Measuring Success:
A Guide to Becoming an Evidence–Based Practice

by Jennifer Fratello, Tarika Daftary Kapur, and Alice Chasan
Vera Institute, Center on Youth Justice

ModelsforChange
Systems Reform in Juvenile Justice

Models for Change

Every young person should have the opportunity to grow up with a good education, get a job and participate in his/her community. Creating more fair and effective juvenile justice systems that support learning and growth and promote accountability can ensure that all of our young people grow up to be healthy, productive members of society.

Models for Change: Systems Reform in Juvenile Justice, a MacArthur Foundation initiative, began by working comprehensively on juvenile justice reform in four states, and then by concentrating on issues of mental health, juvenile indigent defense, and racial and ethnic disparities in 16 states. Through collaboration with the Substance Abuse and Mental Health Services Administration (SAMHSA) and the Office of Juvenile Justice and Delinquency Prevention (OJJDP), *Models for Change* expanded its reach and is now working to replicate and disseminate successful models of juvenile justice reform in 31 states.

Director's Note

Increasingly, community-based juvenile justice practitioners are required to validate that they are engaging in evidence-based practice—proof that they are delivering the services their clients require and that the program yields the desired outcomes for youth in their care. While this expectation is a good one, many practitioners understandably feel overwhelmed as they contemplate the task; they often do not know where to begin or how to lay the foundation. In addition, conducting an outcome evaluation is a resource-intensive task that takes an appropriate amount of funding, planning, and data.

Nevertheless, even when evaluations are not immediately feasible, there are a number of preparatory steps a program can take toward examining its outcomes. For example, data collection, monitoring, and reporting are critical for good program planning and pave the way to developing an evaluation capacity. The Vera Institute of Justice, as part of the MacArthur Foundation's Models for Change initiative, has assisted juvenile-justice practitioners in many settings as they build and monitor their programs. On the basis of our experience in the field, and in collaboration with the Institute for Public Health and Justice at the Louisiana State University Health Sciences Center (the lead entity for the Louisiana Models for Change initiative), we crafted this guide to becoming an evidence-based practice. While it was written in response to the questions of juvenile justice practitioners, its systematic approach to collecting information on goals, treatment methods, and outcomes can benefit other social service providers seeking to measure the efficacy of their interventions.

Annie Salsich
Director, Center on Youth Justice
Vera Institute of Justice

Introduction

In recent years, social service providers of all kinds have felt a growing pressure to demonstrate that what they do is an "evidence-based practice." People who provide community-based services for youth involved in the juvenile justice system are not exempt. They, too, are being asked to provide evidence-based assessments of their work with ever-increasing frequency. Contracts and funding often can depend on a program's ability to produce such evaluations. Not surprisingly, this trend has many service providers in the juvenile justice field wondering what, exactly, it means and how they can qualify.

The simple answer is that you have to be able to point to concrete evidence—hard data—showing that the benefits you claim are tangible and replicable. It isn't enough to say, "I know my program works; I've seen it change lives." For example, a program for at-risk youth may exist to prevent crime and put young people on a positive track toward adulthood. It tries to achieve these goals by working with young people to address substance use issues and help them control impulsive behavior. To qualify as an evidence-based practice, it will need proof that kids emerge from the program with reduced levels of substance use and better impulse control, and that these changes are sustainable over the long term. It should also be able to show that as a result of these changes these kids are less likely to commit new crimes as well. This program's challenge—and yours—is to find a way to collect the necessary information so that funders, fellow program professionals, and others have confidence that it produces the results it claims.

The Vera Institute of Justice, funded by the MacArthur Foundation as part of its Models for Change initiative, assembled this guide in response to questions and requests for help from MacArthur juvenile justice grantees. It describes the process that determines whether a program qualifies as evidence-based and explains how programs can prepare to be evaluated.

Although this guide grows out of and is targeted to juvenile justice practitioners, it is generally applicable to programs in other social service fields as well. It also bears noting that the steps described here are neither simple nor easy. Nevertheless, they are worth undertaking—even if a program does not complete the entire process, any progress along the way is likely to be beneficial.

What is an outcome evaluation?

Program professionals can cite many kinds of evidence about the work they do and the results they achieve. Practitioners, administrators, and directors commonly accumulate anecdotal evidence of their program's impact: stories about individual clients, the challenges they face, and how they responded to the interventions. They may even be able to combine these anecdotes to illustrate a larger phenomenon or descriptive outcome—saying for example, that a certain percentage of the kids they treat graduate from high school. Although these types of descriptive data are valuable, they alone don't yield the information necessary to demonstrate that a program is engaged in evidence-based practice. Such evidence can only be derived from an outcome evaluation.

An outcome evaluation is a formal study that helps to answer the basic question "Is this program working?" Its aim is to find evidence of changes in clients' behavior and, if there are changes, show that they result directly from participants' experience in the program (and not from contact with other programs, other factors, or chance). Imagine an organization for truant youth, for example, that seeks to get participants to attend school by providing them with transportation. An outcome evaluation of this program would collect and analyze data about participants' school attendance rates as well as a number of related issues (such as demographics, academic achievement, etc.). Its goal would be to determine whether participants were in fact attending school more as a result of the program and whether access to transportation—rather, than, say, more vigilant monitoring by parents or school personnel—was responsible for the increased attendance. Ideally, the program could show that these effects were sustainable over a longer period of time—at least six months from the point at which youth exited the program.

As noted earlier, outcome evaluations are formal procedures because they follow a specific method known as a research design. There are two dominant types of research design: experimental (also known as a randomized design) and quasi-experimental.

Experimental designs are considered the gold standard of evaluation research designs, because they eliminate any doubt about the outcomes found and their causes. Experimental designs have three basic elements:

- A treatment group and control group—the former receives the intervention being evaluated; the latter does not.
- A random assignment process—to ensure that the people in the treatment and control groups are as similar as possible.
- Comparative information collected through a set of questions posed to all study participants before they start the program, again after they've completed the program, and ideally some period beyond, to measure changes in attitudes and behavior. This is called a pre/post design.

Experimental designs are expensive to create and carry out. They also may raise ethical concerns, since the people in the control group don't get services they may need. For these reasons, some programs opt instead to use the quasi-experimental approach. Quasi-experimental designs seek to mimic an experimental design by using statistical methods to make up for whatever elements of an experimental design might be missing.

Why do I need an outcome evaluation?

In a time of tight budgets, government agencies, foundations, and other sources of funding want to be sure that the programs they support deliver what they promise. Your organization needs to be able to provide this assurance. You may also want to take a critical look at your program for your own purposes, to learn where it is working well and what changes you may need to make in order to optimize your results. Others in your field have an interest in your program's efficacy as well; everyone committed to better outcomes in a given field is looking for effective practices to adopt. A single, well-done, comprehensive outcome evaluation can serve all of these needs and aspirations.

How do I prepare to do an outcome evaluation?

To prepare for an outcome evaluation, you must first know whether your program is doing what it set out to do. Second, you must choose a research design for the evaluation and gather the appropriate information. Third, you must be ready, once the evaluation is complete, to take the next steps.

CONDUCTING AN OUTCOME EVALUATION

Conducting an outcome evaluation takes focused effort and attention. Few service providers have staff with both the time and the expertise needed for this process. Although some organizations have the capacity in-house, many will have to recruit someone else to carry out the study—usually a consultant, independent organization, or university that specializes in what is often called "measurement and evaluation." Hiring outside evaluators has the benefit of ensuring that the result will be objective, because they don't have a stake in the program's success.

However, undertaking an outcome evaluation with outside evaluators requires financial resources. Fortunately, more and more foundations and government agencies are beginning—through competitive processes—to offer grants to support research and evaluation. When conducting a dedicated fundraising effort, you may want to directly reach out to and learn more about local and national foundations as well as county, state, and federal-level government agencies.

1

Is the program true to its original plan?

The founders of any program had an idea of services they wanted to provide. The first step in preparing for your outcome evaluation is to determine if your current operations are consistent with that original idea, which is usually spelled out in a program plan. A program plan is a consensus statement of your goals, objectives, and process. Usually, it will define who the target population is, the problem the program seeks to address, a set of measurable goals directly related to youth behavior, and the theory of change—the precise element of the program that will cause the desired improvement.

The formal term for comparing the current program's structure and practices to your original plan is a process evaluation (see below). A process evaluation typically precedes an outcome evaluation and is a considerable achievement in itself, because it reveals how closely the program lines up with its declared intentions. It should expose whether you have assembled all of the building blocks for running a sound program, are serving the group you originally targeted, are using the techniques you intended to use for treatment and services, and if you have been measuring your performance. The conclusions you reach about your practices and plan tell you the results you will get when you do an outcome evaluation come from providing the appropriate services to your target population. Evaluating if you've followed your original plan is a crucial step toward figuring out if you need to recalibrate your practices.

PROCESS EVALUATION

A process evaluation is an assessment to ensure that a program is operating the way it was intended to. Typically, someone who is not affiliated with the program and can do it without bias will conduct the process evaluation. He or she will interview staff and management, make site visits, watch the program in action, and compare the findings with the following items:

The program plan goals and objectives. Goals are broad, general statements about what the program expects to accomplish and are usually long term. Objectives are precise actions that move the program closer to achieving its goal. For example, an objective of a program whose goal is to reduce juvenile crime may be to help kids to avoid reoffending.

The original target population. For a program to deliver effective services, it must be clear about the people it intends to serve—its target population. Any outcome evaluation of a program that is unclear about its target population risks having misleading results. For instance, say a program is meant to serve youth with substance abuse issues but is instead serving youth with mental health problems. The outcome evaluation results are likely to show intervention failing to meet its goal, as there will probably be no measurable change in substance abuse.

The service delivery model. Your service delivery model—the method for serving the clients—may have come from the best practices in your field or from conversations with experts. You will review it when you evaluate your program, because it allows you to compare the delivery model your program's founders envisioned to the one being used on the ground.

Performance goals and measures. It is important to define what success for a program would look like, so that everyone involved has the same view. Outcomes refer to changes in knowledge, skills, attitudes, behaviors, and functioning of individuals and families as a result of the program; an indicator is information collected to track whether you've achieved an outcome. For example, a program outcome may be improved behavior of young people in the classroom. The indicator of this change would be a measurable improvement in the student's behavior after completing the program. If the program's designers can identify immediate, intermediate, and long-term outcomes and indicators during the design phase, the program staff can start collecting relevant information before, during, and after the participants receive services.

If you don't already have these elements in a written program manual, now is the time to create one. It's a valuable training tool and a way to make sure that everyone involved in the work is operating with a single understanding of the program.

2 STEP

The elements of an outcome evaluation

Once your project's current operations are shown to be consistent with the program plan, it's time to move on to the next level: the outcome evaluation. To carry it out, the research team will need to define the study group and control group and identify the sources for the data they will collect and analyze.

The study group. Small programs may be able to include all of their clients in an evaluation. Studying the whole population can produce findings that are very accurate and also avoid errors that may occur in choosing a representative sample from the larger population. Evaluators of larger programs, however, may need to select a more manageable portion of the full population. In these cases, they will want to make sure that the sample population is representative of the total program population—that the two are similar in every relevant respect. For people trained in statistical analysis this is not an especially difficult process.

A control group. A control group is a population or sample of a population—that has not been exposed to the program under study. A control group may, for example, be participants from an earlier stage in the program's development, or youth receiving no treatment at all. What is most important is that the control group is similar to the study group in most other respects—such as race, age, risks, or needs—so that comparing the two reveals the program's influence. Again, while the task of finding a control group and doing random assignment may seem daunting, most researchers are very comfortable with these processes and are able to do them effectively.

As noted earlier in this guide, while it's helpful to have a control group, it is not always necessary. Where funding, logistics, or ethics make a control group impractical, researchers will want to use a quasi-experimental method, which uses statistical analysis to produce a control group equivalent.

Data. Most outcome studies will use administrative data, which is drawn from the program itself. Some studies may supplement this with data from other sources. Administrative data is typically quantitative (meaning it can be counted); supplemental data may be either quantitative or qualitative (descriptive).

- **Administrative data.** Most programs register new clients and collect information about them as they manage their cases. They begin by noting the date someone enters the program, biographical information, past treatment, history in the justice system, and the various risks and needs the person presents at intake. Later, staff also keep records about changes in clients' behavior and how they respond to the treatment. In the course of providing drug rehabilitation services, for example, program staff will keep records of a participant's attendance or the results of any required drug tests, as well as a record of assessments or scales that measure substance abuse. This administrative information— collected systematically and uniformly from all participants—usually forms the basis of an outcome evaluation.

- **Supplemental data.** Like administrative data, supplemental data is frequently quantitative. For example, if one of the main outcomes you want to measure is the number of rearrests and re-convictions, you would want access to this data from the police or court. Although it may seem daunting to do so, it is often possible to get such information by making a formal request to the relevant agencies.

Many researchers also find it helpful to collect qualitative data: attitudes, impressions, and opinions gathered through interviews, surveys, or discussion groups. This kind of information provides nuance, texture, and illustrative case studies. It can be very powerful to learn about a program's influence from a client's perspective—for example, when a youth says, "the counselors helped me talk to my parents about why I was skipping school."

WHAT DOES STATISTICAL SIGNIFICANCE MEAN?

Practitioners often hear researchers talk about whether a finding is "statistically significant." But what does that term actually mean? At its most basic, statistical significance is a measure of reliability; it allows you to say, with as much confidence as possible, that research findings are, in fact, real, and not observed by chance, or as a result of differences between the treatment and control groups.

Researchers are responsible for ensuring that the members of a study sample resemble the general population in as many characteristics as possible, in order to be able to assert that what is true of the sample is also true of the whole—that is, to make the findings generalizable.

Researchers also often point to something called the "p-value." This statistic measures the likelihood that a group selected from a larger population would resemble that larger population, as described above. A p-value of .05—meaning the differences between the control and treatment groups are likely the result of chance five times in a hundred—is generally acceptable in social science research. In studies where an entire population is observed (for example,

every youth entering the juvenile justice system), there is no opportunity for sampling error, and statistical significance measures aren't necessary.

Statistical significance is determined by both the magnitude of the differences observed and the size of the sample. Although findings might be important from a programmatic perspective, regardless of their significance level, statistically significant findings carry more weight in the research community.

3

After an outcome evaluation—Next steps

It is important to document an outcome evaluation. You can create a summary document or even a one-page overview of your evaluation that you can use to share your findings with fellow professionals. A full report tracing the steps of your evaluation and describing what was learned would be the best record of what you've done and what you've learned about your program. You can distribute copies to your partners, funders, and other practitioners. If your organization has a website, posting your findings online puts them into the public discussion and brings them to the attention of all interested audiences, including the media. Researchers can use your documentation to assess your research and its findings.

Most audiences for your findings will be interested in whether your program leads to positive outcomes for the clients. Some may have targeted interests, too. For example, funders may be focused on discerning areas in need of further development; other jurisdictions may want to know about specific target populations. When reporting results, stay true to the research and report all findings—both positive and negative. This balanced approach will underscore your program's integrity.

An outcome evaluation can yield a wealth of information about opportunities to improve your program. Use it to fix what doesn't work or could work more effectively. If you've found that your program succeeds with only one segment of the client population, be honest and report its value for that population alone. Likewise, if you've found that changes in youth behavior are immediate, but don't hold up over the long term, report that as well. All of these findings are important and can be used to make your program a more effective intervention for the youth it serves.

You can also use the evidence of an outcome evaluation to seek accreditation for your program. Some organizations, for example the Substance Abuse and Mental Health Services Administration (SAMHSA) [http://www.samhsa.gov/], publish success rate thresholds and, if you meet their criteria, they can give your program their seal of approval. The bar for these measures can be very high, however. But even if your program doesn't meet their standards, it is valuable to be able to point to evidence that you have gathered about your program's impact.

And if you still don't have a program manual, now is the time to produce one, so others can learn from it and consider adopting your program or parts of it.

Conclusion

There are many reasons why your service should aspire to being an evidence-based practice. But as this guide illustrates, there is a great deal of preparatory work leading up to the outcome evaluation on which your designation depends—so much that many programs don't have the capacity and funding to take it on.

Each step in the process is worthwhile for its own sake. A program that conducts only a process evaluation has accomplished a great deal by validating its program plan. People will at least have confidence that it does what it says it does (even if it cannot yet vouch for the outcomes).

Ultimately, understanding how to get to the outcome evaluation stage allows a program to grow intentionally, mindful of the importance of good planning and service delivery, steady program management, and consistent data collection.

For more information on the Center on Youth Justice, please contact its director, Annie Salsich, at (212) 376-3169 or asalsich@vera.org.

The Vera Institute of Justice is an independent nonprofit organization that combines expertise in research, demonstration projects, and technical assistance to help leaders in government and civil society improve the systems people rely on for justice and safety.

ModelsforChange
Systems Reform in Juvenile Justice

www.modelsforchange.net