



Research Corner: *Gauging the Effectiveness of Youth Mentoring*

Dr. Jean Rhodes, University of Massachusetts – Boston, December 2002

Program Evaluation

Although program evaluation is not as natural or spontaneous as this sort of self-evaluation, most programs engage in some form of monitoring. Sometimes it's as simple as asking mentees and mentors about their experiences; in other cases it involves large-scale, rigorous experimental designs.

Of course, programs are more apt to launch the former, less complicated types of evaluation. And for good reason! Such evaluations do not require the same level of expertise, are far less expensive, place minimal burden on participants and staff and can yield useful findings. For example, simple exit interviews can provide staff with important and immediate feedback about programs.

So, you might ask, why not stop there? A primary reason is that funders need more convincing evidence that programs are actually reaching their objectives. Thus, accountability has increasingly involved moving beyond simple descriptions to demonstrating that specific goals have been met. Knowing your options will help you to make informed decisions about the scope and rigor of your design.

Determining the impact of your program

In the following sections, we will describe several options, ranging from the simple to more complicated.¹ we will first outline a strategy that relies on comparing your program to others to determine whether you are having an effect (i.e., Using Benchmarks). Some of the more intensive evaluation approaches (i.e., Quasi-Experimental Designs), on the other hand, might require the expertise of an outside evaluator (which could include a graduate student or professor from a local university). The cost of an outside evaluation tends to vary with intensity, but programs should budget between \$5,000 and \$10,000 for the expertise.

Using benchmarks

Without actually conducting an evaluation, programs can sometimes draw on findings that have been linked to outcomes in similar programs. In other words, findings from other studies can be used as benchmarks against which to gauge a program's relative effectiveness.² This approach is feasible when the two programs:

- Are targeting similar youth;
- Are reasonably similar in terms of relationship structure and content; and
- Have met or exceeded the evaluated program's quality standards.

So, what can we infer from other evaluations?

DuBois and his colleagues³ conducted a meta-analysis of 55 evaluations of one-to-one youth mentoring programs. The analysis summarized the results of each study and calculated effect sizes (the magnitude of impact) across the entire group of studies. Modest effects of mentoring programs were found across fairly diverse programs, but larger effect sizes emerged when:

- **youth** were somewhat vulnerable but had not yet succumbed to severe problems
- **relationships** were characterized by:
 - more frequent contact
 - emotional closeness
 - lasted six months or longer
- **programs** were characterized by practices that increased relationship quality and longevity, including:
 - intensive training for mentors
 - structured activities for mentors and youth,
 - high expectations for frequency of contact
 - greater support and involvement from parents and
 - monitoring of overall program implementation

Since greater numbers of these practices predicted more positive outcomes for youth in mentoring programs, one-to-one programs that have met these criteria can assume positive outcome.

Additionally, my research with colleagues on **one-to-one programs**⁴ has provided two relatively simple benchmarks against which similar, one-to-one mentoring programs can infer that relationships will have positive effects.

Duration

- Because duration tends to imply strong relationships and programs, duration may be the single best benchmark of program effectiveness. Across several studies, longer durations have been associated with stronger effects.

Relationship quality

- Although duration is probably the single best benchmark, we found that the actual quality of a mentoring relationship can predict positive outcomes above and beyond how long the relationship lasts.
- When responses to our questionnaire indicated a positive, non-problematic relationship, that relationship tended to be longer lasting and have more positive effects.

Although using benchmarks can be enormously useful, it may not provide the level of detail or rigor that programs or funders desire. Moreover, at this stage, benchmarks can only be applied to one-to-one programs. Thus, it is often necessary to conduct a structured evaluation.

The nuts and bolts of evaluating mentoring programs

There are two major types of program evaluation: process evaluations and outcome evaluations.

- **Process evaluations** focus on: whether or not a program is being implemented as intended; how it is being experienced; and whether changes are needed to redress any problems (e.g., difficulties recruiting and retaining mentors, high turnover of staff, and high cost of administering the program), etc.
- **Outcome evaluations** focus on what, if any, effects programs are having. Designs may compare youth who were mentored to those who were not, or examine the differences between mentoring approaches. Information of this sort is essential for self-monitoring and can address key questions about programs and relationships.

Process evaluations of mentoring programs usually involve data from interviews, surveys and/or program records that shed light onto such things as:

- Number of new matches;
- Types of activities;
- Length of matches;
- Frequency and duration of meetings; and
- Perceptions of the relationship.

Information of this sort is essential for self-monitoring and can address key questions about programs and relationships.

Despite the importance of such information, outcome evaluations have become increasingly important for accountability. Therefore, we will focus the remaining sections on the issues and decisions involved in conducting an outcome evaluation.

Outcome evaluations of mentoring programs usually involve data from surveys, interviews, records, etc., including:

- Mentees' reports of their grades, behavior and psychological functioning;
- Teacher reports of mentees' classroom behavior;
- Mentors' reports of their well-being;
- Parent-child relationships; and
- High school graduation rate

Although program evaluation is not as natural or spontaneous as this sort of self-evaluation, most programs engage in some form of monitoring. Sometimes it's as simple as asking mentees and mentors about their experiences; in other cases it involves large-scale, rigorous experimental designs.

Of course, programs are more apt to launch the former, less complicated types of evaluation. And for good reason! Such evaluations do not require the same level of expertise, are far less expensive, place minimal burden on participants and staff and can yield useful findings. For example, simple exit interviews can provide staff with important and immediate feedback about programs.

So, you might ask, why not stop there? A primary reason is that funders need more convincing evidence that programs are actually reaching their objectives. Thus, accountability has increasingly involved moving beyond simple descriptions to demonstrating that specific goals have been met. Knowing your options will help you to make informed decisions about the scope and rigor of your design.

Tips for and Traps in Conducting an Outcome Evaluation

1. Measuring outcomes

- Select outcomes that are most:
 - Logically related to (and influenced by) the program;
 - Meaningful to you ; and
 - Persuasive to your funders.
- Be realistic. You are better off building a record of modest successes, which keep staff and funders motivated, than to focus on "big wins," which may be unrealistic and, when not achieved, demoralizing.
- Collect outcome data after the youth and mentors have been meeting for some time, long enough to expect that some changes in the youth have occurred.

2. Determining sources of data

- Obtain information from multiple sources, including reports from mentees, mentors, parents, caseworkers, etc.
- Select multiple criteria rather than just one outcome (e.g., grades, drug use, attitudes).
- Use standardized questionnaires.
 - Questionnaires that have been scientifically validated are more convincing to funders—and provide a better basis for cross-program comparisons—than surveys you might develop on your own.
 - Such surveys are available for public use through tool kits. The Search Institute has one available (What's Working: Tools for Evaluating Your Mentoring Program) for purchase and The Mentor Center links to several free resources online.
 - The Juvenile Justice Evaluation Center provides links to questionnaires that are likely to be of interest to mentoring programs, including questionnaires about delinquency, drug and alcohol use, ethnic identity, peer relations, psychological measures, etc.

3. Selecting an outcome evaluation.

Outcome evaluations generally fall into two major types: single-group and quasi-experimental designs.

- Single-group designs are the simplest and most common types of evaluation. They are less intrusive and costly and require far less effort to complete than the more ambitious methods that we will describe. An example of a single-group evaluation is when a program administers a questionnaire to participants at the completion of the program (post-test only) or administers a questionnaire before and again after the program (pre-test/post-test).
- Quasi-experimental designs help evaluators identify whether the program actually causes change in program participants, using controls to eliminate possible biases. An example of a quasi-experimental design is when a program administers a pre-test at the beginning of a program and a post-test at the completion of the program to both the target mentoring group and to a matched comparison group that does not receive mentoring.

Selecting an outcome evaluation.

Outcome evaluations generally fall into two major types: single-group and quasi-experimental designs.

- **Single-group designs** are the simplest and most common types of evaluation. They are less intrusive and costly and require far less effort to complete than the more ambitious methods that we will describe. An example of a single-group evaluation is when a program administers a questionnaire to participants at the completion of the program (post-test only) or administers a questionnaire before and again after the program (pre-test/post-test).
- **Quasi-experimental designs** help evaluators identify whether the program actually causes change in program participants, using controls to eliminate possible biases. An example of a quasi-experimental design is when a program administers a pre-test at the beginning of a program and a post-test at the completion of the program to both the target mentoring group and to a matched comparison group that does not receive mentoring.

Single-group designs

Post-test only

- Programs commonly use this design to help determine how mentees are doing at the end of a mentoring program. Post-test evaluations can help determine whether the mentees have achieved certain goals (e.g., not dropping out of school) that match the program's implicit or explicit goals. Such evaluations also help discover whether mentors are satisfied with the program.
- Such an evaluation cannot indicate whether the participant has changed during the program, only how the participant is functioning at the end of the program.

Pre-test/post-test designs

- Programs use this design when they want to determine whether or not mentees actually improved while they were in the program. With this type of evaluation, program staff survey how each participant is doing at the time they enroll in the mentoring program and then after they have completed it (e.g. 6- or 12 months after pre-test). By comparing the results of the pre- and post-test, staff can see whether or not the mentee improved.
- This evaluation cannot indicate whether the program caused the improvement. Many viable, alternative interpretations could explain the change, including:
 - **Maturation** - natural change that occurred simply as a result of the passage of time; and
 - **History** - Events that occurred between the time the participants took the pre-test and post-test could influence the outcome.
- Other problems with interpreting findings from this design include:
 - **Self-selection** - The experimental group might differ from the comparison group in some systematic way. For example, quite possibly only the mentees who benefited most remained in the program long enough to take the post-test.
 - **Regression to the mean** - A mentee who is functioning extremely poorly at the program's onset might improve naturally over time. Mentees might enlist in programs when they are most distressed and then naturally return to a higher level of functioning as time passes.

Even if one cannot identify the cause of a mentee's improvement, a pre-test design can be useful in other ways.

- The evaluator can look at differences within the group. For instance, do youth who receive more frequent or enduring mentoring benefit the most?
- The evaluator can determine whether certain mentee characteristics are related to achieving program goals. For instance, do boys benefit more than girls? Do minorities in same-race matches benefit more than those in cross-race matches?

Quasi-experimental designs

Despite their potential benefits, single-design evaluations seldom help evaluators identify whether the program is the cause of change in program participants. To determine that, one needs to conduct evaluations of slightly greater complexity. Such designs are called quasi-experimental because, if carefully planned, they can control for many biases described above. This type of evaluation comes in a variety of types, such as time-series. We will focus on one common type of program evaluation: one that uses a comparison group.

Comparison group designs

- The most direct way to rule out alternative explanations is to observe additional youth who have not been part of the program but are similar in other ways to the program youth. By including a comparison group, evaluators can isolate the effects of the program from the effects of other plausible interpretations of change.
- A comparison group design also helps put in perspective modest improvements, or even unexpected declines. Take, for example, the landmark evaluation of Big Brothers Big

Sisters of America's mentoring program

(http://www.ppv.org/ppv/publication.asp?search_id=7&publication_id=111§ion_id=0)².

Although youth in both the mentored and the control groups showed increases in academic, social-emotional, behavioral and relationship problems over the period of time being studied, the problems of the mentored group increased at a slower rate.

- One vexing problem with comparison group studies is finding a comparison group that is sufficiently similar to the mentor group. Parents who seek out mentoring programs for their children may devote more attention to their kids at home than do parents of non-mentored youth. Similarly, young people who willingly enlist in a mentoring program may differ (in terms of motivation, compliance, etc.) from those who have not enlisted. The BBBS study got around this potential problem by selecting both groups from the organization's waiting list. Unfortunately, many programs either do not keep a waiting list or are not willing to deliberately withhold their program from eligible and motivated participants.

The Bottom Line

People in the mentoring field tend to believe implicitly that mentoring benefits young people and that, therefore, expensive evaluations are an unnecessary drain on precious resources. Given the choice between spending money on evaluation or extending their services, many mentoring programs will gladly choose the latter. Although understandable, such choices may be shortsighted. We should not necessarily assume that all mentoring programs are equally beneficial - and we still have a lot to learn about the many newer types of mentoring programs (e.g., site-based, group, peer, e-mail). Convincing evaluations are needed to assess the effectiveness of both traditional one-to-one mentoring programs and newer approaches. Such work will play an important role in the expansion of high-quality mentoring programs.

Literature Cited

1 Substantial portions of these sections were adapted from Posavac, E. J. & Carey, Program evaluation: Methods and case studies. Englewood Cliffs, NJ: Prentice-Hall, Inc. and Grossman, J. B. & Johnson, A. (1998). Assessing the effectiveness of mentoring programs. In J. B. Grossman (Ed.). Contemporary issues in mentoring. Philadelphia, PA: Public/Private Ventures.

2 Grossman & Johnson, 1998

3 DuBois, D.L., Holloway, B.E., Valentine, J.C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology*, 30, 157-197.

4 Roffman, J., Reddy, R., & Rhodes, J. (2002). Toward predicting successful youth mentoring relationships: A preliminary screening questionnaire. Submitted for publication.