Program

Evaluation

Briefing

Series

#7

Approaches to Assessing Juvenile Justice Program Performance

Juvenile Justice
Evaluation Center

Justice Research and Statistics Association

Office of Juvenile Justice and Delinquency Prevention

The JJEC, which is supported by the Office of Juvenile Justice and Delinquency Prevention, provides evaluation information, training, and technical assistance to enhance juvenile justice evaluation in the states. For more information about the JJEC project, visit our Web site at www.jrsa.org/jjec, or e-mail us at jjec@jrsa.org.

Approaches to Assessing Juvenile Justice Program Performance

This is one of a series of briefings prepared by the Justice Research and Statistics Association under the Juvenile Justice Evaluation Center (JJEC) project. The purpose of this briefing series is to provide juvenile justice program managers with information that will help them evaluate their programs. Each briefing addresses a topic that is of particular interest to juvenile justice program managers who are trying to determine the effectiveness of the programs they operate.

Juvenile Justice Evaluation Center
Justice Research and Statistics Association
777 North Capitol Street, N.E.
Suite 801
Washington, D.C. 20002
(202) 842-9330
www.jrsa.org/jjec

Acknowledgments

This briefing was prepared by Mary Poulin, Project Manager, JRSA Juvenile Justice Evaluation Center Project. Stan Orchowsky, JRSA Research Director, provided oversight, and Nancy Michel, JRSA Director of Publications, provided editing. Eric Peterson, the project's Grant Manager at the Office of Juvenile Justice and Delinquency Prevention, offered valuable support for which we are extremely grateful.

July 2004

This project is supported by Grant No. 98-RN-FX-0112 awarded by the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

Table of Contents

The purpose of this briefing is to focus on different methods of assessing program performance that comply with the JJDPA of 2002.

Introduction

The passage of the Government Performance and Results Act (GPRA) in 1993 brought the issue of assessing program performance to the forefront in the federal government. One of the major requirements of GPRA is that federal agencies must produce annual performance reports that indicate whether their goals have been accomplished. In the last decade, a great deal of time and effort has been devoted to the issue of accountability of federal programs.

In juvenile justice, a focus on "best practices" and "what works" to prevent/reduce delinquency and related problems has been important for a number of years as well. The reauthorization of the Juvenile Justice and Delinquency Prevention Act (JJDPA) in 2002 is a recent example of the federal government's push to focus on outcomes for state programs funded by the federal government. In the requirements for the Title II Part B programs (Formula Grants)[1] in the 2002 reauthorization of the JJDPA, each state is called to:

a) to the extent practicable give priority in funding to programs and activities that are based on rigorous, systematic, and objective research that is scientifically based;

b) from time to time, but not less than annually, review its plan and submit to the Administrator an analysis and evaluation of the effectiveness of the programs and activities carried out under the plan, and any modifications in the plan, including the survey of State and local needs, that it considers necessary; and

c) not expend funds to carry out a program if the recipient of funds who carried out such program during the preceding 2-year period fails to demonstrate, before the expiration of such 2-year period, that such program achieved substantial success in achieving the goals specified in the application submitted by such recipient to the State agency. (P.L. 107-273)

[1] Similar provisions exist for Title II Part C (Delinquency Prevention Block Grants). There are new accountability requirements for Juvenile Accountability Block Grants (JABG) as well.

This represents a significant shift from the 1992 reauthorization of the JJDPA when a state was required to:

> from time to time, but not less often than annually, review its plan and submit to the Administrator an analysis and evaluation of the effectiveness of the programs and activities carried out under the plan, and any modifications in the plan, including the survey of State and local needs, which it considers necessary. (P.L. 102-586)

Though states must submit an annual performance report to OJJDP that includes performance measures, states will develop different approaches to comply with the requirements of the JJDPA. The purpose of this briefing is to focus on different methods of assessing program performance that may be used to comply with the JJDPA of 2002 or other requirements that states and local jurisdictions may face regarding the assessment of program performance.

There are three major methods of assessing program performance: program monitoring, performance measurement, and impact evaluation.

Methods of Assessing Program Performance

There are three major methods of assessing program performance: program monitoring, performance measurement, and impact evaluation. While overlap among the three is considerable, there are important differences in terms of the amount and type of information provided by each and the resources and skills needed to use each. External factors will influence the choice of a method. The review of these differences will, hopefully, provide guidance to program administrators and others responsible for making decisions regarding how to approach the assessment of program performance. Though all three methods are discussed here for comparative purposes, monitoring is not sufficient for meeting JJDPA or GPRA requirements.

Program Monitoring

Program monitoring involves the continuous collection of information about the activities and operation of a program. This information is used to determine what the program is actually doing and whether activities are being accomplished. The focus is on program inputs and outputs. That is, program monitoring tracks what kinds of resources are being expended for the program (inputs) and what products or activities result from the resources expended (outputs). Information on what is actually happening is compared to what is supposed to happen to determine if the program is operating as designed. Outcomes are not part of program monitoring. That is, program monitoring will not produce information on changes in the target(s) (e.g., clients) of the program. Program monitoring can report on, for example, "number of clients served in FY 2003 versus number of clients the program was expected to serve in FY 2003."

Monitoring of inputs and outputs is necessary for accounting purposes and is useful, but not sufficient, when programs are asked to provide information about program effectiveness or whether the program has achieved its goals and objectives. So, because it does not provide information regarding program success, program monitoring does not fulfill all of the requirements of GPRA or the 2002 reauthorized JJDPA.

Performance Measurement

Performance measurement is concerned with collecting information to determine whether a program achieved its goals and objectives. Information from performance measurement is used to improve the operation of the program. Outcome measures, inputs, and outputs are collected and reported. That is, performance measurement includes program monitoring. So, performance measures include both process and outcome measures. Though a much more detailed discussion is needed to understand the requirements of GPRA and the JJDPA, performance measurement does seem to fulfill the requirements.

If an objective of a program is to "improve academic performance of its clients," an outcome measure that could be used to capture whether this objective was achieved is "increase in grade point average (GPA) from the first to the second reporting period." Using an output measure such as "number of tutoring hours provided" does not demonstrate whether there was a change in performance. Though tutoring may be the intervention used to improve performance, the measurement of tutoring is an indication of services received rather than whether the objective was achieved. For performance measurement, an output measure should be used to indicate whether the intervention was provided as planned, but an outcome measure that demonstrates change in the targeted behavior is required to show that the objective was achieved.

The output and outcome measures described above provide information about the entire population served by a program. These statements are useful for making overall assessments about program accomplishments. However, research on what works to prevent/reduce delinquency and other associated problems has firmly established that no one intervention works for everyone in the same manner. That is, what may be useful for one group of individuals to improve school performance may not work for another set of individuals. For this reason, when enough data are available, it is important to disaggregate outcome information by subpopulations that may be differentially affected by the program. Breaking out outcome information by demographics (e.g., gender or race/ethnicity), difficulty of client population (e.g., level of delinquency risk) (Hatry, 1999), or other key differences in the target population will increase our understanding of what works, for whom, and under what circumstances.

For consistency and precision, programs with similar goals and objectives should use the same measure(s) and collect data in the same manner. This is particularly important if the outcomes will ultimately be aggregated by the parent or funding agency. That is, the same definition and the same method of measuring outcomes must be used. This permits the grouping of outcomes (i.e., aggregation) in order to make an overall statement of outcomes for all of the programs and it allows comparison of outcomes across programs. Parent or funding agencies can work with programs and other stakeholders to develop a list of measures from which programs select to measure the performance of their program.

The following example demonstrates the importance of using the same measures and the same method for collecting data. Achievement of the outcome "improvement in academic performance," for example, can be defined in different ways. Change in grade point average (GPA) from the beginning to the end of the program is one way to define improvement in academic performance, but change in completion of homework as reported by the youth weekly throughout the program is another way. Once an outcome has been defined, it is important that it be measured in the same way. GPA and homework completion are two measures of academic performance, but the source of information for each differs: GPA is an official measure while homework completion is a self-report measure. Also, the timing for data collection differs: GPA is tracked at two points in time (beginning and end of program) while homework completion is tracked at multiple points (each week). This will result in differences in how the results are reported. For example, has school performance improved if the youth reports that he/she began doing homework in week three of the program, did it consistently for eight weeks, but by the end of the program stopped doing homework altogether? If programs determine "improvement in academic performance" with different measures and data collection strategies, it becomes difficult to determine how to aggregate outcomes and to compare the accomplishments of one program to those of another.

Impact Evaluation

Performance measurement can provide a great deal of valuable information about program performance, but it cannot determine the impact of the program, i.e., whether the outcomes achieved were due to the program itself and/or factors external to the program. That is, outcomes cannot be attributed to the program itself using only performance measurement. In some situations, such as when a program was not implemented as designed, it may be clear by using performance measurement that the poor program implementation is the cause of or reason for not achieving program objectives. However, in other situations determining the cause is quite difficult. Some performance measurement designs do a better job than others with regard to explaining why outcomes were accomplished. For example, designs that measure change(s) occurring during the program for particular subpopulations (e.g., boys versus girls), which consider the impact of intervention dosage (e.g., program completers versus noncompleters), or examine the relationship between available resources on outcomes may help clarify why outcomes differed by participants. Nonetheless, to determine whether a program caused the outcomes, an impact evaluation must be conducted.

Good performance measurement can facilitate evaluation. It can provide data to use in the evaluation, help raise questions for the evaluation to answer, and serve as a pilot test to ensure data collection "kinks" are resolved. Any relationships developed with stakeholders in the performance measurement process will be helpful when designing and implementing the evaluation. The data collected (i.e., performance measures) may be the same, but the design for data collection and the data analyses will change from performance measurement to evaluation.

Impact evaluation has several distinguishing characteristics relating to focus, methodology, and function. Impact evaluation (1) assesses the effectiveness of an ongoing program in achieving its objectives, (2) uses methods to determine whether achievements can be attributed to the program or other factors, and (3) aims at program improvement through a modification of current operations.

Though evaluation designs vary widely, impact evaluation is markedly different from performance measurement in its ability to control for factors external to the program that may be influencing its achievements or outcomes. Taking the example used above, let us imagine that a program provides tutoring in an effort to improve school performance at the same time that the students' school begins to offer incentives for students who receive A's on their report card. If an increase in grade point average occurs for those students in the program, how do we know if the increase was due to the tutoring, the incentives, both of those factors, or something else? A well-designed and implemented impact evaluation can use methods such as comparison groups to control for external factors and attribute observed changes to the program/intervention.

The definition of impact evaluation used here differs from those offered elsewhere. Often evaluation is used as an umbrella term that includes program monitoring and performance measurement. Each of these approaches can be thought of as evaluative in nature. On a continuum of evaluative activities, program monitoring provides the least amount of evaluative information while impact evaluation provides the most (and the best) amount of evaluation information. It should be noted, however, that the quality of any of these approaches can vary substantially, such that a well-designed and well-implemented performance measurement may be preferable to a poorly designed and implemented impact evaluation.

Summary

This section considered the key differences in the information provided by each of the three methods of assessing program performance. When one moves from program monitoring to performance measurement to impact evaluation, the type and amount of information provided increases. Program monitoring provides information on outputs. Performance measurement provides information on outputs and outcomes. Impact evaluation provides information on outputs and/or outcomes, but also controls for at least one factor external to the program in order to determine whether the program itself is responsible for the goal(s) and objectives being reached.

Available resources—time, money, or knowledge—and one's information needs will be the major factors to consider when choosing a particular method.

Factors That Influence the Selection of an Assessment Method

The introduction to this briefing presents a critical reason for measuring program performance—a funding source may require such an assessment. However, other factors will play a role in which assessment method or combination of methods is selected. The key factors that influence the choice of a method are time, money, funding requirements, information needs, and the level of knowledge needed to effectively employ each method.

Information Desired

The decision about which program assessment method to use should be based upon the questions that one would like to answer about the program or, practically, the questions that one can answer given program constraints. Knowing appropriate questions to ask for each method helps identify the kind of information that each can provide. The complexity of questions increases from monitoring to evaluation, indicating that the most performance information can be provided by evaluation. The following are appropriate questions to ask for each method:

Program Monitoring

- Does the program have the appropriate number of staff?
- Are the activities being performed as scheduled?
- What youths are being served?

Performance Measurement

- All of the questions appropriate for program monitoring
- Were the behavior changes sought by the program achieved?
- Were the system efficiency changes sought by the program achieved?
- Are short-term and intermediate outcomes being achieved?

Impact Evaluation

- All of the questions appropriate for program monitoring and performance assessment
- Did the behavior changes occur as a result of participation in the program?
- Did the observed system efficiency changes occur as a result of the program?
- Did one intervention have a greater impact on outcomes than other interventions?

Resources and Skills Required

With an increase in the question complexity comes a corresponding increase in the resources and skills required to be able to answer the question. So, as one moves along the continuum from program monitoring to evaluation, more resources and skills are needed.

Program Monitoring. Compared to other methods of performance assessment, program monitoring is the easiest to conduct. Monitoring has been a requirement of funding agencies for a long time and grantees are familiar with how to engage in it. Resources to conduct monitoring are usually built into program budgets and requirements are generally well developed by funding agencies. Though program monitors do not usually have the training and background of evaluators, they can be very helpful in working with evaluators to collect data.

Performance Measurement. Performance measurement requires some training in evaluation, data collection, and data analyses. Setting up a good performance measurement system will take several months. It will take time to select performance measures, design a system for data collection, and train individuals in how to collect data. Program duration and timing for data collection will determine how much time it will take before the first outcomes are reported. However, six months to one year after data collection begins is likely the minimum amount of time for first outcomes to be reported.

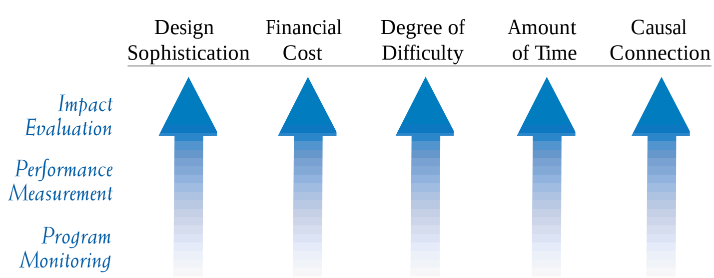
With appropriate training, program staff can conduct performance measurement; however, they may find it useful to consult with a professional evaluator in the course of selecting measures and in the development of a data collection system. In light of the complexity of collecting data, decisions must be made in advance about who will collect data and conduct analyses. Once these decisions are made, training is required for those collecting data to ensure valid and reliable data are collected. Though it will take time, it is a good idea to include stakeholders in the development of the performance measurement system to ensure their support and involvement. This will ultimately produce a more useful performance measurement system.

Evaluation. To conduct an evaluation well, advanced training in research methods and evaluation is required. Typically professional evaluators have completed graduate programs in the social or behavioral sciences. In terms of finances, it is recommended that 5-10% of a program's budget be devoted to evaluation. Unless a program or jurisdiction has an in-house evaluator, someone from outside the agency will have to be hired to conduct the evaluation. This contributes, obviously, to budget costs. Similar to performance measurement, substantial planning for design and data collection is required before the evaluation begins. Unlike with many performance measurement systems, data collection for an evaluation often extends beyond program completion or to persons or entities outside of the program. (For example, follow-up of clients is often used to determine sustainability of program effects.) Finally, data analysis for an evaluation typically takes longer due to the additional data collection and complexity of analyses. Though additional knowledge/training, money, and time are needed for an evaluation, the key benefit of a well-designed and implemented evaluation is the ability to determine whether the program *caused* the outcomes.

Summary

As one moves along the continuum from program monitoring to impact evaluation, the design sophistication, financial cost, degree of difficulty, amount of time, and the ability to make connections between outcomes and the cause(s) of the outcomes generally increase. Available resources, whether they are time, money, or knowledge, and one's information needs will be the major factors to consider when choosing a particular method. The table below presents these differences visually. As a general rule, movement is from lowest to highest in each of these categories, as indicated by the darkening of the arrows in the figure below. However, the differences are relative and approximate.

Factors in the Program Assessment Continuum



Preparing for a Performance Assessment

A program must be evaluable to participate in performance measurement or evaluation.

A program must be stable and well organized in order to participate in performance measurement or evaluation. That is, it must be evaluable. A program is evaluable if there is a sound and formal program design or model in place, the program serves the population it is intended to serve, the resources required to operate the program are in place, the activities are being implemented as designed, and the program has the capacity to provide data for the evaluation or performance measurement. Though no program will be perfect, substantial problems in any one of these areas will produce information of questionable value. Conducting an evaluability assessment is one way to determine if a program is ready for performance measurement or evaluation. For more information on evaluability assessments, see the JJEC Briefing *Evaluability Assessment: Examining the Readiness of a Program for Evaluation* (JRSA, 2003).

States are expected to assess actual performance against the goals and objectives for which funding was provided.

Using Performance Measurement or Impact Evaluation to Fulfill OJJDP Requirements

The 2002 JJDPA contains two new major evaluation-related requirements: one pertains to the funding of scientifically-based programs and the other to the funding of programs that have demonstrated substantial success. When making decisions about which programs to fund, states should develop a funding process that selects programs with demonstrated efficacy or, at least, programs that can reasonably be expected to work based on accumulated research and evaluation evidence. Information on evidence-based programming is available in a number of places, including:

- JJEC Online
  http://www.jrsa.org/jjec
- Center for the Study and Prevention of Violence: Blueprints for Violence Prevention
  http://www.colorado.edu/cspv/blueprints/
- Development Services Group: Model Program Guide
  http://www.dsgonline.com/WebEffects/dhtml_slide_tree/pepg_int.htm
- Substance Abuse and Mental Health Services Administration: Model Programs
  http://www.modelprograms.samhsa.gov/

Once a state has decided what program(s) to fund, performance measurement or evaluation can be conducted to determine if the program(s) achieved substantial success.

According to OJJDP, to determine whether or not a project has achieved "substantial success," a state is expected to require that applicants for subgrants provide clear and achievable goals and measurable objectives and performance measures that will indicate the degree to which goals and objectives have been achieved. States must also require applicants for subgrants to identify data sources to be used to assess the performance measures. In determining whether subgrantees have achieved "substantial success," states are expected to assess actual performance against the goals and objectives for which funding was provided.

Once it is clear what information the state would like to produce and what resources and skills are available, the development of a system can begin.

Summary

This briefing has introduced the concepts of performance measurement and impact evaluation with a particular emphasis on how they relate to the new requirements in the JJDPA concerning the demonstration of program effectiveness. States must make a number of decisions in order to develop a strategy to fulfill these requirements. Desire for information, resources, and available skills will help inform these decisions. There are a number of other decisions to make when creating a performance measurement or impact evaluation system. However, once it is clear what information the state would like to produce and what resources and skills are available, the development of a system can begin.

Selected Bibliography

Barnoski, R., Aos, S., & Lieb, B. (2003). *Recommended quality control standards: Washington state research-based juvenile offender programs*. Olympia, WA: Washington State Institute for Public Policy.
http://www.wsipp.wa.gov/

Center for the Study and Prevention of Violence. (2002). *Blueprints model programs selection criteria.*
http://www.colorado.edu/cspv/blueprints/model/criteria.html

Coalition for Evidence-Based Policy. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user-friendly guide.* Washington, DC: U.S. Department of Education, Institute for Educational Sciences, National Center for Education Evaluation and Regional Assistance.
http://www.excelgov.org/displaycontent.asp?keyword=prppcHomePage

Government Performance and Results Act. U.S. Public Law103-62. 103rd Congress. August 3, 1993

Hatry, H. P. (1999). *Performance measurement: Getting results*. Washington, DC: Urban Institute.

Horn, J. (2001). *A checklist for developing and evaluating evaluation budgets.* Evaluation Checklist Project.
http://www.wmich.edu/evalctr/checklists

Justice Research Center. (2003). *2003 outcome evaluation report.* Tallahassee, FL: Author.
http://www.djj.state.fl.us/djj/research/statsnresearch/index.shtml

Juvenile Justice and Delinquency Prevention Act of 2002. U.S. Public Law 107-273. 107th Congress. November 2, 2002.

Juvenile Justice Evaluation Center. (2003). *Identification of Excellent Juvenile Justice Program Evaluations.*
http://www.jrsa.org/jjec/programs/rating-criteria.html

Maxfield, M. (2001). *Guide to frugal evaluation in criminal justice.* Washington, DC: U.S. Department of Justice, National Institute of Justice. http://www.ncjrs.org/

U.S. Agency for International Development Center for Development Information and Evaluation. (1996). *Selecting performance indicators.* Performance monitoring and evaluation tips. (No. 8). Washington DC: Author.

U.S. General Accounting Office. (1998). *Performance measurement and evaluation: Definitions and relationships.* Washington, DC: U.S. Author.

U.S. General Accounting Office. (1998). *The results act: An evaluator's guide to assessing agency annual performance plans*. Washington, DC: Author.

U.S. General Accounting Office. (2002). *Strategic plan: 2002-2007.* Washington, DC: Author.

Program Evaluation Briefing Series

#1 Juvenile Justice Program Evaluation: An Overview

#2 Hiring and Working With an Evaluator

#3 Strategies for Evaluating Small Juvenile Justice Programs

#4 Cost-Benefit Analysis for Juvenile Justice Programs

#5 Incorporating Evaluation Into the Request for Proposal (RFP) Process

#6 Evaluability Assessment: Examining the Readiness of a Program for Evaluation

#7 Approaches to Assessing Juvenile Justice Program Performance

Juvenile Justice Evaluation Center
Justice Research and Statistics Association
777 North Capitol Street, N.E.
Suite 801
Washington, D.C. 20002
(202) 842-9330
www.jrsa.org/jjec