Acknowledgements
We would like to thank the many nonprofit organizations and leaders who helped us frame the scope and content of the Guide through their willingness to respond to surveys, serve on our Practitioners Advisory Group, review chapters and provide other expertise or assistance as needed. In particular, we would like to thank Martha Breunig, Ray Considine, Deborah DeBare, Diane D’Errico, Joanne Donoghue, James Earley, Jim Haskell, Karen Jeffreys, Kathy Jellison, Bruce MacDonald, Denise K. Maguire, Francine Mantak, John Pearson, Charlotte Ryan, Bob Sable, Surl Silberman, Joyce Strom, Jon Thompson, and Judith E.Wise, CGSE.

Finally, we would like to thank Marilyn Anderson Chase and Carlos Martinez of the United Way of Massachusetts Bay for making this Guide possible.

Deborah Linnell
Zora Radosевич
Jonathan Spack
Program evaluation has become an academic field of research and investigation most often practiced by professional evaluators. The field of evaluation, which is actually a rather young social science, has developed a terminology that has become so scientifically and academically based that it is often difficult for non-practitioners to understand. Scientific, academic style evaluation has its place and can be quite useful in some situations. But the cost, level of scientific rigor, and academic terminology has made evaluation frightening, off-putting, or simply boring to the majority of action-oriented, program-focused nonprofit managers and staff. And it has caused us to lose sight of the fact that any ordinary person, including program participants, can engage in figuring out whether or not a program is effective, a system is working, or an agency is having an impact. Done well, a simple, yet relevant, evaluation practice can significantly improve a program’s effectiveness.

Why Evaluate?
Accountability is the primary reason for evaluation. Many small and mid-sized nonprofits conduct formal evaluations because it is required by their funders. Government agencies, foundations, United Way, and even individual donors are scrutinizing nonprofits more closely for accountability and effectiveness. Program participants, boards, staff, and the public are also demanding greater accountability. Nonprofit organizations are finding that it makes sense to place new emphasis on evaluation in order to:

- prove they are worthy of the public trust and the dollars given to them;
- show contributors that their social investment is working;
- ensure that the people for whom the organization exists are receiving tangible, real benefits;
- better understand staffing patterns and program delivery from a management perspective;
- plan; and
- do a better job.

Who Should Evaluate?
Almost anyone can evaluate. Participatory evaluation enables users of services to take part in designing and implementing the evaluation of those services. The style and values of the organization and the programs being evaluated will determine who participates in the evaluation process. A flatter, more inclusive organization is likely to include more people in the evaluation process. Different types of programs or activities will call for different types of evaluations and different levels of involvement.

Generally, the board of directors is accountable for ensuring that evaluation is taking place, whereas the executive director is responsible for managing the evaluation process. Executive directors have the role of cultivating an atmosphere of openness, continuous learning, and skill at hearing constructive criticism at all levels of the organization. An executive director must have these skills and competencies to successfully lead an organization that continually learns, innovates, and improves on behalf of its clients and its mission.

Leadership must demonstrate a spirit of belief in the importance of evaluation. If evaluation is dismissed by the executive director as an annoyance forced upon the agency from the outside, staff will feel the same way. The executive director must also under-
stand the level of evaluation that is appropriate. It does not make sense to burden program staff with evaluation plans that are so time-consuming they don’t have time to run the program. Be aware of how ready your organization is to create and implement new systems and structures.

It is not appropriate for the executive director to be the evaluator, but he or she needs to provide impetus, resources, and attention to evaluation. The task is to demonstrate a management commitment that speaks to the importance of evaluation by using evaluation itself to coach staff to new heights, involve clients in services, create better programs, increase publicity, and improve funding for the organization.

Organizational Transparency

Organizational transparency is the state an agency reaches when it can communicate clearly and openly about itself at all levels to all stakeholders, including the staff, board, clients, funders, and community constituents. Evaluation will help to create the clarity needed and lay a fact-based foundation for telling the organization’s story. Anecdotal stories serve a purpose, but they do not “prove” the organization’s relevance to the community with the same credibility that formal evaluation does.

Nonprofits are ultimately accountable to the public. As the information age demands more data and open communication, organizational transparency will continue to increase. Key stakeholders need to know that the organization is holding true to its mission through meaningful programs supported by sound management practices. Evaluation is the tool for demonstrating accountability.

Framing Questions for Evaluation, or What Do You Need to Consider?

It is important to be very clear about what you are evaluating, why you are evaluating, and for whom you are evaluating. Paying attention to these questions will help you shape your evaluation into a more useful tool.

What Are You Evaluating?

Think about what you are evaluating. Is it a start-up initiative? Ongoing services? Outcomes for program participants? The effectiveness of management functions such as internal communication, public relations, fund development, or governance? You will use different types of evaluation and different tools depending upon what you want to evaluate.

Why Are You Evaluating?

Why are you evaluating certain activities? Being clear about why and what you are evaluating will help you figure out how to evaluate it.

• Is it to find out if the activities are helping clients?
• To inform future planning?
• To understand better how an initiative got off the ground?
• To find out if activities are having an overall impact?
• To understand the impact of the activities on the organization?

Each of these questions gives rise to a different type of evaluation.

Who Is Your Target Audience?

How you package the findings of the evaluation will change depending upon the audience it is intended to reach. Is the evaluation:

• Internal to a team or department?
• For the staff of an organization?
• For the management team?
• For clients?
• For the entire agency?
• For the funders?
• For the general public?

Evaluation measures merit and value, but it is also a communication tool, one of the most effective a nonprofit can have.
Evaluation Basics

Types of Standard Evaluation

Needs Assessment: A needs assessment is a way to explore the extent and depth of an issue or problem in such a way that a group can determine whether a case can be made for addressing the need. A needs assessment can determine the number and characteristics of individuals or institutions that would constitute the targets of a program. It can also be used to help design new programs or justify the continuation of existing programs.

Monitoring: Monitoring of activities produces regular, information that answers questions about whether a program or project is being implemented as planned. Monitoring helps identify problems and facilitate resolution in a timely way.

Formative Evaluation: Formative evaluation answers questions about how to improve and refine an ongoing program or a program that is being developed. Formative evaluation usually is undertaken during the initial or design phase of a project. It can, however, be helpful in assessing the activities of an established program. Formative evaluation puts data collection systems into effect early in the life of the program, capturing data that might not be available later on. Formative evaluation may include process and impact studies. Typically, the findings from a formative evaluation are provided as feedback to the program staff after an initial start-up period to help them make corrections during the course of the program's life.

For example:
A needs assessment has shown that a certain number of girls in high school have experienced dating violence. You decide to create a program to address this problem. Formative questions might include:

- What is the number of girls who can be served by this program?
- How many of those girls are being reached?
- Of those being reached, how many are choosing to participate in the program?
- Which approach for reaching girls is working and why?
- Can this approach be extended to reach more girls?

The answers to these questions can lead to program refinements that could increase the number of girls participating.

A summative evaluation might include similar questions, but it would take place at the end of the program, in order to judge the program's effectiveness retrospectively.

Outcome Evaluation: Outcome evaluation assesses the effectiveness of a program in producing change. Outcome evaluation focuses on difficult questions that ask what happened to program participants and how much of a difference the program made for them. Impact or outcome evaluation is useful when it is important to know whether and how well the objectives of a project or program were met.

For example:
Outcome questions for a smoking program might include:

- Did the program succeed in helping people to stop smoking?
- Was the program more successful with certain groups of people than with others?
- What aspects of the program did participants find gave the greatest benefit?

Process evaluation: Process evaluation documents the process of a program’s implementation. Process evaluation helps stakeholders see how a program outcome or impact was achieved. The focus of a process evaluation is on the types and quantities of services delivered, the beneficiaries of those services, the resources used to deliver the services, the practical problems encountered, and the ways such problems were resolved.

For example:
Process evaluation questions might include:

- What specific interventions were put into place by the program in order to tackle the problem?
- How many communities were selected for pilot programs, and how were they selected?
- What were the kinds of problems encountered?
- How were the problems resolved?

Information from process evaluation is useful for understanding how program impact and outcome were achieved and for program replication.

**Participant-oriented Evaluation:** Participant-oriented evaluation, although often described in a variety of ways, has several universal characteristics.

- Participant-oriented evaluation emphasizes the human element of an activity, focusing on describing and interpreting what is taking place in the program and the context in which it is taking place, rather than attempting to yield conclusive judgments.
- Participant-oriented evaluation is responsive to the flow of the program, rather than rigidly conforming to predetermined evaluation plans.
- Participant-oriented evaluation allows for differing values or viewpoints to be expressed, rather than holding that a particular value is the correct one.

Participant-oriented evaluation employs a variety of methods, including interviews, observation, and document review. It is especially useful in programs where it is important to capture feelings, reactions, and behavioral changes that cannot be easily quantified.

Proponents of participant-oriented evaluation believe that it has a greater potential than other evaluation approaches for learning about a program’s effects because of its openness, flexibility, and sensitivity to human issues.

**For example:**
Participant-oriented evaluation is appropriate for programs designed to bolster self-esteem. Few standardized instruments, such as scales, surveys, and self-measures, could adequately assess changes in self-esteem after exposure to a program. A participant-oriented evaluator would observe the program in operation, review documents such as records from group meetings, and would note if participants showed changes that are difficult to quantify, such as having more hope for the future, seeking adequate housing, or, making life-changing decisions as a result of the program.

**Selected Methods of Evaluation**
Listed below are some of the most often used evaluation methods:

- **Quantitative evaluation methods:** This method uses data collection and statistical methods to produce numerical values to describe the outcome of a program or project.
- **Qualitative evaluation method:** This method is used to produce narrative descriptions of activities, processes, and outcomes that are generally based on observation and interviews.
- **Sampling:** Sampling draws information from a representative or random portion of a total population to make judgments about the whole. Sampling is used with the survey technique or for select observations from a compilation of records.
- **Survey:** A survey involves the preparation of a series of questions, specifically tailored to evaluate a program, which are then asked of program participants or other program stakeholders. A survey can be performed through face-to-face or telephone interviews, or self-administered questionnaires. It can be distributed randomly, to a representative sampling, or to all program participants, depending on the purpose of the survey.

**About surveys:** A survey should
- be as short as possible, written in simple, neutral language, and accompanied by clear directions;
- use closed-end (yes or no) questions as much as possible and guard against asking leading questions; and
- leave as little as possible to interpretation by the survey reader.

The survey tool must be valid for the evaluation itself to be valid. In developing a survey, it is important to seek help from books, classes, or experts in the community.

- **Testing:** Testing measures an individual’s acquired skills. It is most often used in learning-related programs. Pre- and post-testing can be used to more accurately gauge the learning that can be attributed to the program being evaluated. Questions on pre- and post-tests should be simple and close-ended, and should leave little to interpretation.
**Observation**: This method depends on using an objective third party to observe an activity. The observer can use quantitative methods of counting or categorizing behaviors or qualitative methods of observing activities and reporting on progress in a more general way.

**Document analysis**: This type of analysis involves the review and compilation of existing records and materials within files, such as intake and exit interviews, and other documents.

**Anecdotal method**: This method involves informal assessments by a group, such as program staff, that has begun to see or feel a trend that has not yet been documented. Anecdotal evidence is often used to determine when to do a needs assessment.

**Outcome Measurement**

Many funders, most notably United Way of America, have simplified evaluation requirements to avoid overwhelming grantees with evaluation projects that are beyond the scale of the programs being evaluated. This outcome measurement approach has two basic goals:

- To see if the program really makes a difference in the lives of people and
- To help organizations improve services.

The *outcome measurement approach* is an accessible evaluation approach, intended for use by people without a background in evaluation. Outcome measurement focuses on the benefits to the program participant. It is attractive to community-based nonprofits because it allows them to develop their own outcomes with input from agency stakeholders instead of having outcomes imposed on them by funders, regulatory bodies, or other groups outside the organization.

The outcome measurement process is a more participatory form of evaluation in that stakeholders, including consumers of services and other constituents, board members, staff, and community people can all be involved.

**A Step-by-Step Outline of the United Way Model**

The outcome measurement approach promoted by United Way of America includes the following main steps:

- Get ready.
- Choose outcomes.
- Specify indicators for the outcomes.
- Prepare to collect data on your indicators.

Other steps involve trying out the outcome measurement system, improving any problems, analyzing data, and deciding how to report findings.

1. **Getting Ready**

The executive director’s first steps are to:

- Choose an outcome measurement work group. The group should be representative of various perspectives within the organization, from frontline workers to management staff.
- Ensure there is a common understanding of terms. Following is a brief glossary of useful terms.
  - **Inputs** are resources dedicated to the program and include money, staff and staff time, facilities, equipment, supplies, and volunteers and their time.
  - **Activities** are what the program does with its inputs to fulfill its mission. Activities are the strategies that the organization undertakes to provide its services in a meaningful way.
  - **Outputs** are the products of program activities. Outputs can be measured in terms of work accomplished, using factors such as number of clients reached, number of classes taught, or number of materials distributed.

---

2 United Way of Massachusetts Bay can provide its affiliates and other nonprofit organizations with more comprehensive information on outcome measurement than appears in this chapter.

3 The majority of the following section is taken from *Measuring Program Outcomes: A Practical Approach*, a publication of United Way of America, by Harry Hatry, Therese van Houten, Margaret C. Plantz and Martha Taylor Greenway.
Outcomes are the benefits or changes achieved by individuals or populations during or as a result of participating in program activities. Outcomes relate to behavior, skills, knowledge, attitudes, values, condition, and status. Outcomes are changes in behavior, knowledge, thought, ability, or condition following a program.

Indicators are the pieces of information that tell you how well a program is doing regarding an outcome. For example, if the outcome for attending prenatal classes is assumed to be a healthy baby, then the indicators will be factors that demonstrate that a baby is healthy. Birth weight is considered a standard measure of the health of a baby, so birth weight becomes one of several indicators for determining a healthy baby outcome for prenatal classes.

Once the outcomes measurement team understands the basic concepts and evaluation goals and has received the support of the executive director and board of directors, the team must then:

- decide which program to evaluate first;
- develop a timeline;
- develop ways to engage stakeholders in defining program outcomes;
- choose which indicators will be used;
- develop the data collection system to quantify the indicators;
- run a pilot evaluation and tweak it before going ahead with the full scale effort; and
- decide the format of reporting and disseminating information.

2. Choosing Outcomes
Choosing outcomes is the most participatory aspect of outcome measurement. Reach wide and ask for input from past and current program participants, constituents, program staff, and volunteers. Conduct focus groups. Talk to funders. Ask programs that deliver similar services what they feel outcomes might be; ask about both short-term and long-term outcomes.

It is very important to have fairly specific outcomes to measure. Long-term outcomes like saving the world (or even one individual) are too broad and far-reaching. Too many factors influence an individual’s life for any one program to claim success for completely changing a person’s life. For example, a smoking cessation class may have a long-term outcome of helping an individual quit smoking, but it can not reasonably claim to save that person’s life.

A case could be made, however, that the class decreased the individual’s chances of getting lung cancer – as long as the capacity exists within the agency to gather that data.

3. Specifying Indicators
Since indicators represent information showing that an outcome is being met, indicators must be specific, observable, measurable characteristics or changes that will represent achievement of the outcome. They must also identify the statistic (number or percentage) that will summarize the program’s performance on the outcome.

For example, for a smoking cessation class, the key outcome is that participants stop smoking. Indicators could include:

- the number and percentage of participants who report that they have quit smoking by the end of the course;
- the number and percentage of participants who have not relapsed six months after the program was completed; and
- the number and percentage of participants who have not relapsed one, two or five years after the program was completed.
Comparison of Major Data Collection Methods

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Review of Program Records</th>
<th>Self-Administered Questionnaire</th>
<th>Interview</th>
<th>Rating by Trained Observer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>Low</td>
<td>Moderate</td>
<td>Moderate to high, depending on how it is administered</td>
<td>Depends on availability of low-cost observers</td>
</tr>
<tr>
<td>Amount of training required for data collectors</td>
<td>Some</td>
<td>None to some, depending on how it is distributed</td>
<td>Moderate to high, depending on survey complexity and collectors’ previous experience</td>
<td>Moderate to high, depending on complexity and subtlety of observations</td>
</tr>
<tr>
<td>Completion time</td>
<td>Short, depending on amount of data needed</td>
<td>Moderate to long, depending on how it is distributed</td>
<td>Long</td>
<td>Short to moderate</td>
</tr>
<tr>
<td>Response rate</td>
<td>High, if records contain needed data</td>
<td>Depends on how it is distributed</td>
<td>Generally moderate to good</td>
<td>High</td>
</tr>
</tbody>
</table>

4. Choosing Data Collection Methods

Major sources for data by which to measure progress on indicators are existing records, interviews, or surveys of individuals, tests, or the use of trained outside observers.

Major data collection methods for use in outcome measurement include:

- review of program records from the organization or other agencies, such as schools, hospitals, referral agencies, courts, law enforcement agencies, or health departments;
- self-administered questionnaires or interviews that reach specific individuals such as program participants, relatives of program participants, staff at referring agencies, organization staff, or volunteers of the agency;
- rating by an observer trained to rate behaviors, facilities, or environments; and
- mechanical tests and measurements such as scales or blood tests.

*4. Taken from Measuring Program Outcomes: A Practical Approach, a publication of United Way of America, by Harry Hatry, Therese van Houten, Margaret C. Plantz and Martha Taylor Greenway.*
Sample Outcome Measurement Framework: Southside Children’s Agency⁵

In this example, we list potential outcomes along with one or more measurable indicators to help determine whether the outcome has been met. We also include data sources and collection methods that can provide the necessary information.

Once indicators have been identified and analyzed, one can further analyze program data to look for factors that may influence how successfully the outcomes have been met.

### Program: Teen Mother Parenting Education

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Indicator(s)</th>
<th>Data Source</th>
<th>Data Collection Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teens are knowledgeable of prenatal nutrition and health guidelines</td>
<td>Number and percent of program participants able to identify food items that are good sources of major dietary requirements</td>
<td>Participants</td>
<td>Self-administered survey after second week in the program</td>
</tr>
<tr>
<td>Teens follow proper nutrition and health guidelines</td>
<td>Number and percent of participants who do not smoke</td>
<td>Participants Teachers</td>
<td>Self-report on daily checklist Observation reported on record weekly</td>
</tr>
<tr>
<td></td>
<td>Number and percent of participants who eat at least four calcium servings and one of each other nutritional group each day</td>
<td>Participants</td>
<td>“Healthy Baby” checklist for recording daily food intake</td>
</tr>
<tr>
<td>Teens deliver healthy babies</td>
<td>Number and percent of participants within proper ranges for prenatal weight gain</td>
<td>School scales</td>
<td>Weekly weigh-ins</td>
</tr>
<tr>
<td></td>
<td>Number and percent of participants who take a prenatal vitamin</td>
<td>Participants</td>
<td>“Healthy Baby” checklist for recording daily food intake</td>
</tr>
<tr>
<td></td>
<td>Number and percent of newborns weighing 5.5 pounds or above or scoring 7 or above on Apgar scale</td>
<td>Hospital records</td>
<td>Contact hospital for birth records</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Influencing Factor</th>
<th>Data Source</th>
<th>Data Collection Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother’s age</td>
<td>Program records</td>
<td>Record review</td>
</tr>
<tr>
<td>Mother’s household income</td>
<td>Program records</td>
<td>Record review</td>
</tr>
<tr>
<td>Number of parenting classes mother attended</td>
<td>Program records (based on mother’s self-report)</td>
<td>Record review</td>
</tr>
</tbody>
</table>

⁵Taken from Measuring Program Outcomes: A Practical Approach, a publication of United Way of America, by Harry Hatry, Therese van Houten, Margaret C. Plantz and Martha Taylor Greenway.

5. Putting It All Together

The United Way model recommends that outcome measurement teams pretest their data collection instruments and procedures and conduct a pilot test of their measurement systems that includes analysis and reporting of their findings. They also recommend using the findings internally to change programs, support annual planning, and justify resource allocations, and externally to increase funding and/or enhance the agency’s public image.

The bottom line of outcome measurement, however, is finding out if your program is having any real benefit for participants. It is the most “real” and useful evaluation method, particularly for human
service agencies, to come along and should prove instrumental in pushing the nonprofit sector toward better programs with improved outcomes for their constituents.

**Emerging Trends — Feedback Loops**

Organizations typically do well at setting goals and objectives for programs. They often fall short, however, when it comes to framing:

- **process** questions about how the program is being implemented;
- **formative** questions about how to improve the program mid-course; and
- **outcome/impact** questions about how to measure how well the program did on behalf of its clients.

Some organizations jump into projects without ever justifying the need through a *needs assessment*, simply because a leader thought it was a good idea.

Without solid information about the current environment and how well existing programs are doing, it is difficult to identify the next set of priorities and goals effectively.

*Ongoing evaluation is a key part of providing information about current reality so that organizations can envision the future.*

Smart organizations will begin to implement feedback cycles for individual programs that can facilitate change and innovation more quickly than formal evaluation allows. The same checklist that is used in Chapter 5, *Strategic Planning and Thinking*, to measure an organization’s ability to think strategically, can be used to measure whether the organization is using meaningful feedback.

**Strategically link planning and evaluation;** the two activities are symbiotic. Ask the following questions, from the strategic planning checklist, that relate to evaluation:

- Does the organization *continuously gather data* from and about its external environment?
- Does the organization *learn from the past* and establish open, evolving plans focused on immediate issues?
- Does the organization *set intermediate goals that are measurable and observable*?
- Has the organization designated responsible individuals or teams for implementation?
- Does the organization encourage experimentation (from grassroots up), *measure improvement*, recognize, and reward and institutionalize positive change?
- Does the organization *support deep reflection* in the form of ongoing opportunities for reflection, questioning of underlying assumptions, and searching for deep systematic solutions/interventions to current problems?
- Does the organization *define its learning, planning, and evaluation cycle*?
A feedback loop can be as simple as the one described in the following diagram:

1. **Develop a strategic direction**
2. **Develop goals that we believe will help us move in the strategic direction**
3. **Develop programs that meet these goals, using specific outcome and indicators**
4. **Measure progress toward indicators, using program data collected with specific data collection tools at specific intervals**
5. **Use findings to learn more about our environment or discover new learning**
6. **Introduce findings to the entire organization (and other stakeholders)**
7. **Retool (improve) the current program (or goals or directions), using new learning and feedback from stakeholders**

A feedback loop can be as simple as the one described in the following diagram: