CHAPTER

11 Policy Analysis and Evaluation Research

In the introductory chapter, we addressed the criticism that much criminological and criminal justice research is either common sense or impractical. This chapter focuses on the latter concern: "So what; of what practical use are these research findings?" We will apply what we have learned to the tasks of policy analysis and evaluation research—the cutting edge of government-sponsored criminal justice research today.

POLICY ANALYSIS

Policy analysis

Policy analysis is the "study of whatever governments choose to do or not to do," "the description and explanation of the causes and consequences of government behavior" (Dye, 1995, pp. 3–4). Jones (1977, p. 4) views policy analysis as the study of proposals (specified means for achieving goals), programs (authorized means for achieving goals), decisions (specified actions taken to implement programs), and effects (the measurable impacts of programs). Policy analysis is an

applied subfield of economics, political science, public administration, sociology, law, and statistics. It involves the identification and description of social problems, the development of public policies that may alleviate these problems, and determination of whether these policies work (Dye, 1995, p. 17). Although there are many models, perspectives, and approaches to policy analysis, the policy process could be viewed as a series of political activities consisting of the following:

Identifying problems	Demands are expressed for government action.
Formulating policy proposals	Agenda is set for public discussion.
	Development of program proposals to resolve problem.
Legitimating policies	Selecting a proposal.
	Building political support for it.
	Enacting it as a law.
Implementing policies	Organizing bureaucracies.
	Providing payments or services.
	Levying taxes.
Evaluating policies	Studying programs.
	Reporting "outputs" of government programs.
	Evaluating "impacts" of programs on target and nontarget groups in society.
	Suggesting changes and adjustments (Dye, 1995, p. 21).

Thus the policy process involves identification, formulation, legitimation, implementation, and evaluation.

EVALUATION RESEARCH

Evaluation research is the last stage of the policy process; questions such as the following are asked:

Do the programs work?

Do they produce the desired result?

Do they provide enough benefits to justify their costs?

Are there better ways to attack these problems?

Should the programs be maintained, improved, or eliminated?

Evaluation research

Evaluation research is an applied branch of social science that is intended to supply scientifically valid information with which to guide public policy. Historically, research in the social sciences had its origins in the physical sciences and was oriented toward development of theories and utilization of the experimental model to test those theories. Its concern was much more akin to pure or basic research discussed in Chapter 1—the acquiring and testing of new knowledge.

Evaluation research as a type of applied research has different roots as well as intentions. It evolved from the world of technology rather than science and emphasizes mission or goal accomplishment and product/service delivery rather than theory formation. Evaluation research aims to provide feedback to policy makers in concrete and measurable

terms. Although such an approach has existed informally since early times, the introduction of computer technology in the 1950s and its successful application to "defense systems" and "space systems" have led to the application of evaluation research to "social systems" such as the "criminal justice system." Much of this thinking grew out of the "Planning, Programming, Budgeting Systems" (PPBS) approach originally employed by the U.S. Department of Defense in the 1960s, a method of policy evaluation widely adopted by other government agencies. PPBS attempts to specify (by clearly defining program objectives) and quantify (by developing measures of accomplishments) the output of a government program and to analyze the relative costs and benefits of the program (see Rossi and Freeman, 1993).

As billions of dollars were poured into social programs in the 1960s, the following questions were increasingly asked: Do the programs work or make a difference? Are they cost effective? Are they the most efficient method of providing services? With fewer funds available at the turn of the century, the same questions are still relevant: How can the best use be made of limited resources to accomplish maximum program benefits?

Other than its very practical bent and some relatively esoteric techniques such as cost–benefit analysis, many of the methodological procedures employed in evaluation research have already been covered earlier in this text in Chapters 1–10. Thus, rather than viewing it as a different type of research, readers can confidently assume that they can master the essentials of evaluation research on the basis of knowledge of many of the issues we have already described. Quite simply, *evaluation research can be defined as measurement of the effects of a program in terms of its specific goals, outcomes, or particular program criteria.* Weiss (1972, p. 4) states that the purpose of evaluation research is "to measure the effects of a program against the goals it set out to accomplish as a means of contributing to subsequent decision making about the program and improving future programming." It is essential to this purpose that the research methodology we have discussed be used to measure program outcomes in terms of specifically identified criteria in order to accomplish an applied or practical research objective—better programs. Similar to a scientific experiment, the research methodology is applied to evaluate social action programs to accomplish more efficient programs (Schwarz, 1980).

The National Advisory Committee on Criminal Justice Standards and Goals feels very strongly about the importance of evaluation research:

> A high quality evaluation is expensive and time-consuming. Indeed, it may be many times more expensive than the operational program it is designed to test. Viewed in the context of that single program, such an expenditure may appear absurd. But in the context of advancement of knowledge, this type of concentration of funds is more likely to be fruitful than the same expenditure on a large number of inadequate evaluations would be. Progress does not depend on every program being evaluated; in fact, with limited resources for evaluation, it may be retarded by such a practice. (National Advisory Committee, 1976, p. 52)

Some workers involved in administering applied or action programs in criminal justice may have either little understanding of evaluation, past exposure to poor evaluations, or perhaps little regard for the necessity of evaluation as they are already committed to a particular programmatic strategy. The logic of the National Advisory Committee statement would argue that a few expensive, well-designed evaluations are in the long run more cost-effective in revising or eliminating unnecessary treatments or procedures. The last point—elimination—is perhaps at the

crux of the resistance to evaluations. Similar to early applications of social, scientific, and managerial studies in industry, many of those to be studied obviously have an understandably vested interest in maintaining a favorable image of the current procedures, practices, and staffing of their organizations.

Policy Experiments

Policy experiments

A close link between experimental methods and the assessment of public policy programs has increased dramatically since 1970 (Fagan, 1990, p. 108). **Policy experiments** are applied field experiments that address themselves to immediate practical policy questions. The National Research Council's Committee on Research on Law Enforcement and the Administration of Justice summarized the following steps in designing policy experiments (Garner and Visher, 1988, pp. 7–8):

1. Choose an interesting problem—a policy question that people really care about or an existing procedure that clearly needs improvement.
2. Do some creative thinking to solve legal and ethical issues that may arise.
3. Rigorously maintain the random assignment of persons, cases, or other units into treatment and control groups throughout the experiment.
4. Choose a design and methods of investigation that are appropriate both to the questions to be answered and to the available data.
5. Adopt a team approach between researchers and practitioners and keep working in close cooperation.
6. Put as much into your experiment as you want to get out of it.
7. Use an experiment to inform policy, not to make policy.
8. Understand and confront the political risks an experiment may involve.
9. Insofar as possible, see that the experiment is replicated in a variety of settings before encouraging widespread adoption of experimentally successful treatments.

Before exploring evaluation research more thoroughly, let us first provide an example of a policy analysis program that utilizes evaluation research.

POLICY ANALYSIS: THE CASE OF THE NATIONAL INSTITUTE OF JUSTICE RESEARCH PROGRAM

Although policy analysis and evaluation research in criminology and criminal justice are not restricted solely to government-funded research of primarily government-funded projects, and the National Institute of Justice (NIJ) is not the only agency sponsoring criminal justice research, NIJ does utilize the largest, most ambitious policy-oriented program of its type and has been heralded by the National Academy of Sciences as a pioneer and model for other programs. For this reason, we explore the philosophy, aims, and research program plan of the NIJ.

NIJ Mission Statement

The NIJ is a research branch of the U.S. Department of Justice. The Institute's mission is to develop knowledge about crime, its causes, and methods of controlling it. Priority is given to policy-relevant research that can yield approaches and information that state and local agencies can use in preventing and reducing crime. The decisions made by criminal justice practitioners

and policymakers affect millions of citizens, and crime affects almost all our public institutions and the private sector as well. Targeting resources, assuring their effective allocation, and developing new means of cooperation between the public and private sector are some of the emerging issues in law enforcement and criminal justice that research can help illuminate.

Carrying out the mandate assigned by Congress in the Justice Assistance Act of 1984, the NIJ aims to:

- Sponsor research and development to improve and strengthen the nation's system of justice with a balanced program of basic and applied research.
- Evaluate the effectiveness of criminal justice and law enforcement programs, and identify those that merit application elsewhere.
- Support technological advances applicable to criminal justice.
- Test and demonstrate new and improved approaches to strengthen the justice system.
- Disseminate information from research, development, demonstrations, and evaluations (NIJ, 1994, p. 1).

In establishing its research agenda, the Institute is guided by the priorities of the Attorney General and the needs of the criminal justice field. The Institute actively solicits the views of police, courts, and corrections practitioners as well as the private sector to identify the most critical problems and to plan research that can help resolve them. Recent priorities include:

- Reducing violent crime
- Reducing drug and alcohol-related crime
- Reducing the consequences of crime
- Improving the effectiveness of crime prevention programs
- Improving law enforcement and the criminal justice system
- Developing new technology for law enforcement and the criminal justice system

Studies that involve the use of randomized experimental designs are encouraged, as are multiple strategies for data collection and well-controlled, quasi-experimental designs and equivalent comparison group designs. Qualitative studies, including ethnographic data collection, are also encouraged (NIJ, 1994, p. 2).

NIJ Research Priorities

Some recent research priorities of NIJ (NIJ, 2004) include:

Violence and other criminal behavior

Sex offenders/offenses

Crime and delinquency prevention

Child abuse and neglect

Juvenile delinquency

Policing practices, organization, and administration

Terrorism or counterterrorism

Drugs, drugs and crime/alcohol, and drug testing

Drug treatment

White-collar crime/cybercrime
Transnational crime, organized crime
Justice systems
Courts, prosecution, and defense
Corrections
Offender programs and treatment
Crime mapping and spatial analysis
Other thematic areas

A SYSTEMS MODEL OF EVALUATION RESEARCH

Systems model

Although a variety of terms and competing models of evaluation research exist, the "systems model" is presented here to acquaint the reader with a general evaluation approach. A *model* is a simplified schema that outlines the essential points of a theory. A **systems model** assumes that all parts of an organism, organization, or program are interrelated and could be represented in basic computer language as a system of inputs into an existing system, processing of these inputs, and subsequent outputs (or outcomes). Figure 11.1 presents a systems model for evaluating programs in the criminal justice system.

The project components to be evaluated in this model are inputs, activities, results, outcomes, and feedback (Schneider, 1978, pp. 3,23–3,31):

Inputs — Resources, guidelines, rules, and operating procedures provided for a program, for example, funds for personnel, equipment, operating costs, and authorization to introduce new policies (often an experimental treatment)

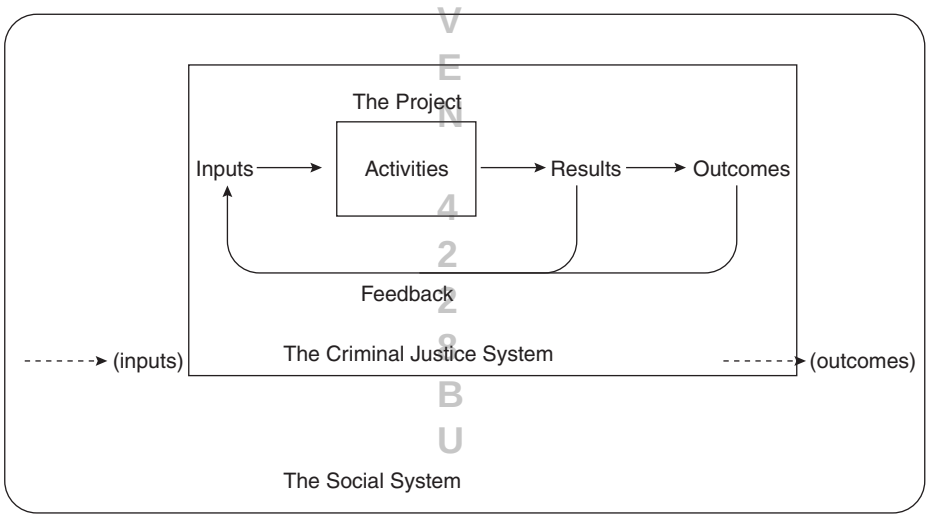

FIGURE 11.1 A Systems Model of Evaluation Research: System and Project Components. *Source*: Schneider, Anne L., et al. *Handbook of Resources for Criminal Justice Evaluators.* Washington, D.C.: U.S. Department of Justice, 1978, pp. 3–24.

Activities	What is done in the project with these inputs (resources), for example, services provided, staffing patterns, and use of materials and human and physical resources (called "process" in many models)
Results	*Specific* consequences of the project activities or the specific objectives of the program, for example, amount of services provided, work completed, production accomplished, or cases closed or cleared (called "output" or "products" in many models)
Outcomes	Accomplishment of broader-range societal goals; these are general consequences of the specific accomplishments (outputs/results) of the program, for example, better justice, health, safety, and education
Feedback	Recycling of results/outcomes into the operation as additional (or modified) inputs; profits may induce a corporation to reinvest in a particularly profitable line, just as losses may lead it to eliminate a less profitable line (also called "feedback loop")

Inputs and process represent specific organizational/program *efforts*, and outputs represent specific organizational/program *results*. Outcomes represent impacts on general, external societal activities. Note this *very* simple illustration:

Input	Grant of $100,000 for a foot patrol program
Process	Two officers assigned to foot patrol in Precinct A for one year
Results	Fifty percent increase in arrests in Precinct A
Outcome	Crime rate declines 10 percent and fear of crime declines 40 percent
Feedback	Allocate $1,000,000 and twenty officers to expanded foot patrol program

To summarize Figure 11.1,

> In this scheme, a criminal justice project is conceived of as a system consisting of *inputs* (resources, guidelines, and operating procedures); *activities* (those things the project and its personnel do); *results* (the initial consequences of the activities); and *outcomes* (the long-range, socially relevant consequence of the project). The system should contain a *feedback* loop through which the results and outcomes of a project impact upon the operation of the project and act as additional inputs. (Schneider et al., 1978, pp. 3–8)

TYPES OF EVALUATION RESEARCH

With the evolution and growth of evaluation research as a field has come a whole lexicon of descriptive tags. Franklin and Thrasher (1976), for instance, mention a variety of research approaches as they relate to evaluation: continuous-versus-one-shot evaluations, "hip pocket"-versus-formal evaluations, policy research, applied research, decision-oriented research, social audits, action research, operations research, discipline-related research, basic research, frontline evaluations, utilization reviews, and continuous monitoring and quality control. Unfortunately, many of these terms are used interchangeably by various writers, and there is no consistent agreement on their meaning in the field. Even the terms *policy analysis* and *evaluation research* are often used as synonyms.

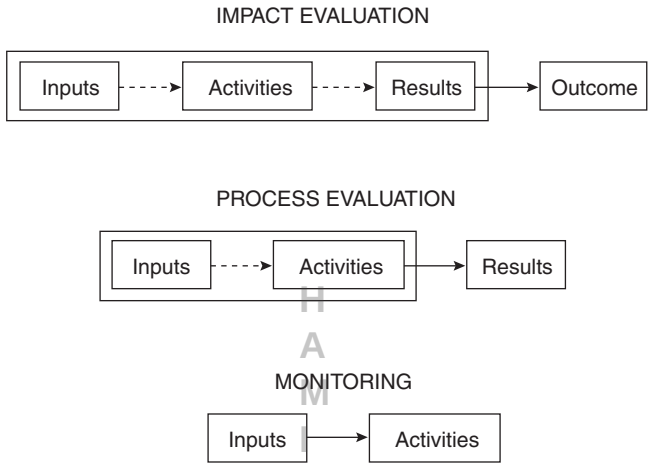

FIGURE 11.2 Types of Evaluation. *Source*: Schneider, Anne L., et al. *Handbook of Resources for Criminal Justice Evaluators.* Washington, D.C.: U.S. Department of Justice, 1978, pp. 3–33.

Evaluation research is different from other types of applied research in that the data are used to make a decision(s) regarding a specific program, rather than simply to represent findings of theoretical interest. Although numerous types have been identified, there are *two general types of evaluation research*: process evaluation and impact evaluation. In most instances, it is the latter term by which "evaluation research" is most often described in references. **Process evaluation** establishes causal relationships between results (such as an increase in arrests) and project inputs and activities (see Figure 11.2).

Process evaluation

Impact evaluation establishes causal relationships between outcomes (such as crime reduction) and inputs, activities, and results of programs.

Impact evaluation

Evaluation research is often confused with two related information-gathering activities: assessment and monitoring. **Assessment** (sometimes called needs assessment) is the enumeration of some activity or resource, for instance, the need for a particular service in some target area. "It is a method of finding service delivery gaps and substantiating unmet needs in a community and is used to establish priorities for addressing problems" (Office of Juvenile Justice, 1978, p. 2).

Assessment

Monitoring is assessment of whether the plans for a project have in fact been realized: Are the activities related to the inputs? Monitoring is similar to an audit, an assessment of program accountability: Is the program doing what it is supposed to be doing (Waller et al., 1975)? A certain portion of the operating budget of an organization might be set aside to fund such a monitoring task.

Monitoring

Evaluation research need not be restricted to solely an analysis of output; it can involve any systematic assessment of various aspects of program review (Suchman, 1967). Effort, efficiency, operation, effectiveness of performance, adequacy of performance, and the like can all be subject to evaluation (Office of Juvenile Justice, 1978, p. 3). *Before an evaluation is undertaken*, it is important that it be decided whether an evaluation can and should be done. According to the Office of Juvenile Justice and Delinquency Prevention (1978, p. 508), *three crucial questions must be answered*:

Will the findings be used?

Is the project evaluable?

Who can do this work?

Will the Findings Be Used?

Evaluation research has an applied quality to it that *requires the active support and cooperation of the agency or program to be evaluated.* Levine, Musheno, and Palumbo (1980, p. 551) put the matter succinctly: "The program administrator's desire to reaffirm his or her position with favorable program evaluations may conflict with the social scientist's desire to acquire an objective appraisal of a program's impact. The end result may be either a research design with low scientific credibility and tainted results, or a credible study that never receives a public hearing because the administrator does not like the results" (Levine, Musheno, and Palumbo, 1980, p. 551). Unless a sincere need for the research has been expressed by the agency administrators and the effort is viewed as something other than a public relations plume, evaluation research may become nothing more than a sham. In discussing problems with contract research, in which the researcher is paid by the contractor, Punch (1986, p. 73) says, "Having paid the piper they want copyright on the tune." This reflects concern that academic reliance on commercial funds may damage academic freedom. Grinyer (1999) suggests that researchers should think through potential questions before agreeing to contract research. These include:

- What happens if the client does not like the research findings?
- What ethical issues are raised by the client becoming the subject of the research?
- If the client objects to the findings, what are the implications for publication?

Is the Project Evaluable?

In asking *whether the project is capable of being evaluated*, the researcher is concerned with the existing design, defined objectives, and other programmatic elements that enable the measurement and assessment of specified criteria. For instance, if the purpose of the program is simply defined as "to do good" and no objectives, records, or other evaluable materials are kept by the organization, much grief can be saved by avoiding an evaluation of this particular organization. The success of the entire evaluation process hinges on the motivation of the administrators and organization in calling for an evaluation in the first place (Schulberg and Baker, 1977).

It should be possible to locate specific organizational objectives that are measurable. "The key assumptions of the program must be stated in a form which can be tested objectively. That is, not only must the outcome be definable, but also the process used to achieve it must be specifiable" (Office of Juvenile Justice, 1978, p. 7). If proper data for evaluation are absent and clear outcomes or criteria of organizational "success" are absent, then a proper evaluation cannot be undertaken. Rutman (1977) refers to this process as "formative research," a reconnaissance operation to determine program evaluability. Wholey (1977) suggests the following steps in **evaluability assessment** (assessing whether the program is evaluable):

Evaluability assessment

1. Bounding the problem or program or determining what the objectives of the program are and where it fits in the service picture
2. Collecting program information that defines its activities, objectives, and assumptions
3. Modeling of the program and the interrelationships of program activities
4. Analyzing plans or determining whether the model and activities are measurable
5. Presenting to management (intended user) or reporting results of evaluation assessment and determination of the next steps to be taken

Rabow (1964, p. 69), in speaking specifically to corrections research, suggests that before any results are attributed to a particular treatment, the evaluation should address important questions, as outlined in the three stages of Rabow's research model.

Stage I is concerned with the population of offenders from which treatment and control groups will be selected.

1. How is the population of offenders from which groups will be selected defined with respect to age, record of offenses, geographical location, or any social or personality characteristics thought to be important?
2. How is selection carried out in order to eliminate bias—by random means or some matching process?
3. When and by whom is selection carried out? What are the mechanics?
4. What steps are taken to demonstrate the lack of bias in selection?

Stage II is concerned with the treatment process and the need to understand what is involved in it.

1. What is the theory of causation upon which treatment is proceeding?
2. What is the intervention strategy utilized in the treatment by which the causation variables will be modified?
3. Can a logical relationship between causation variables and intervention strategy be demonstrated?
4. Can it be demonstrated that the treater is fulfilling role requirements specified by the intervention strategy?
5. Assuming that treatment role requirements are being fulfilled, can it be demonstrated that variables cited in the theory of causation are being modified?
6. How shall any change in the variables be measured?

Stage III involves actual comparisons of groups subsequent to treatment.

1. What are the goals of treatment, that is, how shall success be defined in terms of recidivism, attitudinal change, new social relationships, and personality modification?
2. How is the measurement of these characteristics carried out?
3. Over what period of time are comparisons to continue?
4. How is the cooperation of subjects outlined?

Who Can Do This Work?

In asking "*Who can do this work?"* one must decide on internal or external evaluators. If the evaluation is to be "in-house," that is, to be conducted by the internal staff of the agency to be evaluated, then adequate time and manpower must be allocated to permit a careful and hopefully objective evaluation. Outside evaluators may lend greater objectivity to the evaluation task but, as we will discuss later, require orientation to, and cooperation of, the agency to address the relevant objectives and goals from a policy perspective.

STEPS IN EVALUATION RESEARCH

The actual *steps in evaluation research* do not differ significantly from the basic steps in the research process that were identified in Chapter 1:

Problem formulation

Design of instruments

Research design (evaluation model)

Data collection

Data analysis

Findings and conclusions

Utilization

Only in the last step does evaluation research differ significantly from other research processes. There are, of course, a variety of ways of slicing a pie, but most alternative listings of steps one way or another include the key elements we have identified above. For instance, Albright et al. (1973), in *Criminal Justice Research: Evaluation in Criminal Justice Programs: Guidelines and Examples*, an evaluator's manual prepared on behalf of the National Institute of Law Enforcement and Criminal Justice (now NIJ), focus primarily on the data collection and analysis stages. They outline five essential **steps in evaluation planning** (Albright et al., 1973, p. 5):

Steps in evaluation planning

Quantify the objectives and goals

Determine a quantifiable objective/goal relationship

Develop evaluation measures

Develop data needs considering requirements, constraints, and reporting

Determine methods of analysis

These steps would be assumed or included in the design of instruments, research design, data collection, and data analysis stages that we have discussed throughout this text.

Problem Formulation

Just as in the other types of research we have discussed, evaluation researchers are also often in a hurry to get on with the task without thoroughly grounding the evaluation in the major theoretical issues in the field. Glaser (1974) feels that evaluation research in criminal justice would be more useful were it to differentiate offenses and offenders utilizing causal theory. Without this theoretical grounding, familiarization with past and current literature, and valid operationalization of concepts, many evaluation studies can easily deteriorate into glorious exercises in social accounting.

Glaser (1973) comments on how much of what is regarded as in-house evaluations in correctional agencies has been co-opted and is little more than head counting or the production of tables for annual reports.

The problem formulation stage, to reiterate a point that has been emphasized throughout this text, is the most crucial stage of research.

Design of Instruments

On the basis of problem formulation, review of the relevant literature, and program reconnaissance, a most important element in evaluation research is the identification and operationalization of key components of the program to be analyzed. The National Advisory Committee on Criminal Justice Standards and Goals (1976, p. 113) suggests that professional associations be commissioned to establish standardized definitions based on the following:

> A major problem in research on criminal justice organizations is the absence of standardized definitions for such basic terms as dangerousness, recidivism, discretion, disparity, equity, proportionality, uniformity, individualization, commitment sentence,

> probation, parole and length of follow-up. The confusion over definitions has not only impeded communication among researchers and, more importantly, between researchers and practitioners, but also has hindered comparisons and replications of research studies. R&D-funding agencies, such as the National Institute of Law Enforcement and Criminal Justice and the National Institute of Mental Health, should be sensitive to the way in which the terminology is used in the research studies being supported. Where appropriate, the use of common definitions can facilitate the direct comparison of research findings and, hence, the aggregation of research knowledge. For example, the development of standardized definitions has already occurred in the use of some identically worded questions in victimization surveys.

The greater use of replication of instruments employed by others can contribute to more confidence in the reality and validity of evaluation methodologies, as well as to more useful cross-site comparisons.

Research Design

Ideally, researchers would prefer control over treatment and a classic experimental design, with random assignment of cases to experimental and control groups. Seldom does the evaluation researcher enjoy such a luxury in analyzing ongoing programs. Despite arguments to the contrary (see Boruch, 1976), in many instances, it is very difficult to find organizations that would be willing to undergo experimentation, particularly if it involves the denial of certain treatments (control group) to some clients. Cook, Cook, and Mark (1977) describe some *problems related to the attempt to use randomized designs in field evaluations*:

1. The program planners and staff may resist randomization as a means of allocating treatments, arguing for assignment based on need or merit.
2. The design may not be correctly carried out, resulting in nonequivalent experimental and control groups.
3. The design may break down as some people refuse to participate or drop out of different treatment groups (experimental mortality).
4. Some feel that randomized designs create focused inequity because some groups receive treatment others desire and thus can cause reactions that could be confused with treatments.

Strasser and Deniston (1978) distinguish between preplanned and postplanned evaluations. Although the former may interfere with ongoing program functioning, the latter is less costly, involves less interference in the organization, and is less threatening to the personnel being evaluated. Much of the bemoaning concerning the inadequacy of research design in evaluation methodology in criminal justice has arisen because of an overcommitment to experimental designs and a deficient appreciation of the utility of post hoc controls by means of multivariate statistical techniques (see, for instance, Cain, 1975; Posavec and Carey, 1992).

Logan (1980, p. 36) agrees with this point when he states:

> It may be that more rapid progress can be made in the evaluation of preventive or correctional programs if research designs are based on statistical rather than experimental model. It was noted, above, that one major difficulty in evaluation research is in procuring adequate control groups. Modern statistical techniques can provide a means of resolving this problem by substituting statistical for experimental methods of control.

Data Collection

One principal shortcoming of much evaluation research has been its overreliance on questionnaires as the primary means of *data gathering.* The use of a triangulated strategy of data collection employing multiple methodologies would assure greater confidence in the validity of findings (see, for instance, Fry, 1973). Where possible, agencies' records as outcome measures should be cross-checked against other data sources. Many of the issues discussed previously in this text are, of course, also appropriate to evaluation research. All of the sources of error, particularly in data collection, must be continually checked, to ensure that the findings are true findings and not the result of measurement error. Schwarz (1980, p. 14) presents the issue succinctly:

> In practice, the cup seldom reaches the lip intact. Designs must be compromised. There are mishaps in the field. Expecting both valid results and an impeccable process is overly optimistic. The most that can be expected is that the findings will be valid despite compromise and mishaps. Flaws cannot be avoided.

Although program supporters will jump on methodological or procedural problems in any evaluation that comes to a "negative" conclusion, Schwarz echoes a theme that has been emphasized throughout this text: There is no such thing as research without error. The only way to avoid error is to do no research at all.

MacKenzie and McCarthy (1990, p. 8) indicate that criminal justice researchers should not ignore secondary analysis, nor should they be afraid to reanalyze data previously collected by someone else. Two particularly important sources for such data are the National Archive of Criminal Justice Data (formerly the Criminal Justice Archive and Information Network [CJAIN]) and the National Center for Juvenile Justice (NCJJ). National Archive of Criminal Justice Data databases include many classic and well-known criminal justice studies, as well as data from recent NIJ-sponsored studies. NCJJ archives data on juvenile justice system transactions in about half of the states.

Data Analysis

The choice of appropriate statistical analysis must be based on whether the data meet the assumptions necessary for each technique to be employed. An important additional consideration is pointed out by Glaser (1976, p. 771):

> Some research reports from correctional agencies are not suppressed, but might as well be, for few officials—or even researchers—can understand them. Most notable among such reports are those which describe the use of various types of multiple correlation or multiple association statistical analysis of case data in administrative records to find guides for correctional operations. These reports are submitted to correctional officials who do not understand the statistical terminology and who feel no urgency to learn to understand it since the researchers share with the operations officials the impression that this statistical analysis has little or no practical value at present. Thus these researchers operate in a separate world, inadequately linked either with the university social system which seems to be their reference group, or with the leaders of the correctional system, which they are presumed to serve.

What might be excellent choices of statistical analysis for professional or academic purposes may not be appropriate in form for presentation to a lay audience. Recall in the first chapter of this text the point that, unlike the chemist or physicist, the criminologists must compete with "commonsensical" views and explanations and, unfortunately, must often pitch their evidence toward the lowest common denominator. How, then, can the evaluation researcher in criminal justice resolve this dilemma of treating data with the most appropriate and rigorous statistical methodology they require, however esoteric, yet attempting to communicate these findings so that even politicians would understand? A useful practice is to perform the evaluation and write a report geared for a professional audience and then issue a *report for laypeople*, in which the crucial findings are simplified, summarized, and understood by nonresearch professionals. In writing such reports, the researcher may take license in generalizing findings, but it is exactly this succinct presentation that is usually viewed as most useful by the consumer. Instead of the results of stepwise multiple regressions and intercorrelation matrices, the critical relationships or statistically significant findings could be presented in simple bivariate tables, which are more easily understood by more people. An interesting exercise is boiling down the entire evaluation report to a two-page summary, the type that might be released as a press report. Although, of course, such a brief document does not do justice to the complexity of the analysis, anyone desiring the details can consult the full report.

Utilization

Previous points, particularly with respect to data analysis, have a direct bearing on the *utilization of evaluation findings.*

In discussing the "politicization of evaluation research," Maida and Faucett (1978) point out the increasing political nature of evaluations as they are increasingly used to decide the future of programs. Adams describes the dilemma of the agency administrator who is to be evaluated:

> Part of the administrator's concern about evaluative research comes from the dilemma that research creates for him. The evaluation process casts him in contradictory roles. On the one hand, he is the key person in the agency, and the success of its various operations, including evaluation, depends on his knowledge and involvement. On the other hand, evaluation carries the potentiality of discrediting an administratively sponsored program or of undermining a position the administrator has taken. (Adams, 1975, p. 19)

Factors that limit the utilization of evaluation research findings in criminal justice are much the same obstacles that prevent effective evaluation research.

WHAT WORKS IN CRIMINAL JUSTICE?

If we ran GE, GM, or GTE the way we sometimes run our criminal justice systems, they would all be out of business. Ford would still be making Edsels. A revolution has taken place in criminal justice at the dawn of the twenty-first century. Let us find out what works in criminal justice, what is promising, and what does not work. About thirty years ago, Robert Martinson (1974) rocked the correctional community after reviewing over a hundred programs and concluding that "nothing works." It turns out that Martinson was wrong; some programs do work, but how do we know?

In 1996, the U.S. Congress required the Attorney General to provide a "comprehensive evaluation of the effectiveness" of over $3 billion annually in Department of Justice grants to assist state and local law enforcement and communities in preventing crime. Congress required that the research for the evaluation be "independent in nature" and "employ rigorous and scientifically recognized standards and methodologies." The Assistant Attorney General for the Office of Justice Programs asked the NIJ to commission an "independent review" of over 500 program impact evaluations.

The congressionally mandated evaluation examined hundreds of different strategies used in communities, families, schools, labor markets, places, and police and criminal justice settings (Sherman et al., 1997). It found that very few operational crime prevention programs have been evaluated using scientifically recognized standards and methodologies, including repeated tests under similar and different social settings. Based on a review of more than 500 prevention program evaluations meeting minimum scientific standards, the report (ibid.) concluded that there is minimally adequate evidence to establish a provisional list of what works, what does not, and what is promising. Exhibit 11.1 lists each of these.

EXHIBIT 11.1

Preventing Crime: What Works, What Doesn't, What's Promising

What Works?

- **For infants**: Frequent home visits by nurses and other professionals.
- **For preschoolers**: Classes with weekly home visits by preschool teachers.
- **For delinquent and at-risk preadolescents**: Family therapy and parent training.
- **For schools**:
  Organizational development for innovation.
  Communication and reinforcement of clear, consistent norms.
  Teaching of social competency skills.
  Coaching of high-risk youth in "thinking skills."
- **For older male ex-offenders**: Vocational training.
- **For rental housing with drug dealing**: Nuisance abatement action on landlords.
- **For high- hot spots**: Extra police patrols.
- **For high-risk repeat offenders**: Monitoring by specialized police units.
  Incarceration.
- **For domestic abusers who are** : On-scene arrests.
- **For convicted offenders**: Rehabilitation with risk-focused treatments.
- **For drug-using offenders in prison**: Therapeutic treatment programs.

What Doesn't Work

- Gun "buyback" programs.
- Community mobilization against crime in high-crime poverty areas.
- Police counseling visits to homes of couples days after domestic violence incidents.
- Counseling and peer counseling of students in schools.
- Drug Abuse Resistance (DARE).
- Drug prevention classes focused on fear and other emotional appeals, including self-esteem.
- School-based leisure-time programs.
- Summer jobs or subsidized work for at-risk youth.
- Short-term, nonresidential training programs for at-risk youth.
- Diversion from court to job training as a condition of dismissal.
- Neighborhood watch programs organized with police.
- Arrests of juveniles for minor offenses.
- Arrests of unemployed suspects for domestic.
- Increased arrests or raids on drug market locations.
- Storefront police offices.
- Police newsletters with local crime -formation.

(*continued*)

EXHIBIT 11.1 (Continued)

- Correctional boot camps using traditional basic training.
- "Scared Straight" programs whereby minor juvenile visit adult prisons.
- Shock probation, shock parole, and split sentences adding jail to probation or parole.
- Home detention with electronic monitoring.
- Intensive supervision on parole or probation (ISP).
- Rehabilitation programs using vague, unstructured counseling.
- Residential programs for juvenile offenders using challenging experiences in rural settings.

What's Promising?

- Proactive drunk driving arrests with breath testing (may reduce accident deaths).
- Community policing with meetings to set priorities (may reduce perceptions of crime).
- Police showing greater respect to arrested offenders (may reduce repeat offending).
- Police field interrogations of suspicious persons (may reduce street crime).
- Mailing arrest warrants to domestic violence suspects who leave the scene before arrive.
- Higher numbers of police officers in cities (may reduce crime generally).
- Gang monitoring by community workers and probation and police officers.
- Community-based mentoring by Big Brothers/Big Sisters of America (may prevent abuse).
- Community-based afterschool recreation programs (may reduce local juvenile crime).
- Battered women's shelters (may help some women reduce repeat domestic violence).
- "Schools within schools" that group students into smaller units (may pre-vent crime).
- Training or coaching in "thinking" skills for high-risk youth (may prevent crime).
- Building school capacity through organizational development (may prevent substance abuse).
- Improved classroom management and instructional techniques (may reduce alcohol use).
- Job Corps residential training programs for at-risk youth (may reduce felonies).
- Prison-based vocational educational programs for adult inmates (in federal prisons).
- Moving urban public-housing residents to suburban homes (may reduce risk factors for crime).
- Enterprise zones (may reduce area unemployment, a risk factor for crime).
- Two clerks in already-robbed convenience stores (may reduce robbery).
- Redesigned layout of retail stores (may reduce shoplifting).
- Improved training and management of bar and tavern staff (may reduce violence, DUI).
- Metal detectors (may reduce skyjacking, weapon carrying in schools).
- Street closures, barricades, and rerouting (may reduce violence, burglary).
- "Target hardening" (may reduce vandalism of parking meters and crime involving).
- "Problem-solving" analysis unique to the crime situation at each location.
- Proactive arrests for carrying concealed weapons (may reduce gun crime).
- Drug courts (may reduce repeat offending).
- Drug treatment in jails followed by urine testing in the community.
- Intensive supervision and aftercare of juvenile offenders (both and serious).
- Fines for criminal acts.

Source: Lawrence W. Sherman, et al. *Preventing Crime: What Works, What Doesn't, What's Promising*. Washington, D.C.: Office of Justice Programs, 1997, NCJ 165366.

The clearinghouse for these evaluations had been contracted to the University of Maryland by the NIJ. The reports were intended to be updated regularly (www.preventingcrime.org). A major development has since taken place in attempting to identify "evidence-based" criminal justice interventions (Sherman et al., 2002). These are ones that have been demonstrated to work through replicable, controlled experiments. A strong movement has taken place domestically and internationally to identify "best practice" programs and exemplary programs that might serve as

models in crime prevention. Similar lists of what works have been compiled for juvenile justice programs (Waller and Welsh, 1998) and for international programs (International Center for the Prevention of Crime in Montreal). The list of what works will grow more quickly if the nation invests more resources in scientific evaluations to hold all crime prevention programs accountable for their results.

None of these evaluations or placements as "working" or "not working" is final. Constant replication and reevaluation is required, but a persistent, independent, scientific program of evaluation will go a long way in replacing what we think works or what does not with what does work. Perhaps the field of rehabilitation has overreacted to the previously discussed Martinson report that concluded that "nothing works" in rehabilitation. Marlowe (2006) discusses the danger to researchers who conclude that a program does not work of risking being branded with the "Scarlet M" (for Martinson). The message to researchers is that if they question the value of rehabilitation, they risk their professional reputations.

The Campbell Collaboration (C2)

The Campbell Collaboration is an international research organization founded in 2000 and dedicated to preparing, maintaining, and publicizing systematic reviews of research on the effects of social and educational programs and interventions. Modeled after the successful Cochrane Collaboration in health care, the C2 program is named in honor of Donald Campbell, an innovator in research and experimental designs. In examining "what works," the systematic reviews use scientific and explicit methods to identify, screen, and analyze evaluation studies. The purpose of these reviews is to assist decision makers to better understand the existing research and better inform their decisions using evidence-based research. Various organizations have created a variety of Websites in a number of fields to address evidence-based research. This includes Websites on the blueprints program (Center for the Study and Prevention of Violence), child trends (programs to enhance child development), the Cochrane Collaboration (health care), helping America's youth, programs for justice-involved persons with mental illness, medical-clinical practice, juvenile delinquency prevention, addiction, strengthening families, and alcohol abuse (U.S. Department of HEW, 2008).

The nature of a C2 analysis can be illustrated by Brandon Welsh and David Farrington (2002), who did a meta-analysis of the *Crime Prevention Effects of Closed Circuit Television* (CCTV). An outline or summary of their procedure is instructive. They reviewed forty-six relevant studies from both the United States and Britain on the effectiveness of CCTV according to strict methodological criteria. CCTV had to be the main intervention, and the outcome measure was crime. There had to be measures of crime levels both before and after the intervention, and there had to be a comparable control area. Twenty-two of the forty-six studies met these criteria and were included. They concluded that the best evidence suggested the CCTV reduced crime to a small degree and was most effective with vehicle crime in car parks but had least impact in public transportation and in the center city. The poorly controlled (excluded) studies produced more desirable results than the better controlled (included) studies (ibid.).

Another example of a comprehensive effort to evaluate successful program implementation has been the Blueprints for Violence Prevention program at the University of Colorado (OJJDP, 2004). Figure 11.3 describes the Blueprint Initiative as well as the model and promising programs.

ABOUT THE BLUEPRINTS INITIATIVE

Blueprints for Violence Prevention began at the Center for the Study and Prevention of Violence (CSPV) as an initiative of the state of Colorado, with funding from the Colorado Division of Criminal Justice, the Centers for Disease Control and Prevention, and the Pennsylvania Commission on Crime and Delinquency. The project was originally conceived as an effort to identify model violence-prevention programs and implement them within Colorado. Soon after the creation of Blueprints, the Office of Juvenile Justice and Delinquency Prevention (OJJDP) became an active supporter of the project and provided funding to CSPV to sponsor program replications in sites across the United States. As a result, Blueprints evolved into a large-scale prevention initiative.

The Blueprints for Violence Prevention initiative has two overarching goals:

- Identify effective, research-based programs.
- Replicate these effective programs through a national dissemination project sponsored by OJJDP designed to
  - Provide training and technical assistance (through the program designers) to transfer the requisite knowledge and skills to implement these programs to sites nationwide.
  - Monitor the implementation process to troubleshoot problems, provide feedback to sites, and ensure that programs are implemented with fidelity to their original intent and design.
  - Gather and disseminate information regarding factors that enhance the quality and fidelity of implementation.

IDENTIFYING EFFECTIVE PROGRAMS

Identifying effective programs has been at the forefront of the national agenda on violence prevention for the last decade. Federal funding agencies have increasingly emphasized the need to implement programs that have been demonstrated effective. The focus on research-based practices has stimulated communities to search for the best practices and to determine what types of programs would be most effective and appropriate for their local problems and population.

In recent years, various governmental agencies, and some private organizations, have produced lists of programs that demonstrate at least some evidence of positive effects on violence/aggression, delinquency, substance abuse, and their related risk and protective factors. Taken as a whole, this work has resulted in a large repertoire of research-based programs from which the practitioner community may choose. Although these lists provide a valuable resource for communities, they can be confusing. Some lists are narrow in focus—for example, limiting their descriptions to drug abuse, family strengthening, or school-based programs only. In addition, and perhaps more importantly, the criteria for program inclusion vary tremendously, with some agencies adopting a more rigorous set of criteria than others. In fact, one must be diligent when examining the lists to ensure that at least a minimal scientific standard has been applied; for example, programs should demonstrate effectiveness using a research design that includes a comparison (i.e., control) group. Anything less rigorous than this approach cannot provide sufficient evidence to justify disseminating and implementing programs on a wide scale.

The Blueprints initiative likely uses the most rigorous set of criteria in the field:

- Demonstration of significant deterrent effects on problem behavior (violence, aggression, delinquency, and/or substance abuse) using a strong research design (experimental or quasi-experimental with matched control group).
- Sustained effects at least one year beyond the intervention.
- Replication in at least one other site with demonstrated effects.

FIGURE 11.3 Successful Program Implementation: Lesson from Blueprints *Source*: Mihalic, Sharon, et al. "Blueprints for Violence Prevention." *OJJDP Juvenile Justice Bulletin*, July 2001; Muller, Janine and Sharon Mihalic. "Blueprints: A Violence Prevention Initiative." *OJJDP Fact Sheet*, #110, June 1999; and Mihalic, Sharon, et al. "Blueprints for Violence Prevention Report." Office of Juvenile Justice and Delinquency Prevention, NCJ204274, July 2004.

This high standard is necessary if programs are to be widely disseminated because conducting an outcome evaluation during every implementation effort will be costly, time consuming, and not always possible. Therefore, it is important that programs demonstrate effectiveness, based on a rigorous evaluation, before their widespread dissemination. Programs meeting all three of the criteria are classified as "model" programs, whereas programs meeting at least the first criterion but not all three are considered "promising." To date, Blueprints has identified eleven model programs and twenty-one promising programs.

THE BLUEPRINTS PROGRAMS

The Blueprints for Violence Prevention initiative has identified the following model and promising programs.

MODEL PROGRAMS

Big Brothers Big Sisters of America (BBBSA)
Bullying Prevention Program
Functional Family Therapy (FFT)
Incredible Years: Parent, Teacher, and Child Training Series
Life Skills Training (LST)
Midwestern Prevention Project
Multidimensional Treatment Foster Care (MTFC)
Multisystemic Therapy (MST)
Nurse-Family Partnership
Project Towards No Drug Abuse (Project TND)
Promoting Alternative Thinking Strategies (PATHS)

PROMISING PROGRAMS

Athletes Training and Learning to Avoid Steroids (ATLAS)
Brief Strategic Family Therapy (BSFT)
CASASTART
Fast Track
Good Behavior Game
Guiding Good Choices
High/Scope Perry Preschool
Houston Child Development Center
I Can Problem Solve
Intensive Protective Supervision
Linking the Interests of Families and Teachers
Preventive Intervention
Preventive Treatment Program
Project Northland
Promoting Action Through Holistic Education (PATHE)
School Transitional Environment Program (STEP)
Seattle Social Development Project
Strengthening Families Program: Parents and Children 10–14
Student Training Through Urban Strategies (STATUS)
Syracuse Family Development Program
Yale Child Welfare Project

Descriptions of these programs are available on the Blueprints Web site www.colorado.edu/cspv/blueprints/index.html.

OBSTACLES TO EVALUATION RESEARCH

Obstacles to evaluation research

In its first annual review volume of criminal justice evaluation, the National Criminal Justice Reference Service (NCJRS, 1979) surveyed most of the authors whose works appeared in the volume, members of the editorial board of the volume, as well as a companion volume, *Crime and Justice: An Annual Review of Research* (Morris and Tonry, 1979). In the order of perceived importance, the following dangerous pitfalls were identified by this group of evaluation experts (NCJRS, 1979, p. 370):

Poorly done evaluation design and methodology

Unsound and/or poorly done data analysis

Unethical evaluations

Naive and unprepared evaluation staff

Poor relationships between evaluation and program staff

Co-optation of evaluation staff and/or design

Poor quality data

Poorly done literature reviews of subject area

Focusing on the method not the process

Geller (1997, p. 4) describes impediments to police departments becoming learning organizations:

Skepticism about research as ivory tower and impractical.

Resistance to cooperating with outside researchers because too often they have failed to provide feedback soon enough to assist practitioners.

Distrust of evaluation research because of the blisters that linger from the last time the department was burned by a poorly conducted study.

Skepticism that research findings developed in another jurisdiction have any application at home.

The myth that encouraging critical thinking among the rank and file will undermine necessary paramilitary discipline.

The belief that thinking inhibits doing.

An indoctrination process in most police departments that inhibits employees from contributing meaningfully to organizational appraisal.

A police department that denigrates rank-and-file thinking about the organization's basic business establishes a culture likely to ridicule or demean those who would take time from routine activities (random preventive patrol, etc.), which police have taught themselves, politicians, and the public as constituting real and tough police work.

Reluctance to have cherished views challenged.

Difficulty in engaging in organizational self-criticism while continuing to work with those whose current efforts are criticized.

Insufficient time for employees to reflect on their work and a lack of time, authority, resources, and skills for them to conduct research.

Fear of change.

RESEARCHERS AND HOST AGENCIES

Questions in evaluation research

The National Advisory Committee (1976, p. 133) suggests the following guidelines with respect to relationships between those performing evaluation research and the *host agencies*:

> R&D funding agencies that support studies of criminal justice organizations should be sure that researchers who conduct such studies are sensitive to the needs of the organizations that are part of the study. Such sensitivity will increase the likelihood of completing the project to the satisfaction of the funding agency, the organization that is part of the study (host agency), and the research team.
>
> 1. Before the research begins, clear agreements should be reached between the researcher and the host agency on such issues as: the purpose of the research, duration of effort, data to be collected, plans for protecting confidentiality of sensitive information, resources required of the host agency, extent to which the host agency may be identified by name in publications, form and timing of public disclosure of the results of the study, and any other topic of mutual concern.
> 2. Funding agencies should assist researchers in establishing favorable relationships with host agencies by:
>   a. Assuring that the research design does not necessarily interfere with the host agency's normal operations.
>   b. Arranging for host agencies to receive timely feedback on research progress or results.
>   c. Considering the reimbursement of expense incurred by the host agency in cooperating with the research project.
> 3. Existing educational programs for researchers could be broadened to include relevant courses, on-site projects conducted in cooperation with an operating agency, internships, and exchange programs to make researchers more cognizant of procedures that may improve their relations with criminal justice organizations. These programs should stress the necessity of developing a viable partnership with the host agency during the planning, conduct, and follow-up of a research study.

Summary

Policy analysis is the study of government behavior. It includes proposals, programs, decisions, and effects. The policy process involves identification, formulation, legitimation, implementation, and evaluation. *Policy experiments* are applied field experiments with immediate practical policy implications. *Evaluation research* is an applied branch of social science that evaluates policies and programs to determine whether and how well they work. The NIJ's research program emphasizes policy-oriented programs and attempts to link researchers with practitioners. A *systems model* of evaluation research consists of inputs, activities, results, outcomes, and feedback.

Before an evaluation is undertaken, three crucial questions must be answered: Will the findings be used? Is the project evaluable? Who can do this work? Formative research, or an evaluability assessment, addresses these questions before an evaluation is agreed to be undertaken.

The *steps in evaluation research* are problem formulation, design of instruments, research design (evaluation model), data collection, data analysis, findings and conclusions, and utilization. *Some obstacles or pitfalls in evaluation research are* poor evaluation design and methodology, poor data analysis, unethical evaluations, naive or unprepared evaluation staff, poor relationships between evaluation

and program staff, co-optation of evaluation staff and/or design, poor-quality data, poor literature reviews, and focus on method rather than process. Much evaluation research exemplifies some of these problems, particularly the politics of evaluation.

Of particular importance in effective evaluation is the need for effective relationships between the researcher and the host agency (site to be evaluated). The National Advisory Committee on Criminal Justice Standards and Goals (1976) suggests clear agreements beforehand, assistance from funding agencies in bringing the two parties to suitable agreements, and training programs to acquaint researchers with agency problems and needs.

Key Concepts

Policy analysis *277*
Evaluation research *278*
Policy experiments *280*
Systems model *282*
Process evaluation *284*
Impact evaluation *284*
Assessment *284*
Monitoring *284*
Evaluability assessment *285*
Steps in evaluation planning *287*
Obstacles to evaluation research *296*

Review Questions

1. How does evaluation research fit into the general scheme of policy analysis? Using the NIJ program, what role can research have in public policy debates in criminal justice?
2. Describe the "systems model" of evaluation research. In what way can such a model inform public policy in criminal justice?
3. Evaluation research seldom takes place as planned. Using the discussions in the chapter, elaborate on obstacles to evaluation research in criminal justice.

Useful Web Sites

American Evaluation Association *www.eval.org*
Guide for Writing a Funding Proposal *www.learnerassociates.net/proposal*
Writing Your Thesis or Dissertation *www.learnerassociates.net/dissthes/*
Successful Program Implementation: Lessons from Blueprints *www.ncjrs.org/pdffiles1/ojjdp/204273.pdf*
Resources for Methods in Evaluation *http://gsociology.icaap.org/methods/*
Juvenile Justice Evaluation Needs *www.jrsa.org/pubs/reports/jj_needs_assessment.htm*
Juvenile Justice Evaluation Center Online *www.jrsa.org/jjec/*
Blueprints: Successful Program Implementation *www.ojp.usdoj.gov*
Program Evaluation Toolkit *www.cdc.gov/STD/program/progeval/Ref-PGprogeval.htm*
Basic Guide to Program Evaluation *www.mapnp.org/library/evaluatn/fnl_eval.htm*