U.S. Department of Justice
Office of Juvenile Justice and Delinquency Prevention

135817

OJJDP

EVALUATING JUVENILE JUSTICE PROGRAMS: A DESIGN MONOGRAPH FOR STATE PLANNERS

Technical Reference and Information Series

135817

135817

N C J R S

MAR 5 1992

P. ACQUISITIONS

Evaluating Juvenile Justice Programs
A Design Monograph for State Planners

Prepared by

James C. Coldren
Timothy Bynum
Joe Thome, Project Coordinator

Community Research Associates, Inc.
115 North Neil, Suite 302
Champaign, IL 61820

August 1989

The Assistant Attorney General, Office of Justice Programs, coordinates the criminal and juvenile justice activities of the following program Office and Bureaus: National Institute of Justice, Bureau of Justice Statistics, Bureau of Justice Assistance, Office of Juvenile Justice and Delinquency Prevention, and Office for Victims of Crimes.

PREFACE

Quality and efficiency in programming for young people is a priority for the Office of Juvenile Justice and Delinquency Prevention. The evaluation of programs designed to assist troubled youths and those at-risk is therefore a primary concern as well.

Recently, OJJDP conducted a survey to assess State involvement in evaluating programs funded under the Juvenile Justice and Delinquency Prevention Act's Formula Grant program. The Office found great interest in evaluation, a strong indicator that federal assistance would be extremely helpful in improving quality and consistency in program development.

This monograph therefore represents one aspect of OJJDP's effort to assist the states with program evaluation. We share the states' concerns that programs are responsive to the needs of young people, and designed to efficiently use the dollars allocated to address youths' needs. Sound evaluation research will help planners identify exemplary programs and allow them to share their findings with OJJDP and others interested in addressing similar problems.

Accountability, effectiveness, quality control, and the ability to solve problems all represent benefits of evaluation. With those goals in mind, programs aimed at helping young people should continue to improve. This monograph will help fulfill that mission.

ACKNOWLEDGEMENTS

Contents

Chapter 4 (continued)

Evaluating Juvenile Justice Programs:
A Design Monograph for State Planners

1

"What works?" That simple question seems at times to be the most common query of juvenile justice system administrators and planners. For today's system professionals, who face resource shortfalls daily, the question might also be posed as, "Is my program working?" or "Are our clients getting what they need?"

To answer these types of questions, state juvenile justice specialists and other system administrators must turn to evaluation research. The problems associated with juvenile delinquency are alarmingly sophisticated and intimidating. Doing what is best for the youth of a community and developing programs that take maximum advantage of limited budgets remain priorities. Toward these goals, evaluation research can play an important part.

The purpose of this monograph is to offer evaluation strategies to state juvenile justice specialists, state advisory groups, juvenile program administrators, and others interested in learning more about the processes and outcomes produced by Formula Grants projects. It is a primer designed to provide practical advice on creating or enhancing evaluation programs within the scope of budgetary, staffing, time, and other administrative obstacles.

The Office of Juvenile Justice and Delinquency Prevention (OJJDP) State Relations and Assistance Division (SRAD) sponsored this monograph in response to a

Purpose and Overview

1988 survey of juvenile justice specialists[1] that revealed a strong interest in how evaluation research can provide insight into program activities and outcomes.[2] Respondents asked for assistance in developing and/or refining their evaluation processes.

The U.S. Congress formalized its interest in learning more about the characteristics of state and local programs created using Juvenile Justice and Delinquency Prevention (JJDP) Act Formula Grants Program funds by directing OJJDP to assist state efforts at evaluating, replicating, and marketing JJDP Act programs. Specifically, the 1984 amended version of the JJDP Act states that the Act is intended "to provide for thorough and ongoing evaluation of all federally assisted juvenile delinquency prevention programs" [Sec. 102(a)(1)].

Furthermore, each state that participates in the Act's Formula Grants Program is required to submit a three year comprehensive plan, with annual updates, to OJJDP. Among its other requirements, the plan details how the state will "provide for the development of an adequate research, training, and evaluation capability within the State" [Section 223(a)(11)].

For these reasons, OJJDP's State Relations and Assistance Division (SRAD) commissioned this evaluation monograph.

Benefits to the Reader

This monograph serves as an evaluation primer for State Planning Agency (SPA) staff,

persons and agencies with whom they contract, and for criminal justice analysts involved in juvenile justice program evaluation activities. While the OJJDP evaluation survey indicates that juvenile justice specialists are generally well acquainted with evaluation knowledge and skills, they are not uniformly supportive of evaluation activities, or do not feel they are in a position to conduct meaningful evaluations. Based on these findings, it was felt that a less academic and more practical approach to juvenile justice program evaluation issues was warranted.

This monograph is intended to provide an array of benefits to the reader:

- By explaining the variety and types of program evaluations which can be conducted by juvenile justice specialists, it reveals that their programs are more capable of being evaluated than may have been previously thought.

- It provides suggestions for overcoming obstacles such as a lack of resources, service delivery priorities, and time constraints to get more out of data that is already collected. It offers some insights to measuring program performance with more meaning.

- It provides better communication about evaluation standards between OJJDP and the States, and offers some helpful advice for those who could use it.

- It stands on its own as an informational resource for juvenile specialists, but it can also serve as a reference manual to accompany evaluation training being developed by OJJDP.

Although this manual covers dozens of subjects, it is not inclusive. No single document can cover all there is to understand about evaluation research, and this monograph is not excluded. There are dozens of excellent reference books and articles on the topics included in this monograph (see the final chapter for examples) and the Office's State Relations and Assistance Division provides periodic training on the topic. Additionally, OJJDP's technical assistance programs exist to help persons interested in pursuing some aspects of evaluation. The monograph is therefore intended to be a *primer* on the topic of evaluation. The best way to use this monograph is to supplement the information with additional readings or through the assistance made available by OJJDP.

Monograph Organization

The monograph is divided into seven major sections. The remainder of this introductory chapter provides a reference guide to issues and sections which comprise the majority of the monograph. Chapter Two reviews the incentives and logic of evaluation. Chapter Three discusses approaches to evaluation. Chapter Four addresses issues central to conducting accurate evaluations and reviews their uses. Chapter Five focuses on issues related to organizing or establishing an evaluation program. Chapter Six offers references and resources for those interested in further reading.

The monograph format allows the reader to identify and select for review only those chapters or sections which address issues of greatest concern. Where appropriate, sidebars highlight examples of techniques described in the text or further define or explain specific concepts or methods. Definitions are highlighted on the page where an issue or topic is introduced.

Monograph Guide

Because the monograph contains information, tips, and practical advise on dozens of topics, a handy reference guide is presented here. For each of Chapters Two through Five important points are highlighted and references to the relevant pages provided in parentheses. When using the reference guide in conjunction with the index at the back of the monograph, selecting and findings special topics should be quite easy.

Why Evaluate (Chapter 2)

(1) There are a number of sound reasons to develop a formal evaluation process. A few examples include *(p. 11)*:

- The JJDP Act requires it.
- The Formula Grantee may require it.
- To determine what programs and policies produce.
- To assess project performance.

(2) Developing a flowchart of the state juvenile justice program and goals is crucial to understanding the role which evaluation can play *(pp. 11-12)*.

Approaches to Evaluation (Chapter 3)

(1) The following factors should influence the choice of programs selected for evaluation *(pp. 19-20)*:

- Accessibility of the program.
- Length of time program has been, or will be, in operation.
- Program cost, relative to other programs or juvenile justice initiatives.
- Controversy, or politics surrounding the program.
- Extent of demand for evaluation of the program(s).

(2) Two factors relating to the time frame for evaluation will also influence the decision regarding which program(s) to evaluate *(pp. 20-21)*:

- The timeliness with which evaluation information can be provided to decisionmakers because they must often operate on very short time frames.

- The time span of the program under consideration. Short-lived programs may not be good candidates for evaluation, nor might programs whose intervention takes a long time to be observed.

(3) A good evaluator will first ask "What do I want to learn?" before embarking on an evaluation project. Careful consideration of this question and its ramifications will facilitate research design and prevent problems from arising in mid-evaluation.

(4) The distinction between three classes of evaluation--program monitoring, process evaluation, and outcome evaluation--is clarified (p. 21).

(5) Three types of outcome measures are identified (pp. 22-23):

- Instrumental outcomes--measures of phenomena directly related to program goals and objectives.

- Knowledge production outcomes--generation of new knowledge or understanding about juvenile justice programs.

- Consensus building outcomes--the production of common understanding of juvenile justice issues and programs among various participants.

(6) Six types of evaluation studies are identified; each consists of a different combination of level of evaluation and comparative perspective. The six typologies include program-only and program/comparison versions of monitoring, process evaluation, and outcome evaluation (pp. 23-29).

This typology is explained using a juvenile justice Law Related

Education evaluation example (pp. 24-29).

Critical Issues (Chapter 4)

(1) Measurement in program evaluation covers three general areas-- measurement of input, measurement of intermediate program effects, and measurement of program output. Special attention is given to alternative measures of program impact, e.g., savings to other social service areas result from a new program. Each area carries a special set of concerns regarding the context the program operates in, and the validity of the measures chosen (pp. 32-38).

(2) The validity issue has two aspects-- threats to internal validity and threats to external validity. Threats to internal validity are things that could have caused program results other than program participation, such as client maturation, program history and staff learning curves. Threats to external validity are things that prevent the evaluation findings from being applied to other programs or clients (pp. 38-40).

(3) Common sources of data for program evaluators are reviewed, covering official records, self-reported data, interviews and questionnaires, and surveys. Special attention is given to the development and utilization of surveys (pp. 40-43).

(4) The use of random assignment in evaluation studies, that is, assigning research subjects to treatment/ intervention and control groups on a random basis, is also given special attention because it is the best way to resolve threats to internal validity, which is usually the evaluator's chief concern (pp. 43-44).

(5) A special section is devoted to the use of observation in program evaluations. Periodic visits to programs under evaluation to observe operations first hand, and perhaps talk to staff and participants, will

provide rich anecdotal information that supplements the quantitative data collected *(p. 44)*.

(6) Cost benefit evaluation issues are reviewed. This includes discussion of who benefits from the program, who incurs the program cost, and enumeration of cost and benefit types such as direct versus indirect, alternative costs, and fixed versus variable program costs *(p. 45)*.

(7) Three common errors in program evaluation are reviewed *(pp. 44-46)*:

● Using individual cases as indicators of success. It is tempting to cite individual successful cases as indication that a program is working, because they are powerful motivators for program staff, they can indicate what is working in the program, they provide a good example for other participants, and they serve a public relations purpose as well. But there is danger in implying that an individual success, or even a few successes, proves program success. An overall comparison of statistical and qualitative indicators is required for the evaluation at the program level *(p. 44)*.

● Using expert opinion. The use of experts for consulting purposes, and to review program operations, is recommended. But their reports about program effectiveness or success, based on their experience with other similar programs, should never substitute for thorough evaluation research *(p. 46)*.

● Using an inappropriate comparative base. Before making comparisons, for example between programs or between one program state and another, evaluators and administrators should investigate the basis of the comparison carefully. If the comparison is

inappropriate, e.g., two similar programs are compared, but the educational level of participants varies greatly between them, the evaluation may be discounted *(p. 46)*.

(8) Imagination is recommended as an important component in any evaluation program. When one gets into program evaluation, there are always a number of obstacles, problems, and distractions, both political and research-related. The good evaluator learns to be creative in the research design process, finding proxy measures for important variables perhaps, and in the political process in findings ways to make evaluation results useful. There is no better resource than experience. The new evaluator must consult others in the field, and documented examples of program evaluations abound, as do other resources *(pp. 46-47)*.

(9) Using Evaluations *(pp. 47-49)*.

● The uses of evaluation products extend beyond the immediate need to answer questions about program implementation and success. Evaluation information is an important ingredient in the policy analysis and feedback process at local, state, and national levels. Policy analysis and formulation take place whether or not evaluations are conducted; that is the business of government. The availability of evaluation information can inject an increment of objectivity into the policy process which makes this a primary function of evaluation.

● To insure maximum utility of evaluation findings, products can be geared to audiences in a variety of fashions. For example, separate summary and technical reports on evaluation findings help insure that the appropriate users get the information they are looking for.

- There are a variety of consumers of evaluation information. Those interested in program operations and intermediate effects may be different from those interested in the bottom line indicators of program outcome. Program clients and participants should be considered valid consumers of evaluation information, as should other criminal justice professionals and potential supporters of the program(s) under investigation.

- Products need not be written to be given serious consideration. Video productions, slide shows, and other live presentations will get important points across to some audiences better than a research report.

- Effective and appropriate dissemination of findings should be given serious consideration, especially when knowledge production and consensus building are considered valid outcomes of evaluation projects.

Other Considerations (Chapter 5)

Three important topics are covered in this chapter--building evaluation requirements into program RFPs, preparing evaluation RFPs, and choosing evaluators. These sections include advice from experts who have worked in the evaluation field for many years, and should be considered guidelines that will help planners and juvenile justice specialists make good choices.

(1) Building evaluation into a program RFP. This is a good way to publicly and firmly state your intent to evaluate. The following suggestions for inclusion in juvenile justice program RFPs are offered *(pp. 50-52)*:

- State your plans to evaluate the program at the beginning of a program RFP.

- Include a section describing the evaluation information that program staff must provide.

- Include a section that explains the evaluation plan to potential grantees.

(2) Preparing an evaluation Request for Proposals (RFP). A good RFP for evaluation of a juvenile justice program will have five key sections *(52-56)*:

- A statement of objectives that states clearly and concisely the project to be funded, and the need for the evaluation effort.

- A section containing general information for applicants. This normally contains technical and logistical details that applicants must know.

- A specifications section that explains in detail the entire evaluation project--purpose, timelines, mandatory deliverables, products, and the like. This is the most important section of the RFP.

- A section outlining the detailed information requirements that must be contained in an evaluation proposal.

- A section explaining the RFP evaluation and selection process.

Providing this information to potential evaluators, in as specific a manner as possible, will help insure that the evaluation products meet your needs.

(3) Choosing an evaluator. This can be a frustrating experience. This section reviews the pros and cons of employing outside evaluators, and provides hints for recognizing talent. Good evaluators have three basic characteristics *(pp. 56-57)*:

- Experience. If they have conducted other evaluations, they

should provide references that can be contacted. Generally, experience counts heavily in selecting evaluators.

- Skills. Evaluators should possess, and be able to demonstrate, familiarity with statistics and research design.

- Creativity. Good evaluators should be able to apply knowledge from their experiences and other fields to the problems at hand. This is often the most difficult characteristic to assess.

NOTES

1. The juvenile justice specialist is the staff person for the state agency charged with administering that state's participation in the Juvenile Justice and Delinquency Prevention Act's Formula Grants Program.

2. See the *Survey of State Planning Agency Evaluation Capabilities*, prepared for the Office of Juvenile Justice and Delinquency Prevention by Community Research Associates, Inc. 1988.

2

Why Evaluate?

Program administrators evaluate on a daily basis. Decisions to develop a new policy and procedures manual, hire new staff, purchase new computer equipment, or search for program alternatives all occur because a program manager believes there is a problem which must be addressed. In essence, the program is not providing all that it was designed to accomplish. Evaluation research techniques can be used to assist the juvenile justice program manager in making those decisions, but, they may have been under-utilized.

Evaluation research can take a variety of forms, cover a range of issues and activities, and be conducted in response to diverse concerns. One should not believe that the only meaningful evaluation is that which is time consuming or involves sophisticated data collection. Evaluation should not be thought of only in terms of traditional, restricted definitions. As long as the results are meaningful and accurate, evaluation designs can encompass a variety of formats and approaches.

Regardless of the methods, the purpose for evaluating a program basically remains consistent--to provide timely and useful information for decision-making by staff, funders, and others. Is a program making progress, or creating unintended problems? Is a particular program cost effective? Could a program improve if certain characteristics were changed? Are the appropriate clients being referred to the program? Should the program be continued into a future funding

Getting Started. . .
Incentives and Early Steps

cycle? Depending on the resources available to an evaluator and the design of the evaluation program, questions such as these and many others can be answered.

Evaluation research should address concerns of effect as well as concerns of understanding. A juvenile justice program evaluation might be conducted to measure outcome to determine, as accurately as possible, what happened as the result of the implementation of a particular program. Such research might also be conducted to enhance our understanding of the relationships between the many factors which combine to characterize that program--why do things happen the way they do?

The incentives for building evaluation into program management are numerous. For the juvenile justice specialist, those incentives might include the objectives of the Act, those of the agency administering the Formula Grants Program, those of the state advisory groups, and those of the local program grantees.

For example, the reader may recall that as participants in the JJDP Act's Formula Grants Program, state planning agencies are obligated to develop an "adequate...evaluation capacity (for) the State" [Sec. 223(a)(11)]. Congress intended that participating states demonstrate some type of evaluation capability. Its concern is possibly directly attributable to the large amounts of Formula Grants Program funds distributed under the program since its inception. Participating states and territories have received more than $350 million in this decade alone, justifying

congressional interest in the impact that money has had at the state and local levels.

An annual Performance Report of program successes and problems is also a requirement of participating in the JJDP Act [Section 223(a)(22)]. The ability to provide accurate and informative Performance Reports can reflect the extent of a State's formal evaluation activities.

Beyond Congressional and regulatory requirements, there are a number of ways in which evaluation can benefit the SPA, the state advisory group, and the local program personnel. Generally, evaluations allow an administrator to:

- Judge the worth of ongoing programs,

- Estimate the usefulness of attempts to improve them,

- Increase the effectiveness of program management and administration,

- Objectively assess program performance with the intent of enhancing and improving it,

- Perhaps answer the question, "Did the program work?" If it cannot be answered, the evaluation can still provide useful information on the question, "How does the program work?",

- Address questions regarding program costs,

- Recognize secondary, or unanticipated, program effects from information that simply would not exist if there were no evaluation,

- Identify instances when results suggest it is appropriate to replicate and market programs of interest to the juvenile justice specialist and program contractors in other states, and

- Meet requirements regarding accountability.

These incentives and benefits must be weighed against the perceived or actual problems with evaluations.

The survey of State juvenile justice specialists revealed that the following were considered problems which hindered attempts to evaluate:

- Many of the programs initiated using Formula Grants Program monies are limited in size and aimed at only one section of the at-risk population.

- Relatively low caseloads produce situations where it can cost as much to evaluate a program (using an independent contractor) as it does to run it. Such a scenario scares agencies away from evaluation.

- Potential grantees may be intimidated by the prospect of participating in evaluations and choose simply not to apply for grants which require it.

- Many states are highly dependent on limited Formula Grants Program monies for a large proportion of their juvenile justice system program development. There is often a concern that those monies should be concentrated in service delivery rather than sacrifice some services to conduct evaluations.

- There is a belief that effective evaluations require extreme amounts of time and energy.

Many others believe, however, that the incentives and benefits override the problems associated with evaluation and that with careful design, many of these problems can be overcome or circumvented. As a result, processes are developed to incorporate data collection and analysis into routine program management. In the end, some gauge of whether a program is accomplishing its objectives is achieved.

As resources allow and the advantages which evaluation can provide are understood, evaluation research increasingly becomes a standardized aspect of program development. Justice system problems, as measured by delinquency rates, probation staff caseloads, training school system overcrowding, and the sophistication of delinquent activity, remain a serious concern. It is, therefore, not adequate to identify a problem, develop a responsive program, and let it run its course. The program must be monitored and evaluated to identify lessons and program characteristics which can be extracted and shared with others.

The Logic of Evaluation

Developing an effective, formalized evaluation program at the state level requires proper preparation. Before anything else is done an evaluator must prepare a model, or map, of the state's juvenile justice program plan. This model is laid out in a manner similar to an organizational chart or flowchart, but differs in that it is a systematic diagram outlining factors which define the direction of the state evaluation process. The process of developing such a model is termed mapping by evaluators and the final product is referred to as a system map.

To begin, the map should identify the state's juvenile justice system policies. At the executive level the governor's office and the state advisory group on juvenile justice issues combine to establish state direction and mandates. The state's juvenile justice agency, human service/welfare agency, and corrections agency must work together to apply those policies. The legislature, litigation, and other factors further influence policy direction. The map must account for those directions and mandates.[1]

Identification of state and local goals for service delivery and system improvements is the second factor to be mapped. This aspect

of the process is critical to evaluating program success. The goals developed for the system define the types of data to be collected and the methods to interpret that data. Goal development is the focus of a lengthier discussion below.

Third, the targets of the state strategy should be identified preferably through problem definition and goal development. Those targets may consist of a special client population (e.g., status offenders in need of improved services), a part of the system itself (e.g., an improved monitoring capacity or a service provision organization), or legislation.

Finally, the organizations and individuals responsible for implementing the strategy, and the methods for effecting change, must also be identified.

This brief description of a justice system map may sound familiar to state juvenile justice specialists. The map is essentially a diagrammatic portrayal of the three year state plan and annual plan update submitted to OJJDP as a requirement of participating in the JJDP Act. The format is different, but the groundwork for an evaluation system map has been laid with the development of the annual state plan. A generic illustration of a mapped state system is provided in Exhibit A. Exhibit B illustrates how such an map might look for a hypothetical state. The map should guide all evaluation activities, at least in a general manner. It is also a useful tool for explaining evaluation activities and for insuring that the process is focused and productive.

The goal setting described as a part of the map is integral to overall evaluation program success, since it is the procedure that leads to the desired outcome. Once the goals have been further defined into specific objectives, they become the criteria against which successes can be measured.

The difference between goals and objectives is actually quite distinct. A goal is a broad statement of anticipated accomplishment. The statement, "To reduce reliance on secure detention" is an example of a broad program goal. This goal could be refined further into a series of measurable objectives. "To decrease the use of secure detention by 50 percent" or "to increase the

use of nonsecure alternatives for the supervision of juveniles by 80 percent" are but two examples of specific measurable objectives.

Goal establishment implies that persons involved in a program or system recognize that shortcomings exist and steps must be taken to rectify them. Such recognition is usually a product of needs assessment conducted to locate deficiencies. (The development of a needs assessment is beyond the scope of this monograph. There are many excellent references which detail the specifics of the process. See Chapter Six). However, it should be remembered that the annual state plan alluded to above is a type of needs assessment and can be a critical tool for problem identification.

The process for goal development can be as sophisticated as time allows. The ideal would be to interview or survey all groups of system professionals to identify their objectives. The data obtained could be ranked and organized according to issues of greatest concern, and thus establish priorities.

At the other end of the spectrum, the evaluator can simply examine existing plans or needs assessments to identify goals and establish priorities. This goal definition procedure saves time, but it does not allow for detailed input from all persons or groups affected.

Goal development is important not only to define what officials expect to occur through a program initiative, but also to offer the applicant, grantee, or evaluator some definition of what a program is intended to accomplish.

Both the State (represented in the Formula Grants Program by the Juvenile Justice Specialist) and the grantee (local program administrator) have responsibilities in the evaluation process. For the state, it is the clear delineation of program goals--what it hopes to accomplish by funding an initiative. For the grantee, it is to ensure that the program moves toward those goals and that sincere attempts are made to achieve the desired effects. The evaluation process becomes a central tool for assessing effects, so the understanding of program

Exhibit A
Generic Design for Mapping a Juvenile Justice Strategy

AREA :	Title 1	Title 2	Title 3	Title 4	Title n

(Each area should be listed by title; e.g. Legislation, Corrections, etc.)

PROGRAMS :	Program A	Program A	Program A	Program A	Program A
	Program B	Program B	Program B	Program B	Program B
	.	.	.	.	.

(Each program within a strategy should be listed)

PROGRAM :	GOAL 1	GOAL 1	GOAL 1	GOAL 1	GOAL 1
GOALS & :	GOAL 2	GOAL 2	GOAL 2	GOAL 2	GOAL 2
OBJECTIVES:	.	.	.	.	.

(All goals should be listed for each program area)

	Objective a	Objective a	Objective a	Objective a	Objective a
	Objective b	Objective b	Objective b	Objective b	Objective b
	.	.	.	.	.

(All objectives should be listed for each program area)

TARGETS :	Target a	Target a	Target a	Target a	Target a
	Target b	Target b	Target b	Target b	Target b
	.	.	.	.	.

*(Each target should be listed. Example targets might be Legislators,
Youth programs, JJ staff, and others, depending on the strategy)*

AGENTS :	Agent a	Agent a	Agent a	Agent a	Agent a
	Agent b	Agent b	Agent b	Agent b	Agent b
	.	.	.	.	.

*(Each agent involved should be listed. Agents are those responsible
for fulfilling project goals and objectives. Examples include the
juvenile justice staff, consultants, SAG committees, etc.)*

METHODS :	Method a	Method a	Method a	Method a	Method a
	Method b	Method b	Method b	Method b	Method b
	.	.	.	.	.

*(Each method to be employed by the agents to reach goals and objectives
should be listed.)*

Exhibit B
Hypothetical State Juvenile Justice Strategy Map
Legislative Strategy

PROGRAM	(A) Create Monitoring Inspection Unit	(B) Create a Law Related Education Program
GOALS AND OBJECTIVES	(G1) Create a monitoring inspection unit which will inspect all public and private residential facilities. (G2) Establish an annual appropriation.	(G1) Establish a LRE program within the Department of Youth Services. (G2) Set an appropriation.
	(Obj. 1) Place unit in Department of Corrections so that enforcement mechanism can be built in automatically. (Obj. 2) Establish with 1 director, 1 assistant, 4 staff. (Obj. 3) Seek and obtain judicial, legislative, and executive endorsement prior to committee votes. (Obj. 4) Authorize as inspection arm of Department of Corrections. (Obj. 5) Establish a facility inspections universe and definitions of which facilities to inspect. (Obj. 6) $350,000 annual appropriation.	(Obj. 1) Set program mission statement focused upon delinquency prevention and education of young students. (Obj. 2) Establish a funding level of $200,000. (Obj. 3) Direct DYS to administer and design. Money will provide for 1 director, 2 staff, transportation and printing expenses. (Obj. 4) Establish additional $100,000 appropriation for an evaluation component designed to examine program success.
TARGETS	● J. Stevens--House ● R. Commons--Senate ● Appropriations Subcommittee ● Justice Subcommittee	● House Justice Committee ● Senate Education Subcommittee ● L. Mimmeo--House ● R. Sandberg--Senate
AGENTS	● Juvenile Justice Staff ● Department of Corrections Staff and Director ● SAG Monitoring Subcommittee	● Juvenile Justice Staff ● SAG Delinquency Prevention ● Subcommittee
METHODS	● Solicit/educate Governor ● Visit legislators ● Get on judges conference agenda to explain unit's duties	● Get on state educator's conference in February ● Solicit endorsement of State Chiefs of Police and Sheriffs Associations ● Meet with Subcommittee members

Note: This is an example of how a state staff might prepare a justice system map for an upcoming legislative effort.

goals between the State and the grantee must be developed early in the funding and development process.

To help guarantee that the grantee and state expectations are identical, the Request for Proposal (RFP) announcement should delineate the goals, and establish what is to be accomplished by the program (the evaluable goals). Data collection and evaluation expectations can be portrayed at this process. In a nutshell, the role of the RFP is critical, since it can be used to establish measurable expectations. Because of the importance of RFP's for evaluation, their development is reviewed in greater detail in Chapter Five.

So far we have addressed why one may wish to evaluate, the importance of mapping evaluation programs, and the need to clearly develop goals and objectives. Once the decision is made to evaluate and a map is created, it is time to begin thinking about the approach to be taken.

NOTES

1. The problems which interfere with accomplishing those goals should be considered in the mapping process. Resource allocation problems, service delivery problems, overlapping of responsibilities between agencies, limited data collection efforts, and conflicting goals between state and local agencies are all examples of very real problems which can impede the pursuit of overall system improvements. Often the mapping process will help identify system or planning problems. For example, a map may make it clear that separate programs are performing (or claiming to perform) very similar tasks, but they are not coordinated. Overlaps or gaps in services delivery become clearer when they are mapped as we suggest.

3

Introduction

Any approach to juvenile justice program evaluation will depend on a number of important factors which the specialist must consider.

- A program (or programs) must be selected for evaluation using some priority or sifting criteria.

- Some thought must be given to the time factor--weighing the length of the programs operations against the time it will take to evaluate and observe the program impact or outcome.

- The scope of the evaluation must be defined: what will the evaluation cover in terms of program components, information gathering, and level of evaluation?

- The audience for the evaluation must be considered, and corresponding attention given to what the product (or products) of the evaluation will be.

Once these issues are seriously considered by the specialist and evaluator, the next steps involve the details of evaluation design, data collection, measurement, and analysis. These are covered in Chapter Four. Our concern here is to define and review general issues on approaching, or moving toward, evaluation. A number of concepts and suggestions are provided, and a typology of evaluation types

Approaches to Evaluation

is defined using a Law Related Education project as an illustrative example. After finishing this chapter, the reader will see there are numerous evaluation options, and that there is always room for evaluation in a juvenile justice program plan. The real issue is to decide to what extent evaluation activities are appropriate and can be realistically carried out. The specialist can make those decisions confidently by resolving these few general issues.

Choosing Programs for Evaluation

Assume that you have identified juvenile justice programs of various types--public awareness programs, legislative initiatives, outreach, and intervention or treatment programs. Having decided that you wish to conduct evaluation, and presuming that not all programs can or will be evaluated, the problem becomes that of deciding which program(s) to evaluate. There are practical and political considerations in this selection process, as the following series of questions shows:

- *Which programs are accessible for evaluation?* As the specialist or evaluator, you will find that not all of the programs are accessible for evaluation, for a variety of reasons. Some may be geographically distant, or cover multiple jurisdictions, and so prohibit evaluation because of travel costs and logistics. (In the case of a multi-site program, one or a few sites may be selected for evaluation.) Some programs may even be mobile! Some program

administrators may find evaluation threatening or distasteful and deny access or, if not denying access, make it difficult.

- *Which programs will, or do you hope will, stay in operation for some time into the future?* Since evaluation takes time, and the impact of some programs takes time to be felt, seriously consider the length of time you expect a program candidate for evaluation to be in operation. A program that you know will be short lived, and perhaps may have limited impact, should not be as likely a candidate for evaluation as one that will be in operation for more than a year or two.

- *How expensive are the programs?* More expensive programs will usually justify, or warrant, evaluation, since the grantees and taxpayers have a right to know how their money is being spent. Generally, there will be much more demand for information about an expensive program than for an inexpensive one. On the other hand, the more expensive programs tend to be expensive to evaluate, and sufficient funds may not be available for a thorough job. In almost every case, though, the demand for information overrides the danger of evaluating a program only halfway.

- *Which programs are controversial?* Programs may be controversial in a positive or negative fashion. They

may be new, and generate excitement in the juvenile and criminal justice communities and, for this reason, receive attention in the field and from the media. They may receive bad press, due to politics or sensational cases. In some instances, these circumstances will necessitate evaluation, and in others they will work against it. In any case, controversy affects the evaluation decision.

A program needn't be controversial to be a good candidate for evaluation, however. Some programs may have been operating for years under the assumption that they are working, at least as intended. Such assumptions should always be questioned, and evaluated if possible.

- *Which programs are identified as priorities either by your office or other significant officials?* There are times when the demand for evaluation of a program does not fit into practical considerations, but will be strong enough to win out. For example, evaluation may be mandated by law, executive order, memorandum of agreement, or some other official source. Similar programs in other states may have been evaluated with inconclusive results, and practitioners may turn to you for more data.

Consideration of these issues, program by program and for the program plan in general, can provide ideas for which programs you wish to evaluate. Combining such factors as accessibility, practicality, and the need to know evaluation information will push some programs to the forefront. But there is more thinking to be done. The next two sections consider two issues in detail--the time frame and the scope of evaluation. Thinking along these lines will further focus your evaluation plans.

The Time Frame for Evaluation

Time is a critical aspect of any evaluation plan, and it plays a role in a number of different respects. First and foremost,

perhaps, is the decisionmaker's (be they politicians, funders, program administrators, clients or their parents, judges, probation officers, or any other person whose job decisions are related to juvenile justice programs) need to know evaluation information--how the program works and how it is doing. It is almost banal to state that such information is never timely enough, which makes evaluation efforts meaningless, but that is not really true.

A well conceived and implemented evaluation project will be of value to program executives, administrators, and evaluators. Any program information gathered and presented in an objective manner will be of some value. Quite often, such information is available in the absence of formal evaluation research. Archives, internal program reviews, client summaries, and other records provide the information; the task is one of locating, organizing, and presenting it.

Evaluation information is most useful when it is timely. As you will see in our discussions of outcome measures and levels of evaluation, more often than not timely evaluation information is available (or can be obtained in a timely and cost effective manner), though it may not be the most timely, the most objective, and the most scientifically rigorous data. In some instances, however, even an intensive, expensive evaluation effort will not produce the necessary information on time. In those instances, the specialist/evaluator may decide not to evaluate.

Closely related to the issue of the timeliness is the time span of the program under consideration. How long will it be in operation? If only for a year or less, evaluation (especially if it is costly) should be given a low priority. How long will the intervention or treatment take to administer, and how long will it take to observe its effects? Some programs administer their treatments (education, training, therapy, exposure to various stimuli) in small doses over long periods of time, while others provide heavy doses in short time spans. Neither case should automatically rule out evaluation, but different time frames for administration of treatment will certainly suggest different approaches to evaluation.

In the same vein, different programs will expect to produce impact (behavior changes, attitude changes, test scores) in various time frames; or one program may produce different impacts over time. These considerations, too, will affect evaluation plans and methods.

Consideration of a program's history may affect your evaluation decisions. If a program has been in operation for many years and has been ignored by evaluators (it may be unexciting, it may be controversial and protected, it may be difficult to evaluate), the time may be ripe to approach the subject. It is conceivable also that even a thorough evaluation will not change much in the way of program operations.

A brand new program, on the other hand, may be too young for evaluation. It may be in a learning phase, or a state of flux, making evaluation difficult. Under some circumstances this situation might call for evaluation, but generally it would not. You might not consider a young program important enough to expend scarce evaluation resources, if you are not sure of its future. On the other hand, evaluation efforts, even at a low level, should start early on in a promising program to generate valuable information for later, more comprehensive, efforts. As you can see, there are no easy answers. If you understand that consideration of program history is important in making evaluation decisions, you are thinking appropriately, even if you don't have all the answers!

Scope of the Evaluation

The critical question when thinking about the scope of an evaluation effort is: "What do I want to learn?" Answering this question will make your decisions easier and will help make choices regarding data and methods. It is not a simple question. As with other questions we have reviewed, the answer depends on practical and political considerations, and the trick is to find a reasonable course of action so you can get on with the business of evaluating.

Your total program plan probably includes many programs of different kinds, funded at different levels, with varying goals and methods. Presumably you will not attempt to evaluate them all unless you have vast resources. The scope of your evaluation, then, will be smaller, maybe a few isolated evaluation efforts, maybe a coordinated evaluation of similar programs, or perhaps you will decide to evaluate only one program.

You may also choose to evaluate aspects of one or more large program. A treatment program might include multiple facets or components--therapy, training, community activities--and you could decide to focus on only the training or therapy components, or you may choose to evaluate the training components of a number of programs.

Again, you must ask, "What do I, or what does my evaluation audience, *really* want to know about?" The legislature or state educational association may want to know about all of your training efforts. Your criminal justice constituency may only ask about a particular program's recidivism or failure rate, while you may feel there is more to be considered.

Evaluation need not always consider a single, total program. It also may cover more than one program, a component or components of a single program, a single component of many programs, and so on.

There is an important distinction in the evaluation literature that bears reviewing here, though it will be covered further on: program monitoring versus process and outcome evaluations. The following definitions for these concepts are offered:

Program Monitoring:

Developing and analyzing data for the purpose of counting specific program activities and operations.

Process Evaluation:

Developing and analyzing data to assess program processes and procedures; to assess the connections between various program activities.

Outcome Evaluation:

Developing and analyzing data to assess program impact and effectiveness.

These definitions lend a false simplicity to these concepts, but provide the correct impression that evaluation activities can be distinguished by levels of complexity, difficulty, and cost. In reality, most evaluations comprise some of each of these activities.

When thinking about what you want to learn through evaluation, think in the context of monitoring, process, and outcome evaluation. Evaluations simply cannot proceed without monitoring information, which means answers to such questions as:

- How many different activities does the program operate?

- How many staff, or service deliverers, perform those activities?

- How much work do they do--hours or days spent on activities, clients served, contacts made, administrative duties, etc?

- How do these things change over time?

The volume of work done must be counted, sometimes in very detailed fashion. This is the general nature of program monitoring and it must be done if other evaluation activities are to take place.

Monitoring is, by itself, an evaluation activity. The process will yield information to answer a question such as, "How, or what, is the program doing?" Activity levels can be compared to goals and objectives, and monitoring them over time can provide important feedback to program staff, clients, administrators, and funders.

As soon as you begin asking questions about the relationship between different activity levels, or the sequence of activities and systemic issues (how the activities are related in program procedures), you enter the realm of process evaluation. Sometimes referred to as "formative evaluation," process evaluation is concerned with providing feedback to staff and management to help avoid problems and adapt to changes in the program's internal or external environment.

In monitoring, you may keep track of the number of incoming clients, the staff workloads, and the provision of services to clients. Process evaluation takes these data a step further by analyzing the effect of trends in new clients on existing caseloads (and perhaps the external processes that are affecting program referrals), and on the time required to provide services. You are really building a model of program operations-- identifying the relevant variables and measuring them, and then analyzing their interrelationships. This must be accomplished in some fashion to make the move to outcome evaluation.

An outcome evaluation assesses the success or effectiveness of a program or program component. Having achieved an analytical understanding of how a program operates, through monitoring and process evaluation, the next step is to assess program products, or outcomes. Consider a simple example involving a training program.

The outcome evaluation issues concern how the intended training was provided, the extent to which program activities deviated from the original design, and if the desired effects were achieved (better grades, higher self esteem, better employment, less involvement in crime, etc.). Outcome evaluations may address efficiency or cost-effectiveness issues. They may also uncover unanticipated outcomes.

Conducting outcome evaluation requires adequate monitoring and process evaluation, for you cannot be sure an outcome was achieved by a program unless you can demonstrate a link between program activities (process) and results (outcome).

Varieties of Outcome Measures

In reality, outcome is not an "end process" issue; that is, outcome evaluations need not, and probably should not, be concerned only with what occurs at the end point of a particular process. All program activities have outcomes, whether intended or not. Some are quantifiable and some are not. Some are easily observable and comprehensible and some are not. Additionally, all different types of programs have outcomes, and these same issues apply as much to legislative initiatives as they do to

direct service provision programs. In this section, we review types of outcomes to dramatize this important point. It is important to think of outcome issues in a more comprehensive context. This will enhance evaluation activities generally, and improve the information that goes to policymakers.

Knowledge Production Outcomes

In many instances, for particular programs or for program plans in general, a major goal is the production of knowledge about juvenile justice issues such as juvenile crime, effective prevention and treatment strategies, current legislation, available services, and so on. Evaluation of knowledge production efforts, or identification of knowledge production outcomes, generally receive low priority from evaluators. The frequent assumption is that more instrumental outcome measures--criminal behavior, test scores, and the like--are more desirable. This is not necessarily an appropriate assumption. Knowledge production is a stated goal in the federal Juvenile Justice Delinquency Prevention law. Acceptance of programs at state and local levels depends on the availability of information about the programs and acceptance of them, goals that cannot be achieved without the production of new knowledge in the general and criminal justice communities.

There are ways to monitor and evaluate knowledge production, and they are addressed later in this monograph. The important point here is that knowledge production be considered a valid process and outcome, worthy of evaluation from the start.

Consensus Building Outcomes

A similar argument applies in the case of consensus building outcomes--the need for production of common understanding and efforts among various constituents in the juvenile justice arena, especially where they may not have existed before. Attaining instrumental goals (see below) often depends on significant consensus building around an issue. Jail removal and de-institutionalization are two primary examples. Such programs cannot guarantee success even if they are well-conceived, well-managed, and adequately

funded; they also must enjoy the support of various components of the state and local criminal justice and general communities.

Evaluation efforts tend to ignore this consideration. Consensus building efforts and achievements do not lend themselves to measurement and scientific analysis. However, they provide significant and valid qualitative, or contextual, information, which are worthy of more serious consideration in evaluation efforts. Production of information about the political and social psychological aspects of program implementation would be invaluable.

Instrumental Outcomes

These are the most commonly discussed outcomes in evaluation research. Being directly or indirectly related to a funded program's goals and objectives they, if observed and measured properly, will indicate the program's level of success or effectiveness. Typical areas of measurement include recidivism, educational attainment, self-esteem, and community values or citizenship. These are usually given the highest priority by evaluators and decisionmakers, and for good reason. If a program is not producing the promised results, and evaluation confirms this, then it is time to reconsider program goals, objectives, and methods. Instrumental outcomes are important. They are even more valuable if they are presented with evaluation information about knowledge and consensus building, where appropriate, to give decisionmakers the maximum amount of useful information, in the proper doses.

A Typology of Evaluation Levels

This chapter has presented many concepts and ideas for the specialist/evaluator planning evaluation. To reinforce these issues this chapter offers a typology of evaluation levels using an actual juvenile justice program example--a Law Related Education program implemented and evaluated in Colorado.[1] The multitude of issues and questions raised is not intended to confuse the reader, but to convey the notion that there are many options for evaluating, and many good reasons for making the decision to evaluate. Once the reader accepts the following:

- That evaluation data either do exist, or can be developed,

- That an evaluation effort need not be exhaustive to be useful, and

- That any evaluation information, if logically and carefully developed, will provide decisionmakers with more information than they would have had otherwise, even if it does not provide every single piece of useful information,

then the decision to evaluate will be readily made. This section will demonstrate even more clearly the various types of possible, and useful, evaluations.

Dimensions of the Typology

The typology of evaluation levels relies on the distinction drawn between monitoring, process, and outcome evaluation, and on the level of comparison you intend to achieve, or are able to achieve, given data and resource limitations. Generally, evaluations will be program, or program component, specific, or they will make use of comparisons with other programs or client groups. Combining three levels of evaluation with two general comparative perspectives reveals six evaluation types, as Exhibit C illustrates. As this typology unfolds below, it will become clear that each of these evaluation types has a useful purpose. Choosing one or the other is not a right versus wrong issue. In some instances the resources available and the demand for information will dictate that no more than a basic evaluation should be attempted. In other instances, a basic or comparative process evaluation will satisfy decisionmakers. It is the rare evaluation effort that either can support or, if the financial resources are available, can achieve a true comparative outcome evaluation. Comparative outcome evaluations should be used, however, for the most critical, long-term problems faced by juvenile justice, and for the most promising strategies for addressing those problems.

It is important that the range of possibilities be given serious consideration as you make evaluation plans.

The Law Related Education Program

This program will be used to explain the evaluation typology presented above. It is important to understand that this example--a Law Related Education program--was actually implemented. The evaluation type employed was a comparative outcome evaluation. In this section, five hypothetical evaluations are described to define the other components in the typology, and the comparative outcome evaluation is described from the information produced by the program.

With any evaluation effort it is important to understand a program's goals, objectives, and operations before selecting an evaluation approach and methods. These are reviewed here for the Law Related Education (LRE) program.

LRE Goal:

Provide instruction to students to build a conceptual and practical understanding of law, enforcement, and judicial processes, leading to improved citizenship skills, a desire to work within the legal system to settle grievances and deal with criminal problems, an understanding of the basis for rules, and favorable attitudes towards enforcement and justice.

LRE Objectives:

Provide 30 to 40 semester hours of LRE to school age children (middle to junior high school age) in seven schools.

LRE Procedures:

Each of the seven schools selected a teaching team that was trained in the LRE curriculum. Throughout a semester the team taught law related topics including mock judicial procedures, and utilized legal and law enforcement professionals in class exercises, visits to courts, rides in patrol cars, and home security audits.

LRE Rationale:

The educational activities offered as the LRE program are expected to increase understanding of law, enforcement, and judicial processes because the standard school curricula do not cover such topics.

Exhibit C
AN EVALUATION TYPOLOGY COMBINING LEVELS OF EVALUATION
AND COMPARATIVE PERSPECTIVES

COMPARATIVE PERSPECTIVE

LEVEL OF EVALUATION	Program Only	Program and Comparison
Monitoring	Basic Monitoring	Comparative Monitoring
Process Evaluation	Basic Process Evaluation	Comparative Process Evaluation
Outcome Evaluation	Basic Outcome Evaluation	Comparative Outcome Evaluation

The new educational material will challenge perceptions of these phenomena among school children that are based on television portrayal and popular perceptions among peers. If it is carefully and thoughtfully presented, the LRE curriculum will change these conceptions among the students and foster respect for law abiding behavior.

Basic Monitoring

A basic monitoring evaluation is concerned with answering simple questions about program activities and rationale. A good way to approach the problem is to ask

Monitoring is the process of developing and analyzing data to count and/or identify specific program activities and operations.

the following question: "Who is doing what, when, where, and how often and with what resources?" For the LRE program the following answers are typical:

- Teachers are providing instruction to students, the targets of their

activities, with the aid of law enforcement and judicial professionals.

- The two basic activities are teaching and learning.

- These activities take place at certain times over a school semester, typically one or two classes a week.

- These activities take place in the school setting, and at various other outside locations, such as courtrooms and police stations.

Answering the "How often" question might entail collecting data in the following areas:

- The number of students receiving the instruction:

  -- total number of students taught, and subtotals for various student types--age, sex, ethnicity, family characteristics,

  -- number of students per class, or average number if there is variation, or

-- absences per class.

- Teacher data, including the total number and types of teachers involved.

- Teacher/student ratio.

- Hours of instruction offered, by different teachers, and by course component, e.g., lecturing versus field trips, or school teacher versus other professional.

- Adherence to, or deviation from, the established teaching schedule.

- Program cost information, including salaries, travel, supplies, special materials, and the like.

- Other descriptive program information, such as course plans and outlines, learning objectives, and teacher training.

Basic monitoring information for a program has the following utilities:

- It provides operating and descriptive information for funders and other overseers of the program, for interested outsiders (students, parents, other schools), and most importantly for program staff and administrators. It answers such questions as "What does the LRE program do?"

- It provides the basic information necessary for comparing achievements to stated or desired objectives. If the goal to provide 40 hours of instruction in a school semester was missed by 5 hours, the answer may be found in the scheduling or other descriptive types of information. Such an inquiry may lead to a refinement of the goal, or to scheduling adjustments to avoid the problem in the future. The 5-hour shortfall, a piece of basic monitoring information, may highlight a significant operational problem and lead to its solution.

- It provides educational information for all involved in the program. Evaluations often will be undertaken for new programs, and the program participants or administrators will learn more about the process and impact of the program if monitoring data are available than they would in the absence of such data. They are usually glad to have the opportunity to review them, even if few surprises result.

- Basic monitoring data serve a preventive maintenance function, by tracking indicators of critical elements which, if they deviate too much from the expected norm, signal a program problem. For example, if absentee or student turnover rates reach a high level some immediate action may be required to stabilize the program.

Most program leaders or administrators collect this basic information, or at least some subset of it. Collecting the data, if it is done carefully and reviewed periodically, is evaluation of a basic sort. Comparisons are made between expectations and observed results, or at least data are relied on to set expectations. This is evaluation activity.

It is, of course, not sufficient to simply collect and analyze program data to complete a successful monitoring program. The data and findings should be integrated into the decisionmaking process at the program and/or higher levels.

Comparative Monitoring

In a comparative evaluation effort, basic monitoring data are collected for other similar programs, or for subjects in control groups which do not receive the intervention but are monitored for comparison purposes. The LRE program used both types of comparison in its evaluation. LRE programs were implemented and monitored in seven different schools, and in five of the seven control groups, consisting of students randomly assigned to traditional civics or social science classes, were monitored. In this manner, comparisons were made for schools that implemented an LRE program under slightly different circumstances, reflecting slight variations across schools in

student ages and racial mixtures, and also within schools that did and did not receive the LRE program.

The value of comparative monitoring information lies in the comparative perspective it provides. With basic monitoring information, the only comparison possible is internal to the program, i.e., comparison with program goals and objectives. Comparative monitoring allows such comparisons, but also allows comparing performance with other programs. For example, if one program attains 90% of its planned instruction hours, while other programs achieve more (95%-100%) or less (70%-85%), more has been learned than a simple internal check against program objectives.

More important, however, is the confidence in interpreting findings that the comparative approach provides. When describing and evaluating programs, especially when short or long term outcomes are discussed, it is valuable to consider alternative possible explanations of your findings. It is important to know whether an increase in clients, a change in client behavior, or a response to a program initiative is really the product of the program under study or some other factor, such as increased arrests, client education level, or outside influences. The comparative perspective brings more information to the analyst, and allows control or analysis of factors outside the program. In this manner it increases the ability to distinguish program effects from other influences, and thus gives the evaluator more confidence. Chapter Four addresses this issue in more detail under the section, "Threats to Validity."

Basic Process Evaluation

If the planned outcome for the LRE program consists of certain attitudes and behaviors among the students, or changes in attitudes and behaviors, then there must be some process by which the program activities, as measured by monitoring, produce the expected outcome. By considering the program activities in combination or in some sequence, and by considering the mechanism by which the activities produce the result, you enter the realm of process evaluation.

For example, qualitative and quantitative measures of student-teacher relations or interactions might provide a process measure that is more predictive of program success than simply counting the number of hours spent in class. The effect of time on the program, as indicated by turnover or absenteeism, is another valuable process measure. Qualitative assessments of the program's link with other school activities,

Process evaluation involves developing and analyzing data to assess program processes and procedures, esp., determining the connections between various program activities.

may also prove valuable. For example, did the field trips interfere with other classes or extra-curricular activities, or of the program's acceptance by the school administration. Test and quiz grades may also be good interim measures of program success.

It is these kinds of information that help explain how the various program activities operate together. They produce short-term outcome measures which, if positive, can be expected to produce positive results in the long run as well. Basic process evaluation is valuable for other reasons, including:

● It enables a better understanding of monitoring issues. In examining process, you achieve a better understanding of the formal and informal processes that operate within a program.

● Process information provides important feedback that may allow interim adjustments before a program goes too far in the wrong direction.

● For programs that do not undergo long-term outcome evaluations, process evaluations provide assessments of short-term performance.

Comparative Process Evaluation

Comparative process evaluation employs the same measurements and assessments used

in basic process evaluation, but for comparable or control programs. When comparisons are used for process evaluation, the benefit for evaluators is even greater. Principally, the comparative perspective at the process level provides much more confidence in the findings because the number of cases (programs, or students within programs) increases, and because information from different programs introduces different perspectives and controls into the evaluation.

Consider the LRE program. Process evaluations were conducted in six of the seven schools (one school received minimal administrative support). The LRE programs were similar in each school, but the schools varied in their student types--the grade level ranged from junior high to middle school, and one was a multi-level grade school; one was more racially mixed than the others. The schools also varied in their implementations of the LRE program. This produced variations in quantitative and qualitative information that broadened the entire project's understanding of LRE and its potential, more than a single case study would have done.

Additionally, comparative process measures provide relative measures; that is, short-term performance measures that can be compared with measures from other programs. Relative measures permit comparisons of marginal performance differences across programs, and also allow evaluators to address a variety of policy-related questions--What if we taught more hours? What if there was a diversity of students in the class? What if fewer field trips were taken? These questions can be answered with more confidence when enough cases are present to produce variation in the variables of interest.

Basic Outcome Evaluation

With basic outcome evaluation, the logical sequence from program activities, to program processes, to program outcomes is made for a single program. Such analysis cannot be attempted without the antecedent monitoring and process evaluation that outcome evaluation suggests. In the case of the LRE program, two outcome measures were chosen, and they were taken before and

after the implementation of the program. The measures were:

(1) Student scores on scales measuring correlates of law abiding behavior.

(2) A self-report survey on various criminal and delinquent behaviors.

For this evaluation within a single school, program success is defined as whether, or to what extent, student attitudes and behaviors changed in the expected directions, as determined by comparisons of time series data. Do the scores and other

Outcome evaluation involves developing and analyzing data to assess program impact and effectiveness.

indicators change in expected directions following (and perhaps during) the LRE program? Are these short-term outcomes, measured soon after classes ended, predictive of longer-term behavior, which might be measured by follow up studies?

Other approaches to basic outcome evaluation might include comparisons within a program. Such activities might involve comparing LRE curriculum to other curricula, or comparing different LRE curricula, predicting "high risk" students at the outset of the program and focusing follow-up efforts on them, studying early and intermediate indicators of successful outcome, or measuring outcome at various points during the program.

The value of a basic outcome evaluation such as this lies in providing the best information possible about program performance. In a single school, with this evaluation design, a finding that LRE students scored about the same or worse on post-program measures in comparison to pre-program measures would have hurt the overall program. If nothing else, the findings would have stimulated reconsideration of the program's goals and procedures. There would have been no information showing that it made a difference. In this case, though, differences were observed in the expected directions. Had the outcome

findings been inconclusive the evaluators would turn to the process and monitoring data to explore the reasons, and probably would have found some helpful clues. The LRE program did just that and found other benefits such as favorable feedback from parents, and improvements in police officer handling of juveniles.

A basic outcome evaluation that uses comparisons within a program, as the LRE project did, allows the researcher and program administrator to address the question "Did the program make a difference?" While it doesn't explain what would happen if the program was not implemented, it provides information regarding program impact and program effects.

Comparative Outcome Evaluation

In a comparative outcome evaluation, long-term outcome measures are collected for more than one program, usually for the program under inquiry and a control group of programs, but they may be collected for multiple programs and control groups. This was the research design for the LRE program evaluation. The outcome measures described above were collected for LRE students and control student groups, before and after the program was implemented in five different schools. With the exception of collecting even longer-term measures, e.g., follow-up examination of attitudes and criminal behaviors after one or more years, replicating the research design for one program over multiple programs provides valuable evaluation information. This is especially true when comprehensive monitoring and process evaluation data have been collected in the course of program implementation.

The benefits of comparative outcome evaluation often include all of the benefits of lower levels of evaluation since those kinds of information are necessary to support it. The benefits of comparative outcome evaluation also include:

- Increased confidence in the evaluation findings regarding any single program, due to the benefit of controlled program comparisons.

- A better understanding of the possible range and varieties of outcomes that an LRE program might produce. This is especially helpful for those interested in replicating the program.

Comparative outcome evaluations, then, deserve the highest degree of confidence, especially if a pre/post comparison and a comparison with other controls is employed. Often it will not be possible to design and implement a thorough comparative outcome evaluation. Pre/post only designs, or comparison with controls only, will provide evaluators with good information regarding program performance.

Summary

This chapter has reviewed various approaches to evaluation. Before actual research planning takes place, a number of issues must be considered to help focus the evaluation and to prepare the research design. Good program candidates for evaluation should be identified, since all programs can be evaluated but available resources will not allow it. The program issues of accessibility, length of operation, history, expense, nature of controversy surrounding it, and external pressures to evaluate should be considered in making the selection.

The scope of evaluation is important, too. It can be narrow or broad for both research design and resource considerations. Evaluations can be simple and low-level monitoring or complex and high-level comparative outcome evaluations. While more valuable information is gained at the higher levels, each of the six evaluation levels described provides useful information for various audiences.

If you give careful consideration to these and other issues you will usually find that (1) evaluation is not as difficult or esoteric as it seems, (2) you have been doing it in some fashion already and may as well take credit for the good work, (3) that there may be a broader audience for the information being produced, especially if the manner and format of presentation are adjusted a bit, and (4) providing objective data about juvenile

justice programs will be appreciated by many in the field.

Now, having decided that evaluation can be accomplished, you face decisions about how to conduct them. The next chapter will review the basics of evaluation design and other relevant research issues.

NOTES

1. The evaluation examples presented in this section were derived from a report entitled, "Using School-Based Programs to Improve Students' Citizenship in Colorado," by Grant Johnson and Robert M. Hunter of the Action Research Project at the University of Colorado. It was published by the Colorado Juvenile Justice and Delinquency Prevention Council in October of 1987. Their permission to use their materials is gratefully acknowledged.

4

Introduction

In this chapter, some of the technical and research design-oriented aspects of evaluation are addressed. In this context, "critical issues" means concepts and problem areas you should understand well enough to distinguish between a good or bad evaluation, or evaluation proposal. There also may be ideas or approaches to evaluation research that are new to you, or on which you might need refreshing. Therefore, a presentation on evaluation uses concludes this chapter.

Prior to concluding, we will discuss five critical issue areas in evaluation research, based on the experiences of seasoned program evaluators. They are:

(1) Measurement--measures of juvenile justice program input, intermediate effects, and outcome.

(2) Validity--definitions of, and threats to, internal and external validity.

(3) Special Topics in Program Evaluation--including the varieties of data sources to consider, random assignment in evaluation research design, cost benefit analysis, and the use of observations in evaluation research.

(4) Common Errors in Evaluation--pitfalls in interpreting data.

Critical Issues

(5) Creativity in Evaluation--a section that stresses the need for resourcefulness and ingenuity in conducting evaluations.

Throughout the presentation of these issues, please remember the overall role and purpose of evaluation. It is common to assume that evaluation can address questions such as "Does the program work?", "Has the project been successful?", and "Should the project be continued?" However, it is inappropriate to expect an evaluation, and the evaluator, to address such questions. They represent issues which involve value judgments requiring project and policy worthiness, both of which are beyond the role of the evaluation. Determining project success and continuation involves examinations of resources, priorities, and politics, a process outside the evaluator's task. The role of the evaluation is to provide objective information concerning project activities and their outcomes to program administrators, policy makers, and funding agencies who will make the determination of the worthiness of the project and decide its future.

Measurement Issues

Perhaps there is no more critical issue in evaluation than defining and measuring the variables to be used. The validity of a study depends on appropriate measures of project activities and outcomes. Indeed, the final judgment of the program may depend upon how the program operations are conceptualized and measured. The choice of measurement, and design of the evaluation

will, to a large degree, determine if the evaluation is to be believed by evaluation consumers.

There are no hard and fast rules in the choice of measurement; what is most critical is that the measures are appropriate for the context for which they are intended. For example, in evaluating a juvenile diversion program, one may want to measure a youth's family relationship as a factor that might influence his or her ability to avoid further contact with the court. Obviously in assessing the impact of a positive peer culture program in a juvenile institution this type of information would be less relevant. Other considerations involve the definition of measures of success. If recidivism is defined as a police contact, a different rate of project success will be obtained than if it is defined as adjudication or incarceration. None of these measures are wrong, they are just measuring different things. Thus it is important to be clear about the meaning of measures chosen for the evaluation.

There are three types or categories of measurement integral to any juvenile justice evaluation: measures of program input, program processes, and program outcomes. The discussion of each of these will frame the remainder of this chapter.

Measurement of Program Input

While the popular conception of evaluation focuses upon program outcomes, remember that there is considerable variation in program inputs which can affect the results of the intervention. Often in juvenile

justice there are broad program types such as diversion, education, and family therapy. While there are commonalties among programs within each of these categories, programs with a similar overall description may involve distinct intervention strategies with dissimilar clients. For example, two diversion programs may be oriented to reducing the level of commitment to juvenile court. However, one may involve diversion at the police level for status offenses, while the other may involve screening at the court or prosecution stage to divert minor property offenders to a restitution program. Before we can say what works it is necessary to say what we are doing.

There are several aspects to measuring program input. First consider how the goals and objectives of the program are translated into practice. What facets are to be emphasized through commitment of resources toward the project objectives? What are the major project activities and how do these relate to the anticipated outcomes? Why and how is the program supposed to work? What are the underlying reasons the intervention is presumed to be effective?

These questions all relate to the theory of the program. Although it is often presumed that theory is irrelevant to juvenile justice practice, nothing could be further from the truth. Indeed, the program theory is a statement of the mechanisms through which the intervention is to work. Most importantly from an evaluation standpoint, it tells us what variables and concepts to measure.

Thus, one of the first tasks of evaluation is to obtain a clear explication of the theory behind the program. What should the program change that will result in reduced delinquency? As a prevention program, is it oriented to improving the individual's self concept, attachments to family, school performance, or opportunities? Each potential area of program emphasis implies a different causal process designed to reduce delinquency. Although program designers and administrators may not always claim to have employed theory in the creation of the program, theory is implicit in all forms of delinquency intervention. It can remain the

evaluator's task to clarify the reasoning behind the intervention.

While it may appear quite straightforward to specify what the program is, e.g., the reduction of probation officer caseload size, specifying the content, what is actually going on, may be more difficult. There may be substantial variation in operational procedures, and among program personnel. The more complex the project and varied the components, the more difficult and crucial this task is.

Why be concerned about theory and program content? After all, isn't the issue to measure the effect and impact of the intervention? True, but if the evaluator doesn't know what the program is, he or she may fail to ask the appropriate questions regarding program impact, the wrong variables may be measured, or appropriate measures be omitted. Most importantly, the evaluator will not be able to attribute changes observed to program components or activities. Since a principal reason for evaluation is to replicate successful programs, it is imperative to know precisely what was done in order that the desired components, procedures, and activities can subsequently be implemented.

Another important virtue for these evaluation input measures is to clarify project activities beyond those presented in the funding application. Often at the time of application the specifics of program operation are not finalized, yet many evaluations use the wording of funding proposals as a reflection of what goes on in the program. If the evaluator simply accepts the program statement as fact, the danger exists of making incorrect statements about its effects. In fact one could wind up evaluating a program that does not actually exist.

For example, although the planners of a juvenile diversion program may have designed an intervention for youths who commit criminal offenses, the staff operating the program might decide that a more appropriate intervention, given staff resources and expertise, is to divert youths who have family problems and are principally status offenders. Without an understanding of such a shift, the evaluator may inappropriately

make conclusions regarding a different form of intervention than actually took place. In the words of Carol Weiss (1972:44), "the evaluator has to discover the reality of the program rather than its illusion".

There are two forms of data which can be considered as input measures: data on the program itself, and data about the program participants. Program specific data would include the purposes of the program taken from program statements as well as staff descriptions, resource allocation, methods of operation, day to day procedures, staffing patterns, location, size of program, management structure, and inter-organizational relationships. Plus every evaluation should carefully document the content, duration, and intensity of treatment involved in the intervention.

The second type of input data concerns characteristics of program clients. This would include demographic and personal characteristics such as age, gender, education, employment, and family economic status. In addition, depending upon their relevance to the program, theoretical significance, or use in prior research, one may also wish to collect data on the attitude and perspective of program participants on a variety of issues that may be related to his or her

There are two forms of data which can be considered as input measures: data on the program itself, and data about the program participants.

performance in the project. These areas might include the youth's relationship to family and peers, attitude about the program, motivation for participation, perception of sanctions and deterrence, social responsibility, and self concept. While the project may be reasonably expected to alter some of these factors, others such as gender, are quite impervious to change. It is useful to collect data on these and other control variables to determine the types of clients who are more likely to be successful in the program.

Measuring Intermediate Program Effects

Beyond an accurate reflection of program inputs and content, a thorough evaluation should contain an analysis of the attainment of mid-range goals. Almost every juvenile justice program contains both mid-range and long range objectives. For example, a juvenile diversion program may have the ultimate objective of keeping youth referred to juvenile court from committing subsequent offenses. But there are probably intermediate steps that are believed to lead to this goal. It may be that those diverted are to make restitution; if so intermediate measurements need to determine if the youth in fact did so. While one could collect outcome data on subsequent offenses, and proclaim the program a success or failure, this process would obviously be in error if one could not ascertain that this intermediate, and presumably causal, step had not taken place.

Although it seems obvious that one cannot say a restitution program was successful unless restitution was made, this type of error is quite common in less obvious situations. A program may involve drug treatment as a method of reducing delinquency. While subsequent delinquency may or may not be affected by program participation, it is imperative to consider the impact that the counseling program has upon drug use independent from delinquent activity. It is certainly conceivable that the program may be effective in reducing drug use even though delinquent activity remains unchanged. On the other hand it may be possible that the program has no effect on drug use, and thus any conclusions relative to the program's impact on delinquency through drug treatment are inappropriate.

These two types of outcomes are referred to as theory failure and program failure. If the program works but the outcome criterion is unchanged, i.e., if drug use is reduced but there is no effect on delinquency, then theory failure has occurred. While the program has achieved the desired intermediate effect, our theory about delinquency being a result of drug use may be flawed. Such a situation would require a reformulation of the theory and restructuring of the intervention.

On the other hand, if the program is not observed to affect the intermediate goal, i.e., if drug use is not affected by program participation, then no conclusions can be made regarding the overall impact of the program on delinquency. The relationship between drug use and delinquency has not been adequately tested. Since drug use has not been altered, any changes in delinquency cannot be attributed to drug use patterns. This is an important distinction because although drug use has not changed, delinquency involvement may have changed. If drug use is not measured, then changes in delinquency may be falsely attributed to the treatment program.

This situation also confirms the necessity for adequate input data. Although the changes in delinquency may not be a result of reduction in drug use, there may be other aspects of the program that have resulted in this change. For example, the establishment of a positive relationship with the counselor may have resulted in delinquency reduction independent of drug use. Having these data can aid in the redesign of the program to focus upon relationships that may be more productive in reducing delinquency.

The consideration of intermediate effects has an additional benefit. The statement of intermediate steps forces a clarification of the project, and forms a conceptual model of the processes through which the effects are presumed to be caused. Such explication not only clarifies what is expected to occur, but may serve as a guide for replication and revision of the project after evaluation results have been obtained.

Measuring Program Outcomes

Measuring outcomes is popularly viewed as the essence of evaluation. In spite of the critical nature of measuring inputs and intermediate effects, program planners and administrators still need to address the question: "Did it work?" As we have observed there is often not a direct answer to this question, and in many cases the most straightforward answer may be "It depends." How success is defined and measured will often determine the degree to which the program is viewed as effective.

At first glance determining the success of juvenile justice programs would not appear to be problematic: program participants either commit new offenses or they don't. Unfortunately, program success is generally not so directly determined.

Rather, juvenile justice success is commonly measured through the concept of recidivism. While this concept has a universal meaning of correctional failure, its operational definition is anything but universal. There are a number of dimensions of the concept that must be specified before a working definition is obtained. First the threshold of recidivism must be established. What are the specific criteria that indicate program failure? Is subsequent police contact sufficient, or is it more appropriate to count arrests? Should there be a formal referral to juvenile court or must there be a formal adjudication to indicate failure? Some may argue that commitment to an institutional program after release is the appropriate measure of recidivism.

Obviously the statistics measuring program outcome will be greatly influenced by this criteria decision. If adjudication is the criteria, then youths who have committed subsequent delinquencies will not be counted as failures unless the system responds with a formal adjudication. On the other hand, if one indicates program success as a lack of police contact, then youths who have not committed subsequent delinquent behaviors could be counted as failures, since they may have contact as a result of being known to the police.

In each of these situations the real outcome of the program is the same, but it will appear much different due to the variation in definition. This difference may be substantial. Waldo and Chiricos (1977) in evaluating a work release program noted that program success may vary from 20% to 70% depending on the definition of recidivism.

In defining recidivism, the evaluator must choose the measure that is most appropriate given the scope and objective of the intervention. Generally, the best measure of recidivism is the one that is closest to the behavior itself, either self-reported delinquency or police contacts. The further

a measure is from the individual's behavior, the more it is measuring the influence of organizational behavior and decision-making rather than commission of delinquency acts. The police decision to refer to court and the court decision regarding adjudication may be influenced by a number of factors other than the youth's behavior. This is most apparent in the use of measures of recidivism involving return to the program. If one is evaluating an institutional program and an area of concern is postinstitutional behavior, the focus should be on the individual's performance in the community, not on return to the institution. The offender's return to the institution may be due to a range of factors unrelated to subsequent behavior or delinquency. Thus, measures of program return should be avoided in recidivism studies.

Another important question in defining recidivism is how serious must delinquent behavior be to constitute failure? In evaluating an intervention program aimed at violent juvenile offenders, should a subsequent court referral for a status or minor property offense be considered as failure? There are no hard and fast rules to govern this. However, in many of these decisions greater information can be collected at little or no additional cost. Where possible, data should be presented on the type of subsequent offenses rather than forcing a dichotomous success or failure decisions.

Another issue in defining recidivism is the length of the follow up period. One correct but somewhat unhelpful maxim is the longer the better. Longer follow up periods have the obvious advantage of better testing the lasting effects of the intervention. However, given the need for timely feedback in a public policy environment, long follow up periods are often not feasible. The time span will also affect the appearance of success. The longer an individual is followed, the more likely we are to discover some wrong doing (except for the most saintly clients). Thus major differences in the impact of the program may be observed from a 3-6 month follow up compared to a 2-3 year follow up period. Another complicating factor involves the need for continuing follow up in adult records for youths reaching their age of majority. For

longer follow up periods this becomes a critical issue. Generally, in the evaluation of juvenile justice programs a six month follow up would be viewed as minimal with a one year period desirable.

One of the most common evaluation errors concerns the use of this follow up period. A one year follow up period means that data on the legal status of each participant during the 12 months following program completion will be collected. The important aspect of this definition is that every person has the same time at risk after the program. Too, often the status of offenders is reviewed as of a certain date, e.g., one year after the program began.

Generally, in the evaluation of juvenile justice programs, a six-month follow-up would be viewed as minimal with a one-year period desirable.

Results may then show that after a year of program operation, a certain number of youth have completed the program and a percentage, presumably small, have been rearrested. In this situation some offenders have had lengthy periods at risk while others would have had only a few days in which to fail. While it is not necessary that each participant have the same period of follow-up, it is imperative that the evaluator collect these data and consider it in the analysis.

Although these are the most common problem areas in the definition of recidivism, there are a number of other issues that deserve careful consideration. Included in these are the concerns of revocation policy for those on community supervision status, i.e., technical violations versus new offenses as reasons for failure. Also, recidivism studies are based on the assumption that the intervention will be effective in influencing the participants' delinquent activities. Thus, a complete offense history, including the dates, offense type, and disposition, should be obtained. From this information the evaluator can control for the seriousness of prior delinquent behavior. It is important that the pre and post program data be collected from the same source since different processes and definitions may be

used in collecting various data sets. If multiple data sources are used for the pre and post program measures, then a finding regarding program effect may actually be an artifact of the differences in the manner in which data were collected.

Alternative Measures of Program Impact

Although recidivism is often a measure of outcome, you should not overlook alternative measures of program impact and effectiveness. For example, in measuring the impact of a juvenile diversion program, changes in the level of court referrals would be a valid outcome measure. Similarly, in evaluating a community service program, one might measure the hours worked to compute public cost savings, as well as the attitudes and opinions of participants regarding their responsibility to the community.

You would be well advised to create multiple outcome measures for several reasons. First, it is quite rare that the impact of intervention is observed in only one area. Almost all juvenile justice programs have a range of goals and potential effects. Many purport to benefit clients (better treatment), the organization (greater efficiency), as well as the larger community (lower crime). Measuring recidivism alone does not include the multidimensional aspects of these programs. Multiple measures increase the reliability of the evaluation, and may increase the acceptability of the findings, thereby adding to overall validity and credibility.

Second, it's not wise to place all the outcome eggs in one basket. When judgments are being made regarding program continuation, it is better to have a greater amount of information on performance than to simply rely on one measure such as recidivism, which may be greatly influenced by factors beyond the program's control.

Finally, attention should be paid to less tangible measures of program outcome, such as the consensus building and knowledge production aspects mentioned in Chapter Three. A service delivery or client-oriented program, which might appropriately be evaluated using traditional outcome measures, will have other products, or by-products, worth measuring or assessing in a qualitative way. Consider, for example, the LRE program in Chapter Three. Interviewing participants revealed that the police handled youths differently after participating in the LRE program. This was recognized as a valuable outcome, and may be considered both a consensus building and a knowledge production outcome.

A comprehensive juvenile justice program plan may contain other programs for which recidivism or other quantitative measures are inappropriate as evaluation tools. Legislative initiatives, or standard setting programs fall in this category, as does the creation and implementation of policy or issue review boards. These programs are often directly aimed at consensus building and knowledge production, or some other system-oriented-- versus client-oriented--goal. Such programs are worthy of and amenable to evaluation, and should get serious consideration. Evaluating them will provide alternative measures of programs and initiatives.

Threats to Validity

The evaluation design section noted the importance of constructing the evaluation so as to rule out alternative explanations of the findings. The validity of a study reflects to the accuracy of the results. How confident are we that what we have seen is what is really happening? Can we actually attribute the changes observed to participation in the program?

While the issue of validity can be technical and highly complex, the principal concerns of validity are straightforward and must be considered in every evaluation. Often the issue of validity is broken down to the question of what else, other than program participation, may have caused these results, known as internal validity, and how general or representative are these findings to other groups or jurisdictions, which is external validity. Although in some situations the validity question must be handled empirically, there are several well known threats to validity of which you should be aware.

Threats to Internal Validity

Different research designs are susceptible to various types of validity threats. For

example, the common pre/post design, in which there are measures taken prior to the initiation of treatment and similar measures taken at the conclusion of the treatment or follow-up period, are vulnerable to issues involving how the subjects may have changed from non program effects during the project. Similarly, evaluations which are based on a comparison group design, i.e., nonrandomly selected "similar" group, face validity problems due to potential selection biases.

Maturation Effects

Pre/post designs often are invalid because of what is known as a "maturation" effect. Changes that may naturally occur due to the passage of time, such as becoming older, smarter, or gaining experience, are maturation effects. If these changes are related to the variable under study then a false, or invalid, picture is obtained. This is a particularly problem in evaluating juvenile justice programs, where many youths cease committing delinquent acts as they grow older, independent of any formal intervention. Without an adequate comparison group, which is presumably maturing at the same rate, these changes may mistakenly be attributed to the intervention project.

History Effects

Another common threat to validity is known as a history effect. While maturation refers to natural changes in the participants, history refers to changes in the environment outside of the project that could produce changes in the variables under study. For example, during the course of a diversion project aimed at high risk youth there is a heinous crime committed by a juvenile offender with a corresponding outcry for tougher responses to juveniles. This situation may alter the types of youth referred to the program and presumably affect the results. Attitude surveys are particularly subject to this influence since opinions may largely be influenced by recent events and media presentation of topical issues.

Selection Effects

A third area threat to validity is the effect of selection of program participants. There is an understandable desire to choose individuals who are the most amenable to treatment, who would most likely benefit from participation, and who are the best risks for community treatment. After all, program continuation may be based on the performance of the initial participants, meaning a natural desire to select those who have the best chance of succeeding.

However, this group may be those offenders who will do well regardless of program participation. Furthermore, in many cases this hand picked group is not from the project's stated target population, producing biased results regarding the effectiveness and impact of the program.

For example, a program may be created to divert property offenders from adjudication. In screening potential clients the staff selects very minor offenders, petty shoplifters, who have a positive home situation, since these offenders are most likely to be good risks for diversion. However, it is unlikely that these offenders were being adjudicated prior to the implementation of the program, and it is less likely that they would commit subsequent offenses compared to a more serious delinquent population that would formerly have been adjudicated. Compiling such superficial results and comparing them to regular court probation programs makes it appear that the program has been very effective. But this appearance is likely the result of selection bias and not the effect of the program. Given this common situation, monitoring should be conducted in all juvenile justice evaluations to be sure the appropriate target population is being reached. An adequate comparison group is necessary to indicate if the effect would have occurred with a similar population without the program.

Mortality Threats

Just as selection of program clients can be a source of bias so can the differential dropout rate or mortality among participants. While there is a strong temptation to present results on only those youths who successfully complete the program, this also will result in a biased group for comparison. While program completion and obtaining the full treatment effect are important inputs to an

evaluation, they should not cloud the comparison with the performance of a control group. Many evaluations take pains to insure that equivalent groups are available for comparison. If a comparison is made with only those that complete the program these groups are no longer equal. There are most likely qualities that distinguish those completing the program from those who drop out. To the degree that these qualities are related to delinquency, there will be significant bias as a result of this inappropriate comparison.

Threats to External Validity

The threats to validity of history, maturation, selection, and mortality concern internal validity; that is, the validity of the findings of the evaluation itself. External validity refers to the degree to which the findings can be generalized to other groups or jurisdictions. If there is a relationship between the kinds of youth in the program and performance, or the characteristics of the jurisdiction make it unique from other jurisdictions in ways that may be related to delinquency, then the ability of the program to be replicated successfully is limited. For example, if a status offender intervention program is found to be effective in an upper-middle class jurisdiction there is little reason to believe that a similar program would be effective in a lower class area given the dynamics of status offending and referral process to juvenile court in these areas.

There are a number of additional threats to the "accuracy" of program evaluations. Think through the program procedure and research design and ask what else could produce biased results and threaten the validity of the evaluation findings. Make modifications in the evaluation design to address as many of these pitfalls as practically possible.

Special Topics in Program Evaluation
Sources of Data
Official Sources of Data

In many situations the most convenient sources of data are official records maintained by juvenile or criminal justice agencies. Arrests, juvenile court referrals, and adjudications are examples of frequently

used official data in the evaluation of juvenile justice programs. When recidivism is employed as an outcome criterion it is most often measured from official sources. Although these data are useful and often readily available, you must exercise caution in their use. Remember that the principal reason for which the data were initially

. . . official justice records should always be viewed as an underestimate of the actual amount of criminal or delinquent behavior.

collected is as an accounting of the activities of criminal justice agencies. Arrests are more a measure of police activity than of criminal behavior. However, arrest data are a preferred method of measuring recidivism and constitute the best data available from official sources, since they have been least affected by the filtering process of the juvenile justice system.

In addition to measuring subsequent delinquent behavior, official data sources include the behavior of agency members themselves. If an arrest is made of a juvenile offender it will be represented in the official police records. However, if a youth engages in subsequent delinquent behavior it may not come to the attention of the police or, if it does, the officer may choose not to make an arrest. In either case, it would not be reflected in official police records. For this reason, official justice records should always be viewed as an underestimate of the actual amount of criminal or delinquent behavior. Also, changes in organizational activities or policy can have an effect on official data which should not be mistaken for changes in crime and delinquency. As long as the evaluator is aware of the potential pitfalls of these data and represents them in the report, official records are a valuable source of evaluation data.

Self Reported Data

Instead of relying on criminal or juvenile justice agencies to tell us about the behavior of youth many researchers advocate asking the youths themselves about their delinquent activities. While this process may seem incredible to some, self reported procedures

have repeatedly been found to be valuable and reliable in delinquency research (see Hindelang, et. al., 1981). In this procedure, generally the youth is asked to complete a questionnaire indicating the frequency of his/her involvement in specific types of delinquent activities. While there is general agreement between self reported data and official statistics in identifying the most serious and persistent offenders, self reported instruments offer more accurate and precise measurement of the numbers and variety of delinquent activities.

Interviews and Questionnaires

One of the most valuable sources of data is directly asking program participants, staff, or other individuals questions pertinent to the evaluation. While these approaches may involve somewhat different methods, they are similar in that the evaluator is attempting to elicit information directly from those knowledgeable about, or involved with, program activities. Measurement of self reported delinquency is this type of activity which can be administered either using an interview or questionnaire format.

The choice of format depends upon the specific context of the study, the study population, and the type of information to be elicited. For example, if one is studying an incarcerated population in which the participants may have a low reading ability, conducting interviews may be a more valid procedure than administering a questionnaire. Similarly the evaluator may be attempting to elicit sensitive information from program administrators or asking questions that do not lend themselves to categorized responses. In these situations an interview approach may also be more

Questionnaires have the advantage of being easily administered to a large group of subjects.

appropriate. While interviews often have the advantage of being able to probe the meaning of responses they can be more costly and more time consuming than a questionnaire. Furthermore, an interview strategy requires trained personnel to convey the meaning of questions, establish rapport with the subject, and probe responses. Questionnaires have the advantage of being easily administered to a large group of subjects, and the fact that each person is answering identical questions increases the reliability of responses.

Regardless of which of these methods is employed, the evaluator must exercise great caution in the selection of items to be included. It is imperative that questions be chosen that are clear and simple, elicit only one response, do not have a double meaning, and measure what they are intended to measure. Insuring that these conditions are met increases the validity of the study.

Surveys

Public opinion polls and surveys seem to dominate our modern life. The widespread use of surveys may give the impression that designing and conducting surveys is an easy task. True the administration of a survey is straightforward, which is one of the main advantages of this method. But, producing meaningful information through surveys is a much more complex and difficult task. Surveys that contain ill chosen questions and poor response options, produce information that is likely to be invalid and not reliable, not to mention misleading and potentially damaging. While volumes have been written on survey design and analysis, the following sections will provide some basic guidelines for developing survey items that may be most useful in juvenile justice program evaluation.

Basic Guidelines for the Development of Survey Items

- *Select the appropriate group(s) and numbers of individuals to respond to the survey.*

  If there is an interest in understanding what victims feel about their participation in a restitution program, then a survey of such victims would be most appropriate. If, on the other hand, there is an interest in obtaining viewpoints from victims and nonvictims concerning the acceptability of restitution as a sanction, then surveys of the general population should be used.

One of the principal advantages of surveys is the use of sampling. Sampling allows surveying a small group that is representative of the larger population yet is large enough to generalize to the whole group. The key word is representative. There are many forms of sampling that can produce adequate samples in different situations, but professional advice is well advised.

● *State survey questions in a simple, clear, and direct manner.*

To elicit meaningful information from a survey all respondents must answer the same question. This cannot happen if questions require a respondent to interpret what it is asking. Questions should be stated in as simple and direct manner as possible. The use of double negatives can result in substantial confusion over what is being asked and the appropriate response. Consider the following question, for example:

> "Do you approve or disapprove of the juvenile court not allowing status offenders to be placed in secure detention?"

As phrased, the question is confusing. The respondent may not understand that to disapprove of the statement is to favor secure detention of status offenders. Instead, the question should be stated in the affirmative, such as:

> "Do you favor the use of secure detention for status offenders?"

Avoid stating questions as double negatives and the use of confusing phrases and implicit negative words which require positive responses for a negative opinion. Instead of asking "Do you oppose gun control?", the more direct and positive question "Do you favor gun control?" is less confusing and thus preferable.

Avoid also double-barreled questions. These are single questions that ask for responses about two or more different things. For example:

> "Do you favor community juvenile justice programs such as diversion and restitution?"

Responses to this question could be either opinions about diversion or about restitution. Similarly, the question:

> "How satisfied are you with the police and juvenile court response to delinquency?"

requires the respondent to assess both the police and juvenile court with a single response. In such cases two separate questions should be used to clarify what is being asked, or only one concept should be included in the question.

> "Do you favor diversion as a form of community juvenile justice program?"

would be a better phrasing of the survey item.

● *Survey items should be as specific as possible.*

Better information and more focused responses are obtained from specific rather than general questions. But there needs to be agreement between what the evaluator is asking and what the respondent thinks he or she is answering, because in some situations there can be different definitions of a concept. For example, if a question asks if the respondent was physically abused as a child, the individual may answer "no" because he or she doesn't consider the treatment received to be child abuse. Similarly, asking respondents if they have been crime victims (or offenders) may not elicit accurate responses if they do not consider the behavior under investigation to be criminal. In these situations you should ask specific questions about the actual conduct in question. Instead of asking respondents if they have been crime victims, a series of questions reflecting specific criminal behaviors should be posed. For example:

> "Have you had anything taken from you by force or the threat of force?"

This would reveal if the respondent had been robbed without requiring him to define robbery.

Moreover, specific questions can pinpoint the source of opinions. A general question, "Do you feel that the juvenile court is doing a poor job, a fair job, a good job, or an excellent job?", will indicate the respondent's overall rating of the juvenile court but not the source of, or reasons for, that rating. An alternative approach would present several questions regarding court operations to obtain an evaluation of a range of services.

● *Use forced choice forms of questions and avoid agree/disagree items.*

One of the most common forms of survey questions is the agree/disagree statement. For example, "Do you agree or disagree with the statement that juvenile offenders should be provided with due process rights just as adults?" Studies show that questions stated in this form tend to elicit a positive response ("agree") regardless of their content. Respondents will even agree with contradictory statements due to this tendency. A more appropriate form for of this question might be, "Do you feel that juvenile offenders should have the same due process rights as adults, fewer due process rights than adults, or greater due process rights than adults?"

● *Measure intensity of opinion as well as the position held.*

Learning that a high percentage of respondents agrees with a particular statement tells us little about how strongly they feel about it. Individuals can support a certain statement but not feel strongly about the issue, or those that support a position can be more intense in their views than those who oppose it. Asking responses on the familiar "strongly agree--strongly disagree" continuum confuses the issues of support and intensity. Generally, follow up questions such as "How strongly do you feel about that position?" should be asked to determine the intensity of the respondent's viewpoint.

● *Conduct a pretest of the survey instrument.*

This procedure is one of the most important yet most frequently ignored stages in survey development. The purpose of the pretest is to ensure that the survey is measuring what you think it is measuring, and that if administered a second time it would obtain similar responses. Which proves the responses are not a function of the instrument itself. Pretesting is not an obscure science, instead it is part of the ongoing process of instrument development. Like much of evaluation research it involves common sense to determine if the new creation is performing as expected during one or more dry runs. In conducting a pretest, it is beneficial to debrief respondents (who should be from a similar target population as those to be involved in the actual study) about specific questions to determine how they interpreted them, the reasons they responded as they did, and how they might have responded if a question were presented in a different manner.

While surveys have numerous potential pitfalls, the relatively low cost and ease of administration make them an attractive research tool. Careful design, administration, and analysis can overcome these difficulties and produce valuable evaluation data.

Use of Random Assignment

A major concern in any program evaluation is that something other than the program itself might be the cause of the results. Proper evaluation design can help eliminate many of these alternative explanations. As noted in Chapter Three, an experimental or comparative design provides the most definitive response to this question. Through experimental design the researcher obtains a comparative base that should be equivalent to the treatment group except that its members have not participated in the program under study. However, the use of experimental design in program evaluation is often hampered by the political, ethical, and pragmatic aspects of the public environment in which these projects are conducted. In many situations, program administrators or potential participants may object to the

concept of random assignment, favoring instead selection based on need or order of application, i.e., first come-first served. At other times administrators may feel public and organizational pressure not to withhold treatment from a group of needy and worthy individuals. Furthermore, the costs of conducting an experimental design are often viewed as prohibitive.

While in many situations these factors may be valid and an experimental procedure would be impractical, in others they can be overcome through the persistence of the evaluator and the support of the administrator. The strength of these barriers is often presumed to be greater than it actually is. It is common to limit the initial size of a new program or the number of jurisdictions in which it will be implemented. In other situations program space may naturally be limited to where demand exceeds the ability to accommodate all who would like to participate. These situations create a natural environment for implementing an experimental design.

When program resources are scarce the fairest method of allocation is a random one. Using random assignment, administrators would not be subject to criticisms of political bias or other forms of favoritism in selecting those to receive treatment. The cost of an evaluation using an experimental design should not be appreciably greater than constructing comparison groups in a less rigorous manner. A major cost of experimental studies comes from the length of follow up, a similar cost would be incurred from studies using a post hoc comparison. Additional costs are not a function of the experimental nature but of the data collected. Also, if there is additional cost in this approach it pales in comparison to the cost of widespread implementation of an ineffective, and even potentially damaging, program.

With careful planning and explanation, experimental designs can be utilized in juvenile justice evaluation. The increased quality, rigor, and potential impact on public policy is well worth the effort.

The Use of Observation in Program Evaluation

With the emphasis on empirical data the valuable contribution that observation can make to evaluation is often overlooked. Observation can be a particularly unobtrusive method of collecting information that is less likely to be distorted than that collected through a questionnaire or interview.

This type of data can be collected in a structured and systematic manner through the use of trained observers. In evaluating a guardian ad litem program, for instance, observers in the courtroom could record the type of interaction among the participants.

In addition, observation data may be collected in a systematic manner by the evaluator during participation in project development and routine staff meetings. Notes should be made immediately after these events and a journal kept for later analysis. This form of data is particularly helpful in conducting process evaluation.

A third valuable source of data comes through observing details of the project operation. In the course of interaction, a number of seemingly insignificant details may be important in assessing the program's impact. Consider the evaluation of an education program in a juvenile detention center. If, on a site visit, the evaluator observes that the texts and workbooks are in perfect shape and appear unused he or she may reasonably suspect that the program has not been adequately implemented. While such data are not hard evidence, they can sensitize the evaluator to issues to be further investigated in a more formal manner.

Common Errors in Evaluation

Presented below are a few errors common in program evaluation. These errors most often result from not thinking through the questions to be addressed and avoiding a systematic approach to evaluation.

● *Individual Cases as Indicators of Success*

One widespread and misleading practice is to use individual cases as examples of success instead of statistical evidence.

A Few Words About Cost Benefit Analysis

One of the most common, difficult, and misunderstood approaches in program evaluation is cost benefit analysis. There is a deceptively simple aspect to the main questions posed here: How much did the program cost? What were the benefits? Were the benefits worth the costs? While these questions are straightforward, their definition and answers are not. Determining the costs and the benefits depends on the approach, orientation, and values of the evaluator. While not intended as a primer on cost benefit analysis, the following critical issues should be considered in this approach to program evaluation.

● *Who benefits from the project?*

Juvenile justice programs often have multiple goals. Consequently, there are likely to be several potential beneficiaries. Often, there are goals to increase the efficiency of the organization, and a second set to improve the services delivered. If the project increases the efficiency of the agency, administrators are likely to define the project as beneficial; if, on the other hand, it increases services but does not increase efficiency, other perspectives may find this more beneficial. Cost benefit analysis requires that the evaluation measure all of the potential benefits of the program, not just those pertaining to agency efficiency.

● *Who incurs the cost of the project?*

Sometimes an agency that provides direct funding for a project may not be the agency that is operating it. Further, if cost savings are an issue, who is the beneficiary? Take as an example a project that aims to reduce the use of state incarceration of youth by encouraging local alternatives. While there is an overall cost reduction through fewer youth being incarcerated, the local unit has to provide primary support for the alternative program, which translates to a cost increase for the local jurisdiction. This is often the problem with statements

indicating that incarceration is cheaper than probation. Once there is a commitment to a state facility, the costs to the local jurisdiction cease.

● *What are the types of benefits and costs?*

Costs and benefits of a project can be either direct or indirect. In addition to the direct benefits related to the project goals other, perhaps unanticipated, benefits may occur such as increasing employee productivity or morale. Similarly, there may be other forms of costs to the project. What are the alternative costs? How else could the funding have been spent? What has to be given up in order for this project to operate? How are program costs influenced by the level of fixed costs of the agency, e.g., agencies that are well funded may be absorbing part of the cost for operating the program?

● *How are benefits and costs to be identified and measured?*

Value differences among evaluators or between evaluators and administrators, as well as different perspectives regarding the goals and objectives of the project, can result in substantial disagreement about how costs and benefits are to be measured.

While these issues point to some of the difficulties of cost-benefit statements, they are not meant to eliminate cost considerations in juvenile justice evaluation. Cost assessments, particularly including alternative and comparative costs, can add significantly to an evaluation. Given the difficulties of conducting a direct comparison of costs to benefits, it is often more feasible to conduct comparative cost studies. In such approaches you would measure resources consumed by comparable programs and a set of objective indicators of outcome for each program. Costs per unit for various types of outcomes could then be computed. This yield some valuable information for decisionmakers.

Every program invariably has individuals who do extremely well, and it is common to attempt to demonstrate the success of the program through these individuals. Such sagas of how the program has turned the life of an individual around and saved him from a life of crime can have a dramatic impact on both the general public as well as funding agencies. Unfortunately, they have little to do with the success of the program. The overall assessment of a program is routinely expressed in rates, percentages, correlations, or other forms of statistical comparison. Accounts of individual performance serve a valuable purpose in illuminating certain aspects of the program, and can provide rich information for understanding program operation, but in no way can program success be based on these individual testimonials.

● *Use of Expert Opinion*

As with individual accounts, expert opinion is also misused as a measure of success. Here again the error lies in the avoidance of statistical data to the preference of an individual account. Expert opinion is based on general knowledge of program types rather than experience. Unfortunately, the basis upon which these judgments are made is often unclear. Expert opinion is useful in clarifying and refining program operations, but it should not be viewed as a substitute for the collection and analysis of summary evaluation data.

● *Use of Inappropriate Comparative Base*

A third common evaluation error is using the wrong basis of comparison in making evaluative judgments. For example, often one hears statements to the effect that of adults who commit crime, a certain--usually high--percentage had been adjudicated as juveniles. The presumption is that the juvenile justice system is not working since so many adult offenders had juvenile offense histories. This is an inappropriate comparison about the effectiveness of the juvenile court. We know, for example, that almost all adult offenders drink soda pop, yet we don't blame soda pop for

their criminality. Regardless of how many adult offenders have juvenile histories, the only way that one can make appropriate conclusions about the impact of a juvenile record is to base comparisons on what happens to all juvenile offenders, not on the characteristics of adult offenders. The appropriate comparison is the percentage of juvenile offenders that are later arrested as adults. Unfortunately, this mistake is widespread and, unlike the others discussed, not easily detected. The only check against such misleading conclusions is diligence in considering and/or constructing your comparisons.

The Evaluation Imagination

There is nothing more valued in conducting evaluation than creativity. Often we tend to think of evaluation as a routine and sterile enterprise. Once a program has been designed we simply apply standard routine procedures, collect and analyze data, and produce an evaluation report that will tell the world if the program worked or not. This is a cookbook view of evaluation, and nothing could be further from the actual state of affairs.

It is true that there are accepted practices and methods that produce higher quality data and should be followed to obtain a credible evaluation. These aspects, many of which have been discussed here, could be termed the science of evaluation.

On the other hand, evaluation is as much art as it is science. In spite of controls and regardless of experimental conditions evaluation is not conducted in a controlled environment. These are not laboratory experiments in which the same technique can be employed regardless of jurisdiction or administrative variation. While this is often the most frustrating aspect of social program evaluation, it also can prove to be the most interesting and challenging. Each time a program is implemented there will be differences that must be considered in determining the role of the evaluation and the evaluator, the design of the study, the variables of interest and outcome measures, the criteria of success, the method of data collection, and the manner of presenting findings.

This monograph has sought to provide some overview of these issues, and introduce you to the complexity and consequences of choices that must be made. However, these choices are far from being routine and mechanized. What is appropriate in one circumstance may be totally inappropriate in others. Because of its individual characteristics, each program presents a new challenge. In addition to being a competent scientist, the evaluator must also be an artist and craftsman. He or she must choose from available tools, or perhaps create new ones to design the best method of determining the impact of the program under study. Nothing is more important in this task than a skilled knowledge of these techniques, as well as the vision or imagination on how to creatively apply them. With this in mind, some suggestions regarding the use of evaluations are provided below.

Using Evaluations: Audiences and Products

Evaluation projects, and their products, are important elements in the overall policy analysis and feedback process. New juvenile justice programs, or continuations of those already established, represent policy action and choices, as does the decision not to create new programs. If evaluations of the kind described in this monograph are not undertaken, policy analysis and feedback will still take place. Program funders, backers, critics, and beneficiaries will receive information about the program in various forms, mostly informal, anecdotal, and subjective, and will make decisions about the program that will affect its future. With evaluations a certain, though sometimes small, amount of objectivity is injected into the policy analysis process, and program decisions begin to be made in light of this more scientific evidence. This is the primary function of evaluation.

Sometimes, even though the value of evaluation for policy analysis is recognized, its potential is not fully realized. This occurs, for example, when the multiple users or consumers of evaluation information are not considered and the evaluation products do not reach them. Often, the evaluation product is one or more reports read by the funders of the evaluation effort and/or the funders of the program being evaluated, and

the policy analysis stops there. But there may be other parties with a need and desire to know evaluation results, and it may be in your interest to inform them.

Others who have interest in evaluation information include:

- Program clients or potential clients. Reading the results, if presented from a lay person's perspective, may improve their self image or their image of the program which would be a positive development.

- Other criminal justice professionals. More often than not the program you evaluate exists in some form somewhere else, or a very similar one exists, and the people associated with that program would surely read your findings with great interest. It is likely that similar methods are being tried with adults in other places, indicating an even greater audience for your findings.

- Potential supporters of the program. Thoughtful portrayal of evaluation findings can help build community support and consensus that is needed for program survival. The public education function of evaluation information should not be overlooked.

With a little thought at the front end of an evaluation effort, audiences such as these can be identified for your evaluation results, doing so helps identify the type of evaluation to be conducted, the specific types of information to collect, and the appropriate products that should be developed.

It is the extreme case in which evaluation data stand on their own; they are almost always presented (though not tabulated) with their recipients in mind.

For example, assume that an 18-month evaluation of a jail removal project has been completed and the authors are preparing to present their findings to the funders. More often than not at least two reports of the findings will be prepared. One is an "Executive Summary" or some other non-technical review of the findings that is

brief, to the point, and geared towards the political and administrative issues at hand. The other is a full, or technical report that explains in detail the study design, methods, limitations, and findings. Both reports are valid, reflecting the awareness that there are different audiences, or consumers, of the evaluation information.

In another example, a new program may be viewed with skepticism by school or law enforcement officials, especially if it utilizes an unfamiliar method such as play therapy or psychodrama. One product should be a detailed description of the program itself, with the law enforcement and school official in mind. Once these officials understand the program they will be less skeptical. In the meantime, a good monitoring or process evaluation effort can be accomplished that meets other goals as well. The program description might be excerpted from the final report and published in a tasteful manner.

In a final example, a comparative evaluation effort, might produce some important products for potential clients, above and beyond the final report. A video outlining the pros and cons of the program, or that compares responses and opinions from juveniles in experimental and control programs, might be produced. This would be of great value for potential funders and clients.

When the three types of outcomes are considered--knowledge production, consensus building, and instrumental--a variety of possible audiences and products are suggested, depending on the nature of the evaluation. For example:

- *Program descriptions.* The evaluation may produce the first and only detailed description of program goals, objectives, and operations. This is a valuable product in itself, and will usually have a receptive audience.

- *Process feedback.* Outside of the published evaluation findings, there may be other information uncovered during the course of the study that is not appropriate for a final product, but that has value for program staff and funders. Most

evaluators will provide an informal debriefing to cover such matters. To those people this information may be as important, if not more so, than the published findings.

- *Good news.* In addition to presenting favorable findings to funders, evaluation products can be geared toward interested outside audiences whose concern is not with the specific program, but with perhaps replicating it or portions of it, or selling their constituencies on the idea of implementing similar efforts. If handled appropriately, evaluation information can provide useful public or marketing information.

- *Bad news.* Most evaluations will uncover potential "bad news," that is, information that suggests there are problems with the program under investigation, or that findings will be contrary to the expectations of program funders or managers. Evaluators should be comfortable in presenting this evaluation information too. Generally, three rules should guide the presentation of potential bad news:

  (1) Present the information as soon as you can, but only when you are confident that your information and sources are correct; don't present hunches or preliminary indications as confirmation of program problems. You may find after further inquiry that your initial supposition was incorrect.

  (2) Present the information to an audience limited to program managers or administrators (and perhaps to project funders, depending on how closely they are involved in your work). At the start of an evaluation project, the appropriate individuals who will receive preliminary information and briefings should be identified, and they should be the only persons to receive potential bad news. Then they

can make the decisions regarding how to use the information you provide.

(3) Present the potential bad news in as positive a manner as possible. Program managers should be glad to receive information that helps them solve problems, or that directs their decisions regarding program modification. If your potential bad news comes early enough, they may be in a position to make corrections, lessen the impact, and improve the program. You will often find that your information will be well received, and perhaps that the news is not perceived by the recipients to be as bad as you thought it might be.

As evaluation audiences and products are important to consider in using evaluations, so too is product dissemination. It is not just what you produce and who you produce it for, but how you get it to the intended audience that can be the difference in making your evaluation useful.

If the intended target does not normally read long evaluation reports, and is one that you have not traditionally targeted for your product, then dissemination strategy is critical. In some instances this amounts to a marketing and product presentation strategy.

If the intended audience would benefit most from a brief and direct presentation--of program operations, evaluation findings, preliminary findings--then a road show might be warranted. This could include slides, videos, presentations by former clients, posters and/or product demonstrations, and might prove more effective in delivering findings than a written report. Such a product requires careful consideration of resources and priorities.

In another instance, the final evaluation report might be delivered to a limited audience, say 10 to 20 funders, board members, legislative committee members, but a brief monograph or pamphlet describing the program and the most pertinent findings might go to a broader audience of criminal justice professionals or community members. Again, this may do more good for the program than the detailed report.

Program funders, managers, and evaluators can be creative and broad-minded in their consideration of these issues, and they can influence evaluation products to meet many needs. The most effective way to do this is to specify the products and audiences in program and evaluation requests for proposals (RFPs). These are discussed in Chapter Five.

5

Introduction

This chapter contains suggestions and references that will prove helpful in the evaluation process. Two issues concerning requests for proposals (RFPs) are considered--building evaluation into juvenile justice program RFPs, and designing program evaluation RFPs. Identifying people with evaluation skills and appropriate perspective is in a section on choosing an evaluator. Finally, a list of reading and professional references is provided.

Building Evaluation Into a Program RFP

One of the best ways to insure that evaluation information will be available for funded programs, especially information that is useful to you and others, is to require production of that information as a "deliverable." If you know you want to evaluate the program, then you know that you want to collect basic monitoring information. Why not include a requirement to collect such information in the published program RFP or as a condition of program funding? While you may believe that programs collect that information anyway, and don't want to place unnecessary bureaucratic burdens on your program personnel, there are some good arguments for being more aggressive for producing evaluation information in the RFP and funding processes. These include:

- Potential grantees or funding recipients need to know that you are

Other Considerations

serious about evaluation; that the information they produce is important to you and others. Often, they rightly complain not about producing evaluation information, but about producing information that they feel will not be used by anyone.

- Placing evaluation information needs in a program RFP forces you to think seriously about evaluation design and information needs. Once you resolve the issues regarding research questions, data, and research design for yourself, they will be more easily communicated to program administrators and evaluators.

- Documenting your information makes the collection of the information routine, and strengthens your ability to demand it. It is difficult to introduce requests for information to program administrators or evaluators midway through a project. Likewise, it is hard to ignore information needs that are in writing and that have been agreed to. Once a data collection effort is started it quickly becomes a natural organizational process.

- It is important for program administrators to know that the program will be evaluated if you intend to do so. Explaining this in the RFP puts them on notice and eliminates the element of surprise

for staff that can occur when outside evaluators come knocking.

Even if you are not accustomed to documenting evaluation plans in program RFPs, or to requesting evaluation information from your grantees, these points argue for doing so. By making your plans known, and being as specific as possible about your expectations, you increase your chances of conducting and completing a useful program evaluation.

Including evaluation plans in a juvenile justice program RFP is not difficult. The following suggestions are offered as a means of doing so.

- State your plan to evaluate the program at the outset. Most RFPs begin with a summary of the project or program to be funded. This statement should clearly state that program evaluation will take place and who will conduct the evaluation, and that certain information requirements will have to be met as part of the program activities.

- Stress that program goals must be clearly stated in the proposal, as well as objectives or action plans for achieving the goals and a demonstrated link between program operations, goals, and objectives. While it may not be the case that program "success" will be measured by goal attainment (often this need not be the case), to begin any evaluation the evaluator must have some notion of what the program

developers had in mind in the way of goals, objectives, and operations at the time of the program's creation. This information must be made explicit in the program proposal.

- Include a section on evaluation information. To the extent that you can be specific, identify in a separate section the types of evaluation program monitoring information you expect to be collected. This may be as brief as a list of information items, and a stipulation that the details will be worked out soon after the award is made, or as complete as a list of information items with definitions, formats, data sources, reporting schedules, and so on. The important thing is to identify in as much detail as possible, the data collection responsibilities the grantee assumes. This serves to stress the seriousness of the evaluation effort and to put potential grantees on notice.

- Include a section on evaluation plans. Regardless of the extent program administrators involvement in the actual evaluation or whether you plan to use outside evaluators, it is important to discuss the ways in which the evaluation information provided will be used. For example, if you are monitoring for your information and for the SAG's information, explain that the data will not be widely circulated. If you plan on conducting a comparative monitoring or evaluation effort, explain that others will be collecting the same information, and that the feedback will be useful to program administrators. Similarly, if you plan to produce a widely circulated program description monograph, let that be known. It is also important to delineate the many values of the evaluation information. It will help funders and program administrators make decisions, it will help educate the public and legislature and perhaps reduce myths surrounding the program, it will help others replicate their work, and so on.

If you plan on involving an outside evaluator, there are two additional important pieces of information to include in the RFP:

(1) The statement that an outside evaluator will be involved.

(2) The statement that the evaluation effort will be a collaboration between the program, your office, and the evaluator. It is important for the staff to know that they will have an opportunity to educate the outside evaluator and assist in the interpretation of data, and that your office is interested enough in the evaluation to be involved. Even in the most routine programs, the presence of an outside evaluator often requires the involvement of someone at your level to keep things on a positive note and to smooth out any potential trouble areas. The expected evaluation products should be discussed to provide potential grantees with a clear understanding of why the information is needed and how it will be ultimately used.

Preparing an Evaluation RFP

If you are going to use an outside evaluator, more than likely you will need an evaluation Request for Proposals (RFP) to solicit proposals for the project. Preparing evaluation RFPs is much like preparing any other RFP--it must be detailed and specific regarding the activities and products you will pay for, schedules, responsibilities, and the like. In this section, we will review some important guidelines for preparing an evaluation RFP, paying particular attention to the qualities that set them apart. We recommend, however, that the reader also consult a more thorough reference on preparing requests for proposals, and obtain a good example to follow.

A comprehensive request for proposals will contain at least five key sections:

(1) A statement of objectives that states clearly and concisely the project to be funded and the need for the evaluation effort.

(2) A section containing general information for applicants. This section normally contains technical and logistical details that applicants must know.

(3) A specifications section that explains in detail the entire evaluation project--purpose, timelines, mandatory deliverables, products, and so on. This is the most important section of the RFP.

(4) A section outlining the detailed information requirements that must be contained in a proposal.

(5) A section explaining the RFP evaluation and selection process.

In most states and other jurisdictions, there are a host of sections and subsections that are required for an RFP. These outline specific information requirements e.g., financial disclosure statements, affidavits of various kinds, minority business enterprise information, required clauses for termination of contracts, and the like. Too numerous to mention here, they can be obtained from the appropriate procurement office in your state or local government. Again, obtain a copy of a successful RFP that has passed the procurement office standards if possible. Following is a review of the five sections mentioned above.

Objectives of the RFP

This section serves as the introduction for the reader of the RFP, usually potential applicants or grantees. It states your intent to solicit proposals for the evaluation of one or more specific programs, provides the RFP schedule--when proposals are due, when the evaluation process begins and ends, and such. It is usually brief, and may include other sections such as a brief statement that a formal proposal evaluation process will take place, and the makeup of the proposal evaluation committee, information about a pre-proposal conference in which potential applicants meet to obtain other information from your office, if applicable, and a statement that final approval of the award is subject to review from a higher authority such as the state budget office, perhaps, or the procurement office.

General Information for Applicants

This section normally contains many subsections with logistical and other information that is required in all RFPs. The following is a list of the types of information that should be required. It is not comprehensive since these requirements vary by jurisdiction. You may even wish to present them as section topics under "General Information for Applicants."

(1) Names and phone numbers that applicants can call with questions and inquiries about the RFP.[1]

(2) A statement that your office reserves the right to amend or cancel the RFP process at any time, and that all applicants will be notified of such changes.

(3) A statement that your office reserves the right to require any applicant to make an oral presentation to clarify the proposal.

(4) A statement that your office will not assume any of the proposal preparation costs.

(5) Suggest a method for submitting multiple proposals, or limit each applicant to only one.[2]

(6) Requirement of financial disclosure by applicants, according to the relevant statute.

(7) Requests for certifications of various kinds required by law such as anti-bribery affidavit, non-collusion certificate, public information act notices, procurement affirmation, and minority business enterprise certification.[3]

(8) A statement explaining the conditions under which your office may terminate any contract made under this RFP.

(9) A reference to the relevant statutes and procedures regarding disputes or protests that arise.

(10)Reference to any federal laws or guidelines that apply.

Most of this information is standard for RFPs, though the language and detail varies. They often appear tedious, bureaucratic, and irrelevant, but they serve important purposes--they protect you and your superiors from a number of liabilities, and they provide objectivity and fair competition for the procurement process.

Specifications

This is one of the most important sections of the RFP. In it you explain the program(s) to be evaluated, and the evaluation you would like to have completed by the grantee. This section should contain the following subsections:

- *Background information.* A review of the program(s) to be evaluated, including the following information--historical detail, goals and objectives, program activities and administration, relevant findings and statistics to date (if applicable), the federal initiative (if applicable), and reference to similar programs.[4]

- *Type of evaluation.* Unless you are suggesting that the applicant devise the entire evaluation plan, you should explain the type and level of evaluation that you desire whether monitoring, comparative monitoring, process or outcome evaluation. Again, reference to similar evaluations might be wise.

- *Information requirements.* The type of evaluation desired will automatically suggest certain evaluation information requirements whether monitoring information, process measures, and/or outcome measures. These should be explained in detail.

- *Evaluation products.* This section should explain how the evaluation information should be packaged. You may specify a final report (or two reports--a non-technical summary, and a technical document), a video, a training seminar, a

presentation to relevant audiences, and so on.

- *Timelines.* This section should contain detailed information about important dates and time-specific components relating to the evaluation--program start date, periodic review dates, important meeting dates, funding period, the treatment process time, follow-up periods, etc. You should also mention the time period of acceptance, the time after delivery of final products that you will review the products before making final payment.

- *Product review.* This section should mention the review process that the final products delivered will undergo--the persons involved, the schedule, the criteria for acceptance.

Two important points bear stressing here: (1) The quality of the proposals you receive will depend on the quality of this section. A well thought-out evaluation process will usually receive good proposals, while one that is poorly thought-out will not provide solid enough information, resulting in proposals that are not responsive to your evaluation needs. (2) Applicants will be

A well thought-out evaluation process will usually receive good proposals, while one that is poorly thought-out will not provide solid enough information, resulting in proposals that are not responsive to your evaluation needs.

judged on how they respond to the issues presented in this section. If program and evaluation details are not clearly stated, the review process may steer you in the wrong direction. You may find yourself forced to choose a proposal you know is not the best.

Finally, a word on mandatory versus desirable specifications. A good way to judge proposals and to get the most from applicants is to specify mandatory specifications which are elements that an

evaluation plan must have to be considered at all and desirable specifications which are elements that you would like to see developed as part of the evaluation plan, but that you leave up to the applicant.

For example, you may specify a process or short-term outcome evaluation as mandatory, but specify an end process outcome evaluation plan as desirable. This may result in the applicant giving the extra effort in terms of design and deliverables. Or, you may specify a final report as mandatory, and cite other deliverables as desirable such as a small public information brochure, a formal presentation to staff and the SAG, and so on. If the applicant knows the proposal will be evaluated on the extent to which desirable elements are delivered, you increase your chances of getting the most for your money.[5]

Information Required in the Proposal

This section is nearly as important as the previous one. It specifies for the applicant exactly the information that should be contained in the proposal; it may even go as far as providing a suggested outline. It should be stated clearly that a proposal that does not contain all of the information required will not be considered.

In addition to a detailed explanation of the evaluation program according to the Specifications section, the following should be required from applicants:

- A detailed discussion of the evaluation management plan, listing the people who will work on the project and including their updated resumes, project management, and the responsibilities of each person.

- A detailed budget should be required, accounting for every dollar by spending category including salaries, fringe, overhead, indirect, supplies, contractual, travel, and so on.

- Applicants should be required to submit examples of prior work in program evaluation, and references to past employers.[6]

Evaluation and Selection

It is important, and fair, to provide applicants with precise information on how their applications will be evaluated. This section should outline the proposal evaluation process, the persons who will evaluate them, the evaluation criteria, any weights or mathematical calculations that will be applied, and how and when the final selection will be announced.

Some considerations in this area include:

- Will technical criteria, such as the logic and complexity of the research design be weighted the same as cost criteria (how much the evaluation will cost), or will technical criteria dominate?[7]

- How much influence, will the delivery of desirables versus mandatory specifications have on the evaluation process?

- Will the evaluation be qualitative or quantitative? You may devise an elaborate scoring scheme, with appropriate weights, and have the resulting scores drive your decision, or you may have each of the evaluation committee members read the proposals and make written recommendations with reasons provided. You may combine a qualitative and quantitative approach. Procurement officials tend to like quantitative approaches, since they avoid subjectivity in the evaluation process, but you may find them restrictive.

An evaluation RFP process is recommended if you plan on spending more than a nominal amount of money on evaluation, if competition for such business is high, and especially if you want to maintain an environment of objectivity regarding your evaluation program. You may have no choice in the matter if your state or jurisdiction mandates an RFP process.

Additionally, the RFP should provide an indication of the dollar amount you plan to allocate for the project. Otherwise the

applicants may propose evaluation projects that vary widely in scope and methods. If you wish to spur competitiveness in the RFP process without revealing how much money you plan to spend on evaluation, you may provide information regarding program costs, or suggest a cap on evaluation project costs and indicate that project cost will be a significant factor in the proposal evaluation[8].

Choosing an Evaluator

Regardless of whether you follow an RFP process for your evaluation, you must choose an evaluator. Sometimes this choice will be made simple by the restrictions you face--not enough money to hire an outside evaluator, you have competent evaluators in your office, or self-evaluation by the funded program is deemed appropriate. Other times you will be faced with a dilemma--for example, hiring an outside evaluator when you have the expertise in your office; or imposing your office evaluator on a program with staff capable of conducting evaluations. Selecting an evaluator is a critical decision, one that often hinges on political considerations, or that raises political issues once the decision is made.

In this section, we discuss two aspects of choosing an evaluator: (1) the pros and cons of utilizing outside evaluators, and (2) recognizing a good evaluator when you see one.

Inside Versus Outside Evaluators

There are practical and logistical aspects to this issue. First two types of inside evaluators need to be distinguished. An inside evaluator may include you or one of your staff, if you are evaluating a program funded by your office; or it may mean a staff person(s) from the program itself.

Generally, the more interest exhibited in your program by people and professionals outside of your state or local environment, the more you will want to consider an outside evaluator.

Depending on the situation, or on one's perspective, your staff may be viewed as outsiders. If your office provided the program funding, however, you are not likely to be viewed as an outsider in terms of evaluation.

On the practical side, it is almost always less expensive to use an inside evaluator than to contract for an outside evaluator, but there are exceptions. Using program personnel as evaluators is probably the least expensive, but produces the least objective results. Contracting with a university or college may prove to be a cost effective means of obtaining outside evaluators. Sometimes, based on cost considerations alone, using inside evaluators is the only option for conducting a program evaluation. In many instances, it is a perfectly logical and appropriate thing to do.

You may wish to use an outside evaluator because someone with broad experience in program evaluation is needed, or because the need for an objective perspective is great, which is often the case for expensive and controversial programs. You may, perhaps, intend to replicate a juvenile justice program evaluation conducted elsewhere and need the same evaluator, or at least an outside evaluator to be consistent. Wherever it is essential to be as objective as possible about the evaluation, you should consider an outside evaluator. Generally, the more interest exhibited in your program by people and professionals outside of your state or local environment, the more you will want to consider an outside evaluator. You cannot escape the fact that evaluations performed by insiders run the risk of being criticized as lacking objectivity.

To repeat, it will not always be fiscally possible to hire an outside evaluator. But, well designed and documented evaluation research effort will almost always provide sufficient and reliable evaluation information.

Recognizing a Good Evaluator

Evaluators are made (often self-made) and not born, so it is possible to identify what they are made of. There are three critical qualities of a good evaluator--experience, skill, and brains. Each will be reviewed here.

Program evaluators learn by doing, and the key to conducting good evaluations is

knowing the ins and outs of the political and
logistical aspects of program evaluation.
Because, there are few academic programs in
the country that have a program evaluation
curriculum, a good evaluator is usually one
who has done a number of them, and whose
references will vouch for the work done.

Evaluation skills refers to knowledge
about research design, methodology, and
statistics. These are taught in academic
programs, and they can also be easily
identified in a review of written materials
provided by a prospective evaluator. Like
experience in evaluation, research design and
methods skills are refined in practice. They
will most often be found in who has utilized
those skills in prior program evaluations.

Brains refers to the thought process in
program evaluation research. The best
evaluations are often found in unique or
creative applications of research skills to the
particularities of the program being
evaluated. This may in a special sampling
strategy, utilization of a measure from
another discipline, creative use of archival
records, or an effective explanation of
statistical methods to non-technical readers
that indicates competence and confidence in
the subject matter.

You have three sources of information
that will help you determine the experience,
skills, and brains of an evaluator. They are
(1) the written examples of past work
performed by the evaluator, (2) the resumes
and references provided, and (3) the actual
evaluation plan submitted for the program at
hand. Careful consideration of each of these
with further discussions with the potential
evaluator when you feel it is necessary, will
help you in the selection process.

NOTES

1. In most instances, regulations will require that these be documented carefully for review purposes. Sometimes it is required that all questions and responses be documented and shared with all applicants.

2. In some instances, you may allow applicants to submit "Alternative Proposals." These are proposals to accomplish the evaluation task you defined in what the applicant feels is a better approach to the problem.

3. These generally are requirements that applicants certify that they are honest business enterprises, that they abide by state laws, or that they meet the legal obligations for contractors in your state or jurisdiction. In most instances, applicants will be more familiar with these requirements than you are, and will have little trouble providing the information.

4. It is perfectly acceptable to review the program(s) in this section and include references to other reading material with the details (and expect the applicants to read them). You might consider, in such an instance, providing the referenced material to all applicants.

5. Of course, you run the risk of having an applicant promise far more than can be delivered. To guard against this you must be a good judge of evaluators, a topic we address below.

6. It is acceptable to exclude from consideration any applicant with no experience in social service program evaluation, or in juvenile program evaluation if the field of applicants leaves you that option.

7. You may avoid this problem by announcing the amount of money available for the evaluation project up front. Most applicants will meet the dollar figure or come very close, and then technical merits can dominate your decision.

8. It should be noted that as of 1988, projects which are funded in part or in whole with Federal money must note in the RFP the percentage of the total cost of the project which will be funded with Federal dollars and the dollar amount of Federal funds designated for the project. Contact your State Budget Office or Federal funding source for details on the law--Section 8136 of the DOD Appropriations Act.

6

This chapter provides the reader with references to two important sources of assistance and expertise in the area of juvenile justice program evaluation:

(1) Suggestions on how to find agencies, organizations, and persons involved in juvenile justice research or program evaluation, and

(2) Bibliographic references to standard evaluation research sources and research design texts.

References and Resources

The suggestions fo· locating help will
lead to expert agencies and organizations
which have been involved in juvenile justice
program evaluation as consultants, teachers,
program administrators or practitioners.
They will prove helpful in any of these
capacities, or as pointers to additional
resources in their areas of expertise.

The bibliographic references cover
juvenile justice program evaluation sources
that publish evaluation research, as well as
general evaluation research texts.

Juvenile Justice Program Evaluation Resources

Finding Agencies and Organizations

A number of potential resources exist to assist those interested in learning more about evaluation. They include:

- The OJJDP Source Book lists all the programs funded though each of the Office's divisions. Many of the programs have built-in evaluation components. You can check the Source Book to determine if the Office has current evaluation activities in program areas of interest. If you do not have a Source Book, one can be ordered from the National Criminal Justice Reference Service in Rockville, Maryland.

- Most states have a Statistical Analysis Center (SAC). Although the goals of the SACs vary between states, most are heavily involved in state criminal justice program and policy research. Their access to data and research trends can be a valuable resource for others planning on conducting evaluations. A list of SAC contacts for each state is maintained by the Criminal Justice Statistics Association. CJSA's headquarters is in Washington, D.C. and can be found through directory assistance.

- The State Relations and Assistance Division of OJJDP is developing a training curriculum on evaluation as it pertains to the types of programs which could be funded using Formula Grants money. The training will be offered through the division's annual Juvenile Justice Training workshop. Look for announcements on times and dates regarding the session or contact SRAD for more information.

- The National Criminal Justice Reference Service Juvenile Justice Clearinghouse (NCJRS-JJS) maintains an information archive on thousands of related articles, books, and other publications. Customized annotated searches on dozens of topics can conducted for a small fee. Contact NCJRS at the number listed above.

- Many state, local, and private universities and colleges can assist with evaluation designs. Contact the department of sociology, social work, education, psychology, or criminology for leads.

There are dozens of private companies and consultants who can assist in evaluation research. Check with other state juvenile justice specialists to learn about who they may have used on similar programs in the past.

Many universities and colleges have research and survey units on campus. Call around to determine the level of assistance available and associated costs.

Keep in mind that the costs associated with the services, the areas of expertise, e.g., corrections, community based programs, detention, etc., and the types of assistance available, e.g., phone consultation, publications, on-site assistance, research design, etc., varies between agencies. Contact the agency directly to determine if they can assist you.

Related Publications

If you were to check a nearby research or state library for references regarding evaluation or related topics, you would undoubtedly find hundreds of listings. Journals, texts, primers, and user guides are all available on the subject. For those interested in learning more about evaluation, a short list is provided below. It is not exhaustive, but it does provide a starting point for those who wish to check library holdings or bookstores to find basic materials on the topic.

The references were selected because of their widespread availability. Most can be found in small and medium size libraries and many college bookstores, although all can be ordered from the publisher. Because presentation style and the types of materials covered varies, some will do a better job of addressing your specific questions and concerns then others. Therefore it is best to borrow a specific book through a library or inter-library loan before ordering it. Sample issues of periodicals can often be obtained directly from the publisher.

Note also that the references cited represent general information designed to give the reader a broad overview of the topic. They are not findings from actual evaluations. Because hundreds of program evaluations are conducted and summarized each year, it would be best to search through a reference services such as NCJRS to locate published works.

Other bibliographic references, such as the *Social Sciences Citation Index* and *Criminal Justice Abstracts* are also useful tools for finding evaluations which have been conducted on specific types of programs. Both can be found in university and college libraries.

The references listed below have been classified into the following categories:

- Evaluation issues and guides. This category includes publications on evaluation issues in general and those related to criminal, delinquency, and/or justice issues in particular.

- Survey research and research design.

- General statistics and guides to statistical programs for the computer. The guides contain succinct explanations of various statistical tools.

- Journals. This category includes journals which cover evaluation issues in general and those which concentrate on delinquency, crime, and/or justice issues, but include articles on evaluation topics or evaluated programs.

Evaluation Issues

Adams, S. *Evaluative Research in Corrections: A Practical Guide.* Washington, DC: U.S. Government Printing Office. 1975.

Alkin, M., *A Guide for Evaluation Decision Makers.* Beverly Hills, California: Sage Publications. 1985.

Empey, L. *Model for the Evaluation of Programs in Juvenile Justice.* Washington, DC: U.S. Government Printing Office. 1976.

Forester, J. *The Practice of Evaluation and Policy Analysis.* Berkeley, California: Institute of Urban and Regional Development. 1975.

Glazer, D. *Routinizing Evaluation: Getting Feedback on Effectiveness of Crime and Delinquency Programs.* Washington, DC: National Institutes of Mental Health, Government Printing Office. 1973.

Hatrey, H., R. Winnie, and D. Fisk. *Practical Program Evaluation for State and Local Governments.* Washington, DC: The Urban Institute Press. 1981.

Herman, J. *The Program Evaluation Kit.* Beverly Hills, California: Sage Publications. 1987.

Hoole, F. *Evaluation Research and Developmental Activities*. Beverly Hills, California: Sage Publications.

House, E. *Evaluating With Validity*. Beverly Hills, California: Sage Publications. 1980.

Klein, M., and D. Teilmann (eds.). *Handbook of Criminal Justice Evaluation*. Beverly Hills, California: Sage Publications. 1980.

Kleugel, J. (ed.). *Evaluating Juvenile Justice*. Beverly Hills, California: Sage Publications. 1983.

Kollasch, S. and P. Luc_ore. *Guide to Implementation and Evaluation of Juvenile Justice Programs*. Montgomery, Alabama: University of Alabama. 1974.

Moos, R. *Evaluating Correctional and Community Settings*. New York: John Wiley and Sons. 1975.

Reicken, H. *Social Experimentation: A Method for Planning and Evaluating Social Intervention*. New York. 1975.

Rossi, P. and H. Freeman. *Evaluation: A Systematic Approach*. Beverly Hills, California: Sage Publications. 1982.

Rutman, L. *Evaluation Research Methods: A Basic Guide*. Beverly Hills, California: Sage Publications. 1984.

Weiss, C. *Evaluation Research: Methods for Assessing Program Effectiveness*. Englewood Cliffs, New Jersey: Prentice-Hall. 1972.

Survey and Research Design

Bailey, K. *Methods of Social Research-2nd ed*. New York: The Free Press. 1982.

Campbell, D. and J. Stanley. *Experimental and Qausi-Experimental Designs for Research*. Geneva, Illinois: Houghton Mifflin Co. 1966.

Converse, J., and S. Presser. *Survey Questions: Handcrafting the Standardized Questionnaire*. Beverly Hills, California: Sage Publications. 1986.

Finsterbusch, K., and A. Bender Motz. *Social Research For Policy Decisions*. Belmont, California: Wadsworth. 1980.

Hakim, C. *Research Design*. London. 1979.

Kalton, G. *Introduction to Survey Sampling*. Beverly Hills, California: Sage Publications. 1983.

Kidder, L. *Research Methods in Social Relations-4th ed*. New York: Holt, Rinehart, and Winston. 1981.

Lakner, E. *A Manual of Statistical Sampling Methods for Corrections Planners*. Champaign, Illinois: University of Illinois at Urbana-Champaign. 1976.

Morris, K. and C. Braukman. (eds.). *Behavioral Approaches to Crime and Delinquency: A Handbook of Application, Research, and Concepts*. New York: Plenum Press. 1987.

Moursund, J. *Evaluation: An Introduction to Research Design*. Monterey, California: Brook/Cole Publishing Company. 1979.

Nagel, S., and M. Neef. *Policy Analysis in Social Science Research*. Lanham, Maryland: University Press of America. 1985.

Rossi, P., J. Wright, and A. Anderson (eds.). *Handbook of Survey Research*. New York: Academic Press. 1983.

Sechrest, L. and Rosenblatt, A. "Research Methods." From *Handbook of Juvenile Delinquency*. New York: John Wiley and Sons. Pp. 417-450. 1987.

Sudman, S. and N. Bradburn. *Asking Questions*. San Francisco, California: Jossey-Bass. 1983.

Trochin, W. *Research Design for Program Evaluation*. Beverly Hills, California: Sage Publications. 1984.

General Statistics and Guides to Statistical Programs for the Computer

Blalock, H. *Social Statistics*. New York: McGraw-Hill. 1979.

Burstein, L., H. Freeman, and P. Rossi, (eds.). *Collecting Evaluation Data*. Beverly Hills, California: Sage Publications. 1985.

Cohen, L. and M. Holliday. *Statistics for Social Scientists*. London: Harper and Row. 1982.

Helwig, J. *Statistical Analysis System: User's Guide: Basics*. Cary, North Carolina: SAS Institute Corporation. 1985.

Hoaglin, D., R. Light, B. McPeek, F. Mosteller, and M. Soto. *Data for Decisions*. Lanham, Maryland: University Press of America. 1982.

Nie, N., et. al. *Statistical Package for the Social Sciences*. New York: McGraw-Hill Book Company. 1975.

Reid, S. *Working with Statistics: An Introduction to Quantitative Methods for Social Scientists*. Cambridge: Polity Press. 1987.

Rosenthal, R. and R. Rosnow. *Primer of Methods for the Behavioral Sciences*. New York: John Wiley and Sons. 1975.

Tashman, L. and K. Lanborn. *The Ways and Means of Statistics*. New York: Harcourt Brace Jovanovich. 1979.

Journals and Periodicals

Crime and Delinquency. Beverly Hills, California: Sage Publications.

Criminal Justice Abstracts. Buffalo, New York: Willow Tree Press.

Criminal Justice and Behavior. Beverly Hills, California: Sage Publications.

Criminology. Columbus, Ohio: American Society of Criminologists.

Evaluation and Program Planning. Elmsford, New York: Pergamon Press.

Evaluation Practice. Beverly Hills, California: Sage Publications.

Evaluation Quarterly. Beverly Hills, California: Sage Publications.

Evaluation Review: A Journal of Applied Social Research. Beverly Hills, California: Sage Publications.

Handbook of Evaluation Research. Beverly Hills, California: Sage Publications.

Journal of Criminal Justice. Elmsford, New York: Pergamon Press, Inc.

Journal of Evaluation and Program Planning. Elmsford, New York: Pergamon Press, Inc.

Journal of Research in Crime and Delinquency. Beverly Hills, California: Sage Publications.

Juvenile and Family Court Journal. Reno, Nevada: National Council of Juvenile and Family Court Judges.

Law and Society Review. Amherst, Massachusetts: Law and Society Association.

Monograph References

References Cited in the Text

Hindelang, M., T. Hirschi, and J. Weis. *Measuring Delinquency.* Beverly Hills, California: Sage Publications. 1981.

Johnson, G., and R. M. Hunter. *Using School-Based Programs to Improve Students' Citizenship in Colorado.* The Colorado Juvenile Justice and Delinquency Prevention Council. October 1987.

Waldo, G., and T. Chiricos, "Work Release and Recidivism: An Empirical Evaluation of a Social Policy," in *Evaluation Studies Review Annual,* Vol. 2, pages 623-644. Beverly Hills, California: Sage Publications. 1977.

Weiss, C. *Evaluation Research.* Englewood Cliffs, New Jersey: Prentice Hall. 1972.

Glossary

GLOSSARY

Consensus building outcomes
The production of common understanding regarding juvenile justice issues and programs among various participants.

Cost benefit analysis
An investigation designed to assess the relationship between program monetary costs and outcomes.

Evaluation research
The application of social research methods to assess the activities, processes, and outcomes of intervention and/or treatment programs.

External validity threats
Factors which may reduce program's findings transferability to other groups or jurisdictions.

Formula grantee
The state agency designated to receive and administer the Formula Grants Program monies of the Juvenile Justice and Delinquency Prevention Act.

Formula Grants Program
A provision of the Juvenile Justice and Delinquency Prevention Act which offers federal money to state agencies for pass-through to local jurisdictions. The program is designed to assist efforts aimed at complying with Act mandates and reducing juvenile justice and delinquency problems in the states.

Goals
General statements regarding the desired impact of a program intervention strategy.

History effects
A threat to internal validity in which changes to the participants over time can produce changes to the variables under investigation.

Instrumental outcomes
Measures of phenomena directly related to program goals and objectives.

Intermediate program effects
The short term effects of program intervention which will impact the overall goals of the program and often have a causal effect on the long term effects or outcomes.

Internal validity threats
Factors other than program participation which may impact the results or findings.

Juvenile Justice and Delinquency Prevention Act
As amended in 1988, this Federal act requires participating states to meet certain mandatory requirements regarding the processing of juvenile offenders and nonoffenders, and provides money for programs designed to improve state and local juvenile justice systems and delinquency prevention efforts.

Juvenile justice specialists
The Formula Grantee staff persons who oversee their states' participation in the Juvenile Justice and Delinquency Prevention Act.

Knowledge production outcomes
The generation of new knowledge or understanding about juvenile justice programs.

Mapping
The process of formalizing a state's overall juvenile justice plan by identifying target areas and programs, establishing goals, objectives, and timelines, and identifying resources, and more. The final product is used to identify areas where evaluation research will play a role.

Maturation effects
A threat to the internal validity of an evaluation in which observed outcomes are a result of natural changes of the program participants over time, rather than because of program impact.

Monitoring, basic
Developing and analyzing data to count and/or identify specific program activities and operations.

Monitoring, comparative
A monitoring process in which data is observed for an intervention program/population and a control or comparison program/population.

Mortality threats
A threat to the internal validity of an evaluation in which the effects of the program on participants who withdraw or dropout prior to program conclusion are not measured.

Objectives
Specific, measurable statements regarding the desired outcome of an intervention program.

Office of Juvenile Justice and Delinquency Prevention
An agency within the federal Department of Justice which is responsible for oversight of the Juvenile Justice and Delinquency Prevention Act.

Outcome evaluation, basic
Developing and analyzing data to assess program impact and effectiveness.

Outcome evaluation, comparative
An outcome evaluation in which long term outcome measures are collected for an intervention program/population and a control or comparison program/population.

Performance report
A report submitted annually by state Formula Grantees to the Office of Juvenile Justice and Delinquency Prevention which summarizes the state's progress towards the goals of the Juvenile Justice and Delinquency Prevention Act and the extent to which programs funded with Formula Grants money have contributed to reductions in justice system and delinquency problems.

Process evaluation, basic
Developing and analyzing data to assess program processes ad procedures, esp., determining the connections between various program activities.

Process evaluation, comparative
A process evaluation in which data are collected for the intervention program/population and a control or comparison program/population.

Program failure
A program shortcoming in which the outcome criterion are not effected by participation of the subjects in the program, i.e., the program does not accomplish its objectives.

Random assignment
Placement of study subjects into an experimental treatment or program group and a control group, using a random, or unbiased, assignment methodology.

Recidivism
The repetition of criminal or delinquent-type behavior.

Request for Proposal (RFP)
An open solicitation to potential grantees or contractors inviting them to compete for money available to develop or evaluate programs.

Selection effects
A threat to the internal validity of a program in which program participants are not properly selected or statistically representative.

Self-reported data
Information used to assess program processes or outcomes in which the program participants generate the information themselves.

State Relations and Assistance Division
The division within the Office of Juvenile Justice and Delinquency Prevention which is responsible for oversight of the Juvenile Justice and Delinquency Prevention Act's Formula Grants program.

Survey
The collection of information from a common group through interviews or the application of questionnaires to a representative sample of that group.

Theory failure
A program shortcoming in which the
intermediate program effects succeed as
planned but the outcome criterion
remains unchanged.

Validity
A finding or observation regarding the
accuracy of an evaluation, esp., the
assurance that alternative explanations for
the findings can be discarded. *See also*
external validity threats *and* internal
validity threats.

Index

Index