AFTERSCHOOL EVALUATION 101:

How to Evaluate an Expanded Learning Program

Erin Harris
Harvard Family Research Project

Harvard Family Research Project

Afterschool Evaluation 101

Table of Contents

Introduction

This toolkit is intended to help program directors and staff of afterschool, summer, and other out-of-school time (OST) programs develop an evaluation strategy.

THE IMPORTANCE OF EVALUATION AND DEVELOPING AN EVALUATION STRATEGY

What is an evaluation?

Evaluation helps your OST program measure how successfully it has been implemented and how well it is achieving its **goals**. You can do this by comparing:

The activities you intended to implement	▶	*The activities you actually implemented*
The outcomes you intended to accomplish	▶	*The outcomes you actually achieved*

For example, if you have created an OST program that highlights recreation and nutrition education, with the goal of improving children's abilities to make healthy choices, an evaluation can help you determine if your program has been implemented in a way that will lead to positive **outcomes** related to your goals (e.g., whether activities offer opportunities for physical activity or learning about better nutrition). In addition, an evaluation can help determine whether participants are actually making healthier choices, such as choosing recess games that involve physical activity, or selecting fruit as a snack rather than cookies.

Who should conduct an evaluation?

Every OST program can and should collect at least some basic information to help evaluate its success. Even a new program that is just getting up and running can begin to evaluate who is participating in the program, and why. A more established program, meanwhile, can evaluate outcomes for participants related to program goals. And a program that is going through a transition in its goals, activities, and focus can use evaluation to test out new strategies.

Why conduct an evaluation?

Information gathered during an evaluation helps demonstrate your program's effectiveness and provides valuable insight into how the program can better serve its population. Many programs use evaluations as an opportunity to identify strengths, as well as areas that need improvement, so they can learn how to improve services. In addition, grantmakers often require evaluations of the programs they invest in to ensure that funds are well spent. Positive outcomes for program participants (such as improved academic performance or fewer discipline problems) found during an evaluation can also help "sell" your program to new funders, families, communities, and others who may benefit from, or provide benefits to, your program.

What is an evaluation strategy?

An evaluation strategy involves developing a well-thought out plan for evaluating your program, with the goal of incorporating the lessons learned from the evaluation into program activities. As part of this larger strategy, evaluation is not viewed merely as a one-time event to demonstrate results, but instead as an important part of an ongoing process of **learning and continuous improvement**.

Why create an evaluation strategy?

Creating an evaluation strategy can help you to create a plan for evaluation that can both serve the funders' requirements and also inform efforts to improve your program. An evaluation strategy can help your staff recognize the evaluation as a beneficial process, rather than as an added burden imposed by funders. Even if the evaluation results suggest room for improvement, the fact that the program is collecting such data indicates a commitment to learning and continuous improvement and gives a positive impression of the program's potential.

How do we conduct a program evaluation?

There is no single recipe for conducting a program evaluation. OST programs use a variety of evaluation approaches, methods, and measures—both to collect data for program improvement

and to demonstrate the effectiveness of their programs. It is up to leaders at individual program sites to determine the best approach: one suited to the program's **developmental stage**, the needs of program participants and the community, and funder expectations. One common challenge is developing an evaluation approach that satisfies the different interests and needs of the various **stakeholders**: Funders may want one thing, while parents and program staff want another. Regardless of the approach, your evaluation strategy should be designed within a larger question of how the data you collect can be used to shape and improve your program's activities.

HOW TO USE THIS TOOLKIT

This toolkit stresses the need to create a larger evaluation strategy to guide your evaluation plans. This toolkit will walk you through the steps necessary to plan and implement an evaluation strategy for your OST program.

This toolkit is structured in a series of nine steps:

- Step 1 helps you to determine the overall purpose of your evaluation.
- Step 2 outlines how to create a **logic model**, which is a visual representation of your program strategy that can guide your evaluation.
- Step 3 describes how to think through what resources you have available (staffing, etc.) to actually conduct an evaluation.
- Step 4 discusses how best to focus your evaluation, based on your program's needs, resources, and developmental stage.
- Steps 5 and 6 cover selecting the evaluation design and data collection methods that are best suited to your program.
- Steps 7, 8, and 9 contain information about what to do with the data once you have it, including how to conduct and write up the analysis and, perhaps most importantly, how to use the data that you have analyzed.

Note that each step is designed to build on the last; however, you may prefer to skip ahead to a specific topic of interest. This toolkit also includes a set of discussion questions related to each section that you and others involved in the evaluation may want to consider. In addition, there is a **Glossary** of evaluation terms (words are denoted in **bold blue text**) included in the **Appendix** at the end of this document.

OTHER EVALUATION RESOURCES

This toolkit offers the basics in thinking about and beginning to plan an evaluation strategy for your OST program; you may also want to consult additional resources when it is time to implement your evaluation. Beyond this toolkit, Harvard Family Research Project has several other resources and publications that you may find helpful:

- *The Out-of-School Time Research and Evaluation Database* allows you to search through profiles written about evaluations and **research studies** conducted of OST programs and initiatives. You can search the database by program type to find programs similar to your own that have conducted evaluations, or by methodology to see examples of various types of evaluation methods in practice.

- *Measurement Tools for Evaluating OST Programs* describes instruments used by OST programs to evaluate their implementation and outcomes. This resource can provide ideas for possible data collection instruments to use or adapt for your program.

- *Detangling Data Collection: Methods for Gathering Data* provides an overview of the most commonly used data collection methods and how they are used in evaluation.

- *Performance Measures in Out-of-School Time Evaluation* provides information about the **performance measures** that OST programs have used to document progress and measure **results** of academic, youth development, and prevention outcomes.

- *Learning From Logic Models in Out-of-School Time* offers an in-depth review of logic models and how to construct them. A logic model provides a point of reference against which progress towards achievement of desired outcomes can be measured on an ongoing basis through both performance measurement and evaluation.

These and other related HFRP resources are referenced within the relevant sections of this toolkit for those who want additional information on that topic. This toolkit also contains a list of key resources for evaluation tools developed by others that we have found especially helpful. You will find a link to this resource list on each page.

RELATED RESOURCES FROM HFRP.ORG

- *A Conversation With Gary Henry*
- *Why Evaluate Afterschool Programs?*
- *Why, When, and How to Use Evaluation: Experts Speak Out*

Evaluation Tip:

If you are conducting an evaluation for accountability to funders, consider incorporating it into a strategy that also allows you to use the data you collect for learning and continuous improvement, rather than just for the purpose of fulfilling a funder mandate.

Evaluation Tip:

Having a discussion with everyone involved in your program to clarify goals together can help everybody, including staff, create a shared understanding of your program's purpose.

STEP 1: Determining the Evaluation's Purpose

Programs conduct evaluations for a variety of reasons. It is important for your program to determine (and be in agreement about) the various reasons that you want or need an evaluation before beginning to plan it. There are four questions that are essential to consider in determining the purpose of your evaluation:

- Why should we evaluate our program?
- What is our program trying to do?
- What information do our funders expect?
- What questions do we want to answer?

1. WHY SHOULD WE EVALUATE OUR PROGRAM?

In planning for an evaluation, start by thinking about why you want to conduct an evaluation. What are the benefits and how will it be used? An evaluation should be conducted for a specific purpose, and all parties involved should have a clear understanding of this purpose as they start the evaluation planning.

OST programs usually conduct an evaluation for one, or both, of two primary reasons:

- *To aid learning and continuous improvement.* Rather than being merely a static process where information is collected at a single point in time, an evaluation can become a practical tool for making ongoing program improvements. Evaluation data can help program management make decisions about what is (and isn't) working, where improvement is needed, and how to best allocate available resources.
- *To demonstrate accountability.* Evaluation data can be used to demonstrate to current funders that their investments are paying off.

There are also two secondary purposes that may drive evaluations:

- *To market your program.* Evaluation results, particularly regarding positive outcomes for youth participants, can be used in marketing tools—such as brochures or published reports—to recruit new participants and to promote the program to the media as well as to potential funders and community partners.
- *To build a case for sustainability.* Evaluation results can illustrate your program's **impact** on participants, families, schools, and the community, which can help to secure funding and other resources that will allow your program to continue to operate.

Be clear from the beginning about why you are conducting the evaluation and how you plan to use the results.

2. WHAT IS OUR PROGRAM TRYING TO DO?

One of the initial steps in any evaluation is to define program goals and how services aim to meet these goals. If you are creating an evaluation for an already-established program, chances are that the program's goals, inputs, and outputs are already defined (see more detailed descriptions of **goals**, **inputs**, and **outputs** in Step 2). If not, or if you are starting a new program, these elements will need to be determined. Many established programs can also benefit from revisiting their existing goals, inputs, and outputs, and tweaking them as necessary to ensure that the program is as effective as possible.

As you will learn in Step 2, your inputs, outputs, and goals should all have a logical connection to one another. So, for example, if your program aims to improve academic outcomes, your activities should include a focus on academic instruction or support, such as homework help or a program curriculum that is designed to complement in-school learning.

A useful approach to goal-setting is the development of a logic model. Step 2 of this toolkit defines this term, and explores how to develop a logic model and also how to use it for your

evaluation. Whether or not you choose to develop a logic model, it is crucial that your program have clearly articulated goals and objectives in order to determine the evaluation's purpose and focus.

3. WHAT INFORMATION DO OUR FUNDERS EXPECT?

Many programs enter the evaluation process with specific evaluation requirements. Funders may require programs to collect information about participants, their families, and the services they receive. Some funders also impose specific time frames, formats, and dissemination procedures for reporting results. For example, a funder may tell you that you need to produce a 10-to-15-page evaluation report covering the first year of funding that describes program implementation successes and challenges, and that the report should be made publicly available in some way (such as posting it on your own or the funder's website).

The following strategies can help programs negotiate with funders about evaluation requirements:

- ***Work with funders to clarify what information they expect and when they expect it.*** Maintain a continuing dialogue with funders about the kinds of information they are interested in and would find useful. Ask funders how evaluation results will be used—for example, an evaluation of the first year of a program's operation might be used only to establish a **baseline**. Find out if the evaluation is intended to be **formative** (providing information that will strengthen or improve your program) or **summative** (judging your program's outcomes and overall effectiveness).

- ***Allow for startup time for your program before investing in an evaluation***. A program must be established and running smoothly before it is ready for a formal evaluation. In most cases, a program will not be ready for a full-blown evaluation in its first year of implementation, since the first year tends to involve a lot of trial and error, and refinements to program strategies and activities. These refinements may even extend to a second or third year, depending on the program. Startup time can be used for other evaluation-related tasks, however, such as conducting a **needs assessment** (see Step 4) and collecting background data on the population targeted by your program.

- ***Think collaboratively and creatively about effective evaluation strategies.*** Funders can often provide valuable insights into how to evaluate your program, including suggestions on how to collect data to help with program improvement.

- Work with the funder to set realistic expectations for evaluation based on how long your program has been in operation—that is, the program's **developmental stage**. Be explicit about reasonable time frames; for instance, it is unrealistic to expect progress on long-term participant outcomes after only one year of program participation.

- ***Try to negotiate separate funds for the evaluation component of your program.*** Be sure to note any additional staffing and resources needed for evaluation.

- ***Work with funders to create an evaluation strategy that satisfies their needs, as well as your program's needs.*** Evaluation should be seen as part of a larger learning strategy, rather than just a one-time activity demonstrating accountability to funders. If the funders are truly supportive of your program, they should welcome an evaluation plan that includes collecting data that can be used for program improvements.

If your program has multiple funders that require you to conduct an evaluation, negotiating an evaluation strategy that meets all of their needs is likely to be an additional challenge. Addressing the issues outlined above when planning your evaluation strategy can help you to better navigate diverse funder requirements.

4. WHAT QUESTIONS DO WE WANT TO ANSWER?

Before beginning an evaluation, you must decide which aspects of your program you want to focus on in order to ensure that the results will be useful. Answering the questions below can help you to form evaluation questions that are realistic and reasonable, given your program's mission and goals. Ask yourself the following:

> **Evaluation Tip:**
>
> Set ambitious but realistic expectations for the evaluation. Do not give in to funder pressure to provide results that you think you are unlikely to be able to deliver just to appease the funder.

Evaluation Tip:

Use what you already know to help
shape your evaluation questions. If
you have done previous evaluations
or needs assessments, build on what
you learned to craft new evaluation
questions and approaches. You can
also use existing data sets, such as
those from Kids Count and the U.S.
Census Bureau, as well as other
programs' evaluations and existing
youth development research. See *Key
Resources* for more information about
existing data sets.

• What is our motivation for conducting an evaluation? That is, are we primarily motivated by a desire for learning and continuous improvement, funder accountability, marketing, building a case for sustainability, or some combination of these? (As outlined in *Why should we evaluate our program?*)

• What data can we collect that will assist learning and continuous improvement within our program?

• Are we required by a funder to follow a specific reporting format?

• What is our time frame for completing the evaluation?

• How can we gain support for the evaluation from program staff, schools, and others with an interest in our program, and how do we involve them in the process?

• How can we provide information that is useful to our stakeholders?

All programs should at least examine program **participation data** to determine the demographics of those who participate, how often they participate, and whether they remain in the program. Step 6 provides guidance on how (and why) to start collecting participation data.

Depending on your responses to the above questions, evaluation questions you may want to consider include:

• Does our program respond to participant needs?

• What are the costs of our program?

• Who staffs our program? What training do they need?

• What services does our program provide? How can we improve these services?

• What is our program's impact on youth participants' academic, social, and physical well-being?

• What roles do families and the larger community play in the program, and how does the program benefit them?

STEP 2: Developing a Logic Model

In designing an evaluation, it is important to have a clear understanding of the goals of the program to be evaluated, and to be realistic about expectations. A logic model is a useful tool that will assist you in defining program goals and figuring out the focus of the evaluation. This section addresses the following questions:

- What is a logic model and why should we create one?
- What does a logic model look like?
- How do we define our goals?
- How do we define our inputs and outputs?
- How do we identify our outcomes?
- How do we develop performance measures?

1. WHAT IS A LOGIC MODEL AND WHY SHOULD WE CREATE ONE?

A **logic model** is a concise way to show how a program is structured and how it can make a difference for a program's participants and community. It is a one-page visual presentation—often using graphical elements such as charts, tables, and arrows to show relationships—that displays:

- The key elements of a program (i.e., its activities and resources).
- The rationale behind the program's service delivery approach (i.e., its goals).
- The intended results of the program and how they can be measured (i.e., the program's outcomes).
- The cause-and-effect relationships between the program and its intended results.

A logic model also can help identify the core elements of an evaluation strategy. Like an architect's scale model of a building, a logic model is not supposed to be a detailed "blueprint" of what needs to happen. Instead, the logic model lays out the major strategies to illustrate how they fit together and whether they can be expected to add up to the changes that program stakeholders want to see.

Creating a logic model at the beginning of the evaluation process not only helps program leaders think about how to conduct an evaluation, but also helps programs choose what parts of their program (e.g., which activities and goals) they want to evaluate; it is one of the key methods used to assist organizations in tracking program progress towards implementing activities and achieving goals. Since creating a logic model requires a step-by-step articulation of the goals of your program and the proposed activities that will be conducted to carry out those goals, a logic model can also be helpful in charting the resources your program needs to carry out the evaluation process.

2. WHAT DOES A LOGIC MODEL LOOK LIKE?

Although logic models can be put together in a number of different ways, the following components are important to consider when constructing your logic model:

Goals—what your program ultimately hopes to achieve. Sometimes organizations choose to put their goals at the end of the logic model to show a logical progression. However, your goals should drive the rest of your logic model, and for that reason, you may want to consider putting the goals right at the beginning.

Inputs—the resources at your program's disposal to use to work toward program goals These resources include such supports as your program's staff, funding resources, and community partners.

Outputs—the services that your program provides to reach its goals. These services will primarily consist of the program activities offered to youth participants, although you may also offer other

RELATED RESOURCES FROM HFRP.ORG

- *Eight Outcome Models*
- *An Introduction to Theory of Change*
- *Learning From Logic Models in Out-of-School Time*
- *Evaluating Complicated—and Complex—Programs Using Theory of Change*
- *Using Narrative Methods to Link Program Evaluation and Organization Development*

RELATED RESOURCES FROM HFRP.ORG

- *Logic Models in Out-of-School Time*
- *Logic Model Basics*
- *Project HOPE: Working Across Multiple Contexts to Support At-Risk Students*
- *Theory of Action in Practice*
- *Logic Models in Real Life: After School at the YWCA of Asheville*

services—such as activities aimed at families and communities—that also will be part of your outputs. As part of this step, it is important to have a clear picture of the specific target population for your activities (e.g., girls, low-income youth, a specific age range, youth living in a specific community).

Outcomes—your program's desired short-term, intermediate, and long-term results. Generally, short-term outcomes focus on changes in knowledge and attitudes, intermediate outcomes focus on changes in behaviors, and long-term outcomes tend to focus on the larger impact of your program on the community.

Performance measures—the data that your program collects to assess the progress your program has made toward its goals. These data include:

- **Measures of effort**, which describe whether and to what extent outputs were implemented as intended (e.g., the number of youth served in your program, the level of youth and parent satisfaction with the program).

- **Measures of effect**, which convey whether you are meeting your outcomes (e.g., improvements in youth participants' skills, knowledge, and behaviors).

The table below provides a sample logic model for an OST program that focuses on providing academic support to youth.

TABLE 1: Example of a logic model for an academically focused OST program

Goals	Inputs	Outputs	Outcomes	Performance measures
Improve the academic development and performance of at-risk students.	• Program staff • Funding • School & community partners	**Activities:** • Academic enrichment • Homework help/tutoring **Target population:** Children in the local community classified as "at risk for academic failure" based on family income and poor school performance.	**Short-Term:** • Greater interest in school **Intermediate:** • Improved academic grades and test scores **Long-Term:** • Higher graduation and college attendance rates	**Measures of effort:** • Number of youth served in the program • Number of sessions held • Level of youth and parent satisfaction with the program **Measures of effect:** • Changes in participants' academic grades • Test scores • Graduation rates • College attendance rates

For examples of specific programs' logic models, see:

- *Project HOPE: Working Across Multiple Contexts to Support At-Risk Students*

- *Theory of Action in Practice*

- *Logic Models in Real Life: After School at the YWCA of Asheville*

3. HOW DO WE DEFINE OUR GOALS?

While the goals are the results that your program aims to achieve, they must be considered from the beginning since all of the other pieces of the logic model should be informed by those goals. To determine your goals, ask yourself: What are we ultimately trying to achieve with our program?

For established programs, the goals may already be in place, but it is important to ensure that the goals are clear and realistic, and that there is a common understanding and agreement about these goals across program stakeholders. Revisiting existing goals can allow time for reflection, as well as possible refinement of the goals to better fit the program's current focus and stakeholders' current interests.

4. HOW DO WE DEFINE OUR INPUTS AND OUTPUTS?

Once you have a clear set of goals for your program, you can start to think about the elements of your program—that is, the inputs and outputs—that help your program to achieve its goals. These inputs and outputs should have a direct link to your goals. For example, if your primary goal relates to academic achievement, your program should include outputs or activities that are directly working to improve academic achievement (e.g., tutoring or enrichment).

In defining your **inputs** (resources), consider the following:

- What resources are available to our program—both program infrastructure, such as staffing and funding, and existing community resources, such as social service supports for families— that we can use to work toward our goals?

- Are there additional inputs that we need to have in place in order to implement our program?

In defining your **outputs** (activities and target population), consider the following:

- What activities should our program offer (tutoring, sports, etc.) to best meet our goals?

- Can we implement these activities with the inputs (resources) available?

- Whom do we hope to serve in our program? What ages, demographics, neighborhoods, etc., do we target?

- Should we serve families and/or the larger community in addition to youth? If so, how?

- Does our target participant population align with the demographics of the local community?

5. HOW DO WE IDENTIFY OUR OUTCOMES?

While goals express the big-picture vision for what your program aims to accomplish, outcomes are the "on-the ground" impacts that your program hopes to achieve. For example, if your goal is to improve the academic development and performance of at-risk students, your short-term intended outcome might be to increase students' interest in school, with the long-term anticipated outcome of higher graduation and college attendance rates. As such, you should select outcomes that are SMART: **S**pecific, **M**easurable, **A**ction-oriented, **R**ealistic, and **T**imed. The smarter your outcomes are, the easier it will be to manage performance and assess progress along the way.

Specific. To be useful and meaningful, outcomes should be as specific as possible. For example, an outcome of "improved academic achievement" is somewhat vague—what does this mean? Clearer outcomes for academic achievement could include improved test scores, grades, or graduation rates.

Measurable. Without the ability to measure your outcomes, you have no way of really knowing if, or how much, your program has an impact on outcomes. Outcomes that are specific are also more likely to be measurable. For instance, outcomes in the example above—improved test scores, grades, and graduation rates—all have data collected by schools that can be used to track progress toward these outcomes.

Action-Oriented. Outcomes are not passive by-products of program activities—they require ongoing effort to achieve. If your program is not pursuing activities aimed at producing a specific outcome, it is not reasonable to expect that outcome to result from your program. So for example, an outcome of increased family involvement in children's learning should be accompanied by a program component in which program staff actively seek out parent support and engagement.

Realistic. Outcomes should be something that your program can reasonably expect to accomplish, or at least contribute to, with other community supports. The primary consideration in identifying realistic outcomes is how well the outcomes are aligned with and linked to your activities so that there is a logical connection between your efforts and what you expect to change as a result of your work. For example, a program that does not have a goal of increasing academic achievement should not include outcomes related to participant test scores or grades.

Timed. Logic models allow the designation of short-term, intermediate, and/or long-term outcomes. Short-term and intermediate outcomes for OST programs tend to focus on the changes

Evaluation Tip:

Consider the intensity and duration of your activities in determining whether outcomes are realistic. For example, a once-a-week physical activity component is unlikely to show dramatic decreases in youth obesity.

that you can expect to see in program participants after a year (or more) of participation, such as improved grades or changes in knowledge or attitudes. Long-term goals tend to involve larger changes in the overall community that is being served, or changes in participants that may not be apparent right away, such as increased high school graduation rates or changes in behavior.

6. HOW DO WE DEVELOP PERFORMANCE MEASURES?

The final stage of logic model development is to define your program's **performance measures**. These measures assess your program's progress on the implementation of inputs and outputs. While outcomes lay out what your program hopes to accomplish as a whole, performance measures should be narrower in scope to provide measurable data for evaluation.

There are two types of performance measures:

• *Measures of effort* assess the effectiveness of your outputs. They assess how much you did, but not how well you did it, and are the easiest type of evaluation measure to identify and track. These measures address questions such as: What activities does my program provide? Whom does my program serve? Are program participants satisfied?

• *Measures of effect* are changes that your program—acting alone or in conjunction with partners (e.g., schools or other community organizations)—expects to produce in knowledge, skills, attitudes, or behaviors. These measures address questions such as: How will we know that the children or families that we serve are better off? What changes do we expect to result from our program's inputs and activities?

Strong performance measures should:

• *Have strong ties to program goals, inputs, and outputs.* There should be a direct and logical connection between your performance measures and the other pieces of your logic model. Ask yourself: What do we hope to directly affect through our program? What results are we willing to be directly accountable for producing? What can our program realistically accomplish?

• *Be compatible with the age and stage of your program.* Performance measures should be selected based on your program's current level of maturity and development. For example, a program in its first year should focus more on measures of effort than on measures of effect to ensure that the program is implemented as intended before trying to assess outcomes for participants.

• *Consider if the data you need are available/accessible.* Performance measures should never be selected solely because the data are readily available. For example, if your program does not seek to impact academic outcomes, it does not make sense to examine participants' grades. That said, you should think twice before selecting program performance measures for which data collection will be prohibitively difficult and/or expensive. For example, do not choose performance measures that require access to school records if the school will not provide access to these data.

• *Yield useful information to the program.* Consider the question: "Will the information collected be useful to our program and its stakeholders?" The answer should always be a resounding "Yes." To determine whether the data will be useful, consider the purpose of your evaluation and what you hope to get out of it, as outlined in Step 1.

For more information on performance measures used by afterschool programs, see our OST Evaluation Snapshot: *Performance Measures in Out-of-School Time Evaluation*.

STEP 3: Assessing Your Program's Capacity for Evaluation

Once you have determined why you want to conduct an evaluation and have developed a logic model for your program, you should consider your program's capacity for evaluation.

Questions addressed in this section include:

- Who should be involved in the process?
- What resources must be in place?
- Who will conduct our evaluation?
- How do we find an external evaluator?

RELATED RESOURCES FROM HFRP.ORG

- *Pizza, Transportation, and Transformation: Youth Involvement in Evaluation and Research*
- *Issues and Opportunities in OST Evaluation: Youth Involvement in Evaluation & Research*

Evaluation Tip:

Pay particular attention to the needs of youth participants and their families in planning for evaluation. As the program's "end users," they provide an especially valuable perspective on the program.

1. WHO SHOULD BE INVOLVED IN THE PROCESS?

The input of all program stakeholders is critical in planning an evaluation strategy. Stakeholders are those who hold a vested interest in your program. They include anyone who is interested in or will benefit from knowing about your program's progress, such as board members, funders, collaborators, program participants, families, school staff (e.g., teachers, principals, and superintendents), college or university partners, external evaluators, someone from the next school level (e.g., middle school staff for an elementary school-age program), and community partners.

Recognizing stakeholders' importance to the evaluation process right from the start is important for three reasons. First, such recognition can increase stakeholders' willingness to participate in the evaluation and help address concerns as they arise. Second, it can make stakeholders feel that they are a part of the project—that what they do or say matters. Lastly, it can make the final product richer and more useful to your program.

2. WHAT RESOURCES MUST BE IN PLACE?

Allocating resources to and funding an evaluation are critical to making evaluation a reality. Considering evaluation options involves assessing tradeoffs between what your program needs to know and the resources available to find the answers. Resources include money, time, training commitments, external expertise, stakeholder support, and staff allocation, as well as technological aids such as computers and **management information systems (MIS)**.

Resources for evaluation should be incorporated into all program funding proposals. Evaluation costs can be influenced by the evaluation's design, data collection methods, number of sites included, length of evaluation, use of an outside evaluator, availability of existing data, and type of reports generated. In requesting evaluation funding, you should have a clear idea of what your program plans to measure, why you chose particular data collection methods, and how progress will be monitored. The process of requesting funding for evaluation also helps programs determine what resources are needed for evaluation.

Many organizations can cover evaluation costs with resources from their program budgets. When program resources cannot support evaluations, organizations must find creative ways to obtain resources for evaluation. These resources can come from a range of interested community entities such as local businesses, school districts, and private foundations, and can include cash, professional expertise, and staff time. Universities interested in research can also make great partners and can provide expertise around design issues and assist with data collection and analysis.

Evaluation Tip:

You can always start small and tailor your evaluation to available resources. A small evaluation often allows for quick feedback, so you can use the results in a timely fashion. As programs expand and resources increase, you can also expand your evaluation activities.

3. WHO WILL CONDUCT OUR EVALUATION?

Sometimes program staff and other stakeholders have the skills and experience necessary to design and implement an evaluation. At times, however, programs need the design and analysis expertise of an outside consultant. Further, some funders strongly recommend or require that the evaluation be completed by an objective outsider.

There are pros and cons to consider when hiring an outside consultant. Issues to consider include the following:

Costs. Using an external evaluator is likely to be more costly than using in-house staff. However, an external evaluator can also save money if an in-house evaluator is inefficient or lacks time and expertise.

Loyalty. As an unbiased third party, an external evaluator is likely to be loyal to the evaluation process itself rather than particular people in the organization. Yet this lack of program loyalty can sometimes create conflict with program stakeholders, who may see the evaluator as an outsider who does not understand the program or its needs.

Perspective. An external evaluator may provide a fresh perspective with new ideas. However, this fresh perspective can also result in the external evaluator's inadvertently focusing on issues that are not essential to your program.

Time. External evaluators have the time to focus attention solely on the evaluation. Because they are not connected to the program, however, external evaluators may take a long time to get to know your program and its people.

Relationships. The external evaluator's outsider perspective can be beneficial in managing conflict resolution, but may also result in a lack of the trust necessary to keep lines of communication open and effective.

4. HOW DO WE FIND AN EXTERNAL EVALUATOR?

The use of an external evaluator should be considered in terms of your available resources: Do you have access to knowledgeable evaluators? How much time can program staff allocate to evaluation responsibilities? Does your program have the staff resources to develop in-house forms for evaluation purposes? What type of expertise will the evaluation really need, and how much can your program do itself?

If you do decide to hire an external evaluator, start by researching those recommended by people you know. The **American Evaluation Association** is also a good resource for identifying reputable evaluators. From these recommendations, identify those who give free consultations, have a reputation for the type of evaluation you want to do, and are able to work within your time frame and your budget. Finally, always ask for a proposal and a budget to help you make your decision and to be clear about what the evaluator will do. The following are some questions to ask evaluator candidates:

- How long have you been doing this work?

- What types of evaluations do you specialize in?

- What other organizations have you worked with?

- Can you show me samples of final evaluation reports?

- What role, if any, do you expect the program staff and administration to play in helping you to shape the evaluation process, including the evaluation questions, methods, and design?

- Have you ever worked with youth or families?

- Will you be the one working with us, or do you have partners?

- How do you communicate challenges, progress, and procedures?

- Do you conduct phone consultations or follow-up visits? If so, are extra costs involved?

- Can we review and respond to reports before they are finalized?

A good external evaluator can be expected to observe your program's day-to-day activities at length; be sensitive to the needs of participants, families, and staff; communicate readily and effectively with your program; inspire change and assess processes to implement change; identify program needs; and promote program ownership of the evaluation.

STEP 4: Choosing the Focus of Your Evaluation

Once you have determined why you want to do an evaluation, have assessed your capacity to do it, and have developed a logic model to identify what to evaluate, the next step is determine the evaluation's focus.

We recommend using an evaluation approach called the **five tier approach**. This section addresses the following issues related to choosing a starting point for evaluation:

- What is the Five Tier approach and why use it?

- How do we conduct a needs assessment (Tier 1)?

- How do we document program services (Tier 2)?

- How do we clarify our program (Tier 3)?

- How do we make program modifications (Tier 4)?

- How do we assess program impact (Tier 5)?

1. WHAT IS THE FIVE TIER APPROACH AND WHY USE IT?

As its name suggests, the Five Tier approach describes the evaluation process as a series of five stages, outlined in the table below.

TABLE 2: The Five Tier Approach to Evaluation

Evaluation Tier	Task	Purpose	Methods
Tier 1	Conduct a needs assessment	To address how the program can best meet the needs of the local community.	Determine the community's need for an OST program.
Tier 2	Document program services	To understand how program services are being implemented and to justify expenditures.	Describe program participants, services provided, and costs.
Tier 3	Clarify your program	To see if the program is being implemented as intended.	Examine whether the program is meeting its benchmarks, and whether it matches the logic model developed.
Tier 4	Make program modifications	To improve the program.	Discuss with key stakeholders how to use the evaluation data for program improvement.
Tier 5	Assess program impact	To demonstrate program effectiveness.	Assess outcomes with an experimental or quasi-experimental evaluation design.

The tier on which you begin your evaluation is largely determined by your program's developmental stage. For example, a new program should start with a Tier 1 evaluation, while an older, more established program might be ready to tackle Tier 5. Regardless of what tier you begin your evaluation on, you should first create a logic model (see Step 2) to assist you in defining program goals and figuring out the focus of the evaluation.

While there are many approaches to program evaluation, the Five Tier approach has several important benefits. First, all programs are able to do at least some evaluation using one of the five tiers—for instance, Tier 1 is something that every program can and should do to help ensure that your program is positioned to meet an identified need in the community. Second, the type of data that a program needs can change over time; therefore, the evaluation approach must be flexible enough to allow for this. Third, evaluation is an ongoing, cyclical process—feedback from one tier of the evaluation can be used to shape the next phase.

RELATED RESOURCES FROM HFRP.ORG

- *Evaluation Theory or What Are Evaluation Methods For?*

- *Mindset Matters*

2. HOW DO WE CONDUCT A NEEDS ASSESSMENT (TIER 1)?

The main task of Tier 1 is to do a needs assessment, which is an attempt to better understand how your program is meeting, or can meet, the needs of the local community. For new programs, a needs assessment can help in developing the program in ways that best fit the community's needs, and can also protect against the temptation to just provide services that are easy to implement rather than those services that children and their families actually need. For older programs, a needs assessment can serve as a check to be sure that your program is adequately addressing the community's needs, and can help build a case for program and service expansion.

A needs assessment can help answer the following questions:

- What services are already offered to the children and families in the community? Where are the gaps?
- What does our community want from an OST program?
- Does our target population match the local demographics?
- What are the potential barriers to implementing our program in the local community?

Community needs can be identified through conducting **surveys** of or **interviews** with OST stakeholders (community organizations and leaders, families, businesses, school leaders, etc.). You can also make use of existing statistics, research, and other data (e.g., census data or department of education information, evaluations or research studies of similar programs) to identify needs.

3. HOW DO WE DOCUMENT PROGRAM SERVICES (TIER 2)?

Tier 2 evaluation involves documenting the services your program provides in a systematic way, also called **program monitoring**. Program monitoring has two basic functions. One is to be able to track what your program funding is being spent on—many funders require this type of data. The other function of program monitoring is to describe the details of your program activities, including information about their frequency, content, participation rates, staffing patterns, staff training provided, and transportation usage.

Programs can use their program monitoring data to see if they are reaching their intended target population, to justify continued funding, and to build the capacity needed for program improvements. The data can also be the foundation on which later evaluations are built.

Documenting program services with Tier 2 evaluation helps to answer critical questions:

- What services or activities does our program offer?
- Is our program serving the intended population of children and their families? Are services tailored for different populations?
- Who staffs our program, and in what capacity? What additional staffing needs do we have?
- How are our funds spent?

Program services can be documented through intake forms that describe participant characteristics, forms that record program activities and participation rates, and records that track staff and their training.

4. HOW DO WE CLARIFY OUR PROGRAM (TIER 3)?

This tier of evaluation determines if your program is doing what it set out to do, as outlined in your logic model. Specifically, this tier involves examining what your program looks like in "real life" and how it operates on a day-to-day basis, and whether or not that matches up with how your program was envisioned.

These "real life" data can then be used to make adjustments to your program goals and activities to best meet the community's needs. For example, you may find that the target population of

children that your program aimed to serve are not the ones who can most benefit from the program, so you may want to rethink the activities that you are providing to better reach your target population (or rethink whom you should be targeting). The data can also provide feedback to program staff members on what they are doing well, and what needs improvement. As you look at the data, you may want to revise your logic model based on what you are learning.

In Tier 3 evaluation, the following questions are addressed:

- What were our program's intended activities? Were all activities implemented?
- Are the services offered appropriate to our program's targeted youth participants and their families? Are some youth excluded?
- What do participants think about program offerings? How will their feedback be used?
- How can our program do a better job of serving children and their families?

There are many ways to compare what a program intended to provide with what it actually provides. These methods include:

- Comparing program operations with logic model goals and objectives.
- Using self-assessment or an external evaluator to observe and rate program activities.
- Asking staff members and participants to keep a journal of their experiences with the program.
- Enlisting participating families to give feedback in small group sessions.
- Developing a participant/parent satisfaction survey.

5. HOW DO WE MAKE PROGRAM MODIFICATIONS (TIER 4)?

Having examined whether or not program services match intended program goals (Tier 3), you can begin to fill in the gaps revealed in your program and fine-tune program offerings.

Tier 4 evaluation can help address the following questions:

- Are our short-term goals realistic? If so, how can we measure progress toward these goals?
- Do our program inputs and activities have a direct connection to our intended outcomes?
- What is the community's response to our program?
- What have we accomplished so far?

At this stage, you can discuss your program's evaluation data with staff and other key stakeholders and brainstorm with them for ideas about how to use the data to make program improvements.

6. HOW DO WE ASSESS PROGRAM IMPACT (TIER 5)?

After conducting evaluation Tiers 1–4, some programs are ready to tackle the complex task of determining program effectiveness (Tier 5). Formal evidence of program effectiveness can be collected for the program overall or for specific program components. Programs that can provide convincing evidence that they benefit their youth participants are more likely to get continuing financial and public support to help sustain and even scale up program activities.

In this tier, the following questions are addressed:

- Does our program produce the results we hoped it would?
- Does it work better for some participants than others?
- In what ways has the community benefited from our program?
- How can our findings influence policy decisions?

Evaluation Tip:

A necessary part of evaluation is determining who actually uses the program's services. Consider the following questions: Are we reaching the group(s) that our program targets? Are some groups over- or under-represented? Is the program serving groups it did not expect to attract? How are resources allocated among different types of participants?

RELATED RESOURCES FROM HFRP.ORG

- *After School Programs in the 21st Century: Their Potential and What it Takes to Achieve It*
- *Ask the Expert: Karen Walker, Public/Private Ventures*
- *Does Youth Participation in OST Activities Make a Difference?*
- *Investigating Quality: The Study of Promising After-School Programs*

Evaluation Tip:

Prior to starting Tier 5, consider whether your program is sufficiently developed to begin this process. Many small-scale programs do not have the time, resources, and expertise needed to design an evaluation to verify program effectiveness. And, even with adequate resources, a new program should set realistic expectations about its ability to demonstrate program effectiveness in its first year of operation.

In this tier, you should use an **experimental** or **quasi-experimental design** for your evaluation if possible (see Step 6 for details on different types of evaluation design) to be able to build a credible case that the outcomes observed are the result of program participation. For most programs, this type of evaluation requires hiring external experts who know how to use the specific methods and conduct statistical analysis to demonstrate program impacts.

STEP 5: Selecting the Evaluation Design

Different types of evaluation design are appropriate for different evaluation purposes. This section addresses the following questions related to selecting your evaluation methods and design:

- What type of evaluation should we conduct?
- What evaluation design should we use?
- What type of data should we collect?
- What is our time frame for collecting data?

The table below can serve as a reference as you go through this step and the next to help inform your decisions in selecting the evaluation design and the data to be collected.

TABLE 3: What are the major issues we need to consider in conducting an evaluation?

What is our purpose for evaluation? (see Step 1)	To aid learning and continuous improvement	To demonstrate accountability
What type of performance measures should we focus on? (see Step 2)	Measures of effort (focused on outputs)	Measures of effect (focused on outcomes)
What evaluation Tier should we start on? (see Step 4)	• Needs assessment > Tier 1 • Document program services > Tier 2 • Clarify the program > Tier 3 • Make program modifications > Tier 4	Assess program effectiveness > Tier 5
What type of evaluation should we conduct? (see Step 5)	Formative/process	Summative/outcome
What type of data should we collect? (see Step 5)	• Qualitative data • Limited quantitative data for descriptive purposes, especially numerical data related to participation	Quantitative data for statistical analysis
What evaluation design should we use? (see Step 5)	Descriptive	• Pre-experimental • Quasi-experimental • Experimental
What time frame should we consider for data collection? (see Step 5)	Collect data throughout the program year or as available/needed	Collect data at the beginning (pretest), middle (midtest), and end (posttest) of the year
Whom should we collect data on? (see Step 6)	Program participants	Program participants and a comparison/control group of nonparticipants
What data collection methods should we consider? (see Step 6)	• Case study • Document review • Observation • Secondary source/data review • Interviews or focus groups • Surveys	• Secondary source/data review • Surveys • Tests or assessments

Evaluation Tip:

Programs often design studies that include both formative/process and summative/outcome elements so that they can examine which program components might be linked to specific outcomes.

1. WHAT TYPE OF EVALUATION SHOULD WE CONDUCT?

There are two main types of evaluations:

- **Formative/process** evaluations are conducted during program implementation to provide information that will strengthen or improve the program being studied. Findings typically point to aspects of the program's implementation that can be improved for better participant outcomes, such as how services are provided, how staff are trained, or how leadership decisions are made. As discussed in the next section, formative/process evaluations generally use a **non-experimental evaluation design**.

- **Summative/outcome** evaluations are conducted to determine whether a program's intended outcomes have been achieved. Findings typically judge the program's overall effectiveness, or "worth," based on its success in achieving its outcomes, and are particularly important in deciding whether a program should be continued. As discussed in the next section, summative/outcome evaluations generally use an **experimental** or a **quasi-experimental evaluation design**.

Selection of the evaluation type depends on the tier being evaluated (refer to Step 4). In general, programs evaluating Tiers 1–4 should conduct a formative/process evaluation, while programs evaluating Tier 5 should conduct a summative/outcome evaluation.

2. WHAT TYPE OF DATA SHOULD WE COLLECT?

There are two main types of data.

- **Qualitative data** are **descriptive** rather than numerical, and can help to paint a picture of the program. This type of data is subjective and shows more nuanced outcomes than can be measured with numbers. For example, qualitative data can be used to provide details about program activities that are offered, or feedback from participants about what they like (and dislike) about the program. This type of data is generally used for formative/process evaluations, but can also help to flesh out and explain summative/outcome evaluation findings—for example, to provide specific details about how participants' behavior has changed as a result of the program. Qualitative data can be collected through such methods as **observations** of program activities, **open-ended** survey and interview responses, and **case study** data.

- **Quantitative data** are countable information, including averages, statistics, percentages, etc. These data can be used descriptively as formative/process data for evaluation—for instance, participant demographics (percentage of participants of various ethnicities, gender breakdown of participants, or average age of participants, etc.). However, these numbers are more commonly used as summative/outcome evaluation data—for instance, demonstrating improvements in participants' test scores over time. When a quasi-experimental or experimental design is used, quantitative data are usually examined using a statistical analysis. Quantitative data can be collected through such methods as surveys and **tests/ assessments**.

3. WHAT EVALUATION DESIGN SHOULD WE USE?

There are four main types of evaluation designs, which fall into two categories: **causal** studies, which include experimental and quasi-experimental designs, and non-experimental studies, which include descriptive and **pre-experimental** designs.

I. *Causal studies* use measures of quantitative data (numerical averages, statistics, percentages, etc.). These measures are used to attempt to show a causal relationship between the program and its outcomes. Causal means that you can make a reasonable case that your program had a direct impact on the outcomes. (This is different from a **correlation** relationship, which indicates that two things occurred at the same time, but the case cannot be made that one caused the other.) These experimental and quasi-experimental designs are used to conduct summative/ outcome evaluations. They include several defining features:

- Data are collected before (**pretest**) and after (**posttest**) the program to show whether there were improvements in participants' outcomes that can be attributed to program participation.

- Data are collected on a group of program participants (**treatment group**) and a group of similar nonparticipants (**control/comparison group**) to compare the outcomes between the two. (You can also have multiple control/comparison groups—for example, one group of youth who do not participate in any afterschool activities, and another group who participate in afterschool activities offered by a different program.)

- Statistical analysis is used to compare the outcomes for the treatment group and control/comparison groups, both in terms of overall outcomes and improvements over time. Statistical analysis can be used to determine the probability that any differences that occurred are meaningful versus the probability that they happened by chance.

  a. *Experimental designs* all share one distinctive element: **random assignment** of study participants into the program group or non-program group. Random assignment requires a specific selection procedure in which each individual has an equal chance of being selected for each group. Several steps are required:

    - Define the population that the program is targeting (e.g., all of the students in grades 3–6 attending a given school).

    - Develop a system to ensure that each member of this population is equally likely to be selected for the program—this is what is called random assignment. One common way is to get a randomly-ordered list (e.g., not organized alphabetically, by age, or any other organizing criteria) of everyone in your target population (e.g., an unsorted list of all students in grades 3–6 attending the target school), and pick every other name for the program (the treatment group), and assign the rest to the control/comparison group. Another easy way to do this is to pick names out of a hat. The key factor is that the process must be random to avoid the chance that other factors involved in the selection would affect the groups, even factors that would seemingly have no impact. For example, if participants are chosen alphabetically based on their last names, you may get a disproportionate number of siblings who are program participants, which may affect outcomes.

    - Collect outcome data from both groups at pretest (e.g., at the beginning of the program year), to establish a baseline for the two groups, and then again at posttest (e.g., after the program group has participated in the program for a period of time, or at the end of the program year). You may also want to collect data at some points in between, to track changes over time.

  This design allows you to make a strong argument for a causal relationship, since it minimizes **selection bias**, that is, the chance that the two groups are different from each other in ways that might affect their outcomes, based on how they were selected for each group: the youth who attend the program may over- or under-represent certain characteristics of the overall population of interest, meaning that the treatment group and control/comparison group may not be starting on equal footing. Random assignment lessens the possibility of pre-existing differences.

  However, OST programs rarely have the luxury of randomly assigning youth to be in a program. There are also ethical considerations against denying services to a control/comparison group. Programs that have more eligible participants than available program slots are best equipped to implement an experimental study, since random assignment can be seen as a fair way to choose who gets into the program.

  b. *Quasi-experimental designs* are used to try to establish a causal relationship between program activities and outcomes when experimental design is not possible. They are similar to experimental designs except the treatment and control/comparison groups are not randomly assigned. Instead, existing program participants (the program or treatment group) are compared to a control/comparison group of similar non-participants (e.g., their peers attending the same schools). These designs frequently include an attempt to reduce selection bias by matching program participants to program non-participants, either individually or as a group, based on a set of demographic criteria that have been judged to be important to youth outcomes (school attended, grade/age, gender, etc.). For example,

Evaluation Tip:

Combining experimental or quasi-experimental elements with non-experimental elements can provide for richer data than one method alone: You can use non-experimental data to provide a descriptive picture of your program, backed up by the outcome data provided from an experimental or quasi-experimental design.

RELATED RESOURCES FROM HFRP.ORG

- *A Review of OST Program Quasi-Experimental and Experimental Evaluation Results*

- *Avoiding Unwarranted Death by Evaluation*

- *Learning from Small-Scale Experimental Evaluations of After School Programs*

- *What is the Campbell Collaboration and How Is It Helping to Identify "What Works"?*

for individual matching, a 10-year-old female program participant may be matched to one of her 10-year-old female classmates who is not participating in the program. For group matching, the aim might be to get both samples to have the same (or very similar) gender and racial make-up: so if the program group is 60% male and 30% Hispanic, then the comparison group should be too. However, despite your best effort to match these groups, they may still differ in unanticipated ways that may have a major effect on outcomes. For example, if participants are selected into the program based on youth's interest, the youth who are more interested in coming might tend to have better grades or have more supportive families than those who choose not to participate, both of which may affect their outcomes.

II. *Non-experimental designs* include studies that lack statistical comparative data to allow causal statements about a program's impact. While this type of design is most often used to collect data for formative/process studies, it can also be used to collect summative/outcome data when conditions do not exist to allow for an experimental or quasi-experimental design.

a. *Descriptive designs* are used primarily to conduct formative/process evaluations to explain program implementation, including characteristics of the participants, staff, activities, etc. Unlike causal studies which report evaluation findings as statistics or numbers, descriptive designs tend to tell a story about the program. The data are usually qualitative, although some quantitative data may be included as well, such as counts or percentages describing various participant demographics. Non-experimental evaluation designs include such types as case studies, monitoring for accountability, participatory or theory-based approaches, and ethnographic studies.

b. *Pre-experimental designs* collect quantitative summative/outcome data in instances when resources do not allow for a causal design to examine outcomes. While the data collected may look similar to an experimental or quasi-experimental study, pre-experimental studies lack a control/comparison group and/or pretest/posttest data collection. These designs include "one-shot case study" designs (i.e., studies examining program participants' outcomes at the end of the program in the absence of comparison data); one-group pretest-posttest design (i.e., studies comparing program participants' "before" and "after" data); and static-group comparison (i.e., studies comparing "after" data from the program group with data from a comparison group at the end of the program, in the absence of pretest, or "before," data). Outcomes measured in these ways may include some statistical analysis, but generally cannot make a strong case for a cause-and-effect relationship between program activities and outcomes.

In choosing a design, consider what type is best suited to help answer your evaluation questions and is most appropriate given your program's developmental stage. Programs that have only been operating for a year or two, or that are still figuring out their program focus and activities, should consider a descriptive design that provides a descriptive picture of the program to ensure that the program is being implemented as intended. More established programs may be ready for an experimental causal study design to provide evidence to stakeholders of the program's impact. As noted above, programs also need to consider which design is most feasible—many programs do not have the ability to randomly assign program and comparison groups, as required for an experimental evaluation.

4. WHAT IS OUR TIME FRAME FOR COLLECTING DATA?

Depending on resources and funding requirements, some programs conduct a one-time evaluation from data collected over the course of a single year. Others conduct evaluations on an annual basis, often comparing the current year's findings to findings from the previous year(s). However, conducting annual evaluations does not mean just collecting the same data year after year. Evaluation strategies should evolve over time to match the needs of your program, and the evaluation design should match the developmental stage of the program, as discussed in Step 4. Adapting evaluation strategies to match the present needs of your program will better ensure that the data collection enables learning and improvement over time.

Experimental or quasi-experimental designs require collecting data at multiple time points (pretest and posttest) to assess improvements attributable to the program. Pretest measures can be

collected at the beginning of the program year, before youth have participated in the program; posttest data can be collected from the same participants at the end of the program year. These studies can also include midpoint check-ins to assess incremental improvements over time.

Some programs are able to conduct evaluations that include long-term (or **longitudinal**) tracking, which involves following a program participant across multiple years. Often extending beyond the individual's program participation, longitudinal tracking examines long-term effects of program participation (e.g., examining high school graduation rates amongst youth who participated in an academically-focused OST program while in middle school). This method is usually chosen by organizations using outside consultants or research organizations to conduct a large-scale, multi-method evaluation.

The length of time needed for your evaluation process will depend on available resources, funder requirements, goals of the evaluation, and how you plan to use the data you collect. Consider the following questions in determining the time frame of your evaluation:

- What do our funders require in terms of data collection and reporting time frames?
- Do we need to include pretest/posttest measures to examine changes in our outcomes?
- What time frame is most feasible given our available resources for evaluation?
- How much time do we need to collect and analyze data that will be useful for our program?

Evaluation Tip:

Consider whether you want data for immediate feedback to make program improvements, or whether you want data to make the case for longer-term outcomes. The former will allow a much shorter evaluation time frame than the latter. However, you may want to examine program improvements on an ongoing basis, which would require a longer-term evaluation plan.

STEP 6: Collecting Data

Once you have established which evaluation tier and design are most appropriate for your program, you need to figure out how to collect the data that will best address your evaluation questions and that are most feasible for your program. This section addresses the following questions related to your data collection:

- How do we select our program sample?

- What data collection methods should we use?

- How do we choose evaluation tools and instruments?

- What ethical issues do we need to consider?

- How do we collect participation data?

- How can we manage our data?

Also see *Table 3: What are the major issues we need to consider in conducting an evaluation?,* in Step 5 to help guide your decisions in selecting the data to be collected.

1. HOW DO WE SELECT OUR PROGRAM SAMPLE?

In collecting data for evaluation, programs need to consider the **sample** upon which data will be based—that is, the specific population about which (and from which) to collect data.

In general, OST programs collect data from and about program participants. These data can be collected about all program participants, or on a **subsample** of participants. Collecting data on all participants will get the most thorough data, but this strategy is not always feasible, nor is it always necessary. Cases where collecting data about a subsample makes sense include the following:

- Your program is very large, and thus has too many participants to realistically be able to track them all. In this case, you may want to randomly select participants to get a **purposive sample**. Similar to random assignment (discussed in Step 5) which involves a system of selecting the program or comparison group in such a way that each individual has an equal chance of being selected into either, **random selection** involves using a system of selecting study participants in such a way that each program participant has an equal chance of being chosen to participate in the study, or not. While this process is the same as the process used for random assignment, the samples and purposes are different. Random assignment assigns individuals into program and comparison/control groups for the purpose of allowing comparisons between program participants and nonparticipants that are not likely to be influenced by pre-existing differences between the two groups. Random selection, meanwhile, focuses specifically on the program group, and selects some members of the program group to be involved in data collection for the evaluation—this method helps increase the likelihood that you will have an evaluation sample that is truly representative the overall program population.

- The evaluation questions focus on outcomes for a subsample of participants—for example, on outcomes for minority youth or girls—in which case, data collection should also focus on this subsample of participants.

Data can also be collected about other groups as part of program evaluation. For example, an experimental or quasi-experimental design requires the collection of data on a comparison group of nonparticipants. You can also collect data for your evaluation from others involved in your program, such as parents of participants, program staff, and school staff, to get their perspectives on your program in terms of both their own involvement and what benefits they have observed in participants as a result of the program.

Regardless of the sample population selected for your program evaluation, it is unlikely that you will be able to collect data on the entire sample. Some participants may not agree to participate or will not be available at the times data are collected. It is important to keep track of how many in your population sample do not, in fact, participate in your data collection. This information is used to calculate response rates, that is, the percentage of your sample on which you are able to collect data. The higher your response rate, the more likely it is that your data will be representative of the overall program sample.

2. WHAT DATA COLLECTION METHODS SHOULD WE USE?

Evaluation data are collected in a number of different ways, often through multiple methods. In considering which methods to use, you will need to determine which methods will best help you assess your evaluation questions, and whether these methods will give you information that will be helpful in accomplishing what you set out to do with your evaluation (fulfilling funder requirements, aiding in learning and continuous improvement, etc.). The data collection methods selected should be able to directly address your evaluation questions. Common methods are described below.

- **Case studies** focus on one individual over a set period of time, taking an intensive look at that individual's program participation and the effect on his or her life. Participants can be chosen randomly or using specific criteria. Case studies can include formal interviews, informal contacts such as phone calls or conversations in hallways, and observations. Program staff often carry out this method, since they already have a relationship with the individual in the study and also have existing opportunities for interaction. This method requires the development of tools to guide the relationship between the evaluator and the case study individual. Case studies are most often used to tell a detailed story about participation in the program.

- **Document review** involves a review and analysis of existing program records and other information collected and maintained by your program as part of day-to-day operations. Sources of data include information on staff, budgets, rules and regulations, activities, schedules, participant attendance, meetings, recruitment, and annual reports. These data are most often used to describe program implementation, and as background information to inform evaluation activities.

- **Observation** involves assigning someone to watch and document what is going on in your program for a specified period of time. If possible, at least two people should be assigned to this task. Observations can then be compared to see how consistent they are (called inter-rater reliability); those observations that are not consistent are likely to have been influenced by individual biases. Before beginning observations, the assigned observers should engage in a process of self-reflection in order to identify individual biases and understand how those biases or insights strengthen or weaken their position as observers. Observations can be highly structured—using formal observation tools with protocols to record specific behaviors, individuals, or activities at specific times—or it can be unstructured, taking a more casual "look-and-see" approach to understanding the program's day-to-day operations. Data from observations are usually used to describe program activities and participation in these activities, and are often used to supplement or verify data gathered through other methods.

- **Secondary source or data review** involves reviewing existing data sources (that is, data that were not specifically collected for your evaluation) that may contribute to your evaluation. These sources can include data collected for similar studies to use for comparison with your own data, large data sets, school records, court records, and demographic data. As with document review, these data are most often used to describe program implementation and as background information to inform evaluation activities.

- **Tests or assessments** include such data sources as standardized test scores, psychometric tests, and other assessments of your program and its participants. These data often come from schools (especially for academic tests), and thus can also sometimes be considered secondary source data. This method is most often used to examine outcomes, often using an experimental or quasi-experimental design.

- **Interviews or focus groups** gather detailed information from a specific sample of program stakeholders (e.g., program staff, administrators, participants and their families, funders, and community members) about program processes and the stakeholders' opinions of those processes. Interviews are usually conducted one-on-one with individuals (although several individuals can be interviewed together) either in person or over the phone. Focus groups generally operate in person (although they can be conducted by conference call or web meeting) and involve gathering individuals to provide feedback as a group. Both interviews and focus groups require a set of questions designed to elicit specific information. The questions are generally **open-ended**, but **closed-ended questions** can also be included

Evaluation Tip:

Digital and social media are emerging as innovative sources for data collection, and youth are particularly adept at navigating these resources. Consider ways that you can use digital and social media tools for data collection while allowing youth to play a more active and participatory role. For example, youth could record their own videos of program activities, or keep blogs documenting their program experiences that could then be used as data sources for evaluation.

(discussed in more detail below). This method is most often used to collect subjective or qualitative data, such as how participants feel about a particular activity, or whether or not they found participating in a certain activity helpful to them in a specific way.

- **Surveys** are tools designed to collect information from a large number of individuals over a specific time period. They are administered on paper, through the mail, by email, or on the internet. Questions on surveys may include both closed-ended and open-ended questions. Surveys are an excellent way to obtain participant background data (e.g., demographic information). Many programs use initial survey forms to obtain information about the interests of individuals in their program. Surveys can also be used to get a sense of the individual progress of participants.

  A closed-ended question is a form of question that is answered using a given set of response options, such as a simple "yes" or "no," a selection from multiple choices, or a rating on a scale. An open-ended question does not limit responses to a specific set of options, but allows the individual to provide his or her own response. Open-ended questions are easy to write and can provide more nuanced data, but require more time and effort to analyze. On the other hand, closed-ended questions are easy to analyze, but it can be difficult for those designing the survey or interview questions to create appropriate response categories (beyond yes/no responses) that will provide meaningful data.

Evaluation Tip:

In selecting methods and instruments, it is important to make sure that the type of data you can obtain from a given method will be useful to your evaluation and what you are trying to measure—you should not select a data collection method or instrument simply because you can easily collect the data. On the other hand, make sure that more time- and resource-intensive methods will give you sufficient returns to justify using them.

3. HOW DO WE CHOOSE EVALUATION TOOLS AND INSTRUMENTS?

No matter what data collection methods you select, you will most likely need specific instruments to collect these data. Rather than re-inventing the wheel, you can start by exploring whether any instruments already exist to measure what you plan to measure. The evaluation instruments used by OST programs take a variety of forms, such as checklists of program components, survey questions measuring self-esteem, assessments of academic skills, and others. The *Measurement Tools for Evaluating OST Programs* resource describes instruments that have been used by OST programs to evaluate their implementation and outcomes. This resource can help provide ideas for possible data collection instruments to use or adapt for your program. These existing tools often have the benefit of already having been tested for reliability and validity, which are necessary considerations in selecting evaluation tools. These terms are defined below:

- **Reliability** refers to the consistency of the data collection instrument: You should get consistent results each time you use the instrument if it is reliable. While the results should not vary wildly from one use of the instrument to the next, repeated uses with the same group of program participants over time will hopefully show positive changes in participant outcomes. In addition, administering the instrument to different groups will likely have some variation in results. Beyond these improvements and normal variations between groups, though, results should be relatively consistent or they will not be meaningful or useful to your evaluation. It can be difficult to know what level of variation is normal without statistical analysis, which is why instruments that have already been tested for reliability using statistical methods can be helpful to adopt.

- **Validity** refers to whether the evaluation instrument is actually measuring what you want to measure. For example, a measure that involves rote memorization would not be a valid measure of analytic skills (although it should be a good measure of how well the individual is able to retain information over a short period of time).

You may be able to adapt existing measurement tools to your purposes by taking only specific questions and pieces from those tools, or even rewording some to be more applicable to your program. You should be aware, though, that making these types of changes may threaten the tool's reliability and validity.

4. WHAT ETHICAL ISSUES DO WE NEED TO CONSIDER IN OUR DATA COLLECTION?

Before beginning to collect any data, it is important to communicate the following information to program participants who will be involved in the evaluation (including program staff, youth, and their parents/guardians, if applicable): the evaluation's purpose, expectations for participants' involvement in data collection (e.g., the time required), potential risks of their involvement, how

the data collected about them will be used and reported, and how their confidentiality will be protected. Communicating this information to participants from the beginning has the added benefit of helping to gain their trust and thus their willingness to cooperate with the evaluation process.

There are several additional requirements regarding the rights of participants that must be followed in the evaluation process:

- If your program is affiliated with an institution that has an ethics committee such as an Institutional Review Board (IRB), you will most likely need ethics committee approval before undertaking an evaluation that involves either individuals or groups of people. Institutions that usually have ethics committees include colleges, universities, federal agencies, and some state and local agencies.

- If the evaluation collects personal information about participants that would identify them (such as their names or social security numbers), you are required to obtain their informed consent to participating in the evaluation. Informed consent simply means that the individual has been informed of the purpose of the study, and that the individual has consented to participate. Consent usually involves the participant's signing a form to this effect.

- Children must have parents'/guardians' consent to participate. This consent may need to be given actively, meaning parents/guardians must provide permission (usually through a signed form) for their children to participate, or passively, meaning parents/guardians must inform the evaluator only if they do not want their children to participate in the data collection process for the evaluation. In addition, children aged 7 or older may be required to agree directly to participate in order for the evaluator to be permitted to collect data on them.

- If you are collecting data from schools or within a school setting, you will probably also be required to follow the school's or school district's policies for collecting data from their students.

5. HOW DO WE COLLECT PARTICIPATION DATA?

Regardless of other data your program plans to collect, all programs can and should collect data about who participates in the program. Data on program participants can be helpful in improving your program itself, as well as serving as a key component of any program evaluation. Depending on what participation data you collect, these data can help you to address some or all of the following questions related to program implementation and outcomes (some of these questions require additional data collection to answer):

- Who participates in our program (and who does not)?

- Why do youth participate (and why not)?

- Does our program reach the target population that it intends to reach? If not, what groups are under- or over-represented?

- How engaged are participants in program activities?

- How do participant demographics relate to outcomes? That is, do some participants have better outcomes than others?

- What level of participation seems to be necessary to achieve positive outcomes?

- Are participants regularly attending our program? If not, are participants dropping out or only attending sporadically?

- Do participation levels vary based on the time of year (e.g., is participation higher in the spring than in the fall)? If so, what might account for that (e.g., competing seasonal sports or extracurricular activities, changes in the weather)?

Using participation data, you can shape specific program strategies to better recruit and retain the youth whom your program aims to serve. These data can also be used to refine or improve activities based on the needs and interests of the group served.

RELATED RESOURCES FROM HFRP.ORG

- *Engaging Older Youth: Program and City-level Strategies to Support Sustained Participation in OST*

- *Selection Into OST Activities: The Role of Family Contexts Within and Across Neighborhood*

- *Engaging Adolescents in OST Programs: Learning What Works*

- *Issues and Opportunities in OST Evaluation: Understanding and Measuring Attendance in OST Programs*

There is no right time to begin collecting participation data—the sooner the better. Participation data can be used to help shape your logic model (see Step 2), since who participates in your program may play a role in the outcomes you can expect. You can also collect participation data after constructing a logic model to see if the targeted participant demographics included in your logic model match your actual program demographics.

Regardless of whether you collect participation data before starting an evaluation, you should collect participation data as part of the evaluation process. For a formative/process study, participation data is an important part of documenting program implementation. For a summative/outcome study, participation data allows examination of key issues such as whom your program is benefitting, and whether or not participants attend your program frequently enough to be expected to show positive outcomes (see Step 5 for more details about formative/ process and summative/outcome studies).

There are two types of participation data that all programs should collect:

- *Program participant demographics* include data such as age/grade in school, gender, race, family income levels, migrant status, etc. The data to be collected should be selected with your program's target population in mind. For example, if you are targeting children from low-income families, it will be especially important to collect data on the income levels of participants' families.

- *Program attendance measures* are not restricted to whether or not youth attend your program. Attendance measures also include **intensity** (how often each youth attends a program during a given period as measured in hours per day, days per week, and weeks per year) and **duration** (the history of attendance in years or program terms).

There are two additional types of participation data that you may want to collect, depending on your program's needs and available resources:

- *Demographics of the schools and communities that your program serves* include the same type of demographic data collected at the participant level, but are collected from the larger school and community populations. At the school and community level, demographic data can be used to assess the characteristics of the communities and schools that your program serves. These data can also be compared to participant demographics to determine whether your program is serving its target demographic. In some cases, a program may want to serve a sample of the youth population that is representative of the school/community. In others, the program may want to serve a group disproportionate to the community/school population—for example, a program targeting the neediest students might seek to serve a disproportionate number of children from low-income families.

- *Feedback from participants on why they attend (or do not attend) and their level of engagement* can help determine what draws youth to your program, whether they are actively involved and interested in program activities, and what prevents them from attending more often. Reasons for attending (or not) include such variables as whether or not youth have friends who also attend, if they find the activities interesting, whether they have scheduling conflicts during program hours, or if they have transportation to and from your program. In addition, attendance patterns may shift with seasonal changes (e.g., youth may go home earlier when the sun goes down earlier or may have other extracurricular activities such as sports that take place only on the spring or fall).

Participation data can be collected through the following:

- *Program participation demographics* can be collected through program application forms that parents complete as a prerequisite to enrolling their child in your program. This intake form should collect all demographic data of interest to your program; at the same time, the form should be as simple as possible to ensure that families will take the time to complete it.

- *Program attendance data* can be collected through daily sign-in sheets for participants or a checklist/roll-call by program staff.

- *School and community demographic data* can be collected through such sources as school records and census data. Getting access to school records can be a challenge, but many programs have been able to partner successfully with schools in the community to gain access to these records. Data on individual students are usually not provided by schools due

to confidentiality issues, although sometimes programs can gain access to these individual-level data when identifying details (e.g., students' names) are removed from the records. Additionally, federal regulations under the Family Educational Rights and Privacy Act (FERPA) allow schools to disclose students' school records, without parental or student consent, to "organizations conducting certain studies for or on behalf of the school." The National Center for Education Statistics Common Core of Data can also be used to collect school-level data. In addition, some schools use school-based online data collection systems, such as Infinite Campus or Powerschool, to which programs may be able to gain access.

- *Feedback from participants* can be collected through surveys or interviews of participants and their parents to get a sense of why they participate (or why not) and their level of engagement in your program.

6. HOW CAN WE MANAGE OUR EVALUATION DATA?

One highly successful method for tracking program data (both for evaluation and for other program needs) is to create a **management information system (MIS)** or database that is maintained electronically (i.e., on a computer or the internet). Participation data are only one part of an MIS, which can be used to organize data on all aspects of your program, from information on the activities implemented to staff demographics and professional development opportunities. Developing and putting into place an MIS may require outside expertise.

A functional MIS should:

- Be user-friendly to allow program staff to update and maintain data on an ongoing basis.

- Provide flexibility, allowing the possibility of adding or modifying the type of data collected as additional program needs are identified.

- Be able to easily run queries on the various types of data collected.

- Have safeguards in place to protect the confidentiality of the data (e.g., password protection with restricted access).

Evaluation Tip:

Consider the following questions in designing your MIS: Are we creating extra work or unnecessary burden for our staff and families by keeping too many records? Are our forms designed to collect the relevant data efficiently? Do we actually use the information we collect? Can the data we collect be used for multiple purposes? Are we collecting enough information to know if our program is implemented as planned? Is our documentation helpful for planning?

Evaluation Tip:

Regardless of you how organize your data, be sure that you save a master copy of your data (both electronically, if possible, and as a hard copy) before you start any analysis. This copy should be preserved and put aside, so that you can refer to it as necessary. Then you can use working copies of the data to mark up, edit, and move information around as you please.

Evaluation Tip:

Statistical significance indicates the likelihood that a real difference exists, but it does not necessarily mean that the difference was large, and thus meaningful. Additional statistics (effect sizes) can be calculated to give a sense of whether the difference is large enough to indicate that the program made a meaningful difference.

STEP 7: Analyzing Data

This section discusses what you do with all your data, once you have collected it. Specifically, these questions are addressed:

- How do we prepare our data for analysis?

- How do we analyze quantitative data?

- How do we analyze qualitative data?

- How do we make sense of the data we have analyzed?

1. HOW DO WE PREPARE OUR DATA FOR ANALYSIS?

Before you can begin analysis, you will need to "clean" the data you have collected, which means checking your data for mistakes. This error-checking is especially important for quantitative data, since mistakes in these numbers can make a big difference in your quantitative results. "Cleaning" qualitative data means ensuring that all data were collected consistently by different team members and collected consistently over time in terms of frequency, level of detail, and type of data collected. For quantitative data, this process mostly involves checking for errors in data entry.

A good first step is to look through your data for any obvious mistakes. For example, if you have items rated on a scale of 1–5, be sure that all of the numbers are within that range. You should also check for any missing data. Then, a closer review can reveal more subtle errors or inconsistencies. In some cases, you may need to do additional data collection to fill in the missing data or replace any incorrect data. Once you have done your cleaning, one final "spot check"— checking a random selection of entries for accuracy and consistency—is a good idea.

2. HOW DO WE ANALYZE QUANTITATIVE DATA?

You first need to tabulate your data into overall total counts for each category. A spreadsheet or statistical software can help with these calculations. All scales and yes/no responses should be recoded as numbers. So, for example, a 5-point scale from "strongly disagree" to "strongly agree" should be recoded on a numerical scale of 1–5, where 1 = "strongly disagree" and 5 = "strongly agree." For yes/no response, "yes" is typically coded as "1," and "no" as "0."

These counts can then be used to compute averages, percentages, and other measures that translate the data into numbers that meaningfully describe program implementation and outcomes. For example, if you collected baseline data and follow-up data, average scores can be created for these two time periods to see if there was a change in the average scores for these data over time. Numbers can be computed similarly for your program and comparison groups to see how the numbers for these two groups differ and if the program group appeared to have better outcomes than the comparison group.

If your study has an experimental or quasi-experimental design (that is, if you are collecting pretest and posttest data using a program group and a comparison group of nonparticipants to measure program impacts), you will need to do a statistical analysis of these counts to see if there were statistically significant differences between the program and comparison groups, or between baseline and follow-up scores for the program group. Statistical significance refers to the probability that these differences could have happened by chance: If the probability that the differences happened by chance is shown by statistical analysis to be small (generally assessed to be a 5% chance or less), then the differences are statistically significant. Statistical significance means that a strong argument can be made that your program was responsible for the observed differences. For example, you may want to examine changes in reading levels from the beginning of the year to the end of the year, to see if program participants had larger improvements on average than did their peers who did not attend. You may find in just looking at the average improvements that program participants overall improved more than nonparticipants. However, without statistical analysis, you have no way of knowing if the differences were statistically significant—actual real differences in reading gains between the groups—or just a fluke, perhaps due to a few nonparticipants having a particularly bad testing day, or a few participants having a particularly good testing day (regardless of their true knowledge or ability).

3. HOW DO WE ANALYZE QUALITATIVE DATA?

Analyzing qualitative data tends to be a lengthier and less straightforward process than for quantitative data. Unlike quantitative data, qualitative data are often subject to interpretation, and there is no one "right" way to tackle the analysis. How you choose to conduct the qualitative data analysis will depend in part on what you hope to get out of the data and on the resources (staffing, time, software, etc.) that you have available to conduct the analysis.

Regardless of analysis method, the first step is to go through all of the qualitative data and look for any themes that emerge. For example, you might see data emerging around the theme of "staff relationships with participants" or "parent engagement in program activities." Once you have identified these themes, you can begin to categorize the data by theme. This process tends to be iterative: New themes emerge and are added to your general list of themes as you get further into the data. A common way to conduct this analysis is to "code" the data into themes. Such coding can be done by hand, marking up a printout or computer document, or by using computer software that helps to code qualitative data.

4. HOW DO WE MAKE SENSE OF THE DATA WE HAVE ANALYZED?

Once you have analyzed the data, you will need to organize it in a way that tells a story about your program. As part of your evaluation planning, you should have developed a set of questions that you want to answer with your evaluation (see Step 1). These same questions can help guide the organization of the data that you have analyzed. Whether or not you not choose to organize your final report by your initial evaluation questions, organizing your data in this manner at first will allow you to see how well your data address your evaluation questions.

The next section addresses how to present the data that you have collected in an evaluation report.

STEP 8: Presenting Evaluation Results

Once the data are collected, you need to think about how to present your program's evaluation results, and to whom. This section of the toolkit offers some practical advice on how to present the results to offer maximum utility to program stakeholders. The following questions related to presenting your evaluation are addressed in this section:

- How do we communicate results to stakeholders?

- What level of detail do we report to stakeholders?

- What do we include in an evaluation report?

- How can we make our evaluation findings interesting and accessible?

1. HOW DO WE COMMUNICATE RESULTS TO STAKEHOLDERS?

How do programs talk about or disseminate the results of their evaluation? The answer to this question connects to Step 1 of this evaluation toolkit: determining the purpose of the evaluation. If your program conducts an evaluation for the purpose of program improvement, results will most likely be communicated to administrators, staff, parents, and participants in various ways, many of them informal. However, if your program is conducting a formal evaluation for funders, it must consider more formal ways to communicate results, such as formal reports or presentations. Evaluation results can be presented in the following ways:

- Presentations at:
  - staff meetings for management and staff within the organization.
  - regional meetings or national conferences in the fields of out-of-school time, education, youth development, or public policy.
  - luncheons or seminars for external stakeholders (e.g., school and community partners).

- Comprehensive reports for funders and other stakeholders who want concrete documentation of program impacts.

- Executive summaries or full reports posted to program or organization websites.

- Email or newsletter updates to families and others in the community who are interested in your program.

- Press releases and other media coverage.

2. WHAT LEVEL OF DETAIL DO WE REPORT TO STAKEHOLDERS?

Sharing results does not necessarily mean sharing every single bit of data collected. As with any presentation, deciding which information is most important to which audience is the key to successfully communicating evaluation findings. Consider the level of detail that your various stakeholders are interested in. Do they want full information on the statistical analysis of all the findings, or do they just want the headlines of the major findings? Here are some issues to consider regarding specific audiences when determining how much detail to include:

- *The local community* needs enough detail so that someone who knows nothing about your program will have sufficient information to understand the evaluation and the program's role within and impact on the community. On the other hand, you do not want to include so much detail that the information seems overwhelming or inaccessible.

- *School staff*, such as teachers and principals, are likely to be most interested in findings that have implications for school-day academic achievement and classroom conduct—that is, outcomes related to children's learning and behavior.

- *Program practitioners* often have little time to devote to poring over evaluation results; results presented to this audience should be "short and sweet," and contain only the most important details. However, because results impact areas for improvement, practitioners may want to know much more, and such additional information in the form of a more comprehensive report should also be available to them.

- *Funders* tend to want fairly detailed reports to show exactly how their money was used and the outcomes your program contributed to. However, funders may also impose certain reporting requirements, including adherence to specific page or word limits. These limits will affect the level of detail you can and should include.

- *The program board*, if applicable to your program, will likely want only the major headlines, but they may want more details as well, depending on who is on your board and their level of involvement. Results-in-brief should be provided, but you should also offer access to the full comprehensive results for those who are interested.

- *Parents and youth* will tend to be most interested in findings that relate specifically to youth participation; youth and their parents want to know whether your program is having a positive impact on them. These findings should be kept brief, with a focus on youth outcomes and any other implications of the evaluation for youth and their families.

3. WHAT DO WE INCLUDE IN AN EVALUATION REPORT?

A comprehensive evaluation report should have the following parts:

- An executive summary of key findings.

- A description of the program that was evaluated, including program goals, activities, staffing characteristics, participant characteristics, location, and hours/days/seasons of operation.

- Discussion of research-based evidence—from a literature review or other existing statistics—on the need for or value of the type of program being evaluated.

- If applicable to your program, information about the overall organization and how the evaluated program fits into that organization's mission.

- A description of the evaluation process:

  - Background on the individuals responsible for conducting the evaluation, including information about areas of expertise and educational background.

  - The purpose of the evaluation, including the evaluation questions.

  - Time frame for the data collection, including whether pretest and posttest data were collected.

  - What type of design the evaluation used (i.e., experimental, quasi-experimental, or non-experimental).

  - Methods used to carry out the evaluation process (e.g., tools, protocols, types of data-collection systems).

  - The sample on which data were collected (e.g., all youth participants or a subsample), including response rates for various data collection methods, and any demographic information on the sample (race, gender, etc.).

- A detailed summary of findings, described one at a time:

  - An explanation of how the findings relate to the overall program goals.

  - Any statistical significance or other specific numeric data (e.g., percentages or averages) related to each finding.

  - Any graphs or charts that can help illustrate the findings.

  - If available and relevant, other research in the field related to the findings.

- Implications of the evaluation findings:

  - Ways the findings can help facilitate program improvement and sustainability.

  - If applicable, recommendations for policy and/or for similar programs.

- Any other information necessary to meet funder and other stakeholder expectations.

Evaluation Tip:

Appendices can be used to include all the nitty-gritty methodological information. That way, the information is there as a reference, but will not compromise the readability of the report.

4. HOW CAN WE MAKE OUR EVALUATION FINDINGS INTERESTING AND ACCESSIBLE

No matter who your audience is, you should think creatively about how to present your evaluation results in a way that is interesting and engaging, and not just the same-old, same-old. Specifically, you may want to incorporate the following elements into the reporting of your evaluation results:

- *Digital and social media.* For instance, you may want to create multimedia presentations that use videos or photos highlighting program activities, or audio clips from youth participants. You can also use sources like Facebook, Twitter, and other social media outlets to publicize your evaluation findings.

- *Descriptive examples.* Such examples can bring the data to life so that data become more user-friendly and accessible to multiple audiences. Descriptive examples can also help attract local media, who often like the "human interest" side of an evaluation. Elements that you can add to increase the appeal of your reports include testimonials from satisfied families of program participants and multimedia materials as outlined above.

- *Visual representations of your data and strategies.* These visuals can help to break up the text. They can also be used to help make the information more accessible and dynamic. These visuals can include charts and graphs highlighting evaluation findings, depictions of logic models, and graphics illustrating complex relationships or systems.

STEP 9: Using Evaluation Data

Aside from the use of evaluation data for accountability to funders, evaluation data can—and should—be used to make improvements to your program, inform future evaluation activities, and market your program to funders and other stakeholders. The following questions are addressed in this section:

- How do we use evaluation to make program improvements?

- How do we use evaluation to inform our future evaluation activities?

- How do we use evaluation for marketing?

- Who needs to be involved in decisions about how to use the evaluation data?

1. HOW DO WE USE EVALUATION TO MAKE PROGRAM IMPROVEMENTS?

Evaluation can help to strengthen programs and ensure their sustainability. What you can expect to learn from an evaluation will depend in large part on whether you conducted a formative/process evaluation or a summative/outcome evaluation, or some combination of the two.

Formative/process evaluation can allow you to determine whether services have been implemented in your program as planned, and whether participants and other stakeholders are satisfied with the services offered. If not, you can then reflect on what program changes are needed so that services are operating as intended, and so that stakeholders are happy with these services.

For example, if evaluation data indicate that youth participation in your program is low, some further questions can be considered: Can we make program activities more appealing to youth? Can we make it easier for families to have their children participate (e.g., offering program hours that are convenient to parents, providing transportation)? Can we coordinate the program schedule and activities so that we are complementing, rather than competing with, other OST activities in the community? You may need to collect additional data to answer some of these questions. For example, if participation in your program is low, feedback from parents in the community (through informal conversations or through more formal interviews or surveys) may be helpful to find out why they are not enrolling their children.

If the evaluation results indicate that your program is on the right track—that it has been implemented as planned and stakeholders are satisfied—then you can start thinking about evaluating program outcomes.

Summative/outcome evaluation can help you to determine whether your program is achieving the outcomes that it set out to achieve. In the process of a summative/outcome evaluation, programs sometimes discover that their intended outcomes are not being achieved as planned. In some cases, unexpected circumstances outside of the program's control can affect outcomes, such as low attendance due to competing programs or decreases in program funding that may limit the program's resources. In many cases, when an evaluation shows less-than-perfect outcomes, that is not a bad thing—in fact, it can help you to identify areas for improvement in your program, and to see where the program needs to focus its efforts.

However, evaluation results suggesting very little, or no, progress on program outcomes may be a cause for concern. In this situation, there are three possibilities to consider and troubleshoot:

- *The program may have set expectations too high*. If your program is relatively new, it may be too early to expect the types of outcomes that the program aimed to achieve. Progress may not be seen on many outcomes until your program has been running for several years. In this case, you should consider your evaluation findings as **benchmark** data upon which to judge outcomes in future years. Evaluation findings from similar programs can offer guidance as to what types of results can reasonably be expected.

- *The program may not be implemented as intended, so the services provided are not leading to the intended outcomes*. In this situation, you should consider taking a step back and conducting a formative/process evaluation to examine program implementation before looking at any further outcomes. A formative/process evaluation can help to uncover possible

problems in the way that your program is implemented that may be preventing the program from achieving its intended goals.

- *The evaluation may not be asking the right questions or examining the right outcomes.* Your program's activities may not be directly tied to the outcomes that the evaluation measured. In that case, you should revisit your evaluation plan to ensure that the evaluation is designed to measure outcomes that are directly tied to program activities.

2. HOW DO WE USE EVALUATION TO INFORM OUR FUTURE EVALUATION ACTIVITIES?

Many programs see evaluation as an ongoing activity, and, as such, conduct evaluations on a yearly basis to help with learning and continuous improvement. Future evaluations should be informed by past evaluations: Subsequent evaluation activities can build on the methodologies from previous years, using the previous year's data as a benchmark for progress. If your evaluation indicates significant room for improvement in the areas assessed, you may want to work on implementing program improvements, and then evaluate your progress on these same measures the next year. Alternatively, data collection plans can be adjusted from year to year: In planning for future evaluations, be sure to assess whether or not the current evaluation findings are useful and meaningful to your program, and adjust your data collection plans accordingly.

Bear in mind that evaluation as a tool for improvement is both powerful and versatile. You may want to examine different issues in future evaluations of your program—for example, you may want to move from conducting a formative/process study, if the results suggest that your program is well-implemented, to a summative/outcome evaluation. Or you might want to delve more deeply into specific issues related to program implementation or outcomes that surfaced in your earlier evaluation—for instance, focusing on the participation and outcomes of specific participant subgroups that your program serves, such as minority students or older youth.

In any case, completed evaluations can be a good place to start in determining the appropriate direction of future evaluation activities. Consider the following questions in planning for future evaluations:

- Are there program areas assessed in the prior evaluation showing significant room for improvement that we should examine in subsequent evaluations?

- Are there findings that were not very helpful that we want to remove from consideration for future data collection?

- Do the results of our formative/process evaluation suggest that our program is sufficiently well implemented to allow us to now look at outcomes?

- Are there specific aspects of our program (e.g., participant subgroups served, subsamples of types of activities) that it would be helpful to focus on in subsequent evaluation activities?

3. HOW DO WE USE EVALUATION FOR MARKETING?

Positive evaluation results can help you to better promote your program: You can highlight the evaluation findings on your website, in newsletters, and in other promotional materials. Using evaluation to "sell" your program can ultimately help your program to gain additional funding and community support, as discussed below.

Funding. While many funders require an evaluation component as part of their program grants, funders increasingly want evidence that a program is of high quality and that it is achieving positive outcomes for youth before they agree to fund it. Evaluation data can provide support for a program's value in grant proposals. Even if the evaluation results suggest room for improvement, the fact that your program is collecting such data indicates a commitment to learning and continuous improvement that gives a positive impression of your program's potential. Evaluation data that indicate existing needs for your program and positive outcomes for youth are especially important to highlight in communicating with potential funders.

Community and school support. Evaluation results can be used to demonstrate to local communities and schools your program's value and success, and thus help to further build

community and school support. This support can be of real value in a number of ways, including:

- Building public support for OST programs.

- Recruiting participants to your program.

- Attracting volunteers from the community to help provide program activities.

- Garnering in-kind and cash donations from local businesses.

- Building partnerships and/or coordinating services with schools and other entities that serve youth in the community (health providers, family support services, museums, churches, etc.).

4. WHO NEEDS TO BE INVOLVED IN DECISIONS ABOUT HOW TO USE THE EVALUATION DATA?

To help determine how the evaluation data should be used, it is important to get the input and buy-in of two key groups of program stakeholders: (1) program staff, both front-line and management, and (2) program leadership and the board.

Staff. To ensure that staff have input into how evaluation data are used, consider providing a planning session to:

- Go over the evaluation results. These results should be conveyed to staff as a learning opportunity, not as a time to place blame or point fingers about what's not working. Positive findings should be highlighted, and credit given to staff where it is due.

- Brainstorm with staff about possible program improvements that can be made in response to the evaluation findings.

- Discuss whether the possible changes identified are realistic, how your program can go about making these changes, and what support/help will be needed from staff to make these changes.

All staff should be involved to ensure that all are working with the same information and that they are given an opportunity to provide feedback and insights. Front-line staff in particular can provide a good check as to what changes can reasonably be made with existing resources (including their time). When program staff are involved in the process, they are more likely to support proposed changes and to play an active role in implementing them. These same staff can also provide real-time feedback on how these changes are working. Staff support can help to ensure that program improvements identified by the evaluation are made successfully and that they will "stick."

Program Leadership and the Board. While the staff planning session should contribute to decisions about how to use your evaluation data, program leadership need to have their own discussion about what changes are possible and what types of improvements should be prioritized. Program leadership should develop a concrete plan for implementing these changes with a time line set for the changes to be accomplished.

In addition, a report of the evaluation findings should be provided to the board with an opportunity for them to offer feedback on the implications for your program. The board should be kept in the loop on plans for program improvements, as any changes to program activities or budget will need their approval. The board should also be informed about the results of the staff planning session. Once plans for changes to your program have been approved by the board, the board should be kept informed about progress in implementing these changes.

APPENDIX A: Key Resources

The resources listed below are key non-HFRP resources to which you may want to refer in planning your evaluation strategy. This list is selective, rather than comprehensive—we tried to select the resources that we feel are most useful in planning an OST program evaluation.

Table 4: Evaluation Resources

Topic	Resources
Finding an Evaluator	The **American Evaluation Association** (AEA) has an **online database of evaluation consultants**. The database is searchable by keyword (such as area of expertise) and by state.
Existing Data Sets	The U.S. Department of Education's **National Center for Education Statistics Common Core of Data** collects descriptive data about all public schools, public school districts, and state education agencies in the United States, including student and staff demographics. The Annie E. Casey Foundation's **KIDS COUNT Project** provides national state-by-state data that measures children's educational, social, economic, and physical well-being.
Logic Model Development	The **W. K. Kellogg Foundation Logic Model Development Guide** provides detailed steps on how to create a logic model for a program, including how to use the logic model in planning for evaluation. The University of Wisconsin Cooperative Extension has an online training course, **Enhancing Program Performance with Logic Models**, that helps program practitioners use and apply logic models.
Visualizing Data	**Many Eyes**, a project of BM Research and the IBM Cognos software group, is a collection of data visualizations that allows users to view, discuss, and share visualizations. **A Periodic Table of Visual Elements** provides a comprehensive overview of the major types of visualizations that can be done with data and other types of information. **Visual.ly** is a collection of the "best examples" of infographics and data visualizations found on the web. It also allows users to create their own visualizations. **Edward Tufte's website** provides information about best practices for designing visualizations of data, and includes numerous examples of well-designed visuals.
Measurement Tools and Instruments	The Forum for Youth Investment offers two relevant resources: *Measuring Youth Program Quality: A Guide to Assessment Tools,* **2nd Edition**, which compares and describes program observation tools, and *From Soft Skills to Hard Data: Measuring Youth Program Outcomes*, which reviews eight youth outcome measurement tools that are appropriate for use in afterschool and other settings. The United Way's **Toolfind** online directory is designed to help youth-serving programs find measurement tools for a variety of youth outcomes. The Northwest Regional Educational Laboratory's *Out-of-School Time Program Evaluation: Tools for Action* **(PDF)** guide offers tools and tips for conducting surveys and focus groups as part of a larger evaluation. The Partnership for After School Education's *Afterschool Youth Outcomes Inventory* **(PDF)** is a comprehensive tool for afterschool practitioners to use in assessing and articulating their programs' impacts on youth.
Collecting Participation Data	*Afterschool Counts! A Guide to Issues and Strategies for Monitoring Attendance in Afterschool and Other Youth Programs* provides practical information on attendance data for program directors.
Examples of Afterschool Program Evaluations	**Child Trends' LINKS (Lifecourse Interventions to Nurture Kids Successfully) Database** summarizes evaluations of OST programs that work to enhance children's development. Child Trends also has a number of **resources on OST program evaluation** that programs can reference in planning evaluation.

In order to keep this information timely and relevant, we plan to update this list periodically. Please let us know if you have recommendations for resources that you have found especially helpful in conducting your program evaluation. Email Erin Harris (erin_harris@gse.harvard.edu) with your suggestions.

APPENDIX B: Discussion Questions

The discussion questions below include issues that your program may want to consider as you work through the steps in this toolkit. These questions should be of most use as you move through the corresponding step of this resource, and as you begin the process of conducting your evaluation. You should refer to information in the relevant sections as you reflect on these questions.

Step 1: Determining the Evaluation's Purpose

- What factors are driving our decision to evaluate our program?
- What do we want to know about our program?
- What do we already know based on existing data that can help the evaluation?
- What do our stakeholders (including funders) hope to get out of the evaluation?

Step 2: Developing a Logic Model

- What are the overarching goals that we hope to achieve with our program?
- What are the specific desired outcomes of our program for youth and families?
- What inputs and outputs are necessary to move toward those outcomes?
- What are possible measures that we can use to track progress toward our outcomes?

Step 3: Assessing a Program's Capacity for Evaluation

- Who are the primary stakeholders for our program and the evaluation?
- How can we best involve our stakeholders in the evaluation process?
- What resources must be in place to conduct our evaluation?
- Who should conduct the evaluation?

Step 4: Choosing the Focus of Evaluation

- Which tier of evaluation is most appropriate for our program, given its current developmental stage?

Step 5: Selecting the Evaluation Design

- Does it make more sense for us to conduct a formative/process evaluation or a summative/outcome evaluation?
- Should we use an experimental, quasi-experimental, non-experimental, or pre-experimental design for our evaluation?
- Should we collect quantitative or qualitative data, or both?
- How long do we have to conduct the evaluation?

Step 6: Collecting Data

- Should we collect data on all participants, or just a subsample? If just a subsample, how do we select that subsample?
- What data collection methods make most sense for us to use in terms of their appropriateness to our program and evaluation goals, and their feasibility?
- What participation data can we collect, and how will these data feed into the evaluation?

Step 7: Analyzing Data

- Do we need to do statistical analysis?
- Do we need additional software or external expertise to help analyze the data?
- Do our data adequately answer our evaluation questions?

Step 8: Presenting Evaluation Results

- What information about our evaluation is most interesting and useful to our various stakeholders, and how can we best communicate the findings to them?
- What creative methods can we employ to reach various audiences with our evaluation results?

Step 9: Using Evaluation Results

- What improvements can we realistically make to our program in response to the evaluation findings?
- How can we use the evaluation findings to inform future evaluations?
- How can we use evaluation findings in our grant applications to promote our program?
- How can we best present evaluation findings to the community to get their buy-in to our program?

APPENDIX C: Glossary

Accountability. The process in which an organization enters into a contract with a public or private agency or funder, where the organization is required to perform according to agreed-on terms, within a specified period, and using specified resources and standards.

Baseline data (also known as pretest data). Data collected before the program is implemented. These data are used as a starting point for making comparisons.

Benchmark. (1) an intermediate target to measure progress in a given period, or (2) a reference point or standard against which to compare performance or achievements.

Case study. A data collection method that focuses on one individual over a set period of time, taking an intensive look at that individual's program participation and the effect on his or her life. Case studies can include formal interviews, informal contacts such as phone calls or conversations in hallways, and observations.

Causal. Relationships in which a reasonable case can be made that a specific program or activity directly led to a given outcome. To establish a causal relationship, an experimental design must be implemented to rule out other possible factors that may have contributed to the outcome.

Closed-ended question. A form of question that is answered using a set of provided response options, such as a "yes" or "no," a selection from multiple choices, or a rating on a scale.

Correlation. An indication of some relationship between two events: for example, as program participation increases, academic outcomes improve. However, unlike causation, a strong argument cannot be made that one event caused the other.

Control/comparison group. Used in experimental or quasi-experimental studies—generally called a control group for an experimental study, and a comparison group for a quasi-experimental study. A control/comparison group consists of a set of individuals who do not participate in the program, but who are similar to the group participating in the program, to allow a comparison of outcomes between those who have gone through the program, and those who have not. For an experimental study, individuals are randomly assigned to the control group. For a quasi-experimental study, the comparison group is selected, based on a specific set of criteria, to be similar to the program group, (e.g., students in the same grade who attend the same schools).

Developmental stage. How far along a program is in its implementation. Developmental stage is a function of the amount of time the program has been in operation, as well as how well-established its activities, goals, inputs, and outputs are.

Descriptive design. An evaluation design used primarily to conduct formative/process evaluations to explain program implementation, including characteristics of the participants, staff, activities, etc. The data are usually qualitative, although some quantitative data may be included as well, such as counts or percentages describing various participant demographics.

Document review. A data collection method that involves examining existing program records and other information collected and maintained by the program as part of day-to-day operations. Sources of data include information on staff, budgets, rules and regulations, activities, schedules, participant attendance, meetings, recruitment, and annual reports. Data from document review are most often used to describe program implementation, and also as background information to inform evaluation activities.

Duration. The history of program attendance over time, as measured in years or program terms.

Evaluation. A process of data collection and analysis to help measure how successfully programs have been implemented and how well they are achieving their goals.

Evaluation strategy. A deliberate and intentional plan for evaluation, developed with the goal of incorporating the lessons learned into program activities. As part of this larger evaluation strategy, evaluation is not viewed merely as a one-time event to demonstrate results, but instead as an important tool for ongoing learning and continuous improvement.

Experimental design. An evaluation design with one distinctive element: random assignment of study participants into the program group and comparison/control (i.e., non-program) group. The goal of this design is to rule out possible causes beyond the program that could lead to the desired outcomes.

Five Tier Approach. The evaluation process as a series of five stages: (1) Conduct a needs assessment, (2) Document program services, (3) Clarify the program, (4) Make program modifications, and (5) Assess program impact.

Formative/process data. Collected during program implementation to provide information that will strengthen or improve the program being studied. Findings typically point to aspects of the program's implementation that can be improved for better participant outcomes, such as how services are provided, how staff are trained, or how leadership decisions are made.

Goals. What the program ultimately hopes to achieve.

Impact. A program's effectiveness in achieving its goals.

Inputs. Resources that a program possesses and uses to work toward its goals, including staff, funding resources, community partners, and other supports that the program has to ensure that the activities it undertakes will have an impact on its desired outcomes.

Intensity. How often each youth attends a program during a given period as measured in hours per day, days per week, and weeks per year.

Interviews or focus groups. Data collection methods that involve gathering detailed information from a specific sample of program stakeholders about program processes. These methods require a set of questions designed to elicit specific information. This method is most often used to collect qualitative data, such as how participants feel about a particular activity. Interviews are usually conducted one-on-one with individuals (although several individuals can be interviewed together) either in person or over the phone. Focus groups generally occur in person (although they can be conducted by conference call or web meeting) and involve gathering individuals to provide feedback as a group.

Learning and continuous improvement. How evaluations can be used to inform internal management decisions about what is (and is not) working, where improvement is needed, and how to best allocate available resources. Rather than being merely a static process where information is collected at a single point in time, evaluation becomes a practical tool for making ongoing program improvements.

Logic model. A concise way to show how a program is structured and how it can make a difference for program participants and community. A logic model is a one-page visual presentation—often using graphical elements such as charts, tables, and arrows to show relationships—displaying the key elements of a program (inputs and activities), the rationale behind the program's service delivery approach (goals), the intended results of the program and how they can be measured (outcomes), and the cause-and-effect relationships between the program and its intended results.

Longitudinal data. Data that are collected over multiple years to track changes over time.

Management information systems (MIS). A way to electronically collect, organize, access, and use the data needed for organizations to operate effectively.

Measures of effort. Measures that assess the effectiveness of outputs. They assess how much you did, but not how well you did it. These measures address questions such as: What activities does the program provide? Whom does the program serve? Are program participants satisfied?

Measures of effect. Measures that assess changes that a program expects to produce in participants' knowledge, skills, attitudes, or behaviors. These measures address questions such as: How will we know that the children or families that we serve are better off? What changes do we expect to result from our program's inputs and activities?

Needs assessment. Attempt to better understand how a program is, or can, meet the needs of the local community.

Non-experimental design. Evaluation design that lacks statistical comparative data to allow causal statements about a program's impact. There are two subtypes of non-experimental designs: descriptive designs and pre-experimental designs.

Observation. A data collection method that involves assigning someone to watch and document what is going on in your program for a specified period. This method can be highly structured—using formal observation tools with protocols to record specific behaviors, individuals, or activities at specific times—or it can be unstructured, taking a more casual "look-and-see" approach to understanding the program's day-to-day operation. Data from observations are usually used to describe program activities and participation in these activities, and are often used to supplement or verify data gathered through other methods.

Open-ended question. A form of question that does not limit responses to a specific set of options, but allows the individual to provide his or her own response (i.e., not multiple choice).

Outcomes. A program's desired short-term, intermediate, and long-term results. Generally, short-term outcomes focus on changes in knowledge and attitudes, intermediate outcomes focus on changes in behaviors, and long-term outcomes tend to focus on the program's larger impact on the community.

Outputs. The services that a program provides to reach its goals. These include program activities offered to youth participants as well as other activities offered to families and the local community.

Participation data. Data on program participant demographics (e.g., age, gender, race), program attendance, demographics of the schools and communities that the program serves, and feedback from participants on why they attend (or do not attend) and their level of engagement with the program.

Performance measures. Data that a program collects to assess the progress made toward its goals. These data include measures of effort and measures of effect.

Pre-experimental design. An evaluation design that collects quantitative summative/outcome data in instances when resources do not allow for a causal design to examine outcomes. While the data collected may look similar to an experimental or quasi-experimental study, these studies lack a control/comparison group and/or pretest/posttest data collection. Outcomes measured may include some statistical analysis, but cannot make a strong case for a cause-and-effect relationship between program activities and outcomes.

Pretest data (also known as baseline data). Data collected before the program is implemented to be used as a starting point for making comparisons.

Posttest data. Data collected after program participants have participated for a period of time to demonstrate program outcomes. These data are often compared to pretest or baseline data, to show improvements from before program participation.

Program group. Those who participate in the program. This is the sample from which to collect data related to program implementation and outcomes.

Program monitoring. An evaluation method that involves documenting the services a program provides in a systematic way. Program monitoring tracks how a program is spending their funds and describes the details of the program activities including information about their frequency, content, participation rates, staffing patterns, staff training provided, and transportation usage.

Purposive sample. Randomly selected program participants to participate in the evaluation. This method is generally used when it is not feasible to collect data from the entire program group, due to limited resources and/or a large number of program participants. Similar to random assignment, random selection involves using a system to ensure that each program participant has an equal chance of being chosen to participate in the study or not. This method helps increase the likelihood that the study sample truly represents the overall program.

Qualitative data. Descriptive, rather than numerical, information that can help to paint a picture of the program. This type of data is subjective and shows more nuanced outcomes than can be measured with numbers. This type of data is generally used for formative/process evaluations, but can also help to flesh out and explain findings from summative/outcome evaluation, for example, to provide specific details about how participants' behavior has changed as a result of the program.

Quantitative data. Countable information, including averages, statistics, percentages, etc. These data can be used descriptively as formative/process data for evaluation—for instance, to calculate the average age of participants. However, these numbers are more commonly used as summative/outcome evaluation data—for instance, demonstrating improvements in participants' test scores over time.

Quasi-experimental design. An evaluation design used to try to establish a causal relationship between program activities and outcomes when experimental design is not possible. Quasi-experimental designs are similar to experimental designs except the treatment and control groups are not randomly assigned. Instead, existing program participants (the treatment group) are compared to a comparison group of similar non-participants (e.g., their peers attending the same schools). Quasi-experimental designs frequently include an attempt to reduce selection bias by matching program participants to non-participants, either individually or as a group, based on a set of demographic criteria that have been judged to be important to program outcomes (school attended, age, gender, etc.). However, these groups may still differ in unanticipated ways that may have a major effect on outcomes.

Random assignment. A method of selecting a program and comparison group for an experimental study. It involves a specific selection procedure in which each individual has an equal chance of being selected for each group. This technique eliminates selection bias and allows the groups to be as similar to one another as possible, since any differences between them are due only to chance.

Random selection. A method of selecting a sample of program participants to participate in an evaluation. Like random assignment, it involves a specific selection procedure in which each individual has an equal chance of being selected, but rather than being selected into either the program group or comparison group, random selection involves program participants only (without a comparison group of nonparticipants), who are selected to be included in data collection for the evaluation, or not. It also differs from random assignment in that those who are not selected do not become a comparison group. This method is used to get a representative sample of a program when it is not feasible to collect data on all participants.

Research studies. An attempt to answer specific hypotheses, or research questions, using data that are collected using deliberate methods and analyzed in a systematic way. Evaluations are one type of research study that focuses on a specific program or initiative. Other types of research studies may look more broadly across a number of programs or initiatives to address the hypotheses or questions of interest to the researchers.

Reliability. The consistency of a data collection instrument. That is, the results should not vary wildly from one use to the next, although repeated uses with the same group over time will hopefully show positive changes in participant outcomes, and administering the instrument to different groups will likely have some variation in results.

Results. Findings that are produced from an evaluation to show how the program is implemented and what outcomes it produces.

Sample. A subset of program participants that are selected to participate in an evaluation. Sometimes, a program is able to collect data on all participants, but often this is not feasible nor a good use of resources, so a representative subset is selected. The sample may be selected randomly (in an attempt to avoid any selection bias), or based on specific types of participants of particular interest in the evaluation (e.g., a specific age range, low-income youth). Sometimes, the sample is just one of convenience, based on participants who agree to participate and have permission from their parents.

Secondary Source/ Data Review. A data collection method which involves using existing data sources (that is, data that were not collected specifically for the evaluation) that may contribute to the evaluation. These sources include data collected for similar studies to use for comparison, large data sets, school records, court records, and demographic

data and trends. As with document review, these data are most often used to describe program implementation, and as background information to inform evaluation activities.

Selection bias. The chance that the treatment group and comparison group are different from each other in ways that might affect their outcomes, based on how they were selected for each group. If the two groups were not randomly selected, the youth who attend the program may over- or under-represent certain characteristics of the overall population of interest, meaning that the treatment group and comparison group may not be starting on equal footing.

Stakeholders. Those who hold a vested interest in the program. They include anyone who is interested in or will benefit from knowing about the program's progress, such as board members, funders, collaborators, program participants, families, school staff (e.g., teachers, principals, and superintendents), college or university partners, external evaluators, someone from the next school level (e.g., middle school staff for an elementary school-age program), and community partners.

Survey. A data collection method designed to collect information from a large number of individuals over a specific period. Surveys are administered on paper, through the mail, by email, or on the internet. They are often used to obtain data that provide information on program participants' backgrounds, interests, and progress.

Subsample. A set of program participants selected for analysis in the evaluation. Subsamples may be randomly selected in an effort to represent the entire group of program participants. Subsamples may also be selected based on a set of criteria of particular interest in the evaluation: for example, participants who are seen as most in need of program services (e.g., from low-income families or those who are not performing well in school), or a specific group of interest (e.g., female participants).

Summative/outcome data. Data collected to determine whether a program is achieving the outcomes that it set out to achieve, often using an experimental or quasi-experimental design.

Tests or Assessments. Data collection methods that include such data sources as standardized test scores, psychometric tests, and other assessments of a program and its participants. These data often come from schools (especially for academic tests), and thus can also sometimes be considered secondary source data. This method is most often used to examine outcomes, often using an experimental or quasi-experimental design.

Treatment Group. Another way of referring to the program group when this group is compared to a control/comparison group of nonparticipants.

Validity. Whether an evaluation instrument is actually measuring what you want to measure.