

Part 1:
JUMPstarting Your Program



Measuring
Outcomes

Module 5

Time: 2 hours
Limit: 25 participants

Introduction

Session Goals

Participants will examine approaches to measuring outcomes and explore important ways that findings from outcomes evaluations can contribute to strengthening their programs. By the end of the workshop, they should:

- Know how to define concrete, measurable outcomes for youth in their program
- Understand data-collection strategies
- Know how to measure the quality of the relationships between mentors and youth in their program
- Understand the uses they can make of their outcomes findings

The Basics

1. Outcomes are benefits for program participants during or after their involvement with the program. They involve changes in participants' knowledge, skills, attitudes, values, or behavior.
2. Measuring outcomes does not have to be a complex or expensive process. It should be built into ongoing program operations, and include measuring the quality of the mentor-youth relationship.
3. A program can use outcomes findings to demonstrate to funders and the community that it is a worthwhile investment, and to help itself identify both program strengths and aspects of its operations that need to be improved.
4. In judging outcomes findings, the best comparison for a program is itself. Programs should look at changes in outcomes over time in order to measure their progress.

Agenda

- 1. What Changes Are You Looking For?** (10 minutes)
Participants identify desired outcomes for the children and youth in their mentoring programs.
- 2. What Outcomes Can You Measure?** (15 minutes)
Participants refine their definitions of measurable outcomes.
- 3. Developing a Plan for Measuring Outcomes** (35 minutes)
Small groups develop approaches for collecting outcomes data.
- 4. Measuring the Quality of Mentoring Relationships** (30 minutes)
Participants learn to use the “Youth Survey.”
- 5. Making Use of Your Outcomes Findings** (25 minutes)
Small groups explore ways to use outcomes findings.
- 6. Now What?** (5 minutes)
Participants revisit the goals of this session.

Connections to Other Training Sessions

Some of the information and strategies referred to in this session are covered more fully in these JUMP trainings:

- “Screening Mentors”
- “Making and Supporting the Match”
- “JUMPstarting Your Mentors”

If members of your training group have already attended any of those sessions, you may want to draw on information they have learned there. If they have not yet attended those sessions, you will want to, where appropriate, encourage them to attend in order to reinforce and add to the information that is covered during this “Measuring Outcomes” training.

Preparation

Note: As this session should make clear, measuring some forms of outcomes is relatively straightforward, and the data are relatively easy to analyze. Measuring other forms of outcomes is more complex. This session should help practitioners understand what kinds of outcomes evaluations they can do on their own and when they might want to enlist the services of an outside evaluator. If you do not have experience working with outcomes evaluations, you may discover during the session that your participants are occasionally asking questions or raising issues that you find difficult to answer or address. If this occurs, let them know that you will try to get an answer and follow up with them later, or recommend that they speak to an experienced evaluator about the issue. If the training session is taking place as part of a larger conference, find out if there are people attending the conference who have expertise in this area. Then, if any participants raise questions during the session that you do not feel confident answering, you can try to connect the participants with those experts after the session.

1. Read the three trainer resources. You will need to use the first two (“Identifying Measurable Outcomes” and “Why Programs Should Measure the Strength of Mentor-Youth Relationships”) during Activity #2 and Activity #4. The third trainer resource (“Outcome Measurement: Showing Results in the Nonprofit Sector”) is a very readable article that will provide you with background information for leading this training session.
2. Read the handouts. They contain much of the information you need for leading this session. Note that Handout #4, “Youth Survey: Measuring the Quality of Mentor-Youth Relationships,” is a self-contained, 14-page guide that includes the survey, information for scoring it, and suggestions about how programs can make use of the findings. Allow yourself some extra time to learn the material in this handout, which serves as the basis of Activity #4, “Measuring the Quality of Mentoring Relationships.” Also be prepared to give a very brief (2 or 3 minute) summary of Handout #5, “Tracking the Length of Mentor-Youth Relationships” at the end of the same activity.
3. Visit the Web sites listed on Handout #8, “Resources for Measuring Outcomes,” so you are prepared to describe to participants the kinds of information they can find there.

4. Review the three curriculum modules, listed on the previous page, that include information and strategies relevant to “Measuring Outcomes.”
5. Prepare a set of handouts for each training session participant. (Copy the handout materials onto paper with three-hole punches so participants can keep them in a binder.)
6. Prepare transparencies of the three overheads.
7. Copy the goals of the session onto a flipchart. On a separate sheet of flipchart paper, copy the agenda.



Materials

Trainer Resources

- Trainer Resource #1: Identifying Measurable Outcomes
- Trainer Resource #2: Why Programs Should Measure the Strength of Mentor-Youth Relationships
- Trainer Resource #3: Reading Selection: "Outcome Measurement: Showing Results in the Nonprofit Sector"



Overheads

- Overhead #1: Outcomes
- Overhead #2: The Best Comparison for a Program Is Itself
- Overhead #3: Benchmarks



Handouts

- Session Goals and Basics
- Agenda
- Handout #1: Definitions
- Handout #2: Measuring Outcomes: Developing a Plan
- Handout #3: Worksheet for Developing an Evaluation Plan
- Handout #4: Youth Survey: Measuring the Quality of Mentor-Youth Relationships
- Handout #5: Tracking the Length of Mentor-Youth Relationships
- Handout #6: How Programs Can Use Outcomes Findings
- Handout #7: Why Evaluate? Two Case Studies
- Handout #8: Resources for Measuring Outcomes

You Will Need To Supply

Flipcharts, easels, markers, and masking tape
An overhead projector

Activities

1. What Changes Are You Looking For?

Participants identify desired outcomes for the children and youth in their mentoring programs.



- Introduce yourself to the group.

Display Overhead #1, “Outcomes,” and read the definition.

Then have each participant introduce herself or himself and briefly state *one* desired outcome for the children or youth served by her/his mentoring program.

As participants speak, record each “desired outcome” on the flipchart the first time it is stated. If the same “desired outcome” is repeated by other participants, put a check by it each time it is stated again.

Possible responses include: improved school behavior; improved academic performance; decreased gang involvement; decreased alcohol and other drug use; decreased delinquent behavior; improved anger management; increased self-confidence or self-esteem; increased knowledge of career possibilities; improved employability skills.

Note that programs may be concerned with several kinds of evaluations—for example, determining whether their mentor recruitment strategies are effective, or whether a particular partnership is cost-effective, or whether their mentor training is effective. This training session focuses on one form of evaluation: measuring outcomes for the children and youth in their mentoring programs.

Using the flipchart you have prepared, describe the goals of this training session. Then briefly review the agenda.

- Check to be sure each person has the handouts for this session. Briefly review the table of contents on the cover page of the handouts. Explain that you will be using some of the handouts during this session, and that all of them are intended to be resources that participants can use in their programs.



Refer participants to page 1 of the handouts, “Session Goals and Basics.”

Review the four “basics” that are listed there. This session will be examining those points more fully.

2. What Outcomes Can You Measure?

Participants refine their definitions of measurable outcomes.



- Tell participants that now you want them to think more about which outcomes they could usefully measure and how they would measure them. Refer them to Handout #2, “Measuring Outcomes: Developing a Plan,” and allow a few minutes for them to read it.
- Return to the list of “desired outcomes” you recorded on the flipchart during the previous activity. Select one of the outcomes (perhaps the most frequently mentioned one—the outcome with the most checkmarks by it). Write the outcome on a clean piece of flipchart paper.



Use information on the handout and the examples on Trainer Resource #1, “Identifying Measurable Outcomes,” to help you lead the following discussion.

Ask the group:

Is that outcome “concrete” and “measurable”?

If it is—

What would the “outcome indicators” be? What data would you collect in order to measure the outcome?

How would you collect the data?

When would you collect data?

If it is not—

How can you define the outcome more precisely so it is concrete and measurable?

What would the “outcome indicators” be? What data would you collect in order to measure the redefined outcome?

How would you collect it?

When would you collect it?

Record participants’ responses on the flipchart.

- Then, referring to item #5 on the handout, ask who will collect the data. Do they have the capacity among program staff? Would they need to hire an outside evaluator or recruit a volunteer with experience in this area?

Note that you will be working with item #6 (making use of outcomes findings) later in this session.

3. Developing a Plan for Measuring Outcomes

Small groups develop approaches for collecting outcomes data.

- Tell participants you want them to work in small groups to develop a plan for doing an evaluation of outcomes, applying the information they have just been discussing.



Refer them to Handout #3, “Worksheet for Developing an Evaluation Plan,” which they should use during their small-group work.

Organize participants into small groups of 4 or 5 people. Try to group together people who are interested in measuring the same outcome. (You can refer back to the list of “desired outcomes” you recorded on the flipchart during Activity #1, and have participants organize themselves into groups by choosing which one to work on.)

Each small group should:

1. Select a leader and a recorder/reporter, who will report out later to the full group of participants.
2. Decide on the outcome it wants to measure. (They should not choose the outcome that the whole group just discussed during the previous activity.) Then, using Handout #3 as a framework, the group will develop an approach for measuring that outcome.

Make sure each small group has a flipchart, easel, markers, and masking tape.

Allow 20 minutes for the small groups to complete their work. (Give a “5-minute warning” after 15 minutes so the groups know they have to begin to wrap-up.)

- Have each small group report out to the whole group on the evaluation plan it has developed.

After each presentation, allow time for feedback so participants can ask questions, offer additional ideas, and compare the similarities and differences in the small-groups' approaches and decisions.

4. Measuring the Quality of Mentoring Relationships

Participants learn to use the “Youth Survey.”

- Note that you want to take a step back for a minute.

Suggest that one desired outcome of a mentoring program could be “strong relationships between mentors and youth.” In fact, research has demonstrated that strong mentoring relationships can be seen as an early outcome that is likely to lead to other outcomes for youth, such as improved school behavior.



- Refer participants to Handout #4, “Youth Survey: Measuring the Quality of Mentor-Youth Relationships.” Ask them, for now, to just look at pages 3 and 4, the survey itself. Allow a minute for them to read the questions.

Then ask, “Would a survey like this be useful for your program? What could you learn from it?”

- Now allow a few minutes for participants to look through all of Handout #4, so they can understand what it contains and how it is organized.



Use the information on Trainer Resource #2, “Why Programs Should Measure the Strength of Mentor-Youth Relationships,” to introduce the material in the handout.

Briefly summarize the information contained in the handout about how to score the survey to find the ratings for each of the three qualities of mentor-youth relationships: 1) the extent to which the relationship is centered on the youth’s interests; 2) the youth’s emotional engagement; and 3) the extent to which the youth is dissatisfied with the relationship. (Don’t be intimidated by the number of pages describing the scoring system—it’s actually very straightforward.)

- Tell participants you want them to practice using and scoring the survey.

Organize them into pairs. One person should administer the survey by reading the questions aloud while the other person role-plays a youth and completes the survey, using the “Youth Survey” on pages 3 and 4 of the handout.

Allow 5 minutes for the pairs to complete this activity.

Then ask the pairs to score the completed “Youth Survey” for the quality of “Youth’s Dissatisfaction.”

- Now model the process for finding the average score. (This is described in Section C on the scoring instructions included in the handout).
 1. Tell the group to imagine that all the “youth” who just completed this survey are participants in one mentoring program. Have each pair give its score for “Youth’s Dissatisfaction.” As they do, write each score on the flipchart.
 2. Demonstrate how to find the average.
 3. Explain how to use the average as a benchmark. Note that, for this quality, a lower score is better because the quality being measured is a negative one.

During the discussion of benchmarks, use Overhead #2, “The Best Comparison for a Program Is Itself,” and Overhead #3, “Benchmarks,” to emphasize the point that these and other outcomes findings are an important tool for program improvement. Note that the BBBS ratings included with the scoring information are a good example of how to use external benchmarks—they are outcomes achieved by an “exemplary program” and can serve as a target to strive toward.

- Then model the process for finding the range of scores. (This is described in Section D on the scoring instructions included in the handout).
 1. Use the “Youth’s Dissatisfaction” scores for each pair that you have already recorded on the flipchart.
 2. Demonstrate how to organize the scores into categories and how to calculate the percentage of scores in each category.
 3. Explain how to use the percentages as a benchmark.





- Note that Handout #5, “Tracking the Length of Mentor-Youth Relationships,” contains information to help programs systematically measure the important quality of the relationships. Briefly summarize the information on the handout.

5. Making Use of Your Outcomes Findings

Small groups explore ways to use outcomes findings.



- Ask the group why they might want to use a “Youth Survey” in their programs? How could the findings be useful?

Focus on the idea of using the findings to identify where they could strengthen program elements, particularly screening, matching, and training mentors.

- Refer participants to Handout #6, “How Programs Can Use Outcomes Findings.” Briefly summarize the information, focusing on the bulleted “uses” at the end of the handout.

- Tell participants you want them to meet in groups to develop plans for making use of outcomes findings.

Organize the participants into four small groups. Refer them to Handout #7, “Why Evaluate? Two Case Studies,” and briefly review it.

Assign one of the case studies to each of the groups (so each case study is being addressed by two different groups).

Tell each group to select a leader and a recorder/reporter, who will report out later to the full group of participants. Make sure that each group has a flipchart, easel, markers, and masking tape.

Allow about 15 minutes for the groups to complete their work. (Give a “5-minute warning” after 15 minutes so the groups know they have to begin to wrap-up.)

AS AN OPTION: If you do not have enough time for small-group work, select one of the case studies in the handout and have a whole-group discussion about it. Then move on to Activity #6, “Now What?”

- Have the small groups report out to the whole group, beginning with each of the two small groups that addressed Case Study A.

After both of those small groups has made its presentation, allow time for the full group to give feedback by asking questions, offering additional ideas, and comparing the similarities and differences in the two groups' approaches and decisions.

Repeat this process with the two small groups that worked on Case Study B.

6. Now What?

Participants revisit the goals of this session.

- Tie together the activities from this training workshop by reviewing the goals, written on the flipchart, that you presented at the beginning of the session.
- Note that there are several good resources available for help in planning and conducting outcomes evaluations and using the findings in ways that benefit the program and its participants. (Materials in the handout packet should serve as one such resource.) Direct the group to Handout #8, "Resources for Measuring Outcomes," and review it.





Identifying Measurable Outcomes

The outcomes that programs decide to track should be:

- Based on clearly articulated, **achievable** program goals, and
- **Concrete** and **measurable**

As you lead the discussion about defining measurable outcomes during Activity #2, you may want to use the following points and examples:

1. Outcomes have to be defined concretely in order to be measurable.

Example: “Improved school performance” is too general a term to be measurable. It has to be made more concrete. “Improved school performance” could include a number of things—for example, whether youth have improved school attendance, turned in more homework on time, have better grades, or have been sent to the principal’s office less frequently. Each of these is measurable. Which ones, or other measurable indicators of “improved school performance,” would the program want to track?

Example: “Improved life skills” is another term that is too general to be measurable. Which “life skills” in particular is the program focusing on? More responsible sexual behavior? An improved ability to manage anger? Reduced drug use?

Example: “Reduced gang involvement” could have a number of meanings. Programs working with younger youth might be trying to prevent them from becoming involved with gangs, while programs working with older youth might be trying to move them away from current gang involvement or prevent them from becoming more deeply involved. In addition, “gang involvement” can be viewed as a continuum, ranging from “hanging out where gang members hang out” to “wearing gang colors” to “participating in gang violence.” Programs have to be careful to define the concrete outcomes they are looking for and what the outcome indicators are.

2. Some outcomes are more difficult to measure than others.

Example: Changes in “self-esteem”—an outcome often considered to be positively affected by mentoring—are extremely difficult to measure. Several evaluations that have demonstrated positive impacts of mentoring (including Public/Private Ventures’ evaluation of Big Brothers Big Sisters) were not able to demonstrate that mentored youth’s self-esteem increased more than it would have if they had not been in a mentoring relationship. These evaluations have used respected tools, including Harter’s Global Self-Worth Measure and the Rosenberg Self-Esteem Scale. Thus, while it seems logical that being in a positive mentoring relationship would ultimately increase a youth’s self-esteem, it is difficult to “capture” this increase.



Why Programs Should Measure the Strength of Mentor-Youth Relationships

Handout #4, “Measuring the Quality of Mentor-Youth Relationships” is a self-contained, 14-page guide that includes a survey and scoring information. Activity #4 in the training session is intended to prepare program operators and staff to administer the survey to youth in their programs, to score the survey and perform basic analysis of the scores, and to understand how to use their findings to improve their programs.

During your presentation of the material in the handout, be sure to include the following points:

1. Research—including Public/Private Ventures’ evaluation of Big Brothers Big Sisters—has demonstrated that strong mentoring relationships lead to other outcomes for youth. In fact, these strong relationships could be seen as an early outcome that leads to positive outcomes, such as improved school behavior.
2. Thus, in measuring outcomes, it makes sense for programs, first of all, to be measuring the quality of the mentor-youth relationships. If those relationships are not strong, programs are less likely to be achieving other positive outcomes.
3. It is relatively easy and inexpensive for programs to measure the quality of the mentoring relationships. The handout includes a survey that most programs will be able to use.
4. When a program first uses this survey, its findings can serve as benchmarks showing where it is now in terms of the quality of mentor-youth relationships. The program can then set targets for improving on these benchmarks the next time it administers the survey.
5. Questions on the survey were originally used in Public/Private Ventures’ evaluation of Big Brothers Big Sisters (BBBS), a program with proven, significant positive outcomes for youth. The handout includes information on BBBS scores on the survey. Programs can use the BBBS scores as a gauge against which to compare their own findings.

6. Evaluations of mentoring programs have demonstrated that programs that have a strong infrastructure are more likely to create strong and long-lasting mentoring relationships. This infrastructure includes good practices for screening and matching mentors, training them, and providing ongoing support and supervision. Thus, if programs find, after using this survey, that their mentoring relationships are less strong than is desirable, they should examine and improve their program infrastructure. You can note that other JUMP training sessions—including “Screening Mentors,” “Making and Supporting the Match,” “JUMPstarting Your Mentors,” and “Connecting and Communicating”—focus specifically on strengthening these aspects of program infrastructure.



Reading Selection: Outcome Measurement: Showing Results In the Nonprofit Sector

Like their counterparts in the public sector, managers of nonprofit agencies are aggressively measuring the outcomes of their efforts. This article describes current activities in five important areas, offers 30 useful lessons learned, and discusses seven key challenges still to be overcome. These insights from the nonprofit sector are equally useful to persons working in other settings.

By Margaret C. Plantz and Martha Taylor Greenway
United Way of America
Alexandria, Virginia
and
Michael Hendricks
Independent Consultant
Portland, Oregon

This article summarizes the history of performance measurement in the nonprofit health and human services sector and defines key concepts in outcome measurement. Next, it reports on activities in five key areas and describes 30 lessons the field has learned from those who have led the way. Finally it identifies seven pressing challenges that lie ahead.

Background: Performance Measurement in the Nonprofit Sector

There are approximately 495,000 tax-exempt organizations in the United States (excluding churches) that may receive tax-deductible contributions. Of those with incomes above \$25,000, roughly 55,000 are classified as human service organizations and another 28,000 are health related (Hodgkinson and Weitzman, 1996). In this article, the term “nonprofits” refers to these health and human service organizations. The types of health and human services that nonprofit organizations provide include housing and residential care, youth development, recreation, services to children and families, employment assistance, crime and delinquency prevention, food and nutrition, and substance abuse and addiction treatment (Hodgkinson and Weitzman, 1996).

Nonprofits may receive tax-deductible contributions from individuals, companies and foundations. In addition, they often receive both fees for service and government funds. The latter generally are a mix of federal and state dollars that flow from a federally legislated program. In 1992, for example, government grants accounted for 24 percent of revenues to human service organizations (Hodgkinson and others, 1996).

The nonprofit sector has been measuring certain aspects of performance for 25 years or more. During that period, the scope of performance measurement has expanded to address such issues as:

- **Financial accountability.** The first focus of nonprofit performance measurement was documenting how funds are spent. Early guidelines included *Standards of Accounting and Financial Reporting for Voluntary Health and Welfare Organizations* (National Health Council and the National Assembly for Social Policy and Development, 1964) and *Accounting and Financial Reporting* (United Way of America, 1974).
- **Program products, or outputs.** Shortly after starting to measure how funds were spent, agencies began to measure what the funds generated. Generally, output measures are measures of volume in two categories: products delivered (e.g., numbers of counseling sessions held, brochures distributed, days of care provided), and people served. Twenty years ago, for example, United Way of America's program classification system defined 587 human service categories and suggested product measures for each (United Way of America, 1976).
- **Adherence to standards of quality in service delivery.** Concerns with service delivery issues such as staff qualifications, staff-to-client ratios, specific service delivery practices, record keeping, confidentiality protections and condition of facilities led to the formation of accreditation and certification groups. The Council on Accreditation of Services for Families and Children, formed in 1977, was one of the first independent accrediting bodies.
- **Participant-related measures.** In the 1980s, funders began seeking assurance that agencies provided services to those most in need. This prompted collection and reporting of client data, including demographic characteristics (such as age, income, race/ethnicity, gender, marital status, area of residence) and information about the client's problem or status prior to service.

- **Key performance indicators.** Several public accounting firms created performance measures for nonprofits. These indicators were largely ratios among various categories of inputs, services, outputs, and total costs. Peat Marwick produced a seminal resource in this area (Elkin and Molitor, 1984).
- **Client satisfaction.** Later in the 1980s, accrediting bodies began requiring services to measure participant satisfaction as an additional element of quality assurance. Elements of satisfaction included physical and cultural accessibility, timeliness, courteousness, physical condition of facilities, and overall satisfaction.

By 1990, the nonprofit sector was commonly measuring all of these aspects of performance except client satisfaction (Taylor and Sumariwalla, 1993) and these measures yielded critical information about the services nonprofits were providing. Increasingly, however, there has been recognition that, while such measures show how much effort has been generated for how many individuals, they reveal nothing about whether this effort has made any difference—whether anyone is better off as a result of the service. Outcome measurement responds to this gap.

The measurement of a program's outcomes—the benefits or results it has for its customers, clients, or participants—can and will have a tremendous impact on nonprofit health and human service organizations. Outcome measurement shifts the focus from activities to results, from how a program operates to the good it accomplishes. Information on the extent to which program participants are having the intended outcomes is powerful and useful feedback. Managers of nonprofit organizations that already have implemented outcome measurement report that:

- A clear definition of the program's intended outcomes, in itself, provides focus for the program's work.
- Understanding their current level of outcome achievement provides a barometer to assess progress and direct future activities.
- Outcome measurement provides invaluable information to improve programs and then see if the improvements make the intended difference.
- Outcome information is a powerful motivator of staff, who now can observe the progress they are making with participants in a consistent, tangible manner.

- It becomes a powerful recruitment tool for volunteers who have many other choices for how they spend their time.
- It helps position the agency in the community as a successful organization, which in turn leads to increased promotion and financial support (Hatry, van Houten, Plantz, and Greenway, 1996).

Introducing Outcome Measurement

The recently published United Way of America manual, *Measuring Program Outcomes: A Practical Approach* (Hatry, van Houten, Plantz, and Greenway, 1996), defines program outcomes as “benefits or changes for participants during or after their involvement with a program” (p. 2). Merely attending the program does not represent an outcome. Outcomes are something that the program participant is, has, or does in response to the service provided. This is distinct from program outputs, which are the number of units of service delivered or the number of people served. Said another way, outputs are about the *program*, while outcomes are about the *participants*.

Outcomes usually are benefits or changes in participants’ knowledge, attitudes, values, skills, behavior, condition, or status. Most often, an outcome represents a change for the better, although the outcome for some programs is that participants get worse more slowly than they would have otherwise. Examples of outcomes are that participants:

- Know the daily nutritional requirements for a pregnant woman (knowledge)
- Recognize that school achievement is necessary to future success (attitude)
- Believe that cheating on a test is wrong (value)
- Are able to read at the sixth-grade level (skill)
- Use verbal rather than physical means to resolve conflict (behavior)
- Have improved health (condition)
- Reside in a permanent, independent setting (status)

In many cases there is not just one desired outcome for participants, but a series of outcomes, with one outcome contributing to another. This hierarchy of logically related changes or benefits comprises a series of “if-then” relationships. For example, if a program provides prenatal counseling to pregnant teens, *then* the teens have increased knowledge of good prenatal care. If the teens have increased knowledge of good prenatal care, *then* this leads to changed behavior: the teens eat the proper foods, take a prenatal vitamin each day, and avoid cigarettes, alcohol and other drugs. If the teens follow these practices, *then* the result is that the teens deliver healthy newborns.

This hierarchy of program outcomes describes the logic or theory of how the program brings about benefits for participants. Figure 1 depicts the conceptual chain of influences, with program inputs (resources) leading to activities, which lead to outputs, which lead to a series of outcomes. The figure shows three levels of outcomes—initial, intermediate, and longer-term—although the number and terminology are not as important as the concept of a logically linked series of changes.

Figure 1 - Inputs Through Outcomes: The Conceptual Chain

Inputs → Activities → Outputs → Initial Outcomes → Intermediate Outcomes → Longer-Term Outcomes

The farther beyond a program’s outputs one goes on this “if-then” chain, the less influence the program has on the achievement of the outcome and the more likely it is that other forces will intervene. In the above example, the program can directly influence the initial outcome of the pregnant teens’ knowledge of good prenatal practices. In contrast, the teens’ general health when they became pregnant and their involvement with drugs before coming to the program—neither of which the program can control—may have as much influence as the program on the longer-term outcome of delivering healthy infants.

A recurring, and vexing, issue in outcome measurement is deciding how far out on the outcome chain a program should go in selecting its longest-term outcome. This decision requires a balance between two needs:

- *The longest-term outcome must be far enough out on the “if-then” chain to capture meaningful change for participants and reflect the*

full extent of the program's benefit for them. For the program above, participants knowing what constitutes good prenatal care is an important link in the “if-then” chain, but it is not meaningful as an end in itself and is not all the program aspires to achieve for participants. The program needs to go beyond that initial outcome in measuring its benefits.

- *On the other hand, the longest-term outcome should not be so far out on the “if-then” chain that the program's influence is washed out by other factors.* It must be reasonable to believe that the program can influence the longest-term outcome in a non-trivial way, even though it cannot control it. In the program described above, contact with the teen mothers ends when their infants are born. Extending the program logic chain from the outcome of healthy births to an even longer-term outcome of “the children are developmentally on target at age two” is not sound in view of all the other factors that will influence mothers and babies in the intervening two years.

Recent Developments in the Nonprofit Sector

The past few years have seen an explosion of outcome measurement activity in the nonprofit sector. Several national organizations have set out both to measure program outcomes and to provide resources to help local affiliate agencies measure their own outcomes. The rise of the managed-care industry and the expansion of accreditation and certification criteria to include outcome measurement are lending urgency to some of these efforts. The following examples illustrate five prominent areas of activity.

National organizations are conducting research on outcomes.

Several national nonprofit organizations are supporting studies of outcomes of their various programs. These studies give local agencies a head start on identifying appropriate outcomes to track and may provide measurement methods and tools as well.

Some of the national studies involve experimental or quasi-experimental designs that evaluate the strength of the association between program activities and shorter- and longer-term participant outcomes. Such studies, too costly and time-consuming to be feasible for most local programs, generally are financed by corporate or private foundations. Positive findings provide evidence of a link between the initial and intermediate outcomes that local programs are more likely to

be able to measure and the longer-term outcomes that generally are of more interest to funders but are beyond most local programs' ability to track.

Examples of national nonprofit organizations' studies include the following:

- Big Brothers Big Sisters of America evaluated an elementary school-based intergenerational linkages program. The study found an increase in students' grades across all school subjects and increases in social and emotional growth noted by the parent or guardian, the volunteer, and the teacher (Peterson, 1994).
- Big Brothers Big Sisters' mentoring program was the subject of an evaluation by Public/Private Ventures, a private research and evaluation group. The study found that youth with mentors were less likely to start using alcohol or other drugs, were less likely to hit someone, improved their school attendance and performance, and had improved peer and family relationships (Tierney, Grossman, and Resch, 1995).
- The Child Welfare League of America is beginning to document what types of settings and services are related to what outcomes for what kinds of children and youth. The study will track both shorter-term outcomes, such as placements into less-intensive settings, family reunification, educational achievement, and behavioral problems; and longer-term outcomes, such as employment, contacts with juvenile and criminal justice systems, and use of alcohol and other drugs (Child Welfare League of America, 1995).
- Girl Scouts of the U.S.A. commissioned Louis Harris and Associates to study outcomes of Girl Scouting. Girl Scouts had better grades than a comparison group, were more actively involved in extracurricular activities and were less likely to say they would cheat on a test, engage in sexual activity, and drink alcohol (Brown and Conn, 1990).
- Girls Incorporated examined the effects of four age-targeted adolescent pregnancy prevention programs. The study found that the programs reduced the onset of intercourse, increased use of contraception for girls who were sexually active, and decreased pregnancy (Nicholson, Postrada, and Weiss, 1991).

- Girls Incorporated also evaluated its substance abuse prevention program. Results showed that program participants were more likely to leave situations where peers were using harmful substances and had less favorable attitudes toward drinking alcohol. The program reduced incidence of drinking among participants and delayed the onset of drinking among participants who had not previously drunk alcohol (Jacobs, Nicholson, Plotch, and Weiss, 1993).

National organizations are developing resources for local agencies. A number of national organizations are developing outcome measurement tools to help local program managers monitor the extent to which program participants achieve the changes the program intends. In the ideal case, the outcomes, indicators, measurement approaches, and other materials derive from experimental research linking an intervention to specific outcomes. But in all cases, nationally developed resources save local programs a significant amount of time and effort and often bring to bear a level of expertise most local programs cannot access independently. Recently published resources include the following:

- Boy Scouts of America has produced *Scouting's Positive Impact on the Community* (1996). Developed in response to local United Ways' increasing requests for funded agencies to document program outcomes, the publication gives guidance on presenting the Scouting program within an outcome measurement framework.
- Girls Incorporated has published *Assess for Success* (Frederick and Nicholson, 1991). The manual provides an overview of outcome measurement concepts, as well as instructions and data-collection instruments for observing girls' behavior and surveying girls and parents about girls' outcomes. The instruments are program-specific and age-targeted.
- Goodwill Industries International's report, *Program Evaluation: Guidelines for Development and Implementation* (1994), describes a program evaluation system that incorporates outcome measurement. It includes guidelines for developing outcomes and includes examples of specific outcome indicators.
- United Way of America has published *Measuring Program Outcomes: A Practical Approach* (Hatry, van Houten, Plantz, and Greenway, 1996), a manual for any nonprofit human service

agency seeking to develop an outcome measurement system. It offers a step-by-step process for identifying outcomes, indicators, and data-collection methods, and for reporting and using data. Quotes and vignettes from more than 30 agencies that have implemented outcome measurement offer insights from the field. The manual does not give program-specific outcome indicators or data-collection methods.

Additional resources are in development. For example:

- The American Red Cross is convening representatives of local affiliates to develop program “logic models” with suggested outcomes and indicators for core national programs such as disaster relief.
- Big Brothers Big Sisters of America is developing outcome measurement methods and tools for their affiliates based on the study by Public/Private Ventures described earlier.
- The Girl Scouts of the U.S.A. has engaged a national evaluator who, through focus groups with participants and an extensive literature review, has identified measurable outcomes for Girl Scouting programs and has developed survey instruments and data analysis software for local councils across the country to use.
- The YMCA of the USA is taking a similar approach, working with the Search Institute to develop ten survey modules of aspects of positive youth development.

Managed-care companies are stressing service results. An emerging force in the shift to a focus on outcomes is the growth of managed care, with its emphasis on results-oriented service delivery systems. Increasingly, managed-care companies are requiring nonprofit organizations that seek to be certified for payment to measure participant outcomes as part of the certification process. Affiliates of Catholic Charities USA and Family Service America, for example, are facing this issue in the counseling and home health care areas.

In some cases, managed-care companies are essentially dictating outcomes for programs, sometimes based on little more than intuition about what a desired outcome may be. At this point, there are no standard, managed-care outcomes for human services.

The emergence of managed care as a driver of outcome measurement is a major development that could have far-reaching effects on outcome measurement in the nonprofit sector. As yet, however, the implications of this development are receiving little discussion outside of the program areas directly affected.

Accrediting bodies are considering outcome-related criteria. In part as a response to the interests of managed-care companies, certifying and accrediting bodies increasingly are including standards for outcome measurement in their review criteria. For example:

- The Accreditation Council on Services for People with Disabilities has published *The Outcome Based Performance Measures: A Procedures Manual* (1995). The manual lists specific outcomes for people with disabilities, suggests ways agencies seeking accreditation can assess these outcomes, and describes the independent quality-review process. The Accreditation Council uses the manual to confirm an agency's self-assessment.
- The Council on Accreditation of Services for Families and Children's *Manual for Agency Accreditation* (1992) includes requirements that agencies collect data on program and client service quality and evaluate agency effectiveness, with client outcomes as a necessary component. The manual does not establish what outcomes should be tracked or provide data collection guidance.

Local agencies are operating effective outcome measurement systems. In addition to these national efforts, many local human service agencies—perhaps two to three thousand—have developed sound outcome monitoring approaches using a variety of internal and external resources. While their number is small in comparison to the entire nonprofit human service sector, they are a significant and growing segment of the field.

Lessons Learned in the Nonprofit Sector

All this activity has taught the field much about which approaches are productive and which practices impede success. Offered below are 30 useful lessons in five areas: the value of outcome measurement, agency implementation, the role of funders, effective uses of outcome findings in resource allocation and limitations of outcome measurement.

Lessons About the Value of Outcome Measurement

1. Outcome measurement benefits agencies in multiple ways—ways that counting outputs cannot. It helps them, for example, provide feedback and direction to staff, focus board members on policy and programmatic issues, identify training and technical assistance needs, pinpoint service units and/or participant groups that need attention, compare alternate service delivery strategies, identify partners for collaborations, allocate resources, recruit volunteers, attract customers, set targets for future performance, track program effectiveness over time, increase funding, and enhance their public image.
2. Most programs benefit from simply discussing their intended outcomes. Staff often have varying views of what the program is trying to achieve, and getting everyone focused in the same direction can increase service effectiveness before data collection even begins.
3. The most important reason for implementing outcome measurement is that it helps programs improve services. It also can increase accountability, guide managers in allocating resources, and help funders make better funding decisions, but its value in enhancing service effectiveness should be seen as primary.

Lessons About Effective Implementation by Agencies

4. Outcome measurement is doable. Its nonexperimental design and basic data-analysis requirements make it manageable for even small, grassroots programs.
5. Commitment at the top is essential. Otherwise, the task gets overcome by the other demands of program operation. Before they will commit to outcome measurement, however, agency directors and board presidents must see its value for their agencies.
6. Programs must identify their own outcomes, outcome indicators, and data-collection procedures that are relevant and useful to their own efforts. Outcomes and indicators imposed by outsiders are unlikely to meet these criteria.
7. Creating a written logic model of program inputs, activities, outputs, and outcomes is a helpful way to think through the changes participants experience during and after the program and to check the logic of the “if-then” influences the program intends to set in motion.

8. Agencies should tap many perspectives when identifying program outcomes. Program volunteers, current and past participants (and perhaps family members), persons such as teachers and employers, and other agencies can point out important outcomes that do not occur to staff.
9. Outcome measurement does not always require new data-collection efforts. Agencies often already compile data that reflect on outcomes.
10. Data collection and analysis may pose technical challenges that agencies do not have the in-house capacity to meet. The first time around, guidance on collection and analysis methods from a technical expert will often save time, offer reassurance, and improve results.
11. A trial run of the outcome measurement system is essential and will lead to changes in the system. The trial run must last long enough to encompass all key data collection points and must involve at least a representative group of program participants. Everyone should expect that the trial run will identify problems; that is its purpose.
12. Developing a sound outcome measurement system takes time—to plan, to try out, to adjust, and to implement. It easily could take an agency seven months or more of preparation before collecting any data, and it easily could take three to five years or more before the findings from a program's outcome measurement system actually reflect the program's effectiveness. Rushing the development process decreases the likelihood that the findings will be meaningful.
13. One useful approach to creating an outcome measurement system (Hatry, van Houten, Plantz, and Greenway, 1996) identifies eight steps grouped in three developmental phases: initial preparation (getting ready to begin, choosing outcomes to measure, specifying indicators for the outcomes, and preparing to collect data), a trial run (trying out data-collection procedures and data analysis and reporting methods), and implementation (adjusting the outcome measurement system and using the findings).
14. Once implemented, the outcome measurement system must be monitored and improved continuously. Programs change and programs learn. The system must keep up.

Lessons About Useful Roles for Funders

15. Funders of nonprofit programs, including government agencies, national and local foundations, corporate philanthropic programs, and United Ways, will play a key role in the nonprofit sector's move to a focus on outcomes. To be most constructive, funders should view their role as helping each program develop the outcome measurement approach that provides the most useful information for that program. To the extent that funders impose outcomes, measures, or timetables that do not align with agencies' efforts, they impede successful implementation.
16. Funders serve their own best interests by helping agencies develop capacity for outcome measurement. If agencies do not do well at outcome measurement, funders receive meaningless data and/or must acknowledge that they are supporting ineffective programs.
17. Building capacity requires much more than showing agencies how to complete forms. It involves both hands-on, experiential training and ongoing technical assistance. Before engaging trainers or technical assistance providers, funders should ensure that they have applied experience in local nonprofit settings and are supportive of non-experimental outcome monitoring approaches.
18. Local funders can collaborate with each other very effectively to support agency efforts. They can, for example, pool resources to underwrite training and technical assistance. They also can agree on outcome measurement terminology, methodology, and implementation timetables. Common application and reporting forms go even further in clarifying expectations and reducing the burden of paperwork on local agencies.
19. Funders can help agencies by providing an outside perspective on the reasonableness of agencies' outcome measurement plans and working collaboratively to help improve the proposed approach. Funders should accept outcomes, indicators, and measurement methods established by relevant national organizations and accrediting bodies unless they fail to meet essential criteria.
20. As funders add outcome data as a reporting requirement, they should "walk the talk" of a focus on outcomes and drop existing reporting requirements that do not match this focus. If benefits for people are the critical emphasis, then some reports designed to monitor internal processes (for example, quarterly cash-flow

statements, detailed line-item budgets, salary information for specific staff, staffing structures, minutes of board meetings, internal policy and procedure manuals) should be eliminated. This action also helps offset the added burden for agencies of collecting and reporting outcome data—a real benefit to agencies at a time when resources for the nonprofit sector are shrinking.

Lessons About Using Outcome Findings in Resource Allocation

21. Agency policymakers and funders who want outcomes to guide funding decisions need first to recognize the potential for harm. Done badly, linking outcomes to funding can shift resources from service delivery to measurement with no offsetting benefit to programs, penalize prevention and development programs and others with harder-to-measure outcomes, promote “creaming” (selecting participants who are more likely to succeed), inhibit innovation, punish risk-taking, and discourage interprogram cooperation.
22. Effective funding processes concentrate initially on assuring that outcome measurement systems are sound. Wise fund allocators focus first on whether a program is making a good-faith effort at outcome measurement, not on the outcome findings themselves.
23. Once programs are sure they have specified the appropriate outcomes, fund allocators can consider whether those outcomes align with funding priorities. The issue of alignment is separate from that of program effectiveness and may be considered before outcome data are available. A danger in decisions regarding alignment, however, is that remediation programs may be favored over preventive or developmental programs because, at first glance, outcomes of the latter do not appear to be linked to funding priorities. For example, if preventing teen pregnancy is a priority, the alignment of a program to prevent teen mothers from having a repeat pregnancy is more readily apparent than the alignment of a program that seeks to engage 12- to 15-year-olds in after-school programs to develop competencies and assertiveness skills. Yet over time, the second program may do more to address the priority.
24. Requiring program managers to set outcome targets before they have at least a year of baseline outcome data is counterproductive. Programs with no experience in outcome measurement generally have no basis for setting an appropriate target, and their targets will likely be little more than guesses. If actual performance falls

below uninformed targets, there is a potential for unwarranted negative consequences.

25. Fund allocators do not yet have enough experience with outcome measurement to judge whether a particular level of achieved performance is good, bad, or in between. Arbitrary criterion levels are likely to be unrealistic and therefore not helpful.
26. Comparing seemingly similar programs to reward those with the “best” outcomes is tempting but misguided. Even “similar” programs have meaningful differences in mission, target audience, geographic location, staffing, service methodology, funding level, and myriad other descriptors that must be considered in assessing effectiveness. Fund allocators can use data from other programs to ask questions and probe why the outcome levels are different, but not to determine which program is better.
27. In judging outcome findings, the best comparison for a program is itself. Is the program improving? Is it learning from earlier outcome findings, making adjustments, and having better results?
28. Despite the hope of many fund allocators, outcome findings will not make the allocation decision easier. Decisions about where to direct resources will remain complex and value-based, and funding decisions will always need to consider more than outcome performance. Factors such as target population, service delivery strategy, number served, outcomes intended, and program cost will continue to be important in deciding where to invest limited resources.

Lessons About Limitations of Outcome Measurement

29. There are many things outcome measurement does not do. It does not eliminate the need to monitor resources, activities, and outputs; tell a program whether it is measuring the right outcomes; explain why a program achieved a particular level of outcome; prove that the program caused the observed outcomes; by itself, show what to do to improve the outcome; or answer the judgment question of whether this is an outcome in which resources should be invested.
30. Measuring and improving program-level outcomes does not, by itself, improve community-level outcomes. Except in rare instances, an individual program does not serve enough individuals to affect community-wide statistics, regardless of how successful the program is. In addition, community-level conditions are the

result of a constellation of influences (for example, economic conditions, environmental factors, demographic trends, public- and private-sector policies, cultural norms and expectations) that are far beyond the scope of influence of individual human service programs.

Challenges for the Future

As the previous section shows, the nonprofit sector has learned many important lessons about outcome measurement. One more lesson is: There still is a lot to learn. Perhaps evaluators can help overcome some of these seven key challenges.

- **Measuring harder-to-measure outcomes.** Some programs face special challenges in measuring outcomes. These include programs whose participants are anonymous, those that provide very short-term assistance such as emergency food or shelter, and those involved in prevention, development, public education, advocacy and providing support to other agencies. These programs need creative ideas for using existing records, third-party reports, trained observers, research-based milestones, secondary data, and other data sources and collection strategies. They also need measurable-yet-meaningful approximations of their outcomes they can track while feasible measurement methods are developed.
- **Increasing useful input from research studies.** Wider, more application-oriented communication of the results of experimental and quasi-experimental outcome studies, synthesized to their most actionable elements, would serve several purposes. It would provide research-based links between the initial or intermediate outcomes that programs are able to track and the longer-term outcomes they cannot track, but that are of interest to funders and policymakers. It also would help identify effective practices in human services, demonstrate alternative measurement approaches and tools, and add to the knowledge base regarding appropriate performance targets and benchmarks.
- **Sharing useful outcomes and indicators and successful tools and methods.** Currently, most nonprofit agencies are working on outcome measurement in relative isolation, but it makes little sense for every program to be starting from scratch. Dictating common outcomes for programs dealing with similar issues is a counter-productive approach. However, establishing systems for sharing

information about successful efforts and the context in which they were applied will save the field much time and expense, offering particular benefit for agencies not supported by a national group. Such systems also would contribute to more rapid advancement in the state of the art.

- **Developing computer capacity to store, analyze, report, and track outcome data.** Most agencies and funders have computers, typically used for word processing, but few have the software and/or skills needed to manage outcome data usefully. The field is in immediate need of easy-to-use models built on common, off-the-shelf software; guidance concerning key variables, database structure, and data manipulation requirements; and analysis and reporting examples specifically geared to outcome data management.
- **Guiding reviewers of outcome measurement plans in judging the appropriateness of proposed outcomes, indicators, and measures.** Agency board members, funders' staff, and volunteers and other individuals will be reviewing outcome measurement plans. Often, reviewers have neither hands-on experience nor expertise in the programs whose plans they are reviewing. What criteria do they use to judge whether a proposed plan is sound? On what basis do they ask questions or offer suggestions? Resources to help reviewers offer input that is helpful, rather than distracting, would be of great value.
- **Establishing reasonable ranges for performance targets or benchmarks and identifying appropriate adjustments for different program and participant characteristics.** For most program areas, there is insufficient information about what constitutes "good" performance. In a program working to move homeless families toward self-sufficiency, for example, suppose 20 percent of participating families are in independent, self-supported living situations within a year. Is this an abysmal failure or a tremendous success? A comparative figure would help program managers, as well as funders, evaluate the meaning of the finding. However, the adoption of targets must be done with great care. The same program in a different setting or serving a different population can reasonably have very different results. Identifying key variables that can be expected to relate to different levels of performance is an important safeguard against broad-brush application of criterion levels.

- **Creating methodologies for aligning program-level outcomes with broader efforts to create community-level change.** Measuring program outcomes has value as an end in itself. However, program outcomes also can contribute to, and should be aligned with, broader efforts to effect community-level change. Such change requires a spectrum of activities, including advocating with key players, establishing multi-partner collaborations, supporting volunteer initiatives, influencing public- and private-sector policies, and leveraging funders' money. Effective theories of change (Connell and Kubisch, in preparation) can tie these elements and the outcomes of individual programs into a coordinated whole with the scope needed to affect community issues.

Conclusion

Nonprofit agencies often ask if outcome measurement is just a fad that, if ignored long enough, will go away. The consensus of the field is a resounding “No.” Like earlier aspects of performance measurement that built upon each other to strengthen the management and delivery of nonprofit services, the careful measurement of outcomes is another essential building block that informs us in new and necessary ways about how to improve programs. It is not the last brick to be added. Nor has it developed as far as it can and will. However, outcome measurement is here to stay, and the nonprofit sector and the individuals it serves will be better for it.

References

- Accreditation Council on Services for People with Disabilities. *The Outcome Based Performance Measures: A Procedures Manual*. Towson, MD: The Accreditation Council on Services for People with Disabilities, 1995.
- Boy Scouts of America. *Scouting's Positive Impact on the Community: A Resource for Measuring Program Outcomes*. Irving, TX: Boy Scouts of America, 1996.
- Brown, S., and Conn, M. *Girl Scouts: Who We Are, What We Think*. New York: Girl Scouts of the U.S.A., 1990.
- Child Welfare League of America. *The Odyssey Project: A Descriptive and Prospective Study of Children in Residential Treatment, Group Homes, and Therapeutic Foster Care*. Washington, D.C.: Child Welfare League of America, 1995.

Connell, J. P., and Kubisch, A. C. "Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects, and Problems." In *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Context, Vol. II*. Washington, D.C.: The Aspen Institute, in preparation.

Council on Accreditation of Services for Families and Children. *Manual for Agency Accreditation*. New York: Council on Accreditation of Services for Families and Children, 1992.

Elkin, R., and Molitor, M. *Management Indicators in Nonprofit Organizations: Guidelines to Selection and Implementation*. Baltimore: University of Maryland, School of Social Work and Community Planning, 1984.

Frederick, J., and Nicholson, H. *Assess for Success*. Indianapolis: Girls Incorporated, 1991.

Goodwill Industries International. *Program Evaluation: Guidelines for Development and Implementation*. Bethesda, MD: Goodwill Industries International, 1994.

Hatry, H., van Houten, T., Plantz, M. C., and Greenway, M. T. *Measuring Program Outcomes: A Practical Approach*. Alexandria, VA: United Way of America, 1996.

Hodgkinson, V., and Weitzman, M. S., *Nonprofit Almanac 1996-1997: Dimensions of the Independent Sector*. San Francisco: Jossey-Bass, 1996.

Jacobs, L., Nicholson, H., Plotch, A., and Weiss, F. *It's My Party: Girls Choose to be Substance Free*. Indianapolis: Girls Incorporated, 1993.

National Health Council and National Assembly for Social Policy and Development. *Standards of Accounting and Financial Reporting for Voluntary Health and Welfare Organizations*. New York: National Health Council and National Assembly for Social Policy and Development, 1964.

Nicholson, H., Postrada, L., and Weiss, F. *Truth, Trust and Technology: New Research on Preventing Adolescent Pregnancy*. Indianapolis: Girls Incorporated, 1991.

Peterson, D. *An Evaluation of an Elementary School Based Intergenerational Linkages Program: Mentoring for Academic Enrichment*. Philadelphia: Big Brothers/Big Sisters of America, 1994.

Taylor, M. E., and Sumariwalla, R. D. "Evaluating Nonprofit Effectiveness: Overcoming the Barriers." In D.R. Young, R.M. Hollister, and V.A. Hodgkinson (eds.), *Governing, Leading, and Managing Nonprofit Organizations*. San Francisco: Jossey-Bass Publishers, 1993.

Tierney, J., Grossman, J. and Resch, N. *Making a Difference: An Impact Study of Big Brothers/Big Sisters*. Philadelphia: Public/Private Ventures, 1995.

United Way of America. *Accounting and Financial Reporting: A Guide for United Ways and Not-for-Profit Human Service Organizations*. Alexandria, VA: United Way of America, 1974.

United Way of America. UWASIS H.- *A Taxonomy of Social Goals & Human Service Programs*. Alexandria, VA: United Way of America, 1976.

[Used by permission, United Way of America. This article appeared in *New Directions for Evaluation*, (Number 75, Fall 1997), Kathleen E. Newcomer (ed.), *Using Performance Measurement to Improve Public and Nonprofit Programs*. San Francisco: Jossey-Bass Publishers.]



Outcomes:

**Benefits for participants
during or after their
involvement in a program.**

Outcomes relate to changes in:

Knowledge

Skills

Attitudes

Values

Behavior



In judging outcomes findings, the best comparison for a program is itself.

Programs should use outcomes to measure their progress over time.

Is the program improving?



Benchmarks:

Performance data that are used for comparative purposes.

A program can use its own data as a baseline benchmark against which to measure future performance.

It also can use data from another program as a benchmark.



Handouts

	<i>Page</i>
Session Goals and Basics	1
Agenda	2
Handout #1: Definitions	3
Handout #2: Measuring Outcomes: Developing a Plan	5
Handout #3: Worksheet for Developing an Evaluation Plan	8
Handout #4: Youth Survey: Measuring the Quality of Mentor-Youth Relationships	9
Handout #5: Tracking the Length of Mentor-Youth Relationships	23
Handout #6: How Programs Can Use Outcomes Findings	25
Handout #7: Why Evaluate? Two Case Studies	26
Handout #8: Resources for Measuring Outcomes	28

Measuring

Outcomes



Session Goals

- To know how to define concrete, measurable outcomes for youth in your program
- To understand data-collection strategies
- To know how to measure the quality of the relationships between mentors and youth in your program
- To understand the uses you can make of your outcomes findings

The Basics

1. Outcomes are benefits for program participants during or after their involvement with the program. They involve changes in participants' knowledge, skills, attitudes, values, or behavior.
2. Measuring outcomes does not have to be a complex or expensive process. It should be built into ongoing program operations, and include measuring the quality of the mentor-youth relationships.
3. A program can use outcomes findings to demonstrate to funders and the community that it is a worthwhile investment, and to help itself identify both program strengths and aspects of its operations that need to be improved.
4. In judging outcomes findings, the best comparison for a program is itself. Programs should look at changes in outcomes over time in order to measure their progress.



Agenda

1. What Changes Are You Looking For?

Notes: _____

2. What Outcomes Can You Measure?

Notes: _____

3. Developing a Plan for Measuring Outcomes

Notes: _____

4. Measuring the Quality of Mentoring Relationships

Notes: _____

5. Making Use of Your Outcomes Findings

Notes: _____

6. Now What?

Notes: _____



Definitions

page 1 of 2

Outcomes: Benefits for participants during or after their involvement with a program. Outcomes relate to changes in knowledge, skills, attitudes, values, or behavior. Examples of outcomes include reduced gang involvement, improved school behavior, and more effective responses to conflict. For a particular program, there can be various “levels” of outcomes, with early outcomes leading to longer-term ones. For example, a youth who develops a strong relationship with a mentor may feel more confident about succeeding in school and, thus, may attend school more regularly, which can lead to getting better grades, which can ultimately lead to graduating.

Outcome indicators: The specific items of information that track a program’s success on outcomes. They describe observable, measurable characteristics or changes that represent achievement of an outcome. For example, a program whose desired outcome is that participants improve their school behavior could define “improved school behavior” as better attendance, more homework turned in on time, and fewer trips to the principal’s office.

Benchmarks: Performance data that are used for comparative purposes. A program can use its own data as a baseline benchmark against which to compare future performance. It also can use data from another program as a benchmark. In the latter case, the other program often is chosen because it is exemplary and its data are used as a target to strive for, rather than as a baseline.

Outcome targets: Numerical objectives for a program’s level of achievement on its outcomes. After a program has initially measured outcomes, it can use its findings as benchmarks to set targets for the number and percent of participants expected to achieve desired outcomes in the next reporting period. It also can set targets for the amount of change it expects participants to experience.

Impacts: The extent to which the outcomes are a result of participation in the program. “Outcomes” describe changes. “Impacts” demonstrate how much participating youth changed in these areas beyond whatever change would have occurred naturally. To measure impacts, a program needs to have a control or comparison group—youth who are not in the program but who are comparable to the youth who are. The program needs to collect the same data during the same periods on both groups of youth (for example, school attendance at the beginning and end of a 12-month period). The impact can then be measured by comparing changes in the participating youth to changes in youth in the control or comparison group.

Definitions

page 2 of 2

For example, school attendance by youth with mentors might have *increased* by 3 percent, while school attendance by youth in the control group (similar youth, but without mentors) *decreased* by 7 percent. Without a control group, you would not be able to understand the extent of your program's impact.

(Adapted from *Measuring Program Outcomes: A Practical Approach*. 1996. Used by permission, United Way of America.)



Measuring Outcomes: Developing a Plan

How can a program determine for itself, and demonstrate for others, that it produces positive outcomes for its participants?

In many cases, even small, grassroots programs with very limited resources can do valid and useful outcomes measurements. Programs can follow this process:

1. Examine the content and structure of your program, and identify outcomes the program is most likely to affect.

Then decide which of these outcomes the program wishes to track. The outcomes chosen should be:

- Concrete and measurable
- Based on clearly articulate, *achievable* program goals, such as improved school behavior, improved parental and peer relationships, or more exposure to the world of work

The outcomes should be related to participant behaviors that the program *explicitly* tries to affect—for example, improved conflict-resolution skills or greater knowledge of career options.

2. Identify the kinds of information, or data, you will need to collect.

“Outcome indicators” are the specific items of information that track the extent to which a program is achieving its desired outcomes for children and youth. They describe observable, measurable characteristics or changes.

What will you use as your “outcome indicators”? For example, a program whose desired outcome is that participants improve their school behavior could define “improved school behavior” as: 1) better attendance, 2) more homework turned in on time, and 3) fewer trips to the principal’s office. The program would then collect data in each of these three areas.

3. Decide how you will collect the data.

Will you use questionnaires or surveys? Interviews? A review of records? A combination of several approaches?

For example, to collect school attendance data, you might want to review the school's records. However, if you do not have access to those records, you might choose to collect school attendance data in one of these ways: 1) interviews with the youth; 2) having youth complete a questionnaire; 3) interviews with youth's parents or guardians; 4) having youth show you their report cards.

Programs should also review the kinds of data they are already collecting for other purposes, because you may already be collecting some data you could use for measuring outcomes. JUMP programs, for example, are participating in a national evaluation. In connection with that evaluation, you may be collecting data on, among other items, youth's contacts with law enforcement agencies and youth's grades in school. You might also be able to use those data to help you measure outcomes for youth in your program.

4. *Decide when you will collect the data.*

In most cases, you will want to collect data when youth first enroll in your program. You can then use that data as a baseline against which to measure change.

But when will you collect the follow-up data? How long will it be before you can reasonably expect measurable change in the youth? Obviously, there is no "right" answer to that question—but it is a question to consider as you develop plans to collect data.

Even with something as seemingly straightforward as changes in school attendance, it can be tricky to know when to collect data. Programs might be tempted to collect attendance data for the first quarter of the school year and then again for the fourth quarter of the year in order to see if youth's attendance improved. However, youth generally have better attendance during the first quarter because it is the beginning of the school year and they are not yet bored or frustrated with school. Thus, a program that compares first-quarter attendance with fourth-quarter attendance (after the youth has been matched with a mentor during the school year) is likely to see little or no gains—or even a decrease. Instead, programs should collect attendance data on the fourth quarter of the previous school year to use as their baseline. Then they can collect attendance data for the fourth quarter of the current school year and compare the two in order to measure change.

5. Determine whether you have the resources and experience to do the kind and level of data collection you feel you need in order to measure a particular outcome.

Do you have the in-house experience and capacity to do the data collection or will you need to get help from an outside evaluator?

For example, if you want to collect data on participants' school attendance and you have access to school records, is there available staff time for reviewing those records?

Or you may want to collect data on participants' gang involvement. You might decide to do a baseline survey (when youth are first matched with a mentor) and a follow-up survey (perhaps after they have been in a mentoring relationship for 12 months). Do you have someone on staff who can construct and administer the surveys? Do you have the resources to hire a consultant who can work with you to do this? Or do you have a way to recruit a volunteer—perhaps a graduate student or university faculty member—who can help in constructing the surveys?

6. Decide how you will use your findings.

For example, will you use the information to raise the visibility of your program in the community? To help in recruiting mentors? To identify what kinds of partners might be helpful for improving services? To satisfy funders' requirements or to help in generating new funding? To identify aspects of your programming that need strengthening?



Worksheet for Developing an Evaluation Plan

What outcome are you going to measure?	What "outcome indicators" will you use?	How will you collect the data?	When will you collect the data?	Who will perform the data collection?
	1. 2. 3.	1. 2. 3. <i>Is your program already collecting some of this information for other purposes?</i>	1. 2. 3.	1. 2. 3.



Youth Survey: Measuring the Quality of Mentor-Youth Relationships

page 1 of 14

Several major evaluations have provided convincing evidence that strong mentoring relationships lead to positive outcomes for participants. In fact, a strong mentoring relationship can be seen as an **early outcome** that, in turn, results in **longer-term** outcomes for participants, such as improved school behavior.

The survey and scoring information included in this handout are based on evaluation tools and findings from Public/Private Ventures' (P/PV's) evaluation of Big Brothers Big Sisters (BBBS). The P/PV research examined the effect of BBBS on youth ages 10 to 14 who enrolled in the program at eight study sites between October 1991 and February 1993.

The evaluation of BBBS found that longer-term outcomes for youth in "good" relationships were positive and consistent. Those youth were more likely than similar, non-mentored youth to feel self-confident about doing their school work; less likely to skip school; less likely to start using drugs; less likely to start using alcohol; and had higher grades. (The results of the study are reported in *Making a Difference: An Impact Study of Big Brothers/Big Sisters*. 1995. Joseph P. Tierney, Jean Baldwin Grossman, and Nancy L. Resch. Philadelphia: Public/Private Ventures.)

Thus, measuring the quality of mentor-youth relationships can be a key element in efforts to evaluate a program's outcomes.

This handout provides a complete set of materials for measuring the quality of relationships between mentors and youth in your program. You can use your findings to:

- Determine what your program is achieving now in this area.
- Establish benchmarks so you can measure changes in the quality of your program's mentor-youth relationships over time.
- Gain insight into which of your program elements (such as mentor screening or training) are currently effective and which need to be modified and strengthened.
- Compare your findings about the quality of the mentoring relationships in your program to findings about the quality of the relationships in Big Brothers Big Sisters agencies, programs that have been shown to have significant, positive longer-term outcomes for youth.

Youth Survey: Measuring the Quality of Mentor-Youth Relationships
page 2 of 14

The handout includes three sections:

- 1. The “Youth Survey”**—The survey, which has 19 questions, measures three qualities of mentor-youth relationships: the extent to which the relationship is centered on the youth’s interests; the youth’s emotional engagement; and the extent to which the youth is dissatisfied with the relationship. Use this copy of the survey during the activities in the training session, “Measuring Outcomes.”
- 2. Information on scoring the survey**—This section describes how to calculate scores, and how to find average scores and categorize scores in order to better understand their implications. It includes a sample completed survey.
- 3. A clean copy of the “Youth Survey”**—Programs can use this to make additional copies for administering the survey to their youth.

(Much of the material in this handout is adapted from “Assessing the Effectiveness of Mentoring Programs.” 1999. Jean Baldwin Grossman and Amy Johnson. In *Contemporary Issues in Mentoring*. Edited by Jean Baldwin Grossman. Philadelphia: Public/Private Ventures.)

Turn to the next page to begin the Youth Survey.

Youth Survey: Measuring the Quality of Mentor-Youth Relationships
page 4 of 14

	Not True At All	Not Very True	Sort of True	Very True
4. Sometimes my mentor promises we will do something; then we don't do it.	1	2	3	4
5. My mentor is always interested in what I want to do.	1	2	3	4
6. When I'm with my mentor, I feel excited.	1	2	3	4
7. When my mentor gives me advice, it makes me feel stupid.	1	2	3	4
8. My mentor and I like to do a lot of the same things.	1	2	3	4
9. When I'm with my mentor, I feel sad.	4	3	2	1
10. I feel I can't trust my mentor with secrets—my mentor would tell my parent/guardian.	1	2	3	4
11. My mentor thinks of fun and interesting things to do.	1	2	3	4
12. When I'm with my mentor, I feel important.	1	2	3	4
13. When I'm with my mentor, I feel bored.	4	3	2	1
14. I wish my mentor asked me more about what I think.	1	2	3	4
15. My mentor and I do things I really want to do.	1	2	3	4
16. When I'm with my mentor, I feel mad.	4	3	2	1
17. I wish my mentor knew me better.	1	2	3	4
18. When I'm with my mentor, I feel disappointed.	4	3	2	1
19. When I'm with my mentor, I feel happy.	1	2	3	4

Kids may ask why you are asking about the same thing several times in different ways. Answer: "This is so we can really understand what you think or feel. Like in basketball, if you want to know how well someone shoots, you ask them to shoot several times, not just once."

See the following pages for information on scoring the survey.

Scoring the Youth Survey

This scoring guide has four sections:

Section A describes how to calculate the scores for the survey and describes a way to categorize the scores so you can more easily understand their meaning.

This section also includes the findings from the BBBS evaluation in which these survey questions were originally used. You can use the BBBS findings as an external benchmark against which to gauge your program's findings. Note that the information on BBBS includes two columns: "still-matched" and "ever-matched." If you survey only youth who are currently matched with a mentor, use the "still-matched" BBBS findings as the basis for your comparison. If you survey an entire group of youth who were matched over a period of time (for example, one year), including youth who are no longer in a mentoring relationship, use the "ever-matched" data for comparison.

Section B includes a sample completed survey and the scores calculated from it.

Section C explains how to find the average scores for your program's mentor-youth relationships.

Section D explains how to organize the range of scores for your program's mentor-youth relationships.

A. Calculating and Categorizing Scores

The "Youth Survey" measures three different, but related, qualities of mentor-youth relationships: the extent to which the relationship is youth-centered, the youth's emotional engagement, and the extent to which the youth is dissatisfied with the relationship. When you score this survey, you will calculate a separate score for each of these three qualities. Thus, you will have three different scores for each survey.

The items that measure each of these qualities are deliberately mixed together on the survey. This is because you can get a better understanding of how someone is really feeling when the youth must go anew to a similar question and reconsider her or his response.

1. Youth-Centered Relationship: The research on BBBS demonstrated that youth who feel their mentor takes their preferences and interests into account are more likely to show improvement in their behaviors and attitudes than are youth who feel their mentor is less interested in them.

Youth Survey: Measuring the Quality of Mentor-Youth Relationships
page 6 of 14

This quality is measured through questions 2, 5, 8, 11, 15.

Add the numbers that have been circled for those items. The total will be from 5 to 20. Then divide the total by 5 (the number of items) to get a score that ranges from 1 to 4.

Categories:

4.0 (very youth centered)

3.0 to 3.99

1.0 to 2.99 (not youth centered)

Compare your findings to findings from the evaluation of BBBS:

	Still-Matched	Ever-Matched
4.0 (very youth centered)	50%	40%
3.0 to 3.99	44%	47%
1.0 to 2.99 (not youth centered)	6%	13%

The average BBBS rating was 3.53 for youth who were “ever-matched” and 3.69 for youth who were “still-matched.”

If your ratings are similar to these, it indicates that your mentor-youth relationships are approximately as strong in this area as those in BBBS, a program that has demonstrated positive outcomes for its participants.

- 2. Youth’s Emotional Engagement:** These items measure the degree to which the youth enjoys the relationship and is emotionally engaged in it (for example, whether the youth feels happy, special, mad, or bored). The BBBS research demonstrated that youth who feel better about being around their mentor are more likely to show improvement in their behaviors and attitudes than are youth who feel less positive.

This quality is measured through questions 3, 6, 9, 12, 13, 16, 18, 19.

Add the numbers that have been circled for those items. The total will be from 8 to 32. Then divide the total by 8 (the number of items) to get a score that ranges from 1 to 4.

Categories:

4.0 (highly engaged)

3.0 to 3.9

1.0 to 2.9 (not very engaged)

Youth Survey: Measuring the Quality of Mentor-Youth Relationships
page 7 of 14

Compare your findings to findings from the evaluation of BBBS:

	Still-Matched	Ever-Matched
4.0 (highly engaged)	28%	24%
3.0 to 3.9	60%	57%
1.0 to 2.9 (not very engaged)	12%	19%

The average rating was 3.43 for youth who were “ever-matched” and 3.55 for youth who were “still-matched.”

If your ratings are similar to these, it indicates that your mentor-youth relationships are approximately as strong in this area as those in BBBS, a program that has demonstrated positive outcomes for its participants.

- 3. Youth’s Dissatisfaction:** The BBBS research demonstrated that youth who feel more dissatisfied with their mentor and the relationship are less likely to show improvement in their behaviors and attitudes than are youth with more favorable impressions. Thus, mentoring programs that can create more satisfying relationships are more likely to be effective than are similar mentoring programs that create less satisfying relationships.

This quality is measured through questions 1, 4, 7, 10, 14, 17.

Add the numbers that have been circled for those items. The total will be from 6 to 24. Then divide the total by 6 (the number of items) to get a score that ranges from 1 to 4.

Categories—*note that lower scores are better; they mean the youth was less dissatisfied:*

- 1.0 to 1.49 (highly satisfied)
- 1.5 to 2.49
- 2.5 or higher (highly dissatisfied)

Compare your findings to findings from the evaluation of BBBS:

	Still-Matched	Ever-Matched
1.0 to 1.49 (highly satisfied)	48%	41%
1.5 to 2.49	44%	46%
2.5 or higher (highly dissatisfied)	8%	13%

The average rating was 1.73 for youth who were “ever-matched” and 1.61 for youth who were “still-matched.”

Youth Survey: Measuring the Quality of Mentor-Youth Relationships
page 8 of 14

If your ratings are similar to these, it indicates that your mentor-youth relationships are approximately as strong in this area as those in BBBS, a program that has demonstrated positive outcomes for its participants.

B. Sample Completed Survey and Scores

YC = Youth-Centered Relationship

EE = Youth's Emotional Engagement

YD = Youth's Dissatisfaction

	Not True At All	Not Very True	Sort of True	Very True
YD 1. My mentor makes fun of me in ways I don't like.	①	2	3	4
YC 2. My mentor almost always asks me what I want to do.	1	2	③	4
EE 3. When I'm with my mentor, I feel special.	1	2	3	④
YD 4. Sometimes my mentor promises we will do something; then we don't do it.	1	②	3	4
YC 5. My mentor is always interested in what I want to do.	1	2	③	4
EE 6. When I'm with my mentor, I feel excited.	1	②	3	4
YD 7. When my mentor gives me advice, it makes me feel stupid.	1	②	3	4
YC 8. My mentor and I like to do a lot of the same things.	1	②	3	4
EE 9. When I'm with my mentor, I feel sad.	④	3	2	1
YD 10. I feel I can't trust my mentor with secrets—my mentor would tell my parent/guardian.	①	2	3	4
YC 11. My mentor thinks of fun and interesting things to do.	1	②	3	4
EE 12. When I'm with my mentor, I feel important.	1	2	③	4
EE 13. When I'm with my mentor, I feel bored.	4	3	②	1
YD 14. I wish my mentor asked me more about what I think.	1	2	③	4

Youth Survey: Measuring the Quality of Mentor-Youth Relationships
page 9 of 14

	Not True At All	Not Very True	Sort of True	Very True
YC 15. My mentor and I do things I really want to do.	1	②	3	4
EE 16. When I'm with my mentor, I feel mad.	④	3	2	1
YD 17. I wish my mentor knew me better.	1	2	③	4
EE 18. When I'm with my mentor, I feel disappointed.	4	③	2	1
EE 19. When I'm with my mentor, I feel happy.	1	2	③	4

Scores:

**Youth-Centered
Relationship**

Total for items = 12
Divide by 5
Score = 2.4

**Youth's Emotional
Engagement**

Total for items = 25
Divide by 8
Score = 3.125

**Youth's
Dissatisfaction**

Total for items = 12
Divide by 6
Score = 2

C. Finding the Average Scores for the Mentor-Youth Relationships in Your Program

For **each** of the three qualities that is being measured, total the scores for that quality. Then divide by the number of scores.

Example: You administer the “Youth Survey” to 20 youth who are in mentoring relationships through your program. You then score each of the surveys. The 20 scores for the quality of “Youth-Centered Relationship” are:

3.8	2.4	3.4	3.4	2.8
4	3.2	3	4	1.6
3	3	3.8	2.6	2
4	3.6	3.4	3.2	2.8

The total of these 20 scores is 63.

Divide 63 by 20.

The average rating for “Youth-Centered Relationship” is 3.15.

Youth Survey: Measuring the Quality of Mentor-Youth Relationships
page 10 of 14

If this is the first time you have given this survey to the youth in your program, 3.15 is the benchmark against which you can gauge future program improvement in this area.

You can refer back to scoring information earlier in this handout to compare your findings to an external benchmark: the average score for “Youth-Centered Relationships” in BBBS agencies where this survey was administered to youth.

You can repeat this process to calculate the average scores for each of the other two qualities: Youth’s Emotional Engagement and Youth’s Dissatisfaction.

D. Organizing the Range of Scores for the Mentor-Youth Relationships in Your Program

For **each** of the three qualities that is being measured, count the number of scores that fall into each category. Then divide that number by the total number of scores to calculate the percentage of relationships that falls into each category.

Example: You administer the “Youth Survey” to 20 youth who are in mentoring relationships through your program. You then score each of the surveys. The 20 scores for the quality of “Youth-Centered Relationship” are:

3.8	2.4	3.4	3.4	2.8
4	3.2	3	4	1.6
3	3	3.8	2.6	2
4	3.6	3.4	3.2	2.8

The scores can be organized into three categories:

4.0 (very youth centered)

3.0 to 3.99

1.0 to 2.99 (not youth centered)

There are 3 scores that are 4.0.

There are 11 scores that are from 3.0 to 3.99.

There are 6 scores that are from 1.0 to 2.99.

Youth Survey: Measuring the Quality of Mentor-Youth Relationships
page 11 of 14

Divide each of these numbers by 20 to find the percentage that falls into each category.

<u>Category</u>	<u>Percent of scores in that category</u>
4.0 (very youth centered)	15% (3 divided by 20)
3.0 to 3.99	55% (11 divided by 20)
1.0 to 2.99 (not youth centered)	30% (6 divided by 20)

If this is the first time you have given this survey to the youth in your program, you can use these findings as a benchmark for gauging future program improvement in this area.

You can refer back to scoring information earlier in this handout to compare your findings to an external benchmark: the percentage of scores in each category of “Youth-Centered Relationships” in BBBS agencies where this survey was administered to youth.

You can repeat this process to calculate the percentage of scores in each category of the other two qualities: Youth’s Emotional Engagement and Youth’s Dissatisfaction.

See the following pages for a copy of the Youth Survey that you can use in your program.

Youth Survey

The following two pages contain the survey for measuring the quality of mentor-youth relationships. Feel free to make copies of the survey to use in your program.

Instructions for Administering the Survey

A. It is best to administer the survey to groups of youth rather than individually:

1. Give each youth a copy of the survey.
2. Have a staff member or volunteer read each question aloud. Youth then circle their responses.

When youth complete the survey as part of a group, they are likely to feel more anonymous and, thus, be more honest in their responses.

If you must administer the survey in a one-to-one setting, be sure the adult who is giving the survey is not a caseworker or someone else whom the youth may be trying to please. This could affect the honesty of the responses.

B. While they are taking the survey, youth may ask why you are asking about the same thing several times in different ways. Answer: “This is so we can really understand what you think or feel. Like in basketball, if you want to know how well someone shoots, you ask them to shoot several times, not just once.”

	Not True At All	Not Very True	Sort of True	Very True
6. When I'm with my mentor, I feel excited.	1	2	3	4
7. When my mentor gives me advice, it makes me feel stupid.	1	2	3	4
8. My mentor and I like to do a lot of the same things.	1	2	3	4
9. When I'm with my mentor, I feel sad.	4	3	2	1
10. I feel I can't trust my mentor with secrets—my mentor would tell my parent/guardian.	1	2	3	4
11. My mentor thinks of fun and interesting things to do.	1	2	3	4
12. When I'm with my mentor, I feel important.	1	2	3	4
13. When I'm with my mentor, I feel bored.	4	3	2	1
14. I wish my mentor asked me more about what I think.	1	2	3	4
15. My mentor and I do things I really want to do.	1	2	3	4
16. When I'm with my mentor, I feel mad.	4	3	2	1
17. I wish my mentor knew me better.	1	2	3	4
18. When I'm with my mentor, I feel disappointed.	4	3	2	1
19. When I'm with my mentor, I feel happy.	1	2	3	4



Tracking the Length of Mentor-Youth Relationships

page 1 of 2

The research on Big Brothers Big Sisters (BBBS), described in the previous handout, demonstrated that youth who are matched longer are more likely to show improvement in their behaviors and attitudes than are youth whose matches are shorter. Relationships of short duration probably do not allow adequate time to develop the mutual trust and respect necessary for real growth to occur on the part of the youth. Thus, systematically tracking the length of mentor-youth relationships can help programs know if they are more likely to achieve their desired outcomes for participants.

Collecting Data

Program operators should be able to collect data on the length of mentor-youth relationships (as well as on the frequency of meetings) by examining program records. During their regularly scheduled contacts with mentors, programs should be recording whether relationships are still meeting (and how often the mentor and youth met during the period since the previous check-in).

Gaining Insight from the Data

Systematically measuring the length of mentor-youth relationships will allow you to:

1. Identify what your program is achieving now in this area.
2. Establish benchmarks so you can measure future improvements in the length of your program's mentor-youth relationships.
3. Gain insight into which of your program elements are currently effective and which need to be modified and strengthened. These elements include mentor screening and training; the process of matching mentors and youth; and ongoing program support for mentors.

Each program will want to establish its own benchmark concerning length of relationships, and then use that benchmark to gauge its progress in that area. In addition, the findings from the evaluation of BBBS provide a useful external benchmark. In the study of BBBS, the average length of relationship, 18 months after youth applied to the program, was 11.3 months for youth who were "ever-matched" with a mentor and 12.8 months for youth who were "still-matched" when the data were gathered. (Because the BBBS agencies all had waiting lists, it took several months for youth to be matched after they applied.)

Tracking the Length of Mentor-Youth Relationships
page 2 of 2

The data on length of matches in BBBS were organized into four categories:

Categories	BBBS Levels	
	Still-Matched	Ever-Matched
Less than 3 months	2%	4%
3 to 6 months	8%	15%
More than 6 months to 12 months	29%	36%
More than 12 months	60%	46%

School-based programs, in which the relationship is intended to last approximately 9 months, or the length of a school year, are obviously much less likely to observe matches lasting more than a year. Like all programs, however, they will want to pay attention to the percentage of matches that fail in less than 3 months. These are, in general, the matches that have failed to “jell.” If you have a high percentage of matches in this category, it is a strong indication that key program elements need to be examined and improved.



How Programs Can Use Outcomes Findings

In growing numbers, service providers, governments, other funders, and the public are calling for clearer evidence that the resources they expend actually produce benefits for people. Consumers of services and volunteers who provide services want to know that programs to which they devote their time really make a difference. That is, they want better accountability for the use of resources. One clear and compelling answer to the question, “Why measure outcomes?” is this: *To see if programs really make a difference in the lives of people.* Being able to demonstrate that their efforts are making a difference for people pays important dividends for programs.

Although improved accountability has been a major force behind the move to outcomes measurement, there is an even more important reason: *To help programs improve services.* Outcomes measurement provides a learning loop that feeds information back into programs on how well they are doing. The results show not only where services are being effective for participants, but also where outcomes are not as expected. Thus, it offers findings that programs can use to adapt, improve, and become more effective.

This dividend doesn’t take years to occur. It often starts appearing early in the process of setting up an outcomes measurement system. *Just the process of focusing on outcomes—on why the program is doing what it’s doing and how it thinks participants will be better off—gives program managers and staff a clearer picture of the purpose of their efforts.* Staff often have somewhat different views of what a program is trying to achieve. Getting everyone focused in the same direction can increase service effectiveness before data collection even begins.

Program managers can use outcomes findings for a wide range of purposes, including to:

- Provide direction for staff
- Identify training needs of staff and mentors
- Strengthen less effective program operations and services
- Target effective services for expansion
- Focus board members’ attention on programmatic issues
- Suggest outcomes targets
- Gauge the program’s progress over time
- Recruit talented staff and mentors
- Identify partners for collaboration
- Promote the program to potential participants
- Enhance the program’s public image
- Guide budgets and justify resource allocations
- Support annual and long-range planning
- Retain and increase funding

(Adapted from *Measuring Program Outcomes: A Practical Approach*. 1996. Used by permission, United Way of America.)



Why Evaluate? Two Case Studies

page 1 of 2

Your group will work with *one* of these case studies.

A. Your program provides one-to-one mentoring for youth ages 8 to 12 who all live in a high-crime, low-income community in the southwest section of the city. Mentors meet with youth on weekends, four times a month, for two or three hours. You have done an evaluation to measure the length of the mentoring relationships and have also given a survey to youth to measure the strength of their relationships with their mentors. As a result of this process, you learned two key facts:

- Almost 17 percent of the mentoring relationships stopped meeting within the first three months.
- Of the youth who remain matched with a mentor, 82 percent feel they are in a very strong relationship with their mentor, and the others feel they are in a fairly strong relationship.

You are gratified by the “satisfaction” findings but feel that the percentage of failed matches is much too high.

Your group should discuss the following issues, and be prepared to present its ideas:

1. What do you think may be the cause(s) of the high percentage of failed matches?

Is there additional information you feel you need in order to better understand the cause(s)? How would you collect that information?

2. What changes will you consider making in your program as a result of this finding about the high number of failed matches?

(Examples might include changes in the ways you recruit, screen, match, or train mentors.)

3. How will you make use of the positive findings about the percentage of youth who are highly satisfied with their relationship?

Can you use it to improve program practices? (If so, how?) Recruit mentors? Raise additional funds? Anything else?

B. Your program provides one-to-one mentoring for youth ages 10 to 14 who are at high risk of becoming gang members. Mentors meet with the youth twice a month for three hours to do activities together or just hang out and talk and have fun. You gave a questionnaire to the youth when they first enrolled in your program in order to collect baseline data about their attitude toward gangs and their association with gang members. (For example, you asked whether they hang out in places where gang members hang out, and whether they have any friends who are gang members.) You gave a follow-up questionnaire to the youth 12 months after they began meeting with their mentors. You learned two key facts:

- Almost 52 percent of the youth reported that they now have a more negative attitude toward gangs.
- Only 7 percent of the youth reported less association with gang members. Nearly 10 percent reported increased association; for the others, the degree of association remained about the same.

You feel fairly satisfied with the shifts in youth's attitudes, but you know that means little if the youth are still associating with gang members.

Your group should discuss the following issues, and be prepared to present its ideas.

1. What do you think may be the cause(s) of the high percentage of youth who are continuing to associate with gang members?

Why are your mentors having such a limited effect? Is there additional information you feel you need in order to better understand the cause(s)? How would you collect that information?

2. What changes will you consider making in your program as a result of this finding about gang association?

(Examples might include changes in recruiting, screening, matching, or training mentors; in other services you may provide to youth; in your partnerships with other organizations.)

3. How will you make use of the positive finding about the percentage of youth who report more negative attitudes about gangs?

Can you use it to improve program practices? (If so, how?) Recruit mentors? Form partnerships with other organizations? Raise additional funds? Anything else?



Resources for Measuring Outcomes

Print material

Evaluating Your Program: A Self-Evaluation Workbook for Mentoring Programs (available in early 2000). Office of Juvenile Justice and Delinquency Prevention.

Measuring Program Outcomes: A Practical Approach. 1996. United Way of America. Available through United Way, (800) 772-0008 (ask for item #0989), or www.unitedway.org/outcomes.

Measuring The Difference Volunteers Make. 1997. Melissa Eystad, Ed. Minnesota Department of Human Services. Available through Energize, (800) 395-9800, or www.energizeinc.com.

Some useful Web sites

www.nydic.org

The National Youth Development Information Center—provides information about approaches to evaluation and links to related sites.

www.sagepub.com

Sage Publications, Inc.—publishes books on evaluation methods.

www.unitedway.org/outcomes

The United Way's Outcome Measurement Resource Network—provides information (including articles and excerpts) about measuring outcomes and links to related sites.