *Practical strategies for culturally competent evaluation*. U.S. Department of Health and Human Services. [https://www.cdc.gov/asthma/program\\_eval/cultural\\_competence\\_guide.pdf](https://www.cdc.gov/asthma/program_eval/cultural_competence_guide.pdf)
- CDC. (2018). *Analyzing quantitative data for evaluation*. *Evaluation Brief*. U.S. Department of Health and Human Services. <https://www.cdc.gov/healthyyouth/evaluation/pdf/brief20.pdf>
- CESSDA Training Team. (2017–2020). *Qualitative coding*. <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/3.-Process/Qualitative-coding>
- Deke, J., Sama-Miller, E., & Hershey, A. (2015). *Addressing attrition bias in randomized controlled trials: Considerations for systematic evidence reviews*. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. [https://homvee.acf.hhs.gov/sites/default/files/2019-06/HomVEE-Attrition-White\\_Paper-7-2015.pdf](https://homvee.acf.hhs.gov/sites/default/files/2019-06/HomVEE-Attrition-White_Paper-7-2015.pdf)
- Delve. (n.d.-a). *The essential guide to coding qualitative data*. <https://delvetool.com/guide#:~:text=Qualitative%20coding%20is%20a%20process,themes%20and%20patterns%20for%20analysis>.
- Delve. (n.d.-b). *Deductive and inductive coding*. <https://delvetool.com/blog/deductiveinductive#:~:text=Inductive%20coding%20is%20a%20ground,from%20the%20raw%20data%20itself>

- Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). *Qualitative content analysis: A focus on trustworthiness*. SAGE Open. <https://doi.org/10.1177/2158244014522633>
- Multon, K. D., & Coleman, J. S. M. (2018). Inter-rater reliability. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation*. <https://methods.sagepub.com/reference/the-sage-encyclopedia-of-educational-research-measurement-and-evaluation/i11331.xml>
- Murray, B., Falkenburger, E., & Saxena, P. (2015). *Data walks: An innovative way to share data with communities*. Urban Institute. <https://www.urban.org/research/publication/data-walks-innovative-way-share-data-communities>
- National Center for Education Statistics. (n.d.). *What are independent and dependent variables?* [graphing tutorial]. [https://nces.ed.gov/nceskids/help/user\\_guide/graph/variables.asp](https://nces.ed.gov/nceskids/help/user_guide/graph/variables.asp)
- National Forum on Education Statistics. (n.d.). *Forum guide to collecting and using disaggregated data on racial/ethnic subgroups*. [https://nces.ed.gov/forum/pdf/Disaggregated\\_Data\\_PPT.pdf](https://nces.ed.gov/forum/pdf/Disaggregated_Data_PPT.pdf)
- OPRE (Office of Planning, Research, and Evaluation). (2010). *The program manager's guide to evaluation* (Second ed.). U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/report/program-managers-guide-evaluation-second-edition>
- Peters, B. (2022). *Qualitative methods in monitoring and evaluation: Analyzing qualitative data*. <https://programs.online.american.edu/msme/masters-in-measurement-and-evaluation/resources/qualitative-methods-project-cycle>
- Race Matters Institute of JustPartners. (2019). *The essentials of disaggregated data for advancing racial equity*. <https://viablefuturescenter.org/racemattersinstitute/resources/disaggregated-data/>
- Robert Wood Johnson Foundation. (2006). *Qualitative research guidelines project*. <http://www.qualres.org/HomeAudi-3700.html#:~:text=An%20audit%20trail%20is%20a,was%20done%20in%20an%20investigation>
- Satterlund, T. (n.d.). *Note to self: Writing analytic memos*. UC Davis Center for Evaluation and Research. [https://tobaccoeval.ucdavis.edu/sites/g/files/dgvnsk5301/files/inline-files/Newsletter--Memoing%202.22.12\\_edited.pdf](https://tobaccoeval.ucdavis.edu/sites/g/files/dgvnsk5301/files/inline-files/Newsletter--Memoing%202.22.12_edited.pdf)
- Sharpe, R. (2019). Disaggregating data by race allows for more accurate research. *Nature Human Behaviour*, 3, 1240. <https://doi.org/10.1038/s41562-019-0696-1>
- Statistics Solutions. (n.d.). *Thematic analysis*. <https://www.statisticssolutions.com/thematic-analysis/#:~:text=Familiarization%3A%20This%20is%20the%20process,qualitative%20researcher%20with%20the%20data>
- UCLA. (n.d.). *What does Cronbach's alpha mean? SPSS FAQ*. Advanced Research Computing. <https://stats.oarc.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean/>

University of Melbourne. (n.d.). *Reflexivity*. <https://medicine.unimelb.edu.au/school-structure/medical-education/research/qualitative-journey/themes/reflexivity#:~:text=Reflexivity%20is%20about%20acknowledging%20your,will%20influence%20the%20research%20process>

U.S. Department of Education. (n.d.) *What works clearinghouse*. <https://ies.ed.gov/ncee/wwc/>

U.S. Department of Education. (2021). *Statistical significance and sample size*. National Center for Educational Statistics. <https://nces.ed.gov/nationsreportcard/guides/statsig.aspx>

U.S. Department of Labor. (n.d.). CLEAR: *Clearinghouse for labor evaluation and research*. <https://clear.dol.gov/#:~:text=Homepage%20%7C%20CLEAR&text=CLEAR's%20mission%20is%20to%20make,about%20labor%20policies%20and%20programs>

Wilder Research. (2009). *Analyzing and interpreting data: Evaluation resources from Wilder Research*. <http://www.evaluatod.org/assets/resources/evaluation-guides/analyzing-interpretingdata-8->

# Chapter 8. Share Lessons Learned

## What's Inside?



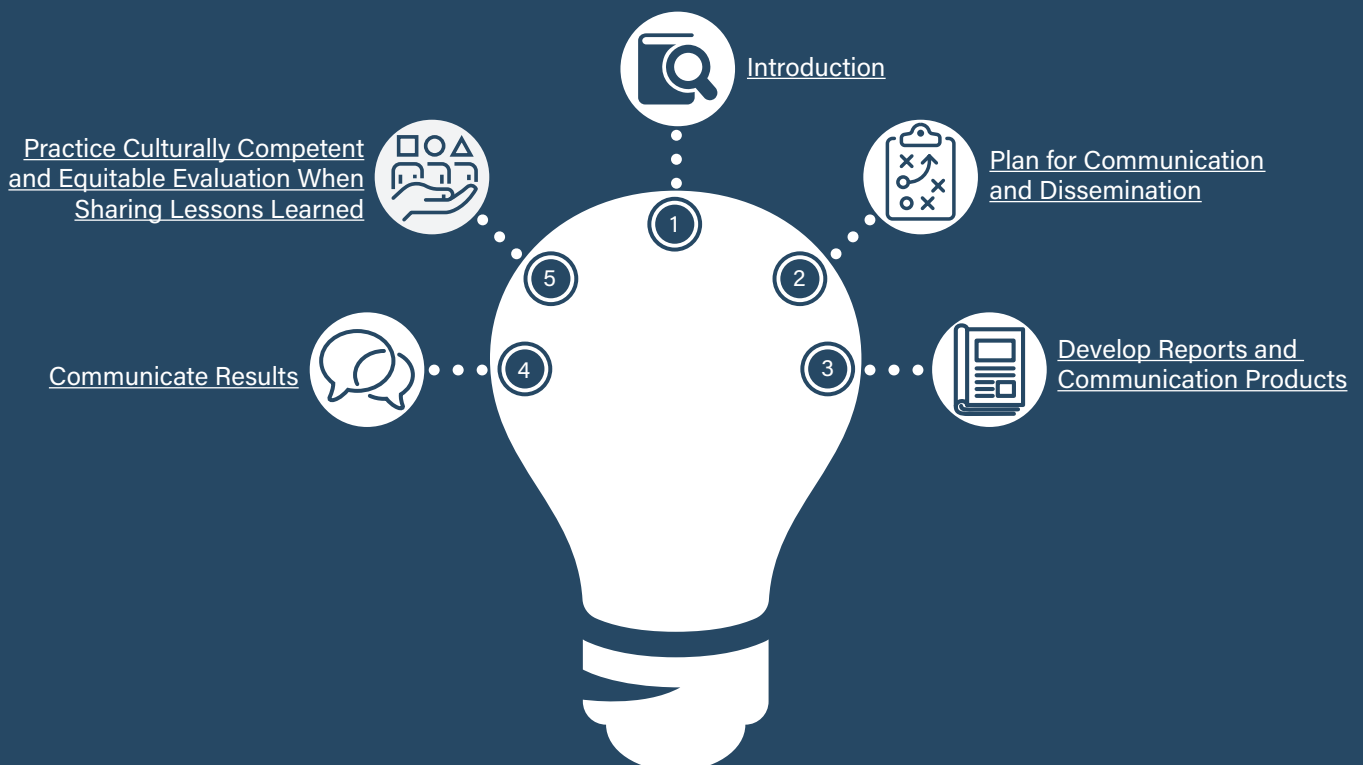
### What this chapter contains

- A description of communication planning
- Recommendations for developing reports and other communication products
- A discussion of communication channels and partners
- Examples of ways to apply culturally responsive and equitable principles when sharing lessons learned

### Who can use this chapter

- Evaluation team members involved in evaluation reporting and communication

Click the links below to view the relevant section



## Introduction

Evaluation is an applied science: It is designed to generate information and practical and actionable knowledge that leads to improvements in the program, activity, or policy being evaluated.

### Transparency

As discussed in [chapter 1](#), ACF's Evaluation Policy (2021) highlights transparency as an important principle. Sharing evaluation findings—whether good, bad, or null—is one way to ensure transparency and in turn, foster positive change.

To increase the chances your evaluation fosters positive change, focus on two elements when sharing the evaluation's lessons learned:

### Element 1

Communicate your lessons learned clearly and succinctly through multiple formats tailored to your target audiences.

### Element 2

Communicate lessons learned widely, so many people, organizations, and decision-makers can access your products.

This chapter provides advice and recommendations to support transparent communication and sharing your evaluation's findings. It describes various ways you can “tell the story” of the program you evaluated.

## Plan for Communication and Dissemination

A communication plan helps you prepare to share your work and findings. While communication and reporting are discussed last in this Guide, don't wait until you've completed the analysis activities described in [chapter 7](#) to start thinking about reporting and communication. Ideally, you will develop a communication plan alongside or shortly after you craft your evaluation plan. Communication is a part of the research process and entails sharing your work and your findings to audiences and engaging reciprocally with them (CWCC Evaluation TA Team, 2021).

### What is the difference between communication and dissemination?

Dissemination involves a one-way broadcast of information from a single origin, while communication involves active exchanges between two or more parties.

Communication plans can use different structures or organizational plans to capture the information needed to share lessons learned. In this chapter, we use the *Evaluation Dissemination Planning Guide: Building Capacity to Evaluate Child Welfare Community Collaborations to Strengthen and Preserve Families* (CWCC Evaluation TA Team, 2021) as a recommended structure.

Broadly speaking, your overall communication plan should be based on three factors:

## Factor 1

### Identify the communication budget

How much money does the evaluator and the program have to devote to communication? When are those dollars available? What restrictions might you have on how you spend those funds? Project budgets should reflect the importance of communication in ensuring that evaluation findings are used by relevant audiences.

## Factor 2

### Establish roles and responsibilities

While the evaluator typically leads communication of evaluation findings, program leadership may consider using findings more directly related to programming efforts (such as updates to the program's website). Program staff may write portions of products or lead their development, provide guidance and feedback on products, co-present at conferences or on podcasts, or lead the outreach and follow-up with communication partners (more on that below). Consider engaging with communication specialists when possible to maximize the reach of your work.

### Evaluator independence

As discussed in [chapter 1](#), ACF's Evaluation Policy highlights evaluator independence as an important principle in product development and communication. While the evaluation team includes program staff, ensure the subject matter expert in the evaluation has the final say about what findings are presented and how they are interpreted. Ensuring the evaluation experts play an independent role in communication adds credibility to your dissemination products.

## Factor 3

### Create strategic communication goals

Your plan should stem from your goals or reasons for distributing and sharing the findings of your evaluation. While most funders require a final report, you will likely have communication goals broader than funding compliance. For each goal consider how you can measure progress. Consider whether there are diagnostic metrics associated with each strategic goal, and develop a plan for how you will establish, analyze, and learn from them. Some examples of common communication goals follow:

- **Advance the field and improve programs and policies.** We have much to learn about how to improve program implementation and respond to implementation challenges. By sharing what you have learned in your evaluation, other organizations can benefit. Policymakers and decision-makers may take your recommendations into account when establishing laws, regulations, or program requirements.

- **Get attention from potential funders.** Reporting on the effectiveness of the program may signal to funders that your program should be continued, expanded, replicated, or scaled up. Your evaluation might also identify specific areas of improvement where future funding could help increase program reach or strength. Finally, your nonexperimental evaluation (e.g., an outcome evaluation without a comparison group) findings could indicate your program is ready to participate in an impact evaluation to generate causal evidence of the program's effectiveness.
- **Increase awareness.** Evaluation results tell your program's story. Your findings may include information about the demographic characteristics and reach of your program participant population, descriptions of program participant experiences, and positive changes associated with your program. Sharing this information can increase awareness of your program generally and build interest among your target population (Palen & Briggs, 2020).

### Evidence review clearinghouses

Numerous fields, such as workforce development, home visiting, child welfare, and education, have large online clearinghouses that review evaluation evidence on how programs or interventions improve outcomes. Many of these clearinghouses provide opportunities for evaluators to submit their studies for inclusion when a program is reviewed. One communication goal might be to submit a clearinghouse-focused report or article for consideration.

### Communicate throughout the life of an evaluation

Do not wait until your final report is complete to start sharing information about your evaluation. You have knowledge to contribute throughout all the steps in this guide:

- During planning, you can share your evaluation design decisions, instrument and measurement development, and strategies for including community representation on the evaluation team.
- During data collection, you can share interim findings and updates on effective study recruitment and retention strategies and ways you are monitoring the evaluation activities.

While the communication goals above are largely focused on the end state of your evaluation, you may also have interim goals tied to each step in your evaluation. For example, during your planning stages, you might share your evaluation design at an evaluators conference to solicit feedback and learn from other evaluation practitioners. Early communication can also help create awareness of and build demand for evaluation findings among the scientific and practitioner community. Later, you may present interim findings at a community town hall to solicit input and perspectives on how to interpret findings.

After you have structured your overall approach to communication—why, how much, and by whom—

you can develop each product and determine how to get that product to your intended audience. The next sections describe that process.

## Develop Reports and Communication Products

Develop a plan for each report or product you create to share lessons from your evaluation. A plan will help ensure your product uses an effective format and approach for reaching the target audience. Product plans typically address the following components:



- **The audience.** What groups are you trying to communicate with? The more detailed your audience definition, the better you can tailor your writing to their information needs and motivations. While you might want to reach many audiences with a single product, designate primary and secondary audiences to keep your document focused and tailored.
- **The message.** What are you trying to say? What knowledge do you want to impart? What actions do you want your audience to take after engaging with your product?
- **The format.** What vehicle will you use to communicate with your audience? For example, will you write a report or record a podcast (see table 8.1)?
- **The channels.** How will you get your message to your audience? For example, will you post your report on your organization’s website or distribute it through your evaluator’s listserv? See [Communicate Results](#).

Audiences, formats, and messages all interact to help you determine the right combination. You might start with an audience such as grant funders and build your product out from there. Conversely, you might try out a new data visualization platform, then focus on what message the data can express, and the audience that should receive that message.

Table 8.1 describes a variety of product formats, the types of messages they typically communicate well, and the audiences that may best connect with them.

**Table 8.1. Communication Product Formats, Messages, and Audiences**

Format	Message Types	Audiences
Interim or final report	Interim or final summation of the evaluation; findings and interpretations	<ul style="list-style-type: none"> <li>■ Current and future funders</li> <li>■ Evidence review clearinghouses</li> <li>■ Project/organizational leadership</li> <li>■ Researchers and evaluators</li> </ul>
Executive summary	Top-level summary of evaluation findings	<ul style="list-style-type: none"> <li>■ Policymakers and decision-makers</li> <li>■ Community leaders</li> <li>■ General public</li> <li>■ Other program managers</li> </ul>
Journal article	Technical report focused on methodology, quality of evaluation, results, and contribution to the field	<ul style="list-style-type: none"> <li>■ Researchers and evaluators</li> <li>■ Evidence review clearinghouses</li> <li>■ Program developers</li> <li>■ Policymakers</li> </ul>
Research brief	Short but detailed report covering methodology, evaluation quality, and findings	<ul style="list-style-type: none"> <li>■ Current and future funders</li> <li>■ Policymakers and decision-makers</li> <li>■ Researchers and evaluators</li> <li>■ Program developers</li> </ul>
Nontechnical brief	Short report focused on findings, interpretations, and implications	<ul style="list-style-type: none"> <li>■ Other program managers</li> <li>■ Community leaders</li> <li>■ Community members</li> <li>■ General public</li> <li>■ The media</li> </ul>

Table 8.1. Communication Product Formats, Messages, and Audiences (continued)

Format	Message Types	Audiences
Fact sheet	Short document presenting simple findings in visually appealing displays	<ul style="list-style-type: none"> <li>▪ Other program managers</li> <li>▪ Community leaders</li> <li>▪ Community members</li> </ul>
Tip sheet	Short document describing actions and next steps based on interpretations of findings or lessons learned through implementation of the evaluation	<ul style="list-style-type: none"> <li>▪ Program developers</li> <li>▪ Other program managers</li> <li>▪ Current and future funders</li> </ul>
Infographic, data visualization	Diagram, illustration, or other visual that presents information in an easy-to-understand way; can stand alone or be part of other formats	<ul style="list-style-type: none"> <li>▪ Other program managers</li> <li>▪ Community leaders</li> <li>▪ Community members</li> <li>▪ General public</li> <li>▪ The media</li> </ul>
Press release	Short announcement about the evaluation set in local or current context	<ul style="list-style-type: none"> <li>▪ The media</li> </ul>
Blogpost	Short, simple web-based content that provides easily digestible information, often with graphics and videos and links to longer products	<ul style="list-style-type: none"> <li>▪ Other program managers</li> <li>▪ Community leaders</li> <li>▪ General public</li> </ul>
Social media post	Text-based or multimedia (image, graphic, video) content for sharing on social media platforms. Social media posts can both contain independently informative news, updates, and information and direct attention to other communication products	<ul style="list-style-type: none"> <li>▪ Community members</li> <li>▪ Community leaders</li> <li>▪ Researchers and evaluators</li> <li>▪ General public</li> <li>▪ The media</li> <li>▪ Other program managers</li> </ul>
Presentation	Virtual or in person, often guided by a slide deck and with time for audience interaction and feedback	<ul style="list-style-type: none"> <li>▪ Researchers and evaluators</li> <li>▪ Other program managers</li> <li>▪ Current and future funders</li> </ul>
Video recording	Visual and audio reporting. Videos can be animated or interview based and can be used to put a human face to research and evaluation work and tell highly engaging, data-based stories	<ul style="list-style-type: none"> <li>▪ Community leaders</li> <li>▪ Community members</li> <li>▪ General public</li> </ul>
Audio recording	Audio reporting that tells a data-based story in an interview or small group discussion format	<ul style="list-style-type: none"> <li>▪ Community leaders</li> <li>▪ Community members</li> <li>▪ General public</li> </ul>

Source: CWCC [Child Welfare Community Collaborations] Evaluation TA Team (2021)

Social media posts and campaigns, and products like briefs and blog posts, can both contain novel evidence and insights and synthesize them across other products. They can also valuably direct attention to other products.

While the materials above are for external audiences, you may also develop reports about the evaluation for internal program use only. For example, you might develop yearly fidelity calculations or track program retention for program management or continuous quality improvement purposes. Such reports should focus on the data rather than the messages. They use the exact terminology program staff use and share other pertinent information such as the precise timeframe the data cover.

## Write a final evaluation report

Most externally funded evaluations are required to develop a comprehensive final report. These reports usually have the funder as a primary audience and may have related documents that distill and translate the findings for more internal use by program staff. Your final report might be your evaluation's most important product because it can affect future funding decisions.

Many funders have specific requirements for your final report's content and the way it's organized. You must meet those expectations to ensure it complies and gives the funder access to your evaluation and its findings. Provide detailed information about the program, the evaluation design and methods, the data quality and your analytical approaches, your findings, and their interpretations and implications. This Guide provides final and interim evaluation report templates in [appendix B](#).

While you must first comply with funder requirements, other best practices for evaluation final evaluation reports follow:

- **Develop a separate executive summary.** This stand-alone two- to three-page document should provide a nontechnical description of the highlights of the evaluation's findings and implications and information about the program and its goals.
- **Place technical information in an appendix.** While you need to provide sufficient information for a reader to assess the quality of your evaluation, most of the audience does not need to read technical specifics. Information such as the model specifications or complex sampling plans should appear in the appendix and referenced in the body of the report. Those interested in that information can access it in a way that does not encumber the flow of the evaluation report.
- **Place instruments in an appendix.** A detail-oriented reader and members of the program's field appreciate you sharing your data collection instruments. Sharing instruments is also an act of transparency that enables your audience to review the exact measures you used to capture the evaluation's data.

### Develop a report for an evidence review

If you have conducted an impact evaluation and want to submit it for an evidence review or ensure you have documented all relevant information, consider the following reporting guides produced by evidence clearinghouses:

- *Title IV-E Prevention Services Clearinghouse Reporting Guide for Study Authors* (Kerns et al., 2021)
- *Home Visiting Evidence of Effectiveness Reporting Guide for Authors* (OPRE, 2020)
- *WWC Reporting Guide for Study Authors* (IES [Institute for Educational Sciences], 2021)

- **Situate your evaluation in the field.** Put your evaluation in context; for example, by providing information about the program rationale and local conditions, rationale for your selection of outcomes, and findings from other evaluations of the same or similar program. Cite materials that provide this contextual information.
- **Provide information that enables readers to assess rigor and quality.** As described in [chapter 7](#), information such as sample size, attrition rates, and amount of missing data all indicate the rigor of your evaluation's findings. Evaluations based on lower quality data should be viewed as more speculative than those with higher quality data.
- **Organize your findings around your evaluation questions.** Well-constructed evaluation questions drive an evaluation from conceptualization to communication. Using these questions to structure your findings section is a good way to maintain consistency and guide your audience through your results.
- **Bring integrity to your reporting.** Many evaluations have unexpected or disappointing findings; however, evaluation reports are informative, not persuasive. Resist the urge to suppress or withhold any findings germane to your evaluation questions. This tension is one reason advance study registration is considered a scientific best practice and requested by many funders. Study registries document analysis plans, and high-profile evaluation reports are often compared with the plans.

### Progress reports

Your funder may also require progress reports. These reports typically focus on programmatic updates but may also require updates on evaluation progress, or even interim findings to date. Make sure your evaluation team is prepared to provide information according to funder expectations.

## Tips for effective writing

While the section above addressed development of your major final evaluation report, this section provides advice relevant to all products you may develop, including your final evaluation report.

### Plain language guidelines

According to [plainlanguage.gov](http://plainlanguage.gov), plain language (also called plain writing or plain English) is communication your audience can understand the first time they read or hear it. This website provides a variety of guides, checklists, and templates to help writers craft accessible and clear products.

can easily comprehend. Strategies to write well include writing concisely with shorter sentences and paragraphs, eliminating words that add no value, using the active voice, varying sentence length, and following plain language guidelines (see textbox).

- **Use simple language.** Flowery, complicated, and jargon-filled language is sometimes incorrectly seen as a sign of education and expertise in a topic. However, such language makes a document hard to understand and distracts from the message. If a writer understands their topic, they can write about it in simple terms the audience
- **Use inclusive language** so you center the evaluation on people rather than their characteristics or conditions (e.g., people in prison rather than incarcerated people). Assess the reading comprehension level of your writing. Most word processing software provide editor functions that assess a document's grade-level reading. While technical writing is acceptable at a 12th grade reading level, target 8th to 10th grade level writing for nontechnical pieces.

- **Write assuming your audience is busy.** One way to boost the chances your audience gets your message is to state your findings or argument several times—in the introduction, in the body of the product, and again in the conclusion. Use a consistent pattern within similar sections. For example, if you describe findings for each evaluation question, describe quantitative findings and then qualitative findings.
- **Remember your audience does not know what you know.** Pinker (2014) argues that most written materials invite confusion and frustration because the writers assume their audience has knowledge they do not. He recommends writers share copies of their drafts with members of their intended audience for feedback, and if time allows, set writing aside for a while, and return to read it with a critical mindset.
- **Be careful with comparative language.** As detailed more in the [Practice Culturally Competent and Equitable Evaluation section in chapter 7](#), think about what race/ethnicity, gender, and religion you center in your subgroup discussions. Why would you or would you not make White, male, Christian, heterosexual the reference category?
- **Follow a writing process.** Writing is a laborious, iterative practice. Use proven techniques to develop your products such as starting from an outline or storyboard, writing a draft and sharing it with others for feedback, and writing subsequent drafts. Seek professional or peer reviews and editing support.
- **Establish authorship.** Determine who gets credit as an author for each product. If more than one person is writing, identify the order people will be listed as authors. Authorship means the person made a significant contribution to the product and will vouch for the data analysis, findings, and veracity of the product. You might identify people who provided support such as reviewers in an acknowledgements section. Establish plans for organizational contributions. Which organizations' logos will appear on your products? How will you document program and evaluation funding?

### Writing resources

Numerous tip sheets and guides can help support strong and effective writing and copy edit processes. Be sure to check out the links at the end of this chapter.

## Ways to engage with your audience

In addition to the words you use in your products, the structure of your products can influence audience engagement. People are inundated with information. You need to invest in making it easy for them to engage in your materials:

- **Make your products visually appealing.** Use easy-to-read fonts, add pictures, and leave plenty of “white space.” Avoid too much formatting, which results in a “busy” and uninviting appearance.
- **Make your products easy to navigate.** Use headers and subheaders. Provide agendas or roadmaps at the start of a presentation. Add page numbers. Use a table of contents if your document is longer than 20 pages. Use bold, bullets, and textboxes to separate and highlight main points.
- **Make your products consistent.** Use the same color for the same concept or outcome throughout the product. For example, if you provide data separately for people by race/ethnicity, make sure Latino and Black, for example, are the same color/shading in each visual. Show data in the same order across each visual.

- **Consider guiding titles for your visuals.** Rather than a bland descriptive title, such as “Changes in parenting attitudes over time for treatment and comparison groups,” try highlighting interesting findings: “Treatment group parents show larger improvements in parenting attitudes over time than comparison group parents.”
- **Use effective data visualizations.** Tables, graphs, charts, and infographics can help tell your story and engage your audience. Data science provides ample recommendations for how to make the best use of visuals and avoid common mistakes (recommendations at the end of this chapter). For a web-based product, consider building interactive data visualizations using products such as Tableau, so your audience can filter and manipulate your data to answer their questions and address their curiosity.
- **Provide practical implications and next steps** based on your findings, such as a textbox titled “Tips for Practitioners.”

## Make your products accessible

Did you know that about a quarter of Americans have a disability? About 13 percent of American adults have difficulty hearing, while 6 percent of Americans aged 12 and over have difficulties seeing (Madans et al., 2021; NIH [National Institutes of Health], 2006). Make all your products accessible to as many people as possible. If the federal government funded your project or you intend to distribute it through federal channels, you may be required to make your products accessible.

The federal government defines accessibility adaptations according to Section 508 of the Rehabilitation Act of 1979. While you will need to obtain formal guidance to adopt all necessary accessibility procedures, the guidance falls into two main categories:

- Your document will need to be ready for reading by the software (e.g., JAWS) that individuals with visual impairments use. Supports are needed such as alternative text describing visuals, use of high-contrast colors, and correct structuring of tables, for example, so individuals with visual impairments can engage with the written or visual product.
- Provide supports such as closed captioning, so individuals with hearing impairments can engage in an oral or audio product.

See Information Gateway/CB (2020) and Office of the Chief Information Officer (2021) for more information.

## Communicate Results

After you have developed a product, get it to your audience members. Some audiences are small and identifiable such as a group of eight program staff who need an internal data update report. Others are more diverse and not known by name or email, such as parents who live in the program's county.

Break down communication into two parts: how you will get your product to your audience and which partners can help you amplify your communication.

Table 8.2 shows common communication channels. You may distribute a product through multiple channels or use different versions of the same channel type to reach different audiences.

**Table 8.2. Communication Channels**

Communication Channel Type	Examples
Websites	<ul style="list-style-type: none"> <li>▪ Program or program's organizational website</li> <li>▪ Evaluator's website</li> <li>▪ Local community websites, such as news sites or community bulletin boards</li> <li>▪ Information clearinghouses and databases</li> <li>▪ Program developers' websites</li> <li>▪ Online journals</li> </ul>
Social and digital media	<ul style="list-style-type: none"> <li>▪ Blogs</li> <li>▪ Social media</li> <li>▪ Newsletters, e-blasts, email marketing</li> <li>▪ Listservs and communities of practice</li> <li>▪ Podcasts</li> <li>▪ Online media</li> </ul>
Events	<ul style="list-style-type: none"> <li>▪ Government meetings</li> <li>▪ Conferences</li> <li>▪ Local, state, and regional organizational meetings</li> <li>▪ Webinars</li> <li>▪ Town halls</li> </ul>
Print	<ul style="list-style-type: none"> <li>▪ Flyers</li> <li>▪ Mailings</li> <li>▪ Newspapers</li> <li>▪ Journals</li> </ul>

Some of these channels or options within a channel are better suited to some audiences and some products. For example, town halls may be an effective means of reaching community members, while webinars through a professional membership organization may be more effective at reaching similar program staff and providers. It is important to identify audience preferences and meet your audiences where they are.

Identify communication partners and influencers who can help amplify your communication through their own networks. For example, you might consider asking a national organization such as the National Head Start Association to retweet you or add language about your report to their newsletter to reach many more relevant audience members. Develop a list of trusted communication partners with the type of audience they can reach and the communication channels. Ensure their approach, philosophies, and reputation in the field are aligned with yours. Communication is inherently reciprocal: Offer to help amplify partner and audience messages when appropriate to do so.

Typical types of communication partners follow:

- Respected community members or individuals in your field who are interested in supporting and amplifying the evaluation
- Local or national organizations related to the topic or service population associated with your program
- Members of your formal or informal networks
- Professional media consultants

Your level of engagement with communication partners can be something as simple as tagging an organization in a tweet to a formal memorandum of agreement for cross-communication efforts.

## Practice Culturally Competent and Equitable Evaluation When Sharing Lessons Learned

### Potential audiences

- Community members who could be past or future participants
- Local government offices, community activists, school administrators, and other organizations who could influence program objectives

When you share findings with your program's communities, you can strengthen relationships among program staff and community members. Communicating findings will demonstrate transparency with data provided by the community and position them as valued program partners. Consider your potential audiences and what information would most interest them when deciding what to share.

In a CREE approach to program evaluation, communication strategies are co-created with communities so findings are relevant to local organizations and residents. Community members' insights on communication strategy decision points can help answer questions such as those listed here.

- Which findings are most relevant to community audiences?
  - ▶ Should you present or report findings across all program communities or just from this community?
  - ▶ Do comparisons across participant groups, such as which groups benefitted the most or least, help with decision-making?
- How can you present findings with the most clarity?
  - ▶ What insider language or jargon do you need to replace?
  - ▶ What do you need to do to ensure any tables, charts, photos, and quotes you share improve your communities' understanding?

### Communicating with the community

The Tribal Early Childhood Research Center presented on their work at the annual OPRE Methods Meeting (Barnes-Najor et al., 2021). Their recommendations for CREE-aligned communication include thinking about a communication strategy throughout the project; sharing findings that contradict existing paradigms; paying attention to different audiences' information needs and concerns; and producing many nonacademic products to reach community members.



- Which communication channels would best reach each audience?
  - ▶ For example, a school board may prefer a presentation, while you may best reach parents through an email from their children's teachers.
  - ▶ If you are trying to communicate to young mothers through social media, which platforms (e.g., Facebook, Instagram) are the most popular, and what groups should you tag?
- Who should the spokesperson be for findings?
  - ▶ How do other factors, such as communication channels, influence who should deliver the message? For instance, social media posts might be best developed by community members, while presentations to the local transportation authority might be best delivered by community members and program staff.
- Can you use communication activities to collect community feedback?
  - ▶ Could community listening sessions share findings and collect information about future program needs?

Individuals and organizations who communicate do more than broadcast: they reciprocally inform, motivate, and learn. Including community voices in the communication of findings to other audiences should align with CREE. For instance, having evaluators and engaged program participants co-present at conferences can help make findings meaningful to the audience. When including community representatives and program participants in presentations and other forms of communication, representation should be more than token. The representatives should help decide what to present, know what each person is contributing, understand the audience and what types of questions might be likely, and have ample time to practice what they will say and learn about any technology that will be used.

### To learn more ...

- [APA Style \(APA, 2022\)](#)
- [Dissemination Planning Tool \(AHRQ, 2005\)](#)
- [Equitable Research Communication Guidelines \(Gross, 2020\)](#)
- [Guidance Note on Developing an Evaluation Dissemination Strategy \(United Nations, 2009\)](#)
- [Plain Writing in One Page \(HHS.gov, 2015\)](#)
- [Six Tips for Making a Quality Report Appealing and Easy To Skim | Agency for Healthcare Research and Quality \(AHRQ, 2019\)](#)
- [The Copy Editing and Proofreading Checklist All Writers Need \(Klems, 2016\)](#)
- [The Value-Added Research Dissemination Framework \(Macoubrie & Harrison, 2013\)](#)
- [Three Ways to Expose Formatting Inconsistencies in a Word Document \(Harkins, 2016\)](#)
- [Disseminating Evaluation Results \(Palen & Briggs, 2020\)](#)
- [Tips for Effective Data Visualization \(Thatte, 2019\)](#)
- [U.S. Government Publishing Office Style Manual \(U.S. Government Publishing Office, 2016\)](#)

## References

- ACF (Administration for Children and Families). (2021). *ACF Evaluation Policy*. U.S. Department of Health and Human Services. <https://www.acf.hhs.gov/opre/report/acf-evaluation-policy>
- AHRQ (Agency for Healthcare Research and Quality). (2019). *Six tips for making a quality report appealing and easy to skim*. <https://www.ahrq.gov/talkingquality/resources/design/general-tips/index.html>
- AHRQ. (2005). *Advances in patient safety: Dissemination planning tool*. U.S. Department of Health and Human Services. <https://www.ahrq.gov/sites/default/files/wysiwyg/professionals/quality-patient-safety/patient-safety-resources/resources/advances-in-patient-safety/vol4/planningtool.pdf>
- APA (American Psychological Association). (2022). *APA style*. <https://apastyle.apa.org/>
- Barnes-Najor, J., Around Him, D., & Cameron, A. (2021). *Community engagement in a federally-sponsored center: The TRC* [Presentation at OPRE Methods Meeting]. <https://opremethodsmeeting.org/wp-content/uploads/2021/09/Case-Study-3-Community-Engagement-in-Fed-Sponsored-Center-.pdf>
- CDC (Centers for Disease Control and Prevention). (2021). *Preferred terms for select population groups & communities*. [https://www.cdc.gov/healthcommunication/Preferred\\_Terms.html](https://www.cdc.gov/healthcommunication/Preferred_Terms.html)
- CWCC [Child Welfare Community Collaborations] Evaluation TA Team. (2021). *Evaluation dissemination planning guide*.
- Gross, E. (2020). *Equitable research communication guidelines*. Child Trends. <https://www.childtrends.org/publications/equitable-research-communication-guidelines>
- Harkins, S. (2016). *Three ways to expose formatting inconsistencies in a Word document*. <https://www.techrepublic.com/article/three-ways-to-expose-format-inconsistencies-in-a-word-document/>
- HHS (U.S. Department of Health and Human Services). (n.d.). *Plain writing in one page*. <https://www.hhs.gov/web/building-and-managing-websites/managing-websites/plain-writing-in-one-page/index.html>
- HHS.gov. (2015). *Plain writing in one page*. U.S. Department of Health and Human Services. <https://www.hhs.gov/web/building-and-managing-websites/managing-websites/plain-writing-in-one-page/index.html>
- IES (Institute of Education Sciences). (2021). *WWC reporting guide for study authors*. U.S. Department of Education. [https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/WWC\\_Author\\_Guide\\_Jul2021.pdf](https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/WWC_Author_Guide_Jul2021.pdf)
- Information Gateway/CB. (2020). *Section 508 Tips for document creation*. U.S. Department of Health and Human Services. [https://www.acf.hhs.gov/sites/default/files/documents/cb/508\\_tip\\_sheet.pdf](https://www.acf.hhs.gov/sites/default/files/documents/cb/508_tip_sheet.pdf)

- Kerns, S., Wilson, S., Brown, S., Weiss, C., & Gubits, D. (2021). *Title IV-E prevention services clearinghouse reporting guide for study authors* (OPRE Report 2021-27). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/sites/default/files/documents/opre/reporting-guide-study-authors-apr-2021.pdf>
- Klems, B. (2016). *The copy editing and proofreading checklist all writers need*. *Writer's Digest*. <https://www.writersdigest.com/write-better-fiction/copy-editing-proofreading-checklist-writers-need>
- Macoubrie, J., & Harrison, C. (2013). *The value-added research dissemination framework*. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/sites/default/files/documents/opre/valueadded.pdf>
- Madans, J. H., Weeks, J. D., & Elgaddal, N. (2021). *Hearing difficulties among adults: United States, 2019* (NCHS Data Brief No. 414). National Center for Health Statistics. <https://www.cdc.gov/nchs/products/databriefs/db414.htm>
- NIH (National Institutes of Health). (2006). *Study finds most Americans have good vision, but 14 million are visually impaired*. U.S. Department of Health and Human Services. <https://www.nih.gov/news-events/news-releases/study-finds-most-americans-have-good-vision-14-million-are-visually-impaired#:~:text=A%20National%20Institutes%20of%20Health,visual%20impairment%2C%20such%20as%20nearsightedness.>
- Office of the Chief Information Officer, U.S. Department of Health and Human Services. (2021). *Accessibility compliance checklist*. <https://www.hhs.gov/web/section-508/accessibility-checklists/index.html>
- OPRE (2020). *Home visiting evidence of effectiveness reporting guide for authors*. U.S. Department of Health and Human Services. [https://homvee.acf.hhs.gov/sites/default/files/2020-12/HomVEE\\_Reporting\\_Guide.pdf](https://homvee.acf.hhs.gov/sites/default/files/2020-12/HomVEE_Reporting_Guide.pdf)
- Palen, L., & Briggs, S. (2020). *Disseminating evaluation results*. U.S. Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Family and Youth Services Bureau. <https://teenpregnancy.acf.hhs.gov/sites/default/files/resource-files/Disseminating%20Evaluation%20Results%20Tip%20Sheet.pdf>
- Pinker, S. (2014). *The source of bad writing*. Dow Jones & Company Inc. [https://stevenpinker.com/files/pinker/files/the\\_source\\_of\\_bad\\_writing\\_-\\_wsj\\_0.pdf](https://stevenpinker.com/files/pinker/files/the_source_of_bad_writing_-_wsj_0.pdf)
- Thatte, S. (2019). *Tips for effective data visualization*. *Towards Data Science*. <https://towardsdatascience.com/tips-for-effective-data-visualization-d4b2af91db37>
- United Nations. (2009). *Guidance note on developing an evaluation dissemination strategy*. *United Nations Development Fund for Women*. [https://www.endvawnow.org/uploads/browser/files/UNIFEM\\_guidance%20note\\_evaluation\\_Dissemination.pdf](https://www.endvawnow.org/uploads/browser/files/UNIFEM_guidance%20note_evaluation_Dissemination.pdf)
- U.S. Government Publishing Office. (2016). *Style manual*. <https://www.govinfo.gov/content/pkg/GPO-STYLEMANUAL-2016/pdf/GPO-STYLEMANUAL-2016.pdf>



# Appendix A. Additional Resources

## Evaluation Design and Planning

Ahonen, P., Geary, E., & Keene, K. (2019). *Tribal TANF-Child Welfare coordination: Theory of change and logic models* (OPRE Report No. 2019-55). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/tribal-tanf-child-welfare-coordination-theory-change-and-logic-models>

This tool provides an overview of key concepts and strategies for creating a theory of change as well as a logic model. It also includes a discussion of strategies for ensuring that programs' outputs and outcomes, two key components of a logic model, are measurable.

Ahonen, P., Keene, K., & Geary, E. (2020). *Communication guide for TTCW grantees: What to consider when sharing program accomplishments* (OPRE Report No. 2021-14). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/communication-guide-ttcw-grantees-what-consider-when-sharing-program-accomplishments>

This resource discusses communicating grant-funded Tribal social service programs to desired audiences.

Atukpawu-Tipton, G., Higman, S., & Morrison, C. (2020). *Qualitative evaluation* (OPRE Report No. 2020-136). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/qualitative-evaluation>

This report describes how to implement strong qualitative evaluations and minimize bias throughout each stage of evaluation.

Atukpawu-Tipton, G., & Poes, M. (2020). *Rapid cycle evaluation at a glance* (OPRE Report No. 2020-152). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/rapid-cycle-evaluation-glance?msclkid=120cc94dd07111ec9faf5902fd75b3ad>

The purpose of this brief is to introduce rapid cycle evaluation and its potential use in Maternal, Infant, and Early Childhood Home Visiting programs.

Bartko, T., Higman, S., & Thomson, A. (2021). *Linking process indicators to outcomes in evaluations of home visiting programs* (OPRE Report No. 2021-54). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/linking-process-indicators-outcomes-evaluations-home-visiting-programs>

This evaluation brief defines measures of home visiting services called process indicators, describes how process indicators link to short- and long-term outcomes in home visiting evaluations, and provides an example illustrating the role of process indicators in evaluations.

Bell, S., Harvill, E., Moulton, S., & Peck, L. (2017). *Using within-site experimental evidence to reduce cross-site attributional bias in connecting program components to program impacts* (OPRE Report No. 2017-13). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/using-within-site-experimental-evidence-reduce-cross-site-attributional-bias-connecting>

This report uses a Cross-Site Attributional Model by Calibration to test the impact of a program.

Breck, A., & Wakar, B. (2021). *Methods, challenges, and best practices for conducting subgroup analysis* (OPRE Report No. 2021-17). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/methods-challenges-and-best-practices-conducting-subgroup-analysis>

This brief aims to describe the feature of subgroup analysis that uses a multiple regression framework and provide an overview of methodological developments and alternative approaches to conducting subgroup analysis.

Center for Supporting Research on CCDBG Implementation. (2019). *A dozen policy questions you can answer with your agency's administrative data: A webinar for Child Care Development Fund lead agencies* [Webinar]. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/dozen-policy-questions-you-can-answer-your-agencys-administrative-data-webinar-child>

This webinar is designed to support Child Care Development Fund lead agency staff and their partners in using existing administrative data to address policy questions posed by state legislators, agency heads, local childcare providers, and others.

Child Care Research and Evaluation Capacity Building Center. (2020). *Using child care provider surveys to inform policy responses to COVID-19* [Webinar]. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/training-technical-assistance/webinar-using-child-care-provider-surveys-inform-policy?msclkid=158da7c4d07311eca19893746725da17>

This webinar provides tips on developing good survey questions and collecting meaningful data for childcare providers.

Clary, E., & Bradley, M. C. (2018). *Strengthening grantee capacity through technical assistance* (OPRE Report No. 2018-99). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/strengthening-grantee-capacity-through-evaluation-technical-assistance>

This report provides a description of evaluation technical assistance for capacity building.

Cody, S., & Arbour, M. (2019). *Rapid learning: Methods to examine and improve social programs* (OPRE Report No. 2019-86). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/rapid-learning-methods-examine-and-improve-social-programs>

This brief and the accompanying presentation provide an orientation to rapid learning methods, including (1) a definition of rapid learning methods, (2) a guiding framework of questions to design an optimal rapid learning approach, and (3) suggested steps federal agencies can take to promote the effective use of rapid learning methods.

Coffey, A., & Isaacs, J. (2019). *Evaluating training and professional development for home-based providers: A brief for CCDF lead agencies and researchers* (OPRE Report No. 2019-11). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/evaluating-training-and-professional-development-home-based-providers>

This summary of past research approaches and tips from research experts aims to provide information on engaging home-based childcare providers.

Deke, J. (2018). *Causal validity considerations for including high quality non-experimental evidence in systematic reviews* (OPRE Report No. 2018-63). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/causal-validity-considerations-including-high-quality-non-experimental-evidence>

This brief describes the need for nonexperimental study designs when a randomized control trial is not appropriate.

Derrick-Mills, T., Reginal, T., & Isaacs, J. (2020). *Procuring research and evaluation services: A guide for CCDF lead agencies and researchers* (OPRE Report No. 2020-89). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/procuring-research-and-evaluation-services-guide-ccdf-lead-agencies-and-researchers>

This is a guide to procuring specialized research or evaluation services.

Gutuskey, L. (2022). *Centering equity in program evaluation* (OPRE Report No. 2022-211). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

This resource aims to help evaluators, program managers, and technical assistance (TA) providers apply an equity lens when designing, conducting, and managing evaluations.

Hansen, D., & Holzwart, R. (2020). *OPRE 2019 methods meeting resource list* (OPRE Report No. 2020-131). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/opre-2019-methods-meeting-resources-list>

This document provides a list of resources for readers who wish to learn more about open science methods.

Haydon, A., & Kendall-Taylor, N. (2015). *Communicating scientific findings about adolescence and self-regulation: Challenges and opportunities* (OPRE Report No. 2015-78). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. [https://www.acf.hhs.gov/sites/default/files/documents/frameworks\\_adolescent\\_self\\_regulation\\_strategic\\_brief\\_opre\\_final\\_0.pdf](https://www.acf.hhs.gov/sites/default/files/documents/frameworks_adolescent_self_regulation_strategic_brief_opre_final_0.pdf)

This report discusses challenges associated with communicating scientific findings about adolescence and self-regulation.

Holman, D., Pennington, A., Schaberg, K., & Rock, A. (2020). *Compendium of administrative data sources for self-sufficiency research* (OPRE Report No. 2020-42). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/compendium-administrative-data-sources-self-sufficiency-research>

This report describes promising administrative data sources for evaluations of economic and social interventions.

Holzwart, R., Sama, H., & Wright, D. (2018). *Understanding Bayesian statistics: Frequently asked questions and recommended resources* (OPRE Report No. 2018-54). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/understanding-bayesian-statistics-frequently-asked-questions-and-recommended-resources>

This brief offers researchers short answers to four common questions about Bayesian methods, along with a curated list of resources (including journal articles, book chapters, online courses, and blogs) for further reading.

Holzward, R., Skinner, R., & Wright, D. (2019). *Understanding rapid learning methods: Frequently asked questions and recommended resources* (OPRE Report No. 2019-89). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/understanding-rapid-learning-methods-frequently-asked-questions-and-recommended?msclkid=384aa3a3d07111ecb807e922b228d6cc>

This document is a guide for readers who wish to understand, employ, or encourage use of rapid learning methods in social service settings.

Holzward, R., & Wagner, H. (2019). *Rapid learning: Methods for testing and evaluating change in social service programs* (OPRE Report No. 2019-57). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/rapid-learning-methods-testing-and-evaluating-change-social-service-programs>

This resource discusses topics related to rapid learning methods.

Holzward, R., & Wright, D. (2018). *Bayesian methods for social policy research and evaluation* (OPRE Report No. 2018-38). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/bayesian-methods-social-policy-research-and-evaluation?msclkid=87e75bbad07211eca0aaf56e1218caa1>

This report provides information on the underlying assumptions, tradeoffs, validity, and generalizability of results in a Bayesian framework.

Hyra, A. (2022). *Engaging community representation in program evaluation research* (OPRE Report No. 2022-169). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

This resource discusses engaging representatives of the community in a program evaluation, including why, the benefits, guiding principles, and recommended additional readings.

Jacob, R. (2016). *Using aggregate administrative data in social policy research* (OPRE Report No. 2016-91). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/using-aggregate-administrative-data-social-policy-research>

This brief provides an overview of how aggregate administrative data can be used in social policy research.

Keene, K., Keating, K., & Ahonen, P. (2016). *The power of stories: Enriching program research & reporting*. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/power-stories-enriching-program-research-reporting>

This report explores opportunities, considerations, and methods for using storytelling to understand and communicate information about social service programs in tribal communities.



Kline, N. (2022). *Using administrative data in social policy research* (OPRE Report No. 2022-163). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

This brief provides a definition and examples of administrative data, basics of administrative data (i.e., access and capacity), exploration of individual and aggregate administrative data, using administrative data in visualizations, and principles of equity in administrative data.

Lin, V., Maxwell, K., King, C., Martinchek, K., & Isaacs, J. (2021). *Working with administrative data in early childhood or related fields: A list of resources* (OPRE Report No. 2021-21). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/working-administrative-data-early-childhood-and-related-fields>

The resource list catalogs materials that explain how to acquire, use, manage, link, and analyze administrative data in early childhood or related fields.

Lyskawa, J., Kirby, G., Caronongan, P., Kelly, A., & Burwick, A. (2020). *Challenges and solutions to conducting intensive studies in early care and education settings* (OPRE Brief No. 2020-96). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/challenges-and-solutions-conducting-intensive-studies-early-care-and-education-settings>

The brief discusses the challenges of recruiting childcare centers and conducting qualitative research, cost analysis, and self-reported data collection with staff in center-based settings and offers potential solutions to those challenges.

Maxwell, K. (2017). *Issues in accessing and using administrative data* (OPRE Report No. 2017-24). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/issues-accessing-and-using-administrative-data>

This brief provides an overview of use and access issues to consider when using administrative data for social policy research.

McCay, J., Meckstroth, A., Akers, L., Resch, A., Derr, M., & Berk, J. (2015). *Learning what works: A guide to opportunistic experiments for human services agencies* (OPRE Report No. 2015-98). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/learning-what-works-guide-opportunistic-experiments-human-services-agencies>

This report introduces human services program operators to randomized controlled trials (RCTs) and provides guidance on how to conduct them.

Morgan-Lopez, A., & Bir, A. (2017). *Unpacking the “black box” of programs and policies: A conceptual overview of mediation analysis* (OPRE Report No. 2017-01). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/unpacking-black-box-programs-and-policies-conceptual-overview-mediation-analysis>

This brief describes mediation analysis and the analytic tools available for conducting mediation analysis.

OPRE (Office of Planning, Research, and Evaluation). (2016). *The Administration for Children and Families. common framework for research and evaluation*. U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/report/administration-children-families-common-framework-research-and-evaluation>

This framework outlines the roles of various types of research and evaluation in generating information and answering empirical questions related to the human services provided by the ACF.

OPRE. (2016). *The promises and challenges of administrative data in social policy research: Roundtable discussion* [Video]. U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/training-technical-assistance/promises-and-challenges-administrative-data-social-policy>

In this video roundtable, government experts and experienced researchers discuss the opportunities and challenges presented when using administrative data for social policy research.

OPRE. (2016). *Using administrative data in social policy research* (OPRE Report No. 2016-62). U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/report/using-administrative-data-social-policy-research>

This brief summarizes OPRE's 2015 Innovative Methods Meeting, which considered the potential benefits and pitfalls of using administrative data for research purposes.

OPRE. (2021). *Administrative data on federal policies and programs that support young children with disabilities: Resource guide for researchers*. U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/report/administrative-data-federal-policies-and-programs-support-young-children-disabilities>

This resource guide provides information for researchers about administrative data collected on federal policies and programs that (in whole or part) support young children with disabilities.

OPRE. (2021). *Designing and conducting home visiting evaluations in tribal communities: Takeaways from the HomVEE review of research with tribal populations—2020*. U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/report/designing-and-conducting-home-visiting-evaluations-tribal-communities-takeaways-homvee>

This brief summarizes findings on designing and conducting early childhood home visiting evaluations in tribal communities and the effectiveness of the models examined for the HomVEE review.

Rohacek, M. (2017). *Research and evaluation capacity: Self-assessment tool and discussion guide for CCDF lead agencies* (OPRE Report No. 2017-63). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/research-and-evaluation-capacity-self-assessment-tool-and-discussion-guide-ccdf-lead>

This tool supports CCDF lead agencies in strengthening their capacity to carry out and use research in decision-making.

Rohacek, M., Coffey, A., Isaacs, J., & Stephens, K. (2019). *Research and evaluation capacity building: A resource guide for Child Care and Development Fund lead agencies* (Rev. 2019) (OPRE Report No. 2019-74). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/research-and-evaluation-capacity-building-resource-guide-child-care-and-development>

This guide provides an annotated list of selected written and online resources to support CCDF lead agencies seeking to build research and evaluation capacity.

Sandstrom, H., & Isaacs, J. (2020). *Tips on developing surveys of child care providers* (OPRE Report No. 2020-114). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/tips-developing-surveys-child-care-providers>

This brief describes best practices for developing and testing surveys of childcare providers.

Steigelman, C. & Gutuskey, L. (2022). *Equity annotated bibliography* (OPRE Report No. 2022-178). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

This resource provides materials on the growing body of literature on applying equity principles to program design, research, and evaluation.

Till, L., & Zaid, S. (2019). *Developing a state learning agenda: The Maternal, Infant, and Early Childhood Home Visiting Program* (OPRE Report No. 2019-14). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/developing-state-learning-agenda-maternal-infant-and-early-childhood-home-visiting>

This brief explains what a learning agenda is, how to develop one, and how to integrate it with programmatic and research and evaluation activities.

Tribal Evaluation Institute. (2020). *Rigorous evaluation in tribal MIECHV: A series of briefs*. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/rigorous-evaluation-tribal-miechv-series-briefs>

Five evaluation briefs share the story of grantees' rigorous evaluations and provide recommendations for those who oversee evaluations with tribal communities or are seeking to support evaluations with tribal populations.

Wood, R., Goesling, B., & Paulsell, D. (2018). *Design for an impact study of five healthy marriage and relationship education programs and strategies* (OPRE Report No. 2018-32). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/design-impact-study-five-healthy-marriage-and-relationship-education-programs-and>

This report presents the design of five impact evaluations of healthy marriage and relationship education services.

## Evaluation Reports

Dworsky, A. (2020). *Supporting college students transitioning out of foster care: A formative evaluation report on the Seita Scholars program* (OPRE Report No. 2020-102). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/supporting-college-students-transitioning-out-foster-care-formative-evaluation-report>

This report describes lessons learned about the Seita Scholars program from formative evaluation activities and shares assessments of whether this program and others like it could be rigorously evaluated.

Hamadyk, J., & Gardiner, K. (2018). *"We get a chance to show impact": Program staff reflect on participating in a rigorous, multi-site evaluation* (OPRE Report No. 2018-123). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/we-get-chance-show-impact-program-staff-reflect-participating-rigorous-multi-site>

This brief summarizes findings from interviews conducted with leadership and staff from eight programs that participated in the Pathways for Advancing Careers and Education Evaluation, a rigorous, multi-site evaluation of "career pathways" programs.

Lee, H., Warren, A., & Gill, L. (2015). *Cheaper, faster, better: Are state administrative data the answer? The Mother and Infant Home Visiting Program Evaluation-Strong Start second annual report* (OPRE Report No. 2015-09). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/cheaper-faster-better-are-state-administrative-data-answer-mother-and-infant-home>

This report details the Mother and Infant Home Visiting Program Evaluation-Strong Start (MIHOPE-Strong Start) process of acquiring administrative vital records and Medicaid data from 20 states and more than 40 state agencies.

Michalopoulos, C., Lee, H., Snell, E., Crowne, S., Filene, J., Fox, M., Kranker, K., Mijanovich, T., Gill, I., & Duggan, A. (2015). *Design for the Mother and Infant Home Visiting Program Evaluation-Strong Start* (OPRE Report No. 2015-63). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/design-mother-and-infant-home-visiting-program-evaluation-strong-start>

This report describes the design of the Mother and Infant Home Visiting Program Evaluation-Strong Start.

Mills, G., McKernan, S., Ratcliffe, C., Edelstein, S., Pergamit, M., Braga, B., Hahn, H., & Elkin, S. (2016). *Building savings for success: Early impacts from the assets for independence program randomized evaluation* (OPRE Report No. 2016-59). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/building-savings-success-early-impacts-assets-independence-program-randomized>

This report describes the findings from a randomized control trial.

OPRE (Office of Planning, Research, and Evaluation). (2015). *Designing an impact study of four selected programs to reduce teen pregnancy*. U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/report/designing-impact-study-four-selected-programs-reduce-teen-pregnancy>

This brief summarizes key highlights from the report Design for an Impact Study of Four PREP Programs.

OPRE, Permanency Innovations Initiative Evaluation Team. (2016). *Using child welfare administrative data in the Permanency Innovations Initiative evaluation* (OPRE Report No. 2016-47). U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/report/using-child-welfare-administrative-data-permanency-innovations-initiative-evaluation>

This brief discusses the use of administrative data in the Permanency Innovations Initiative evaluation.

Werner, A., Loprest, P., Schwartz, D., Koralek, R., & Sick, N. (2018). *Final report: National implementation evaluation of the first round Health Profession Opportunity Grants (HPOG 1.0)* (OPRE Report No. 2018-09). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/final-report-national-implementation-evaluation-first-round-health-profession>

This report provides a summary of findings from the National Implementation Evaluation Descriptive Implementation and Outcome Studies and Systems Change Analysis.

Wood, R., Goesling, B., Zief, S., & Knab, J. (2015). *Design for an impact study of four PREP Programs* (OPRE Report No. 2015-01). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/design-impact-study-four-prep-programs>

This report summarizes the overall design of a random assignment evaluation of four PREP-funded programs.

## Performance Measurements and Indicators

Bailey, R., Barnes, S. P., Park, C., Sokolovic, N., & Jones, S. M. (2018). *Executive function mapping project measures compendium: A resource for selecting measures related to executive function and other regulation-related skills in early childhood* (OPRE Report No. 2018-59). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. Health and Human Services. <https://www.acf.hhs.gov/opre/report/executive-function-mapping-project-measures-compendium-resource-selecting-measures>

This resource provides information about the range of measures available to assess executive function and other regulation-related skills.

Brennan, E., Manno, M., & Steimle, S. (2019). *Using data to understand your program* (OPRE Report No. 2019-90). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/using-data-understand-your-program>

This infographic provides a framework to help organizations think about how the data they may already be collecting or could collect to help answer questions about their program or identify areas for improvement.

Burke, J. G., O'Malley, T. L., Hagen, C. A., Rabinovich, B. A., & Harmon, M. A. (2019). *A theoretical and stakeholder-informed assessment framework for the National Domestic Violence Hotline*. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/theoretical-and-stakeholder-informed-assessment-framework-national-domestic-violence>

This brief describes the effort of a project to develop a theoretical framework to explain how the National Domestic Violence Hotline empowers and supports contactors.

Davis, L., & Tucker, L. P. (2020). *Using continuous quality improvement to refine interventions for youth at risk of homelessness* (OPRE Report Number 2020-03). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/lessons-field-using-continuous-quality-improvement-refine-interventions-youth-risk>

In this brief, local evaluators working with two agencies, Alameda County, California, and the Colorado Department of Human Services, describe how their teams used CQI to learn from the initial implementation of model interventions designed to prevent homelessness among youth and young adults who have been involved in the child welfare system.

Derrick-Mills, T., Winkler, M., Healy, O., & Greenberg, E. (2015). *A resource guide for Head Start programs: Moving beyond a culture of compliance to a culture of continuous improvement* (OPRE Report No. 2015-02). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/resource-guide-head-start-programs-moving-beyond-culture-compliance-culture-continuous>

This resource guide helps those in Head Start and Early Head Start programs understand how data can help them achieve their goals, learn techniques for fostering a culture of learning in their organization, and continuously improve their programs.

Friend, D., Kleinman, R., Hague Angus, M., McInerney, H., Pranschke, L., & Avellar, S. (2020). *Building data capacity in Healthy Marriage and Responsible Fatherhood grantees: Challenges and recommended support* (OPRE Report No. 2020-95). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/building-data-capacity-healthy-marriage-and-responsible-fatherhood-grantees-challenges>

This report seeks to understand data capacity by looking at challenges faced by the 2015 cohort of Healthy Marriage and Responsible Fatherhood (HMRF) grantees.

Friese, S., Lin, V., Forry, N., & Tout, K. (2017). *Defining and measuring access to high quality early care and education: A guidebook for policymakers and researchers* (OPRE Report No. 2017-08). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/defining-and-measuring-access-high-quality-early-care-and-education-ece-guidebook>

This guidebook addresses the development of a common understanding and approach to measuring access to early care and education.

Hagen, C. A., Burke, J. G., O'Malley, T. L., Greene, A. D., Rabinovich, B. A., Kali, J., & Bravo Bueno, J. N. (2020). *Theoretical framework and performance measures for the National Domestic Violence Hotline: Report from the National Domestic Violence Hotline services assessment framework based on theory project* (OPRE Report No. 2020-109). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/theoretical-framework-and-performance-measures-national-domestic-violence-hotline>

This report helps a broad audience (e.g., practitioners, policymakers, academics, researchers, the public) understand the process of developing a theoretical framework for a brief crisis intervention and associated performance measures to inform program performance monitoring and evaluation.

Hagen, C. A., Green, A. D., Burke, J. G., O'Malley, T. L., Kali, J., Rabinovich, B. A., Bravo Bueno, J. N., & Crandall, J. P. (2020). *Theoretically-informed performance measures for the National Domestic Violence Hotline: Summary brief from the National Domestic Violence Hotline services assessment framework based on theory project* (OPRE Report No. 2020-110). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/theoretically-informed-performance-measures-national-domestic-violence-hotline-summary>

This brief provides a summary description of efforts to develop a survivor-centered theoretical framework.

Halle, T., Partika, A., & Nagle, K. (2019). *Measuring readiness for change in early care and education* (OPRE Report No. 2019-63). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/measuring-readiness-change-early-care-and-education>

This brief provides a framework for understanding readiness within the early childcare and education (ECE) field and to share examples of how ECE researchers are currently attempting to capture the dimensions of readiness—and factors that support readiness—using different data collection methods and standardized measurement tools.

Kautz, T., & Moore, Q. (2020). *Selecting and testing measures of self-regulation skills among low-income populations* (OPRE Report No. 2020-138). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/selecting-and-testing-measures-self-regulation-skills-among-low-income-populations>

This report discusses issues related to selecting and testing measures of self-regulation skills in evaluations of employment programs for low-income populations.



Keene, K., Geary, E., & Ahonen, P. (2020). *Tribal TANF-Child Welfare Coordination: Collaboration assessment tool* (OPRE Report No. 2020-40). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/tribal-tanf-child-welfare-coordination-collaboration-assessment-tool>

This tool helps current and future Tribal TANF-Child Welfare Coordination grantees assess their initiatives' partnership performance in a concrete and measurable way.

Klerman, J., Judkins, D., & Locke, G. (2019). *Impact evaluation design plan for the HPOG 2.0 national evaluation* (OPRE Report No. 2019-82). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/national-and-tribal-evaluation-2nd-generation-health-profession-opportunity-grants-1>

This design report presents detailed plans for the Impact Evaluation of HPOG 2.0, to understand what difference the program made.

Malone, L., Knas, E., Cavanaugh, M., & West, J. (2016). *Early care, education, and home visiting in American Indian and Alaska Native communities: Design options for assessing early childhood needs* (OPRE Report No. 2016-49). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/early-care-early-education-and-home-visiting-american-indian-and-alaska-native>

This report describes three potential designs for studies to assess the needs for early care and education and home visiting among American Indian and Alaska Native children and families.

McCay, J., Derr, M., & Person, A. (2017). *Using a "road test" to improve human services programs*. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/using-road-test-improve-human-services-programs>

This brief explains the road test process within the context of a larger systematic and evidence-informed framework for program improvement, provides practical guidance for using this approach in human services programs, and describes concrete examples of road tests.

McCay, J., Derr, M., & Person, A. (2019). *The Learn phase: Creating sustainable change in human services programs* (OPRE Report No. 2019-15). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/learn-phase-creating-sustainable-change-human-services-programs>

This practice brief provides an overview of the first phase of Learn, Innovate, Improve (LI2)—the Learn phase—which is intended to lay the foundation for successful and sustainable program changes.

McCay, J., France, M., Lujan, L., Maestas, V., & Whittaker, A. (2019). *Mobile coaching: Innovation and small-scale experimentation to better engage program participants in rural Colorado* (OPRE Report No. 2019-45). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/mobile-coaching-innovation-and-small-scale-experimentation-better-engage-program>

The brief describes the team's design process and road map for change (the logic model underpinning this creative strategy) as well as their approach to prototyping and testing on a small scale.

McCombs-Thornton, K., & Poes, M. (2021). *Measuring program effects in home visiting evaluation: Improving estimates with propensity score matching* (OPRE Report No. 2021-55). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/measuring-program-effects-home-visiting-evaluation-improving-estimates-propensity-score>

This brief provides an overview of propensity score matching.

National Survey of Early Care and Education Project Team. (2015). *Measuring predictors of quality in early care and education settings in the National Survey of Early Care and Education* (OPRE Report No. 2015-93). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/measuring-predictors-quality-early-care-and-education-settings-national-survey-early>

This methodological report describes how selected predictors of quality can be measured using data from the National Survey of Early Care and Education.

OPRE (Office of Planning, Research, and Evaluation). (2018). *Continuous quality improvement (CQI) toolkit: A resource for Maternal, Infant, and Early Childhood Home Visiting Program awardees*. U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/report/continuous-quality-improvement-toolkit-resource-maternal-infant-and-early-childhood>

The toolkit contains nine modules that cover continuous quality improvement (CQI).

OPRE. (2018). *Measuring self-regulation skills in evaluations of employment programs for low-income populations: Challenges and recommendations* (OPRE Report No. 2018-83). U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/report/measuring-self-regulation-skills-evaluations-employment-programs-low-income-populations>

This report discusses issues related to measuring self-regulation skills in evaluations of employment programs for low-income populations.

Roberts, E., Iannone-Walker, M., Callis, A., Porter, R., Geary, E., & Park, C. (2021). *Supporting data systems improvement in tribal home visiting: Capacity built and lessons learned* (OPRE Report No. 2021-05). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/supporting-data-systems-improvement-tribal-home-visiting-capacity-built-and-lessons>

This brief describes the capacity-building approach of ACF, which helps Tribal Home Visiting grantees strengthen their data systems through technical assistance.

Sarna, M., & Werner, A. (2018). *Targeting higher skills and healthcare jobs: How HPOG grantees set and use performance goals* (OPRE Report No. 2018-122). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/targeting-higher-skills-and-healthcare-jobs-how-hpog-grantees-set-and-use-performance>

This report explores how grantees develop performance projections for the Health Profession Opportunity Grants.

Strong, D., Stange, M., Roemer, G., Avellar, S., & Noonan, B. (2020). *Supporting program progress: Performance measures, data system, and technical assistance for the 2020 Healthy Marriage and Responsible Fatherhood grantees* (OPRE Report No. 2020-64). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/supporting-program-progress-performance-measures-data-system-and-technical-assistance>

This resource is a review to identify potential changes to the performance measures, management information system functionality, and activities that support data collection among HMRF grantees.

Thomson, D., Cantrell, E., Guerra, G., Gooze, R., & Tout, K. (2020). *Conceptualizing and measuring access to early care and education* (OPRE Report No. 2020-106). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/conceptualizing-and-measuring-access-early-care-and-education>

This report crosswalks recent definitions of access in the literature with the multidimensional definition as presented in the Access Guidebook, providing a launching point for future discussion around ongoing and planned efforts to document and improve access.

Xue, Y., Bandel, E., Vogel, C. A., & Boller, K. (2015). *Measuring infant/toddler language development: Lessons learned about assessment and screening tools* (OPRE Brief 2015-52). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/measuring-infant/toddler-language-development-lessons-learned-about-assessment-and>

The brief provides suggestions for factors programs should consider when selecting measures of children's development.

Zeribi, K., Mackrain, M., Arbour, M., & O'Carroll, K. (2017). *Partnering with families in continuous quality improvement: The Maternal, Infant, and Early Childhood Home Visiting Program* (OPRE Report No. 2017-47). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/partnering-families-continuous-quality-improvement-maternal-infant-and-early-childhood>

This tip sheet discusses the potential benefits of partnering with participants and their families in CQI efforts and discusses considerations and strategies that programs can use to do so effectively.

## Program Design and Implementation

Baumgartner, S., Overcash, A., Holcomb, P., & Zaveri, H. (2020). *Pathways-to-outcomes snapshots: Tools for building evidence for responsible fatherhood (RF) programs* (OPRE Brief No. 2020-116). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/pathways-outcomes-snapshots-tools-building-evidence-responsible-fatherhood-programs>

These snapshots provide information for practitioners and researchers involved in designing, improving, or evaluating RF programs.

Behrmann, R., & Brennan, E. (2020). *Inside, outside, round and round: Sustaining engagement in responsible fatherhood programs* (OPRE Report No. 2020-34). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/inside-outside-round-and-round-sustaining-engagement-responsible-fatherhood-programs>

This resource discusses a study that implemented a variety of practices to keep participants engaged.

Center for Supporting Research on CCDBG Implementation. (2020). *Answering more child care policy questions: Pairing stakeholder perspectives with your data* [Webinar]. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/training-technical-assistance/webinar-answering-more-child-care-policy-questions-pairing>

This webinar is designed to support CCDF lead agency staff and partners in understanding how various perspectives can be paired with agency data to help answer more policy questions.

Derr, M., McCay, J., & Person, A. (2019). *The innovate phase: Co-creating evidence-informed solutions to improve human services programs*. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. [https://www.acf.hhs.gov/sites/default/files/documents/opre/li2\\_innovate\\_co\\_creating\\_evidence\\_informed\\_solutions\\_final\\_508.pdf](https://www.acf.hhs.gov/sites/default/files/documents/opre/li2_innovate_co_creating_evidence_informed_solutions_final_508.pdf)

This brief provides an overview of the Learn, Innovate, Improve (LI2) process.

Higman, S., Miller, K., Till, L., Atukpawu-Tipton, G., Zaid, S., & Clark, M. (2020). *Community readiness: A toolkit to support Maternal, Infant, and Early Childhood Home Visiting Program awardees in assessing community capacity* (OPRE Report No. 2020-05). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. [https://www.acf.hhs.gov/sites/default/files/documents/opre/community\\_readiness\\_toolkit\\_jan\\_2020.pdf](https://www.acf.hhs.gov/sites/default/files/documents/opre/community_readiness_toolkit_jan_2020.pdf)

This toolkit helps Maternal, Infant, and Early Childhood Home Visiting Program awardees complete their community readiness assessment as part of their requirement to conduct a state- or territory-wide needs assessment.

Meckstroth, A., Resch, A., McCay, J., Derr, M., Berk, J., & Akers, L. (2015). *Advancing evidence-based decision making: A toolkit on recognizing and conducting opportunistic experiments in the family self-sufficiency and stability policy area* (OPRE Report No. 2015-97). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/advancing-evidence-based-decision-making-toolkit-recognizing-and-conducting>

This report describes in detail how researchers, policymakers, and program administrators can recognize opportunities for experiments and carry them out.

OPRE (Office of Planning, Research, and Evaluation). (2016). *What works, under what circumstances, and how? Methods for unpacking the "black box" of programs and policies* (OPRE Report No. 2016-54). U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/report/what-works-under-what-circumstances-and-how>

The brief considers methods and designs that move beyond questions about whether programs and policies work but also address questions about which particular parts work, under what circumstances, and how.

Whitesell, N. (2017). *Evidence and equity: Challenges in research design* (OPRE Report No. 2017-76). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://www.acf.hhs.gov/opre/report/evidence-and-equity-challenges-research-design>

This brief discusses research disparities between distinct groups and present strategies to address research disparities.

## Professional Associations

Table A.1. Professional Associations and Descriptions

Professional Association	Website	Description
American Evaluation Association (AEA)	<a href="http://www.eval.org/">http://www.eval.org/</a>	AEA is a professional association of evaluators devoted to the application and exploration of evaluation as a profession. AEA has a listing of association members who are available for evaluation consulting.
American Sociological Association (ASA)	<a href="http://www.asanet.org/">http://www.asanet.org/</a>	ASA is the national professional membership association for sociologists and others interested in sociology. ASA members include students, faculty working in a full range of institutions, and people working in government agencies and nonprofit and private sector organizations.
Association for Public Policy Analysis and Management (APPAM)	<a href="http://www.asanet.org/">http://www.asanet.org/</a>	APPAM is dedicated to improving public policy and management by fostering excellence in research, analysis, and education.
National Legislative Program Evaluation Society (NLPES)	<a href="https://www.ncsl.org/legislators-staff/legislative-staff/program-evaluation.aspx">https://www.ncsl.org/legislators-staff/legislative-staff/program-evaluation.aspx</a>	One of nine professional staff associations connected with the National Conference of State Legislatures. NLPES includes employees and state legislative agencies engaged in program evaluation or performance auditing.
Society for Research in Child Development (SRCD)	<a href="https://www.srcd.org/">https://www.srcd.org/</a>	SRCD advances the developmental sciences and promotes the use of developmental research to improve human lives.

# Appendix B. Templates and Examples

## Logic Model

Table B.1. Sample Logic Model: Child Abuse Prevention Program

Inputs	Assumptions	Activities	Outputs	Immediate Outcomes	Subsequent Outcomes	Impacts
Program staff	Overview: Children of parents with substance use disorder (SUD) are at high risk for child maltreatment	Overview: Implement a program that addresses SUD and child maltreatment simultaneously	Overview: Serve 350 families over 3-year period	Overview: Improve overall quality of family functioning	Overview: Reduce SUD and child maltreatment in families	Overview: Eliminate SUD for all families; eliminate child maltreatment among families experiencing maltreatment
Program participants (parents and children)						
Program manuals and guides	The risk for child maltreatment will decrease if parents cease using substances	Provide SUD service referrals Provide home visiting services	285 parents receive at least 5 home visiting services	Parents enroll and engage in SUD services Parents reduce substance use	Parents complete SUD treatment and continue with support groups or other aftercare services	Eliminate SUD among participating parents
Offices and facilities	The risk for child maltreatment will decrease if parents develop effective parenting skills	Provide parent effectiveness education Establish parent support groups Offer child development education	300 parents attend 80% of education sessions and 2 support group meetings	Parents increase parenting skill knowledge Parents increase child development knowledge Parents understand intergenerational trauma	Parents use healthy parenting practices with children	Eliminate child maltreatment among participating parents
Recruitment and service delivery partners	Children of parents with SUD are at elevated risk for SUD	Provide refusal skill building classes to children Offer recreational activities to children	500 children attend 80% of skill building classes and 3 recreational activities	Children increase refusal skills Children form friendships within the group	Children increase in resilience and sense of purpose	Participating children stop intergenerational transfer of SUD
Program funding						

Context: Program is operating in Minneapolis; voluntarily serving children of parents with substance use disorders (SUDs) and their parents, recruited through attendance at SUD support groups. Community is currently experiencing significant increases in child maltreatment rates, partly attributed to opioid use disorder.

Table B.2. Worksheet: Logic Model

Inputs	Assumptions	Activities	Outputs	Immediate Outcomes	Subsequent Outcomes	Impacts

Context: Program is operating in Minneapolis; voluntarily serving children of parents with substance use disorders (SUDs) and their parents, recruited through attendance at SUD support groups. Community is currently experiencing significant increases in child maltreatment rates, partly attributed to opioid use disorder.



## Implementation Objectives Stated in Measurable Terms

Table B.3. Worksheet: Describing Program Implementation Objectives in Measurable Terms

How You Know Planned Activity Occurred	Who Will Do It	What Population You Reach	How Many Individuals You Will Reach

## Participant Outcome Objectives Stated in Measurable Terms

Table B.4. Worksheet: Participant Outcome Objectives Stated in Measurable Terms

Expected Change	How Change Is Expected to Occur	For Whom Expected Changed Will Occur	How You Will Know Expected Change Occurred

## Analyzing Information on Implementation Objectives

Table B.5. Worksheet: Analyzing Information on Implementation Objectives

Implementation Objective	Actual Implementation	Differences? (Yes/No)	If Yes, Reasons for Change	Barriers Encountered	Facilitating Factors

# Final Report Outline

## Sample Outline: Final Evaluation Report

- I. Introduction: General Description of the Project
  - A. Description of program components, including services or training delivered and target population for each service
  - B. Description of collaborative efforts (if relevant), including the agencies participating in the collaboration and their various roles and responsibilities in the project
  - C. Description of strategies for recruiting program participants (if relevant)
  - D. Description of special issues relevant to serving the project's target population (or providing education and training to participants) and plans to address them
    1. Agency and staffing issues
    2. Participants' cultural background, socioeconomic status, literacy levels, and other characteristics
- II. Evaluation of Program Implementation Objectives
  - A. Description of the project's implementation objectives (measurable objectives)
    1. What you planned to do (planned services/interventions/training/education; duration and intensity of each service/intervention/training period)
    2. Whom you planned to have do it (planned staffing arrangements and qualifications/ characteristics of staff)
    3. Target population (intended characteristics and number of members of the target population to be reached by each service/intervention/training/ education effort and how you planned to recruit participants)
    4. Description of the project's objectives for collaborating with community agencies
      - a. Planned collaborative arrangements
      - b. Services/interventions/training provided by collaborating agencies
  - B. Statement of evaluation questions (Were program implementation objectives attained? If not, why not? What were the barriers to and facilitators of attaining implementation objectives?)

### Examples

- How successful was the project in implementing a parenting education class for mothers with substance abuse problems? What were the policies, practices, and procedures used to attain this objective? What were the barriers to, and facilitators of, attaining this objective?
- How successful was the project in recruiting the intended target population and serving the expected number of participants? What were the policies, practices, and procedures used to recruit and maintain participants in the project? What were the barriers to, and facilitators of, attaining this objective?

- How successful was the project in developing and implementing a multidisciplinary training curriculum? What were the practices and procedures used to develop and implement the curriculum? What were the barriers to, and facilitators of, attaining this objective?
  - How successful was the project in establishing collaborative relationships with other agencies in the community? What were the policies, practices, and procedures used to attain this objective? What were the barriers to, and facilitators of, attaining this objective?
- C. Description of data collection methods and data collected for each evaluation question
1. Description of data collected
  2. Description of methodology of data collection
  3. Description of data sources (such as project documents, project staff, project participants, and collaborating agency staff)
  4. Description of sampling procedures
- D. Description of data analysis procedures
- E. Description of results of analysis
1. Statement of findings with respect to each evaluation question

### Examples

- The project's success in attaining the objective
  - The effectiveness of particular policies, practices, and procedures in attaining the objective
  - The barriers to and facilitators of attainment of the objective
2. Statement of issues that may have affected the evaluation's findings

### Examples

- The need to make changes in the evaluation because of changes in program implementation or characteristics of the population served
- Staff turnover in the project resulting in inconsistent data collection procedures
- Changes in evaluation staff

### III. Evaluation of Participant Outcome Objectives

- A. Description of participant outcome objectives (in measurable terms)
1. What changes were participants expected to exhibit as a result of their participation in each service/intervention/training module provided by the project?
  2. What changes were participants expected to exhibit as a result of participation in the project in general?

3. What changes were expected to occur in the community's service delivery system as a result of the project?
- B. Statement of evaluation questions, evaluation design, and method for assessing change for each question

## Examples

- How effective was the project in attaining its expected outcome of decreasing parental substance abuse? How was this measured? What design was used to establish that a change occurred and to relate the change to the project's interventions (such as preintervention and postintervention, control groups, comparison groups, etc.)? Why was this design selected?
  - How effective was the project in attaining its expected outcome of increasing children's self-esteem? How was this measured? What design was used to establish that a change occurred and to relate the change to the project's interventions? Why was this design selected?
  - How effective was the project in increasing the knowledge and skills of training participants? How was this measured? What design was used to establish that a change occurred and to relate the change to the project's interventions? Why was this design selected?
- C. Discussion of data collection methods (for each evaluation question)
1. Data collected
  2. Method of data collection

## Examples

- ▶ Case record reviews
  - ▶ Interviews
  - ▶ Self-report questionnaires or inventories (if you developed an instrument for this evaluation, attach a copy to the final report)
  - ▶ Observations
3. Data sources (for each evaluation question) and sampling plans, when relevant
- D. Discussion of issues that affected the outcome evaluation and how they were addressed
1. Program-related issues
    - a. Staff turnover
    - b. Changes in target population characteristics
    - c. Changes in services/interventions during the project
    - d. Changes in staffing plans
    - e. Changes in collaborative arrangements
    - f. Characteristics of participants

2. Evaluation-related issues
  - a. Problems encountered in obtaining participant consent
  - b. Change in numbers of participants served requiring change in analysis plans
  - c. Questionable cultural relevance of evaluation data collection instruments and/or procedures
  - d. Problems encountered due to participant attrition
- E. Procedures for data analyses
- F. Results of data analyses
  1. Significant and negative analyses results (including statement of established level of significance) for each outcome evaluation question
  2. Promising, but inconclusive analyses results
  3. Issues/problems relevant to the analyses

## Examples

- Issues relevant to data collection procedures, particularly consistency in methods and consistency across data collectors
  - Issues relevant to the number of participants served by the project and those included in the analysis
  - Missing data or differences in size of sample for various analyses
- G. Discussion of results
    1. Interpretation of results for each evaluation question, including any explanatory information from the process evaluation
      - a. The effectiveness of the project in attaining a specific outcome objective
      - b. Variables associated with attainment of specific outcomes, such as characteristics of the population, characteristics of the service provider or trainer, duration, or intensity of services or training, and characteristics of the service or training
    2. Issues relevant to interpretation of results
- IV. Integration of Process and Outcome Evaluation Information
    - A. Summary of process evaluation results
    - B. Summary of outcome evaluation results
    - C. Discussion of potential relationships between program implementation and participant outcome evaluation results

## Examples

- Did particular policies, practices, or procedures used to attain program implementation objectives have different effects on participant outcomes?

- How did practices and procedures used to recruit and maintain participants in services affect participant outcomes?
- What collaboration practices and procedures were found to be related to attainment of expected community outcomes?
- Were particular training modules more effective than others in attaining expected outcomes for participants? If so, what were the features of these modules that may have contributed to their effectiveness (such as characteristics of the trainers, characteristics of the curriculum, the duration and intensity of the services)?

#### V. Recommendations to Program Administrators or Funders for Future Program and Evaluation Efforts

##### Examples

- Based on the evaluation findings, it is recommended that the particular service approach developed for this program be used to target mothers who are 25 years of age or older. Younger mothers do not appear to benefit from this type of approach.
- The evaluation findings suggest that traditional educational services are not as effective as self-esteem building services in promoting attitude changes among adolescents regarding substance abuse. We recommend that future program development focus on providing these types of services to youth at risk for substance abuse.
- Based on the evaluation findings, it is recommended that funders provide sufficient funding for evaluation that permit a long-term follow-up assessment of participants. The kinds of participant changes that the program may bring about may not be observable until 3 or 6 months after they leave the program.



Final Report

OPRE Report No. 2022-208

October 2022

Shereen El Mallah, Lila Gutuskey, Allison Hyra, Amanda Hare, Rachel Holzwart, & Chrissy Steigelman,  
Insight Policy Research

Submitted to:

Kriti Jain, Contracting Officer's Representative  
Harmanpreet Bhatti, Task Lead  
Office of Planning, Research, and Evaluation  
Administration for Children and Families  
U.S. Department of Health and Human Services

This report is in the public domain. Permission to reproduce is not necessary.

Suggested citation:

El Mallah, S., Gutuskey, L., Hyra, A., Hare, A., Holzwart, R., & Steigelman, C. (2022). *The program manager's guide to evaluation (OPRE Report No. 2022-2008)*. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

*Disclaimer*

The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research, and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

[www.acf.hhs.gov/opre](http://www.acf.hhs.gov/opre)

## Connect with OPRE

