National Youth Court Center
Evaluation and Grant Writing Training Seminar

Indianapolis, Indiana
October 23, 2002

EVALUATION SESSION

Faculty
Janeen Buck
Jeffrey Butts

URBAN INSTITUTE
Justice Policy Center
2100 M Street NW
Washington, DC 20037
www.urban.org
# Table of Contents

INTRODUCTION AND OVERVIEW ........................................................................................................... 1

SECTION 1: EVALUATION What Is It, Why Do It, and When?......................................................... 3

SECTION 2: LOGIC MODELS How Do You Decide What to Evaluate?........................................... 7

SECTION 3: EVALUATION STRATEGIES What Are the Basic Approaches?................................. 13

SECTION 4: FINDINGS FROM THE OJJDP EVALUATION OF TEEN COURTS (ETC) PROJECT ............................................................................................................................... 18

SECTION 5: DATA COLLECTION AND MEASUREMENT ISSUES .................................................. 26

SECTION 6: GUIDELINES FOR SELECTING AN EVALUATOR ....................................................... 32

Appendix A
GLOSSARY OF EVALUATION TERMS

Appendix B
EVALUATION STRATEGIES FOR HUMAN SERVICES PROGRAMS:
A GUIDE FOR POLICYMAKERS AND PROVIDERS

Appendix C
TEEN COURTS: A FOCUS ON RESEARCH

Appendix D
EVALUATION WEB RESOURCES

Appendix E
URBAN INSTITUTE PARENT AND YOUTH QUESTIONNAIRE 1

Appendix F
URBAN INSTITUTE PARENT AND YOUTH QUESTIONNAIRE 2

Appendix G
URBAN INSTITUTE YOUTH QUESTIONNAIRE 3
INTRODUCTION AND OVERVIEW
Introduction

Evaluation can be a valuable tool. A well-designed evaluation will provide important information about program performance that can be used to enhance services, improve operations and outcomes, target the use of limited resources and even help attract new resources. A poorly executed evaluation, however, can frustrate participants, waste valuable resources, produce unreliable or misleading data and yield little helpful information.

This workbook is designed to provide a practical, application-oriented introduction to evaluation fundamentals (principles, key terms, and definitions), and to help individuals recognize the difference between a solid evaluation and a shoddy one. The Urban Institute developed this workbook for participants in the Teen Court Evaluation Seminar sponsored by the National Youth Court Center. The materials that follow capture much of what is covered in the seminar, but also feature data collection instruments used in the Urban Institute's recently completed Evaluation of Teen Courts project, a list of recommended evaluation resources, and a glossary of key evaluation terms. Several guides on evaluation have also been included for easy reference. (See Appendices A-G.)

The first section of this workbook provides an overview of evaluation concepts, and discusses the issue of evaluation readiness (i.e., what needs to be in place before an evaluation begins) and how to determine if a program is ready to be evaluated. The second section explains how to use logic models to determine the linkages between program operations and expected outcomes. The third section examines various approaches to evaluation, and what works under different circumstances (or, what works best, when). The topics of measurement and data collection, including the rights and welfare of research participants, are addressed in the final sections of the workbook.

---

1 Points of view or opinions expressed in this document do not necessarily represent the official position or policies of the National Youth Court Center, OJJDP, the U.S. Department of Justice, the Urban Institute, its board, or sponsors.
SECTION 1: EVALUATION
What Is It, Why Do It, and When?
1.0 Evaluation

Every evaluation is designed to answer a question or set of questions. Identifying and defining the questions to be answered are the first steps in developing any evaluation (Harrell et al., 1996). Translating each question into something measurable is the next step, followed by selection of an appropriate research design. This section explores the purposes of evaluation and answers the important questions about why and when to evaluate.

1.1 What is Evaluation?

Evaluation is the process of answering a question or series of questions in a methodical manner in order to inform a decision that must be made. Stated simply, evaluation is the systematic gathering of information and analysis of that information to answer a specific question and inform a related course of action.

There are many kinds of evaluation. This workbook focuses on program evaluation. Program evaluation serves multiple purposes and may answer a variety of questions. Some program evaluations examine the extent to which goals and objectives have been met. Others assess whether a program was implemented as intended, and if not, why not. Many program evaluations are geared toward understanding how well a program functions and how it could be improved. Most are designed to determine the scope and nature of the program’s impact on individuals, the agency and community. Evaluation findings are often used to make decisions about future funding for an existing program.

1.1 Why Evaluate?

Program evaluation is performed for a variety of reasons:

• to improve program operations (what works, what doesn’t work, and why)
• to promote accountability
• to determine whether and how to allocate additional funds
• to determine if and how well the needs of the target population are being met

1.2 For Whom – Identifying the Concerns of Key Stakeholders

Program evaluation can answer a range of questions. Determining which questions to answer depends on the research audience. Program stakeholders are those individuals who have a share or an interest in the program and its success. Key stakeholders often represent the target audience for evaluation findings. Teen court stakeholders may include:
• judges
• criminal justice officials
• law enforcement officials
• school officials
• funding authorities
• policy makers
• community residents
• clients (youth and parents)
• program staff

1.3 When to Evaluate

After identifying the pertinent evaluation questions, the next step is to consider whether an evaluation is “feasible, justified, and likely to produce useful information” (Harrell et al., 1996: 7). The process of exploring and answering these questions is known as an evaluability assessment. An evaluability assessment takes into account:

• **Whether the question is “answerable”**
  
  Can the answer to the question be measured in some way? How? Using what kind of information or data?

• **Availability of information**
  
  Does enough information exist to answer the question? Can that information be collected easily or in a cost-effective manner?

• **When an answer is essential to continued program operations**
  
  Can the program or staff continue to function if this question is not answered? Will funding continue? Will cases still be referred, etc?

• **Do answers already exist?**
  
  Have previous evaluations already answered the question? If so, will this evaluation yield more precise information or address a different aspect of the question that might benefit others?
• **Program maturity – how stable is the program?**

  Is the program sufficiently well developed (i.e., are target population, program procedures and services firmly fixed) and established to sustain an evaluation? Are program components and activities clearly defined?

• **Availability of resources**

  Do sufficient resources (money, time, expertise, interest) exist to support the evaluation?

• **Vision and support for future use of evaluation findings**

  Is there a clearly stated purpose for the evaluation? How will the findings be used?
SECTION 2: LOGIC MODELS
How Do You Decide What to Evaluate?
2.0 Logic Models

Constructing a logic model is the second step in developing an evaluation strategy. A logic model maps the linkages between the object of the evaluation – i.e., the thing about which questions need to be answered (most often a program or intervention) – its characteristics, specified outputs (activities), and expected outcomes (results). Logic models portray how program activities and goals fit together – or, more importantly, when and how they don’t.

Although the function and form of a logic model may differ according to the type of evaluation being conducted (see Section 3 for more information on evaluation design), most logic models generally contain at least four components:

- **Background factors**

  Background factors refer to the environmental conditions (context in which a program operates) or client characteristics that may affect the ability of the program to achieve its goals (Harrell et al., 1996). A logic model for a youth-focused intervention like teen courts could include individual client characteristics such as demographic factors, family factors (composition, conflict/cohesion), and youth factors (peers, delinquency, pro-social attitudes, school attachment). Neighborhood level indicators also may be relevant background factors (level of poverty, crime rate, drug use).

- **Intervention Factors**

  Intervention factors generally refer to the program or intervention that is to be evaluated, and its characteristics (key aspects of the program model) and activities (services or level of services to be delivered – i.e., outputs). For teen courts, intervention factors could include the courtroom model, the range of sanctions imposed on youth, the role of youth volunteers, and the organizational context (setting) in which the program operates.

- **Mediating Factors**

  Mediating factors are conditions that may cause the program to appear more or less successful than it really is. The degree to which a teen court defendant participates in the court process could be a mediating factor. The receipt of other services (counseling, substance abuse treatment) while going through the teen court process could also be a mediating factor. The number and type of mediating factors included in a logic model depends on local context.
• Outcomes

Logic models illustrate the link between background factors, activities and outcomes. They are a helpful tool for determining the extent to which a program's goals and objectives are related to program activities, and whether program activities are accurately and logically tied to anticipated outcomes.

Outcomes refer to the result produced by a program or an intervention. These results may be either positive (e.g., increased school attendance) or negative (e.g., increased drug use). Most programs are expected to produce a variety of results. Logic models typically present the desired outcomes that are believed to result from specific program activities. For example, teen courts are believed to help keep defendants from getting into trouble gain with the law and to enhance their respect for the law. In a logic model, these assertions might be stated as a decrease in re-offending and an increase in pro-social attitudes.

Chronology and causality generally flow from left to right in logic model diagrams. Background factors are recorded in the far left of the diagram. Intervention factors are recorded to the right of the background factors and mediating factors are listed to the right of the intervention factors. The hypothesized outcomes are shown on the far right of the diagram. Arrows typically illustrate the linkages between elements or components of the logic model – i.e., they illustrate the logical flow between activities, outputs and outcomes.

As mentioned earlier, some logic models may include additional details. Many logic models also map intermediate outcomes – i.e., evidence of short-term changes. Harrell et al. (1996) suggest that special attention should be paid to the timing and ordering of events and expected results.
2.0 Logic Models

**Background Factors**

- Client Factors
  - Demographics
  - Family
  - School
  - Peer Influences
  - Pro-social Attitudes
  - Delinquency
  - Substance Abuse

- Neighborhood Factors

**Intervention Factors**

- Program Components and Activities
  - What does the program actually do?
  - How?
  - When?
  - Where?
  - To whom?

**Mediating Factors**

- Factors Not Directly Related to Intervention, But Which Could Influence Outcomes
  - Local context
  - Unanticipated program effects
  - Unexpected client behavior

**Outcomes**

- Intended Result
  1. Intermediate
  2. Long-term
Group Exercise #1 - Logic Model Worksheet

This exercise is designed to help participants apply the concepts learned about program logic models.

Instructions:

Working together as a group, answer the questions in each section of the worksheet. Next, record the group’s best answers for each section in the appropriate column of the blank Logic Model Diagram provided (remember: all elements of the logic model must be feasible and cost-effective to measure). Please be prepared to explain how your group identified these items and the placement of these items in your diagram.

Program Goals (what are the three most important goals of your program?)

1. ______________________________________________________________

2. ______________________________________________________________

3. ______________________________________________________________

Measurable Outcomes (what exactly does your program try to change?)

1. ______________________________________________________________

2. ______________________________________________________________

3. ______________________________________________________________
Intervention Factors (what exactly does your program do to achieve goals?)

1. ________________________________________________________________

2. ________________________________________________________________

3. ________________________________________________________________

Background Factors (i.e., pre-existing factors that affect program impact)

1. ________________________________________________________________

2. ________________________________________________________________

3. ________________________________________________________________

Mediating Factors (what can alter effects of your program on its goals?)

1. ________________________________________________________________

2. ________________________________________________________________

3. ________________________________________________________________
Logic Model Diagram - Group Exercise #1

Background Factors

Intervention Factors

Mediating Factors

Outcomes
SECTION 3: EVALUATION STRATEGIES
What Are the Basic Approaches?
3.0 Evaluation Approaches

There are four basic approaches to evaluation: process, cost-analysis, performance monitoring, and impact. Each is designed to answer a specific set of questions. It is important to understand what each is designed to do and what can be learned from each. This knowledge, combined with what is learned during completion of a logic model can assist in selecting the appropriate evaluation approach.

3.1 Process Evaluation

A process evaluation yields information about how a program or intervention functions and operates. Process evaluations are typically designed to capture the administrative, legal, and programmatic issues that affect the evolution and operation of a program. The information generated by process evaluation provides a context for understanding the results of other research designs.

Process evaluations generally rely on document review, direct observation, and face-to-face interviews with program staff, partner agencies and clients to answer questions such as:

- What is the structure of the program (staff, target population, how does it fit into the broader community, etc)?
- What services are delivered, and how?
- What factors contribute to the success of a program?
- What factors inhibit the success of a program?
- What role does community involvement play in program implementation?
- What influence does the program have on the community, and how does it fit in with other agencies or systems?

The information in this section was adapted from Evaluation Strategies for Human Services Programs: A Guide for Policymakers and Providers by Adele Harrell, Martha Burt, Harry Hatry, Shelli Rossman and William Sabol (1996). The full text publication is located in Appendix B.
• What lessons have been learned that could benefit other jurisdictions planning or implementing similar programs?

Process evaluations employ a variety of techniques including case studies, focus groups, and ethnographies. For more information about these techniques, see Appendix B.

3.2 Cost Analysis

According to Harrell et al. (1996), “cost studies address how much the program or program components cost, preferably in relation to alternative uses of the same resources and to the benefits being produced by the program” (p. 2). A cost analysis will quantify the cost of staff and services provided by the program and the savings associated with these services.

3.3 Performance Measurement

Performance measurement (also called performance monitoring) involves the regular and ongoing process of collecting data about key aspects of program activities and results. Frequent data collection is designed to produce ongoing feedback that can be used to improve program operations and outcomes, and to assess accomplishments.

3.4 Impact

An impact evaluation addresses questions of causality and attribution. Causality is the capacity to show that a specific intervention had a specific effect (i.e., did an intervention cause the intended effect?). Attribution is being able to credit a specific intervention with a specific result. For example, lower re-arrest rates could be attributed to teen courts if this were observed among teen court defendants but not among a group of youth just like teen court defendants but who were not referred teen court. An impact evaluation answers the questions of causality and attribution by measuring program outcomes and comparing those outcomes against what happened without the program (Harrell et al., 1996). There are three kinds of impact evaluations: experimental, quasi-experimental and non-experimental.

3.4a Experimental Designs

An experimental impact evaluation involves the use of treatment and control groups. The treatment group receives the services, program or treatment being
evaluated and the control group does not. Cases are randomly assigned to one group or the other.

Cases or individuals selected to participate in a study that uses an experimental design typically must meet a set of eligibility criteria. These eligibility criteria help ensure that cases are similar to each other. The randomness of group assignment eliminates factors that could affect outcomes such as selection bias (i.e., people who volunteer for a program maybe different from those who don’t and that that difference may account for their outcomes). Random assignment “controls” for any unknown factors that could possibly affect intervention outcomes. This method is a way of testing whether an outcome or result is really the product of the intervention.

Experimental designs are considered the “gold standard” for evaluation because of the definitive nature of the findings they produce. There are, however, several limitations/considerations associated with using an experimental design:

- **Expense** – experimental designs require extensive resources. Random assignment typically extends the time it takes to recruit an adequate number of subjects, and requires sufficient staff to oversee, monitor and manage the screening, recruiting and assignment process.

- **Ethics** – random assignment means that some people are denied services that could potentially benefit them. It is important to consider whether it is justified and ethical to treat people differently.

- **Precision** – experimental designs are able to answer a very precise question. This means the evaluation focus is typically very narrow.

3.4b Quasi-Experimental Design

A quasi-experimental design compares outcomes among groups; however, assignment to these groups is not random as it is in an experimental design (Harrell et al., 1996). Comparison groups are made up of cases that are believed to be similar to cases in the treatment group. For example, evaluators might select comparison group cases by matching the characteristics of cases (sex, race, age, grade, etc.) to those of the treatment group.

Caution must be used when interpreting the differences in outcomes for non-random comparison groups. Without random assignment, many reasons other than the intervention may contribute to the differences between these groups.
3.4c Non-Experimental Design

Non-experimental designs do not use control or comparison groups. Instead, outcomes for a group are measured multiple times and compared over time to detect change and to determine the nature of that change. Common methods include the use of "pre-post", time series, panel, and post-program comparison designs (for more information on these techniques see Appendix B).
SECTION 4: FINDINGS FROM THE OJJDP EVALUATION OF TEEN COURTS (ETC) PROJECT
4.0 Overview of the Evaluation of Teen Courts (ETC) Project

The Evaluation of Teen Courts (ETC) Project was sponsored by the Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice, and conducted by the Urban Institute. This multi-year evaluation, which began in 1998 and concluded in 2002, was the first national study of teen courts (also known as youth courts and peer courts). The ETC Project was designed to measure the effect of handling relatively non-serious, adolescent offenders in youth-operated teen courts rather than in juvenile or family courts. Four sites participated in this study: Anchorage, AK; Maricopa County, AZ; Rockville, MD; and Independence, MO.

The evaluation examined how teen courts work and how they impact young offenders. There were two main components of the study: a process evaluation and an impact evaluation. The process component explored the legal, administrative, and case processing factors that affected the ability of programs to achieve their goals. Individual and group interviews with program staff and key partners, as well as analyses of program documents and observation of teen court sessions were conducted to examine process and implementation issues. The process evaluation was designed to answer the following questions:

- What factors contribute to the success of teen court programs?
- What role does community involvement play in program implementation, and what impact do teen courts have on juvenile courts, schools, or the community?
- To what extent do teen courts contribute to community awareness of youth needs that traditional institutions may not?

The impact evaluation examined whether teen courts were effective in achieving two basic goals—reducing recidivism and fostering positive changes in the attitudes of young offenders. Key research questions included:

- Who are the juveniles participating in teen court, including the nature of their immediate and past offending, if any?
- What is the nature of sentences imposed by teen courts and the level of compliance obtained?
- How do teen courts affect defendant recidivism, pro-social attitudes, and behavior with peers, family, and authority figures?
• How satisfied are defendants and parents with the teen court process and outcomes?

• To what degree do outcomes vary across teen court models and across subsets of offenders?

4.1 Reason for the Study

Youth courts vary greatly in their case handling procedures, courtroom models, and the range of sanctions they use to hold offenders accountable. Some programs use courtroom procedures that are relatively formal; others are quite informal. Some include adults in prominent roles; others do not. Some involve youth attorneys and some involve youth juries, but many do not. Are these differences important? Do they affect the ability of youth courts to reduce recidivism? What are the essential ingredients for an effective youth court process?

Ultimately, these questions must be answered by evaluation research. The youth court research literature is finally beginning to grow but even more research is needed. Only well-designed evaluations can provide the guidance that state and local officials will need as they begin to think about starting or expanding youth court programs. The Office of Juvenile Justice and Delinquency Prevention funded the Evaluation of Teen Courts project to begin the process of answering basic questions about the comparative impact of youth courts and to encourage the further development of effective youth court programs.

4.2 Conceptual Framework

The ETC Project’s conceptual approach to the impact evaluation was portrayed in a logic model (see Figure 1). Chronology and causality flow from left to right in the diagram. Hypothesized teen court outcomes are shown on the far right. The first outcome, type and severity of penalties, is used to assess whether teen courts actually provide youth accountability—did each youth face a consequence and was it a serious consequence? The second outcome is satisfactory completion of the teen court process itself, since making sure that youths complete the program is important to ensure accountability.

Reducing recidivism was a critical outcome in the evaluation because it is a primary goal of teen court programs. Subsequent delinquency is measured by 1) self-report data and 2) police and court records of contacts for the study youths. The final set of expected outcomes includes various measures of attitudes associated with strengthening social bonds through increased positive perceptions of justice and increased commitment to pro-social norms (e.g., school attachment and family relationships).
The ability of teen courts to achieve these outcomes was expected to be influenced by "background" factors (the column on the far left) and by "mediating" factors that shape the impact of the teen court process on individual youths. Background factors are similar to the general predictors of youth problem behaviors. They were drawn from the research literature on high-risk youths. In addition to demographic information, background factors included family risks (substance use, inadequate parental supervision), youth risks (school problems, lack of pro-social norms, association with delinquent peers), and environmental risks (prevalence of poverty, crime and drugs in the neighborhood).

4.3 Methodology

The Urban Institute’s evaluation targeted first-time misdemeanor offenders. Recruitment of the sample was prospective and based on a rolling enrollment that began in July 2000 and concluded in October 2001. Recruitment procedures were developed in partnership with teen court staff in each of the four study sites. Active affirmative consent was obtained for all youth for whom questionnaire and face-to-face interview data were collected. The final study sample included 523 youth.

Data were collected from multiple sources. Information about the legal, administrative, and case processing factors that influenced the operation and goals of these programs was collected from interviews with teen court staff, and through the review of various documents. Urban Institute researchers visited each site once to twice a year over the life of the project. Likewise, the evaluation team periodically scanned local newspapers for articles about the programs. The sites also supplied UI with similar materials.

Data sources for the outcome evaluation included (1) brief self-administered youth and parent questionnaires; (2) face-to-face baseline and follow-up surveys with teen court defendants in one (MD) site; (3) teen court program files; and (4) police and court records.

Youth and Parent Questionnaires. The youth and parent questionnaires measure the respondents’ perceptions of teen court before and after exposure to the program and the impact of teen court on the defendant’s attitudes.

There are five versions of the questionnaire: three for the youth and two for the youth's parent/care giver (see Appendices E-G). Each version was a different color to ensure proper administration of the survey sequence. Youth questionnaires were purple, gray, and blue; parent questionnaires were yellow and green. Together, these five surveys contain 124 items. Youth questionnaires include measures of antisocial behavior peer associations, attitudes toward delinquency and the justice system, and a limited set of demographic items. The parent questionnaires are similar in content and structure to the youth instruments. Items in the parent questionnaire focus on the
adult’s attitude about delinquency and the justice system, and perception of their child’s behavior and peer associations. Parent questionnaires also contain items about socio-economic status.

Each survey was self-administered and took about five minutes to complete. The parent and youth survey was administered prior to the teen court hearing. Parents and defendants completed the second questionnaire immediately after court. The third youth survey was administered following the time in which sanctions were to be completed had ended. Most Q3 surveys were mailed to respondents. A letter from the evaluation's Project Manager, the Q3 survey, and a postage-paid, self-addressed enveloped were sent to each respondent. The letter thanked the youth for his or her participation in the study, discussed the enclosed Q3 survey, and provided instructions for mailing the completed survey back to the Institute.

Face-to-Face Interviews. Two waves of self-report interview data were collected from youth referred to the Montgomery County Teen Court in Maryland and group of youth handled by the traditional juvenile justice system in a neighboring jurisdiction. Both sets of interviews were conducted by Westat, Inc., a survey research firm located in Rockville, MD.

The baseline survey was designed to measure self-reported delinquency, family cohesion and conflict, peer associations, commitment to pro-social norms, maturity of judgment, and perceptions of justice. Such information served as a benchmark against which to compare (1) changes in the attitudes and behavior of defendants before and after going to teen court; (2) characteristics of teen court defendants and comparison group youth; and (3) any differences in outcomes between teen court defendants and the comparison group youth. It took roughly 45 minutes to administer. Respondents received a non-monetary incentive valued at $10 to thank them for participating in this interview.

The follow-up interview closely paralleled the structure and content of the baseline interview with the exception of two sections that were designed to capture the respondents’ receipt of services and level of satisfaction with those services. This survey was administered four to six months after the baseline interview and focused on events and experiences that occurred during that time. It took approximately 60 minutes to conduct. As with the baseline interview, respondents received a $10 gift certificate upon completion.

Informed consent was administered prior to the start of both interviews. This process served to remind youth and their caregiver of what the interview consisted of and to re-confirm both their understanding and willingness to participate in this portion of the research.
**Teen Court Program Data.** Information about the how cases were processed in teen court was collected from program case files. Program records furnished data about several aspects of the youth and the teen court process including the courtroom model used and the mix of sanctions received.

**Official Records Data.** These data were used to measure recidivism outcomes among research subjects. Because identical data elements could not be collected across sources, recidivism was defined as any police arrest or referral to court for a new offense which occurred subsequent to the respondent's teen court referral.

### 4.4 Key Findings

Findings from the ETC Project suggest that teen courts (or youth courts) may be a positive alternative to the normal juvenile justice process, especially in jurisdictions that do not already have a wide variety of intervention options for young, first-time juvenile offenders.

More than 500 teen court cases across all four sites were compared with similar cases handled by the traditional juvenile justice systems in those jurisdictions. The evaluation collected baseline data about the youth and their parents or guardians, and tracked the youth for at least six months, measuring the extent to which official recidivism differed between the teen court youth and those processed within the juvenile justice system. Other measures, such as self-reported recidivism and youth attitudes and opinions, were collected from youth court subjects at three points — just before their youth court hearing, just after their hearing, and 30 to 60 days later.

In three of four study sites, the 6-month recidivism rate for youth court was lower than that of the comparison group. In the Alaska, Arizona, and Missouri, youth courts were compared with the average juvenile justice response in cases involving young, first-time offenders. In other words, young offenders in the comparison group were not offered special services or sanctions. They received whatever was typical in that jurisdiction for first-time offenders, including warning letters, informal adjustments, and outright dismissals. In all three of these sites, the juveniles handled in youth court were less likely to commit new offenses and be re-referred to the juvenile justice system. In two of the sites, this difference was statistically significant.

In Alaska, recidivism for youth court cases was 6 percent, compared with 23 percent of cases handled by the traditional juvenile justice system. In Missouri, the recidivism rate was 9 percent in youth court and 27 percent for the traditional process. In the third site, Arizona, the recidivism rate among youth court cases was 9 percent, compared with 15 percent for cases in the traditional system, but the difference was not statistically significant.
In the fourth site (Maryland), youth court was compared with a more proactive, police diversion program that provided many of the same services and sanctions offered by youth courts. Young offenders were ordered to pay restitution, perform community service, and write apology letters, just as they would in a youth court. The entire process, however, was managed by police officers and a police department social worker. Youth were not required to appear in court. Recidivism among the Maryland comparison group was slightly lower than it was among youth court cases, but reoffending rates were low for both groups (under 10 percent) and the difference between them was not statistically significant.

One could argue that the evaluation design in Maryland was a more rigorous test of youth court effectiveness because it came closest to isolating the impact of the one feature of youth courts that is truly unique, peer-to-peer justice in a courtroom setting. When other aspects of the programs being compared are similar — i.e., when youth court cases and comparison group cases receive similar sanctions and services — there may be little difference in recidivism six months later. What would this suggest about the overall value of youth courts? Perhaps the primary benefit of youth court is its ability to ensure the delivery of early interventions and meaningful sanctions for first-time, delinquent offenders.

4.5 Implication of Findings

The findings of the Evaluation of Teen Courts indicate that teen courts and youth courts may be preferable to the normal juvenile justice process in jurisdictions that do not, or cannot, provide meaningful sanctions for all young, first-time juvenile offenders. In jurisdictions that do provide meaningful sanctions and services for these offenders, youth court may still perform just as well as a more traditional, adult-run program. Furthermore, the fact that youth courts depend largely on volunteer labor and operate with very low budgets suggests that they may be a particularly cost-effective alternative for many jurisdictions, even those that do have a rich array of prevention programs.

This summary originally appeared in the newsletter of the National Youth Court Center (2002). For more information see the web site of the Program on Youth Justice at the Urban Institute at: http://youth.urban.org.
Figure 1. Logic Model for Evaluation of Teen Courts (ETC) Project

Background Factors

Demographic Factors
Family Factors
Conflict/Cohesion
Youth Factors
Delinquency
School attachment
Pro-social attitudes
Peer associations
Perceptions of justice
Maturity of judgment

Intervention

Teen Court Process
Range of Sanctions Used
Quality of Peer-to-Peer Justice

Mediating Factors

Extent of Youth/Family Participation
Youth Involvement with Other Services
Administrative Coherence & Consistency

Outcomes

Recidivism
- Self-reported
- Official

Change in Social Bonds
Enhanced commitment to pro-social norms
Improved perceptions of justice
SECTION 5: DATA COLLECTION AND MEASUREMENT ISSUES
5.0 Data, Measurement and Data Collection Issues

Data can be collected from a variety of sources, for a variety of reasons. At the outset of any evaluation, it is important to have a clear sense of what information needs to be collected and how that information will be used. Most teen court programs are able to access a wide range of useful information. Some information, however, does not exist in a reliable and convenient form. Program staff may need to develop a plan for collecting data they will need to answer specific questions. Whether the information already exists or must be collected, knowing what to collect and how to use that information can be challenging. This section examines issues about data (types and sources), measurement and data collection including instrument design, use, logistics, and legal issues.

5.1 Data

Data can be either qualitative or quantitative. Qualitative data are non-numerical and provide text and narrative. Quantitative data are numeric. Numbers and statistics are used to describe and explain a phenomenon or effect. Impact, outcome, and cost analysis studies generate and rely on quantitative data. Both types of data have advantages and disadvantages. Knowing which to collect and use depends on the goals of the evaluation. Often, the best approach is to collect a variety of data.

5.2 Data Sources and Methods

There are numerous sources of, and methods for, collecting qualitative and quantitative data. Process evaluations often rely on focus groups and semi-structured interviews with individuals. Generally, these methods solicit input from key individuals about program operations and services, as well as broader topics that might affect the ability of a program to achieve stated goals (i.e., collaboration, coordination, information-sharing, policies and procedures). Program documents and reports are another source of important qualitative data. Impact, cost analysis and performance measurement studies provide uniform information that can be translated statistically. Administrative data (i.e., program files, police and court records), as well as survey and questionnaire data are other common sources of quantitative data.

5.2a Administrative Data (program, police, court and other records)

Client records maintained by programs, police arrest records, and court records are all examples of administrative data. Records data usually produce uniform information about events and services. This information can be used to measure critical program outcomes. For example, many teen court programs use police arrest records...
to measure re-offending or recidivism – a central outcome of teen court efficacy. Arrest date, a standard element of police records data, allows placement of that event on a timeline – i.e., an arrest after teen court would constitute re-offending.

The advantage of records data is that it offers uniform or standard information, and it is generally easy to gain access to it (data considered to be public information, such as arrest data, will be easier to access than information considered confidential such as medical or substance abuse treatment records). Although it may be easy to gain access to official records data, it may be difficult to actually obtain the information. In some jurisdictions, antiquated data systems make it difficult to extract the data needed. Sometimes, data must be extracted by hand – which can be expensive, time-intensive, and require complex measures to protect client confidentiality if data are of a sensitive nature. Further, not all data systems are well maintained or monitored for quality. Data may be missing or inaccurate.

5.1b Self-Reported Data (surveys and questionnaires)

Surveys and questionnaires collect self-reported data – i.e., information reported directly by an individual. Both are a good way to collect information about attitudes and behaviors, especially more sensitive information such as personal values or involvement in risky behaviors or illegal activities.

Self-reported data can be gathered using self-administered questionnaires like those developed for the Evaluation of Teen Courts Project, or with the help of a trained interviewer. Interviews may use highly structured and comprehensive survey instruments that include questions about school, involvement in extracurricular activities, peer associations, family cohesion and parent-child relationships, the use of drugs and alcohol, and involvement in crime.

Self-report data also provide a greater degree of accuracy for some measures (i.e., self-reported delinquency may be a more reliable measure than official arrest records for measuring recidivism). Collecting self-report data, however, can be expensive. Considerable staff time may be required to monitor the process, and depending on the method used, to collect the data. In-person interviews tend to be lengthy and require on-going quality control to make sure interviewers administer each survey as designed.

5.2 Measurement Issues

How do you know you’re measuring what you think you’re measuring? What is the “best” way to measure something? These are critical questions.
Deciding how best to measure something can be a difficult process. For example, a primary goal of teen courts is to keep defendants from getting into trouble with the law again. Getting into trouble with the law could mean many things. One way to measure this may be to look at whether defendants re-offend after leaving teen court. Re-offending is a measurable definition. In the Evaluation of Teen Courts Project, re-offending was defined as either re-arrest or re-referral to juvenile court. Two definitions were used because comparable data did not exist across the sites. Re-arrest data are often preferable to court referral data because re-arrest data represent a more complete measure of re-offending (i.e., not all arrests are referred to court but most court referrals are the result of an arrest).

There are many measurement issues to keep in mind when designing and using surveys and questionnaires. For example, survey questions (also called “items) must be mutually exclusive. Wording must be balanced and precise. It is important to stay away from the use of extremes such as “always” or “never” which can bias responses. Experienced survey researchers are often necessary partners in program evaluations.

5.3 Data Collection

Data collection activities typically include case assignment and tracking, questionnaire administration, information retrieval from official records and program files, data entry and other administrative tasks. Data collection may also include the process of informed consent, depending on the population targeted and the type of information collected (see section 5.5).

5.3a Developing a Data Collection Plan

It is important to put together a data collection plan before any data are collected. This plan should reflect the scope of proposed data collection activities and describe the procedures for collecting, maintaining and using all data. The first step in formulating a data collection plan is to identify the data to be collected and the sources from which the data will be drawn. Next, detailed procedures should be developed for collecting data from each proposed source. This step should also identify who will be responsible for implementing and monitoring these procedures. Last, a data collection plan should specify how data will be maintained and used. It is also helpful to think about the potential risks and benefits associated with both collecting the desired data and the proposed procedures. (Handout excerpt from the Evaluation of Teen Courts Project data collection plan.)
5.5 Informed Consent

The subjects of a research study must be fully informed of the nature of the research in which they are participating; what will, or will not, happen to them if they do, or do not, participate in the research; that they are given the opportunity to decline to participate; and that no penalty is invoked if they decline to participate in the research. This is known as informed consent. All federally funded research requires informed consent for studies that involve sensitive populations such as prisoners, children, the disabled or the poor. Though not required for non-federally funded research, federal regulations for the protection of human subjects offer important guidelines that should be applied whenever research is conducted.

Another important aspect of informed consent involves the promise of voluntary participation. Participants must be assured that their participation is voluntary. This means they not only have the right to decline to participate in the research project overall, but that they have the right to decline to answer any particular questions, or provide particular information, during any of the interviews, or to decline to participate in a particular interview at any time.

A copy of the parental consent agreement used by the ETC Project is included in this workbook. Consistent with federal regulations, this form explains the scope and nature of the research including the purpose of the research, the information to be collected, and how the Urban Institute will use this information. The voluntary nature of the research and the promise of confidentiality are also highlighted. The Urban Institute obtained consent for the ETC Project by reviewing the form with parents and youth. These individuals were asked to read the consent form and given the opportunity to ask questions. Consent to participate in the research was requested following this discussion. Both parties (i.e., the youth and the parent) were required to sign the research consent form in order for the consent to be viewed as valid.
Group Exercise #2 – Sample Question Design

The objective of this exercise is to provide an opportunity for participants to have direct experience in writing survey questions.

Instructions:

Working as a group, please compose three questions for each of the three scenarios described on the following pages.

Questions should be worded in such a way that they can be answered using the responses Strongly Agree, Agree, Disagree or Strongly Disagree.

Questions should be as brief as possible, usually no more than 30 words.

Be careful to avoid biased language and use questions that can be answered clearly and without ambiguity.
Situation #1 – Attitudes of mature adults about MTV

A media consultant wants to know how to get the over-60 crowd to watch more MTV. You have been asked to devise three questions for a survey of older adults that explore what it is about MTV that older adults don’t like. (e.g., the repetitiveness, the nudity, violence, just bad music?)

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Question 1

Question 2

Question 3
**Situation #2 – Survey of teenagers’ attitudes toward junk food**

Next, a media client wants you to develop a survey for teenagers (ages 13 to 17) on their attitudes toward junk food. Your client wants to know why teenagers eat so much junk food, and what they like about it.

<table>
<thead>
<tr>
<th>Question 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
**Situation #3 - Survey of first-time parents with infants at home.**

Finally, you have a client who wants to know what obstacles new parents face and which problems seem to trouble them the most.

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

**Question 1**

**Question 2**

**Question 3**
The name of youth court has been selected to participate in a national study of teen court programs. The Urban Institute, a non-profit research firm in Washington, DC, is conducting this evaluation for the Office of Juvenile Justice and Delinquency Prevention (OJJDP), U.S. Department of Justice. The purpose of this study is to determine whether teen court programs help keep children, like your son or daughter, from committing other offenses (i.e., getting into trouble with the law). We are also interested to learn about your child's impressions of teen court, including how fair the process was and how satisfied she or he was with the sentence received.

The Urban Institute would like to invite you and your child to participate in this study. The following things will happen as part of the research study:

- You will be asked to fill out two five-minute questionnaires or surveys. You will receive the first questionnaire when you arrive for name of youth's teen court hearing and will be asked to complete it before the hearing begins. This survey will ask you what you think will happen in youth court and about your thoughts on your child's recent behavior. You will receive the second survey right after the youth court hearing. This survey will ask you about the youth court hearing and what you think about the outcome of the hearing.

- Name of youth will also be asked to fill out three five-minute surveys. Name of youth will receive the same surveys as you, as well as a third survey. Name of youth will receive the last survey a few months after the youth court hearing. This survey will ask name of youth about current activities and impressions of youth court.

- Information will be collected for research purposes about your child from law enforcement and court records. This will include information on your child's contacts with the justice system.

- Information will be provided to the Urban Institute by the name of youth court about your child's participation in the youth court program.

Your help with the research will allow your community leaders to learn more about the problems facing youth in your community and the kind of services that would be useful for them. We promise you the following things:

- **Confidentiality:** Your and your child's responses will be kept confidential. Your name, your child's name, and any other identifying information will not be disclosed to anyone, other than the researchers conducting this study, without your permission. This promise of confidentiality also means that we will not share your child's answers with you. And, we will not share your answers with your child.

  The only exception to the promise of confidentiality is specific information about child abuse or intent to commit a future crime or intent to harm yourself. The Urban Institute and Westat will not ask you or your child about these issues. However, the Urban Institute and Westat
will report this information to the proper authorities if you or your child disclose such information.

The confidentiality of your and your child's responses are protected under Title 28 Part 22 of the Code of Federal Regulations. All Urban Institute employees must sign a Pledge of Confidentiality. I have signed a Pledge of Confidentiality requiring me not to tell anyone outside of the research team anything about you or your child.

- **Voluntary Participation:** You and your child may refuse to provide any requested information whenever you want to. You or your child may withdraw consent to participate in the study at any time. This will not affect your child's ability to participate in *name of teen court program* or the outcome of your child's youth court hearing.

By signing this form, I agree to be part of the study of *name of youth court* as described above. This consent is good for the duration of this study.

Parent/ Guardian Name: ________________________________________________________

(please print)

Parent/ Guardian Signature: ______________________________________________________

I agree to participate in the study described above. I understand that I may refuse to answer any question and that I may discontinue my participation in the research at any time without penalty.

Youth's Name: _________________________________________________________________

(please print)

Youth's Signature: ______________________________________________________________

Date: _________________________________________________________________________

Site I.D.: ______________________________________________________________________
DATA SOURCES

Data to be collected will consist of: (1) interviews with program and partner agency staff in each site; (2) brief self-administered questionnaires to be completed by teen court defendants and their parents, designed to measure respondents’ perception of teen court before and after exposure to the program and the impact of teen court on the defendant; (3) face-to-face baseline and follow-up surveys conducted by interviewers with teen court defendants and comparison group youths in one site, designed to document youth self-reported delinquency, family cohesion and conflict, peer associations, commitment to prosocial norms, maturity of judgment, and perceptions of justice; and (4) police and juvenile court records for data on offense history including the number and nature of prior offenses and any offenses committed after entry to the study.

ADMINISTRATIVE DATA
(police and juvenile court records)

Records data from the police, juvenile court, and where relevant, juvenile probation will be collected about youth in all four sites. Detailed information includes date of birth, gender, race, date of arrest, nature of offense, disposition, and date of disposition.

Protections: These data will be extracted by local agencies and stored on password-protected diskettes for transfer to the Urban Institute. Local agencies will also remove all identifying information (i.e., names, addresses, police or court case numbers) from the data collected about the "blind" comparison group prior to sending these data to the Urban Institute. To minimize the potential for loss of data, diskettes will be sent to the Urban Institute via Federal Express (i.e., this service assigns a tracking number to all parcels and increases the likelihood of retrieving a delayed package).

Risk Assessment: The use of these data to produce statistical summaries for research purposes does not pose a risk to the respondents. No additional effort on their part is required. No information will be made available that could affect how their case is handled. The research will not disclose individual level data.

Potential Benefit: Findings from this segment of the research inform program development and may enhance the services or programs available to these youths.

DATA SECURITY

Password-protected diskettes from local law enforcement and juvenile justice system agencies containing arrest and court disposition data will be stored in a locked file.
in a secure area. Electronic files created for the analysis of data from these sources will be stored a restricted-access, confidential drive. Access to the confidential drive will be limited to only those project researchers who need to work with the data. Implementation of the following measures will ensure that the security of individual client data are protected and maintained:

- Data will only be collected from the individuals for whom a signed consent has been obtained.
- Data identifiable to an individual subject will not be revealed.
- Access to data will be restricted to only those research team members who require access for evaluation purposes and who have agreed, in writing, to maintain the confidentiality of all data.
- Adequate precautions will be taken to ensure the physical and administrative security of personally identifiable data.
- Project findings and reports prepared for dissemination will not contain information that could readily be used to identify an individual participant or the members of his/her family.
SECTION 6: GUIDELINES FOR SELECTING AN EVALUATOR
6.0 Selecting, Hiring and Working with an Evaluator

A final question to consider is whether to conduct an evaluation internally (meaning, program staff are responsible for defining, developing, implementing, monitoring and explaining the evaluation) or to hire an evaluator. This section offers guidelines for selecting, hiring and working with an evaluator.

6.1 Defining objectives and goals

It is important to identify and define the goals and objectives of the evaluation before speaking with an evaluator. Make sure to factor in budget and timing constraints (the upper and lower limits for spending). Defining the evaluation goals and objectives will help clarify what the evaluator is being asked to do, as well as form realistic expectations.

6.2 Issuing a “Request for Proposal” (RFP)

A Request for Proposal (RFP) is a key step in both selecting and hiring an evaluator. The RFP should clearly state the tasks required, the time frame for completion and the amount of money available, and include the pertinent fiscal or reporting policies for your organization. The RFP should also instruct applicants to address each task and to submit a detailed budget.

6.3 Qualifications and other considerations

When selecting an evaluator, it is important to consider the scope and breadth of the individual’s evaluation experience (technical skills and capabilities) including his or her involvement with and access to an Institutional Review Board (most universities and colleges have an IRB which exists to screen research projects for compliance with standards for the protection for the human subjects). An equally important factor is the evaluator’s level of familiarity with teen courts – the theory behind teen courts and how they function – and the juvenile justice system.

---

3 These guidelines were crafted from several materials including Hiring and Working with an Evaluator produced by the Justice Research and Statistics Center (the full text document is included with this workbook).
If possible, review the evaluator’s prior work and ask for client references. Use these resources to gauge individual's evaluation expertise, ability to work collaboratively and to complete work on time and within budget.

6.3 Developing a Contract

A contract should be developed and signed before work begins. In addition to budget issues, this contract should address data rights (who owns the data and data files once the evaluation ends; who may use the data), reporting requirements, and the products and services to be delivered.
Appendices
Appendix A
GLOSSARY OF EVALUATION TERMS
Glossary of Evaluation Terms

This glossary of terms was developed by the Administration for Children and Families, U.S. Department of Health and Human Services, and can be found in the resource titled, Program Manager’s Guide to Evaluation. It is available at http://www.acf.dhhs.gov/programs/core/pubs_reports/glossary_pmguide.html (Accessed October 16, 2002.)

**baseline data** — Initial information on program participants or other program aspects collected prior to receipt of services or program intervention. Baseline data are often gathered through intake interviews and observations and are used later for comparing measures that determine changes in your participants, program, or environment.

**bias** — (refers to statistical bias). Inaccurate representation that produces systematic error in a research finding. Bias may result in overestimating or underestimating certain characteristics of the population. It may result from incomplete information or invalid collection methods and may be intentional or unintentional.

**comparison group** — Individuals whose characteristics (such as race/ethnicity, gender, and age) are similar to those of your program participants. These individuals may not receive any services, or they may receive a different set of services, activities, or products. In no instance do they receive the same service(s) as those you are evaluating. As part of the evaluation process, the experimental (or treatment) group and the comparison group are assessed to determine which type of services, activities, or products provided by your program produced the expected changes.

**confidentiality** — Since an evaluation may entail exchanging or gathering privileged or sensitive information about individuals, a written form that assures evaluation participants that information provided will not be openly disclosed nor associated with them by name is important. Such a form ensures that their privacy will be maintained.

**consultant** — An individual who provides expert or professional advice or services, often in a paid capacity.

**control group** — A group of individuals whose characteristics (such as race/ethnicity, gender, and age) are similar to those of your program participants, but do not receive the program (services, products, or activities) you are evaluating. Participants are randomly assigned to either the treatment (or program) group and the control group. A control group is used to assess the effect of your program on participants as compared to similar individuals not receiving the services, products, or activities you are evaluating. The same
information is collected for people in the control group as in the experimental group.

**cost-benefit analysis** — A type of analysis that involves comparing the relative costs of operating a program (program expenses, staff salaries, etc.) to the benefits (gains to individuals or society) it generates. For example, a program to reduce cigarette smoking would focus on the difference between the dollars expended for converting smokers into nonsmokers with the dollar savings from reduced medical care for smoking related disease, days lost from work, and the like.

**cost effectiveness analysis** — A type of analysis that involves comparing the relative costs of operating a program with the extent to which the program met its goals and objectives. For example, a program to reduce cigarette smoking would estimate the dollars that had to be expended in order to convert each smoker into a nonsmoker.

**cultural relevance** — Demonstration that evaluation methods, procedures, and/or instruments are appropriate for the culture(s) to which they are applied. (Other terms include cultural competency, cultural sensitivity).

**culture** — The shared values, traditions, norms, customs, arts, history, institutions, and experience of a group of people. The group may be identified by race, age, ethnicity, language, national origin, religion, or other social category or grouping.

**data** — Specific information or facts that are collected. A data item is usually a discrete or single measure. Examples of data items might include age, date of entry into program, or reading level. Sources of data may include case records, attendance records, referrals, assessments, interviews, and the like.

**data analysis** — The process of systematically applying statistical and logical techniques to describe, summarize, and compare data collected.

**data collection instruments** — Forms used to collect information for your evaluation. Forms may include interview instruments, intake forms, case logs, and attendance records. They may be developed specifically for your evaluation or modified from existing instruments. A professional evaluator can help select those that are most appropriate for your program.

**data collection plan** — A written document describing the specific procedures to be used to gather the evaluation information or data. The plan describes who collects the information, when and where it is collected, and how it is to be obtained.
**database** — An accumulation of information that has been systematically organized for easy access and analysis. Databases typically are computerized.

**design** — The overall plan and specification of the approach expected in a particular evaluation. The design describes how you plan to measure program components and how you plan to use the resulting measurements. A pre- and post-intervention design with or without a comparison or control group is the design needed to evaluate participant outcome objectives.

**evaluation** — A systematic method for collecting, analyzing, and using information to answer basic questions about your program. It helps to identify effective and ineffective services, practices, and approaches.

**evaluator** — An individual trained and experienced in designing and conducting an evaluation that uses tested and accepted research methodologies.

**evaluation plan** — A written document describing the overall approach or design you anticipate using to guide your evaluation. It includes what you plan to do, how you plan to do it, who will do it, when it will be done, and why the evaluation is being conducted. The evaluation plan serves as a guide for the evaluation.

**evaluation team** — The individuals, such as the outside evaluator, evaluation consultant, program manager, and program staff who participate in planning and conducting the evaluation. Team members assist in developing the evaluation design, developing data collection instruments, collecting data, analyzing data, and writing the report.

**exit data** — Information gathered after an individual leaves your program. Exit data are often compared to baseline data. For example, a Head Start program may complete a developmental assessment of children at the end of the program year to measure a child's developmental progress by comparing developmental status at the beginning and end of the program year.

**experimental group** — A group of individuals receiving the treatment or intervention being evaluated or studied. Experimental groups (also known as treatment groups) are usually compared to a control or comparison group.

**focus group** — A group of 7-10 people convened for the purpose of obtaining perceptions or opinions, suggesting ideas, or recommending actions. A focus group is a method of collecting data for evaluation purposes.

**formative evaluation** — A type of process evaluation of new programs or services that focuses on collecting data on program operations so that needed changes or modifications can be made to the program in its early stages. Formative evaluations are used to provide feedback to staff about the program components that are working and those that need to be changed.
**immediate outcomes** — The changes in program participants, knowledge, attitudes, and behavior that occur early in the course of the program. They may occur at certain program points, or at program completion. For example, acknowledging substance abuse problems is an immediate outcome.

**impact evaluation** — A type of outcome evaluation that focuses on the broad, longer-term impacts or results of a program. For example, an impact evaluation could show that a decrease in a community's overall infant mortality rate was the direct result of a program designed to provide early prenatal care.

**in-kind service** — Time or services donated to your program.

**informed consent** — A written agreement by program participants to voluntarily participate in an evaluation or study after having been advised of the purpose of the study, the type of information being collected, and how the information will be used.

**instrument** — A tool used to collect and organize information. Includes written instruments or measures, such as questionnaires, scales, and tests.

**intermediate outcomes** — Results or outcomes of a program or treatment that may require some time before they are realized. For example, part-time employment would be an intermediate outcome of a program designed to assist at-risk youth in becoming self-sufficient.

**internal resources** — An agency's or organization's resources including staff skills and experiences and any information you already have available through current program activities.

**intervention** — The specific services, activities, or products developed and implemented to change or improve program participants' knowledge, attitudes, behaviors, or awareness.

**logic model** — See the definition for program model.

**management information system (MIS)** — An information collection and analysis system, usually computerized, that facilitates access to program and participant information. It is usually designed and used for administrative purposes. The types of information typically included in an MIS are service delivery measures, such as session, contacts, or referrals; staff caseloads; client sociodemographic information; client status; and treatment outcomes. Many MIS can be adapted to meet evaluation requirements.

**measurable terms** — Specifying, through clear language, what it is you plan to do and how you plan to do it. Stating time periods for activities, "dosage" or
frequency information (such as three 1-hour training sessions), and number of participants helps to make project activities measurable.

**methodology** — The way in which you find out information; a methodology describes how something will be (or was) done. The methodology includes the methods, procedures, and techniques used to collect and analyze information.

**monitoring** — The process of reviewing a program or activity to determine whether set standards or requirements are being met. Unlike evaluation, monitoring compares a program to an ideal or exact state.

**objective** — A specific statement that explains how a program goal will be accomplished. For example, an objective of the goal to improve adult literacy could be to provide tutoring to participants on a weekly basis for 6 months. An objective is stated so that changes, in this case, an increase in a specific type of knowledge, can be measured and analyzed. Objectives are written using measurable terms and are time-limited.

**outcome** — Outcomes are a result of the program, services, or products you provide and refer to changes in knowledge, attitude, or behavior in participants. They are referred to as participant outcomes in this manual.

**outcome evaluation** — Evaluation designed to assess the extent to which a program or intervention affects participants according to specific variables or data elements. These results are expected to be caused by program activities and tested by comparison of results across sample groups in the target population. Also known as impact and summative evaluation.

**outcome objectives** — The changes in knowledge, attitudes, awareness, or behavior that you expect to occur as a result of implementing your program component, service, or activity. Also known as participant outcome objectives.

**outside evaluator** — An evaluator not affiliated with your agency prior to the program evaluation. Also known as a third-party evaluator.

**participant** — An individual, family, agency, neighborhood, community, or State, receiving or participating in services provided by your program. Also known as a client or target population group.

**pilot test** — Preliminary test or study of your program or evaluation activities to try out procedures and make any needed changes or adjustments. For example, an agency may pilot test new data collection instruments that were developed for the evaluation.
**posttest** — A test or measurement taken after a service or intervention takes place. It is compared with the results of a pretest to show evidence of the effects or changes as a result of the service or intervention being evaluated.

**pretest** — A test or measurement taken before a service or intervention begins. It is compared with the results of a posttest to show evidence of the effects of the service or intervention being evaluated. A pretest can be used to obtain baseline data.

**process evaluation** — An evaluation that examines the extent to which a program is operating as intended by assessing ongoing program operations and whether the targeted population is being served. A process evaluation involves collecting data that describes program operations in detail, including the types and levels of services provided, the location of service delivery, staffing; sociodemographic characteristics of participants; the community in which services are provided, and the linkages with collaborating agencies. A process evaluation helps program staff identify needed interventions and/or change program components to improve service delivery. It is also called formative or implementation evaluation.

**program implementation objectives** — What you plan to do in your program, component, or service. For example, providing therapeutic child care for 15 children, giving them 2 hot meals per day, are referred to as program implementation objectives.

**program model (or logic model)** — A diagram showing the logic or rationale underlying your particular program. In other words, it is a picture of a program that shows what it is supposed to accomplish. A logic model describes the links between program objectives, program activities, and expected program outcomes.

**qualitative data** — Information that is difficult to measure, count, or express in numerical terms. For example, a participant's impression about the fairness of a program rule/requirement is qualitative data.

**quantitative data** — Information that can be expressed in numerical terms, counted or compared on a scale. For example, improvement in a child's reading level as measured by a reading test.

**random assignment** — The assignment of individuals in the pool of all potential participants to either the experimental (treatment) or control group in such a manner that their assignment to a group is determined entirely by chance.

**reliability** — Extent to which a measurement (such as an instrument or a data collection procedure) produces consistent results over repeated observations or administrations of the instrument under the same conditions each time. It is also
important that reliability be maintained across data collectors; this is call interrater reliability.

**sample** — A subset of participants selected from the total study population. Samples can be random (selected by chance, such as every 6th individual on a waiting list) or nonrandom (selected purposefully, such as all 2-year olds in a Head Start program).

**standardized instruments** — Assessments, inventories, questionnaires, or interviews, that have been tested with a large number of individuals and are designed to be administered to program participants in consistent manner. Results of tests with program participants can be compared to reported results of the tests used with other populations.

**statistical procedures** — The set of standards and rules based in statistical theory, by which one can describe and evaluate what has occurred.

**statistical test** — Type of statistical procedure, such as a t-test or Z-score, that is applied to data to determine whether your results are statistically significant (i.e., the outcome is not likely to have resulted by chance alone).

**summative evaluation** — A type of outcome evaluation that assesses the results or outcomes of a program. This type of evaluation is concerned with a program's overall effectiveness.

**treatment group** — Also called an experimental group, a treatment group is composed of a group of individuals receiving the services, products, or activities (interventions) that you are evaluating.

**validity** — The extent to which a measurement instrument or test accurately measures what it is supposed to measure. For example, a reading test is a valid measure of reading skills, but is not a valid measure of total language competency.

**variables** — Specific characteristics or attributes, such as behaviors, age, or test scores, that are expected to change or vary. For example, the level of adolescent drug use after being exposed to a drug prevention program is one variable that may be examined in an evaluation.
Appendix B
EVALUATION STRATEGIES FOR HUMAN SERVICES PROGRAMS: A GUIDE FOR POLICYMAKERS AND PROVIDERS
Evaluation Strategies for Human Services Programs
A Guide for Policymakers and Providers

Adele Harrell
with
Martha Burt
Harry Hatry
Shelli Rossman
Jeffrey Roth
William Sabol

The Urban Institute
Washington, D.C.

Contents
Clarifying the Evaluation Questions, 2
Developing a Logic Model, 3
Assessing Readiness for Evaluation, 7
Selecting an Evaluation Design, 8
Identifying Potential Evaluation Problems, 27
Conclusions, 30

EXHIBITS
Exhibit A: Logic Model Used in Evaluation of the Children At Risk Program, 6
Exhibit B: Process for Selecting Impact Evaluation Designs, 18

Evaluation Strategies for Human Services Programs
A Guide for Policymakers and Providers

In the continuing effort to improve human service programs, funders, policymakers, and service providers are increasingly recognizing the importance of rigorous program evaluations. They want to know what the programs accomplish, what they cost, and how they should be operated to achieve maximum cost-effectiveness. They want to know which programs work for which groups, and they want conclusions based on evidence, rather than testimonials and impassioned pleas.

This paper lays out, for the nontechnician, the basic principles of program evaluation design. It signals common pitfalls, identifies constraints that need to be considered, and presents ideas for solving potential problems. These principles are general and can be applied to a wide range of human service programs. We illustrate these principles here with examples from programs for vulnerable children and youth. Evaluation of these programs is particularly challenging because they address a wide diversity of
problems and possible solutions, often include multiple agencies and clients, and change over time to meet shifting service needs.

Steps in Selecting the Appropriate Evaluation Design. The first step in the process of selecting an evaluation design is to clarify the questions that need to be answered. The next step is to develop a logic model that lays out the expected causal linkages between the program (or program components) and the program goals. Without tracing these anticipated links it is impossible to interpret the evaluation evidence that is collected. The third step is to review the program to assess its readiness for evaluation. These three steps can be done at the same time or in overlapping stages. For expositional clarity we will discuss each of them in turn. We will then describe how to select the best design for a given purpose from among the major types of evaluation that exist.

Clarifying the Evaluation Questions
The design of any evaluation begins by defining the audience for the evaluation findings, what they need to know, and when. These questions determine which of the following four major types of evaluation should be chosen:

Impact evaluations focus on questions of causality. Did the program have its intended effects? If so, who was helped and what activities or characteristics of the program created the impact? Did the program have any unintended consequences, positive or negative?

Performance monitoring provides information on key aspects of how a system or program is operating and the extent to which specified program objectives are being attained (e.g., numbers of youth served compared to target goals, reductions in school dropouts compared to target goals). Results are used by service providers, funders, and policymakers to assess the program's performance and accomplishments.

Process evaluations answer questions about how the program operates and document the procedures and activities undertaken in service delivery. Such evaluations help identify problems faced in delivering services and strategies for overcoming these problems. They are useful to practitioners and service providers in replicating or adapting program strategies.

Cost evaluations address how much the program or program components cost, preferably in relation to alternative uses of the same resources and to the benefits being produced by the program. In the current fiscal environment, programs must expect to defend their costs against alternative uses.

A comprehensive evaluation will include all these activities. Sometimes, however, the questions raised, the target audience for findings, or the available resources limit the evaluation focus to one or two of these activities.

Whether to provide preliminary evaluations to staff for use in improving program operations and developing additional services is an issue that needs to be faced. Preliminary results can be effectively used to identify operational problems and develop the capacity of program staff to conduct their own ongoing evaluation and monitoring activities.(1) But this use of evaluation findings, called formative evaluations, presents a challenge to evaluators who are faced with the much more difficult task of estimating the impact of an evolving intervention. When the program itself is continuing to change, measuring impact requires ongoing measurement of the types and level of service provided. The danger in formative evaluations is that the line between program operations and assessment will be blurred. The extra effort and resources required for impact analysis in formative evaluations has to be measured against the potential gains to the program from ongoing improvements and the greater usefulness of the final evaluation findings.
Developing a Logic Model

It is impossible to interpret evaluation findings without a clear understanding of program goals, implementation sequences, and the expected links between them and expected program benefits. Expectations about these linkages are made explicit by developing a logic model. Such a model is developed by discussing with service providers and funders the goals of and rationales behind program organization and content, examining planning documents and program reports, and reviewing research findings on similar programs or problems. The literature review may be particularly helpful in identifying plausible causal links and any factors other than the program which should be considered in the evaluation.

The logic model provides a simplified description of the program, the intended outputs, and the intended outcomes. Program characteristics include the population to be reached, the resources to be used, and identification of the types and levels of service elements. Outputs are immediate program products resulting from the internal operations of the program, such as the delivery of planned services. Examples of output indicators in the area of programs for vulnerable children and youth might include the numbers of children immunized, home visits by case managers, or youth completing a job training program. These program outputs are, in turn, the vehicle for producing the desired program outcomes, for example, decreases in childhood illnesses, decreases in abuse and neglect cases, or increases in youth employment. Careful attention must be paid to when the anticipated outcome should be expected to occur. For this reason it is often useful to divide outcomes into intermediate versus longer term. For example, improved school attendance in early grades might be an intermediate outcome associated with the longer-term outcome of dropout prevention. Care must be given to focusing on outcomes which will occur within the study period.

A classic failure in selecting an outcome that is expected to occur within the time frame of the study occurred in evaluations of the DARE drug prevention program, an educational program for fifth and sixth graders designed to prevent drug use. Evaluation results showed no significant prevention of drug use at the end of the program. This result should have been anticipated, since drug use does not typically begin among youth in this country until the mid-teen years (14 to 17). An age-appropriate intermediate outcome should have been selected as the primary outcome measure, such as improved peer resistance skills and changes in beliefs about the risks of drug use.

The logic model should also include explicit mapping of the conditions present in the program environment or characteristics of the target group or community that may affect the program's ability to achieve its goals. Non-program characteristics of the program organization, community or target population that are likely to influence the outputs and outcomes and/or use of program services are called antecedent variables. Conditions or events in the program, target population, or community that may limit or expand the extent to which program outputs actually produce the desired outcomes are called mediating variables. For example, a drug abuse prevention program may be less effective if the program staff are inexperienced, or if the local community offers fewer recreational alternatives to substance abuse and/or more active open drug markets (antecedent variables). Offering other support services in combination with the program may enhance its impact (a mediating variable).

In impact evaluations the logic model is used to spell out how, and for whom, certain services are expected to create specific changes/benefits. For example, if the program includes parenting classes, the logic model will identify this activity as a key program component and show the types of changes in parenting that will be used to measure program outcomes (e.g., by improving parental assistance with homework or helping parents communicate more effectively with adolescents).

In performance monitoring, the logic model is used to focus on which kinds of output and outcome
indicators are appropriate for specific target populations, communities, or time periods. For example, among indicators of child improvement in school, one might expect attendance to improve in the first semester of a program, but academic test score improvement only after a significant period of program participation—with the timing possibly varying by the age and developmental stage of the children.

In process evaluation, the logic model is used to identify expectations about how the program should work—an "ideal type"—which can then be used to assess the deviations in practice, why these deviations have occurred, and how the deviations may affect program outputs. This assists program managers (and evaluators) to identify differences (including positive and negative unintended consequences), consider possible mechanisms for fine-tuning program operations to align the actual program with the planned approach, or re-visit program strategies to consider alternatives.(2)

Logic models are constructed to show temporal sequences, building left to right, and they typically diagram relationships with arrows. An example of a logic model is shown in Exhibit A. It was developed by the Urban Institute during the planning of the evaluation of the Children At Risk program (CAR). CAR is an intensive intervention program designed to prevent involvement in drugs and crime, and to foster healthy development among adolescents ages 13 to 15 who exhibit serious risk indicators and live in severely distressed inner-city neighborhoods.

The intervention consists of eight required program components:

Case Managers employed by the program make a service plan for all members of the household of participating youth and provide intensive follow-up on referrals to needed services, handling a caseload of 15;

Family Services include parenting skills training for all parents, and referral to other services as needed (intensive family counseling, stress management/coping skills training, identification and treatment of substance abuse, health care, job training and employment programs, housing, and income support services);

Education Services include tutoring or homework assistance for all youth, and referral to other services as needed (educational testing, special education classes);

After-School and Summer Activities for all CAR youth include recreational programs and life-skill/leadership development activities, combined with training or education;

Mentoring is provided by local organizations for youth in need of a caring relationship with an adult. The role of the mentor is to: (a) inform youth about alternative available choices (e.g., activities and goals); (b) familiarize them with strategies available for pursuing those choices; (c) provide training, opportunities for practice, and feedback in the development of skills for implementing particular strategies; and (d) provide relationships through which youth are affirmed, inspired, and encouraged to make healthy choices;

Incentives such as gifts and special events are used to build morale and attachment to the pro-social goals of the program (e.g., gift certificates, trips, and vouchers for pizza, sports shops, movies, and stipends for community service during summer programs);

Community Policing/Enhanced Enforcement is used in all target neighborhoods to create safer environments with less drug activity. Law enforcement activities include out-stationing police in schools and neighborhood locations to maintain order and enhance relationships with community
groups;

**Criminal/juvenile Justice Intervention** involves collaboration between case managers and juvenile court personnel to provide community service opportunities and enhanced supervision of youth in the justice system.

*Antecedent variables* include the levels and types of neighborhood, family, peer group, and personal risk factors for participants as well as their demographic characteristics. These are influences that are present before the program intervention.

*Mediating variables* include exposure to other social or educational services, perceptions of opportunities, and social norms. These are influences that operate at the same time as the program is operating. The program components are designed to achieve the intermediate outcomes—reductions in risk factors and enhancement of protective factors at the end of program participation. These intermediate outcomes, measured at the end of program participation, are hypothesized to be requisite steps towards the desired longer-term outcomes—prevention of drug use, drug selling, delinquency, school failure and dropout, and teen parenthood.

Program outputs, not shown in this diagram, include indicators of performance such as the number of tutoring sessions provided, number of home visits by case managers, and number of times parents participated in program activities.
Assessing Readiness for Evaluation

Evaluability assessment is a systematic procedure for deciding whether program evaluation is justified, feasible, and likely to provide useful information. Questions to be considered in an evaluability assessment include: (3)

Is the program’s logic model plausible given the resources available and guidance from the relevant literature? If program goals are unrealistic or the intervention strategies not well grounded in theory and/or prior evidence, then evaluation is not a good investment.

What kinds of data will be needed, from what number of subjects, and what data are likely to be already available? Evaluations should be designed to maximize the use of available data, as long as these are valid indicators of important concepts and are reliable. Available data may, for example, include government statistics, individual and summary agency records and statistics, and information collected by researchers for other studies. If there are crucial data needs not met with existing data, resources must be available to collect the requisite new data.

Are adequate resources and assets available—money, time, expertise, and community and government support? Are there any factors that limit or constrain access to these resources?

Can the evaluation be achieved in a time frame that will permit the findings to be useful in making program and policy decisions by federal, state, and local officials?

To what extent does evaluation information already exist somewhere on the same or a closely related intervention? The answer to this question can have important implications for action. Any successful previous attempts may yield promising models for replication. Lessons learned from previous unsuccessful attempts may inform the current effort. If sufficient evidence already exists from previous efforts, the value of a new evaluation may be marginal.

To what extent are the findings from an evaluation likely to be generalizable to other communities, and therefore useful in assessing whether the program should be expanded to other settings or areas? Are there unique characteristics of the projects to be evaluated that might not apply to most other projects? Program characteristics that are not generalizable reduce the value of any findings.

Selecting an Evaluation Design

Selection of the evaluation design follows the systematic consideration of these questions. As noted, there are four major types of evaluation: impact, performance monitoring, process, and cost. We discuss each in turn.

Impact Evaluation Designs

Three possible designs are possible for impact evaluations: experimental, quasi-experimental, and non-experimental. They all share the strategy of comparing program outcomes with some measure of what would have happened without the program. Experimental designs are the most powerful and produce the strongest evidence. These are not always possible, however, in which case one of the two other
alternatives must be chosen. (A later section discusses how to make the choice.)

**EXPERIMENTAL DESIGNS**

*Key elements.* Experimental designs are considered the "gold standard" in impact evaluation. Experiments require that individuals or groups, such as classrooms or schools, be assigned at random (by the flip of a coin or equivalent randomizing procedure) to one or more groups prior to the start of services. The "treatment" group or groups will be designated to receive particular services designed to achieve clearly specified outcomes. If multiple treatment groups are designated, the outcomes for the treatment groups may be compared to one another to estimate the relative impact of the different services or the impact relative to a control group. A "control" group receives no services. The treatment group outcomes are compared to control group outcomes to estimate impact. Because chance alone determines who receives the program services, the groups can be assumed to be similar on all characteristics that might affect the outcome measures except the program. Any differences between treatment and control groups, therefore, can be attributed with confidence to the impacts of the program.

*Design Variations.* One design variation is based on a random selection of time periods during which services are provided. For example, new services may be offered on randomly chosen weeks or days. A version of this approach is to use "week on/week off" assignment procedures. Although not truly random, this approach closely approximates random assignment if client characteristics do not vary systematically from week to week. It has the major advantage that program staff often find it easier to implement than making decisions on program entry by the flip of a coin on a case-by-case basis. A second design variation is a staggered start approach -in which some members of the target group are randomly selected to receive services with the understanding that the remainder will receive services at a later time (in the case of a school or classroom, the next semester or month). One disadvantage of the staggered start design is that the observations of outcomes are limited to the period between the time the first group completes the program and the second group begins. As a result, it is generally restricted to assessing gains made during participation in relatively short-term programs.

*Limitations/Considerations.* Although experiments are the preferred design for an impact evaluation on scientific grounds, random assignment evaluations are not always the ideal choice in real-life settings. Some interventions are inherently impossible to study through randomized experiments. Youth curfews, for example, cannot be enforced against a randomly selected subset of children in a community. And "week on/week off" enforcement is likely to breed contempt for both the law and enforcement.

A second consideration is whether random assignment is ethical and acceptable to the community. Public opinion may resist treating similar children differently on the basis of a coin flip or may view random assignment as exploiting vulnerable populations and powerless people. Carefully designed procedures for randomization may be able to overcome such resistance. One strategy is random selection of those to receive services from a list of those who meet eligibility requirements when resources are not available to serve everyone who is eligible. This form of drawing lots is close enough to "first come, first served" to be accepted as fair in many situations. Providing services for some clients at a later time (the next month or semester as described above) may satisfy community concerns about fairness and be consistent with available staff and resources. Sometimes, random assignment can involve relaxing a requirement instead of adding one, which makes randomization less controversial.

Great care needs to be taken to ensure that the control group is not denied essential services they would otherwise have, that the benefits to participants and the community are carefully explained, and that program staff and participants understand and support the research. Many funders require a formal review of the research design by a panel trained in guidelines developed to protect research participants. Even when such review is not required, explicit consideration of this issue is essential.
A third important issue is whether the results that are likely to be obtained justify the investment. Experiments typically require high levels of resources—money, time, expertise, and support from program staff, government agencies, funders and the community. Evaluation planners have to ask themselves whether the answers to the list of evaluation questions—and the decisions on program continuation, expansion, or modification that will be made on the basis of the findings—could be based on less costly, less definitive, but still acceptable evaluation strategies.

**Practical Issues.** Experimental designs run the most risk of being contaminated because of deliberate or accidental mistakes made in the field. To minimize this danger, there must be close collaboration between the evaluation team and the program staff in identifying objectives, setting schedules, dividing responsibilities for record-keeping and data collection, making decisions regarding client contact, and sharing information on progress and problems. Active support of the key program administrators, ongoing staff training and communication via meetings, conference calls, or e-mail are essential.

Failure to adhere to the plan for random assignment is a common problem. Staff are often intensely committed to their clients and will want to base program entry decisions on their perceptions of who needs, or will benefit from, the program. To prevent this pitfall, procedures should be set up so that the evaluator, not program staff, is in charge of the allocation to treatment or control group. Statistical adjustments in the analysis may be needed if there are operational failures to maintain the randomization process(4). And even these may be inadequate to remove the biases thus introduced.

Another potential problem area is noncomparable information for treatment and control group members. Program staff can readily collect data and provide contact information for treatment group members because they have continuing contacts with clients, other agencies, and the community. Collecting comparable data and contact information on control group members can be difficult. If the experiment loses track altogether of more control than treatment group members, the evaluation data will not only be incomplete, it will provide distorted and therefore misleading information on what impacts the program has. The best way to avoid bias from this problem (called differential attrition) is to plan tracking procedures and data collection at the start of the evaluation, gathering information from the control group members on how they can be located, and developing agreements with other community agencies, preferably in writing, for assistance in data collection and sample member tracking. These agreements are helpful in maintaining sample continuity in the face of staff turnover at the agencies involved.

If the program services and content change over time, it may be difficult to determine what level or type of services produced the outcomes. The best strategy is to identify key changes in the program and the timing of changes as part of a process evaluation and use this information to define "types of program" variations in the program experience of different participants for the impact analysis. Other potential problems may be solvable through the use of special statistical techniques. Such problems include insufficient or unequal follow-up periods for treatment versus control,(5) and the risk of events (e.g., failure in school, incarceration, injury, moving) that are more likely to remove some types of members from a sample than others before the end of the planned follow-up period.(6)

**Example.** The evaluation of Project Alert, an eight-week junior high school curriculum for teaching seventh grade students to avoid drug use, used an experimental design(7). Thirty California and Oregon schools were randomly assigned to three groups: 1) students instructed by adult health educators, 2) students instructed by older teenagers, and 3) a no-treatment control group, although four of the non-treatment schools provided other drug prevention instructional programs. To increase the generalizability of the findings, the schools were drawn from eight urban, suburban, and rural communities and nearly a third of the schools had minority populations of 50 percent or higher. To
increase the pre-assignment similarities of the three experimental groups and strengthen the statistical power of the analysis (given the relatively small sample of schools), each experimental group was included in at least one school in each community, and the schools included in the experiment were matched to the extent possible to reduce differences among groups in such characteristics as test scores, language spoken at home, drug use among 8th graders, and ethnic and income composition. These procedures produced substantial pre-experimental similarities in factors related to drug use among the experimental groups. Since schools but not students were randomly assigned, statistical adjustments were used to correct for the clustering of students within schools. Students completed questionnaires about their drug use seven times between grades 7 and 12; those who transferred to other schools or districts completed mail and telephone interviews to minimize sample attrition. Outcome measures included cognitive risk factors associated with drug use: beliefs about consequences of use, norms regarding drug use, peer resistance, self-efficacy, and expected future drug use.

Experimental evaluations are costly. The Children At Risk evaluation, for example, cost $1.5 million. But the rigorous design permitted strong conclusions about the long-term effectiveness of drug prevention education during early adolescence and demonstrated that results are not restricted to middle class communities, but can be used in schools with high proportions of lower income and minority students.

QUASI-EXPERIMENTAL DESIGNS

Key Elements. Like experiments, quasi-experimental evaluations compare outcomes from program participants to outcomes for comparison groups that do not receive program services. The critical difference is that the decision on who receives the program is not random. Comparison groups are made up of members of the target population as similar as possible to program participants on factors that could affect the selected outcomes to be observed. Multivariate statistical techniques are then used to control for remaining differences between the groups.

Usually, evaluators use existing population groups for comparison - those who live in a similar area, or are enrolled in the same school in a different classroom, or attended the same school with the same teacher in the previous year. In some situations, staff (or schools or communities) are willing or trained to try the new "treatment" while others are not, but the same rules for service eligibility are used by all.

Design Variations. The primary variation is to construct a comparison group by matching individuals in the treatment group on a selected set of characteristics. This process for selecting a comparison group is methodologically less defensible(8). The threats to validity are twofold. 1) Matches based on similarities at a single point in time do not always result in groups of individuals who are comparable over time. Thus, the groups may become increasingly different over time independent of the program, 2) Differences in variables not used in the matching may have a substantial effect independently of the program being evaluated.

Quasi-experimental designs vary in the number and timing of the collection of data on program outcome measures. The selection of the number and timing of measurements is based on an assessment of the potential threats posed by competing hypotheses that cannot be ruled out by the comparison methodology. In many situations, the strongest designs are those that collect pre-program measures of outcomes and risk factors and use these in the analysis to focus on within-individual changes that occur during the program period. These variables are also used to identify groups of participants who benefit most from the services. One design variation involves additional measurement points (in addition to simple before and after) to measure trends more precisely. Another variation is useful when pre-program data collection (such as administering a test on knowledge or attitudes) might "teach" youth about the questions to be asked after the program to measure change, and thus distort the measurement of program impact. This variation involves limiting data collection to the end of the program period for some
groups, allowing their post-program answers to be compared with the post-program answers of those who also participated in the pre-program testing.

**Considerations/Limitations.** Use of non-equivalent control group designs requires careful attention to procedures that rule out competing hypotheses regarding what caused any observed differences on the outcomes of interest. In evaluations of programs for vulnerable children and youth, three threats to validity stand out.(9)

The first is the threat of "maturation"--the possibility that age-related processes will contribute to outcomes independently of the program intervention. Among youth, certain outcomes, positive and negative, are strongly tied to age--outcomes such as drug use, delinquency, and early parenthood. It is therefore necessary to be sure that the comparison group is made up of youth at the same developmental stage.

A second threat is that of "history"--the risk that unrelated events may affect outcomes. For example, the rapid spread of crack use among women childbearing age in the United States in the late 1980s greatly increased rates of drug-exposed infants. Thus, a comparison group for an evaluation of a prenatal health care program would need to be drawn from the same years and communities to "control" for the spread of crack. Otherwise, the upward trend in negative outcomes due to crack could obscure the prevention benefits of the program. Similarly, designs need to consider controls for geographic variation in events external to the program. For example, gang crackdowns in some neighborhoods and not others could influence assessments of the impact of a school-based delinquency or drug prevention program. If the crackdown occurred in the "treatment" neighborhood, the program effects might be over-estimated; if it occurred in the comparison neighborhood, program effects might be under-estimated.

A third threat to validity is the process of "selection"--the factors that determine who receives services. Some of these factors are readily identified and can be used as control variables in statistical models, such as living in a specific school district or meeting program eligibility criteria. However, it is unlikely that all factors will be correctly identified and adequately measured. For example, program participants may receive services because they are more motivated, skillful, or socially well connected than nonparticipants. Such differences are not easy to measure during a program evaluation.

**Practical Problems.** Building defenses or "controls" for threats to validity into evaluation designs through the selection of comparison groups and the timing of outcome observations is a challenge. Controls for maturation, history, and selection may involve, respectively, selecting a sample that includes multiple age cohorts, collecting data in similar or nearby localities that lack the program, or applying a statistical model that controls for foreseeable biases in selecting program participants. Even when the comparison group is carefully selected, the researcher cannot be sure that all relevant group differences have been identified and measured accurately. Statistical methods can adjust for such problems and increase the precision with which program effects can be estimated, but they do not fully compensate for the non-random design. Findings need to be interpreted extremely cautiously and untested alternative hypotheses carefully considered.

As in experimental evaluation, plans for quasi-experimental evaluations need to pay close attention to the problem of collecting comparable information on control group members and developing procedures for tracking them. However, the need for close collaboration with program staff is reduced, since the staff are generally neither involved in selecting participants nor in contact with comparison group members.

**Example.** The evaluation of the Teen Age Parenting Program (TAPP) for adolescents divided teen mothers into three groups designed to be similar in age and other characteristics. Each group was
evenly divided among black, Hispanic, and white participants. One group attended an alternative school with child development and parenting classes and a nursery school featuring a parenting-child development curriculum. Another group attended an alternative school without a nursery school. The remaining group received no special services for teenage parents. Services began during pregnancy. Assessments of educational progress, fertility, knowledge, and child development two to four years later were based on interviews and school records. Mothers in the alternative school with the nursery program had completed more schooling and were more likely to still be enrolled in school than the other mothers. Mothers in both alternative schools had more knowledge about parenting and reproduction and more positive attitudes about parenting than those without special services. But there were no significant differences in the groups on child development outcome measures. How to interpret this seeming inconsistency is complicated, because the evaluation design did not have pre-program measures of individual differences and assignment was not random. The education and knowledge differences across the three groups may have been there from the beginning, rather than being attributable to the special services.

NON-EXPERIMENTAL IMPACT EVALUATIONS

Key Elements. Non-experimental impact evaluations examine changes in levels of risk or outcomes among program participants, or groups including program participants, but do not include comparison groups of other individuals or groups not exposed to the program.

Design Variations. The four primary types of non-experimental designs include: 1) before and after comparisons of program participants; 2) time series designs based on repeated measures of outcomes before and after the program for groups that include program participants; 3) panel studies based on repeated measurement of outcomes on the same group of participants; and 4) post-program comparisons among groups of participants.

The first two designs are based on analysis of aggregate data. In before and after comparisons, outcomes for groups of participants (program groups that enter the program at a specific time and progress through it over the same time frame) are measured before and after an intervention and an assessment of impact inferred from the differences. This simple design is often used to assess whether knowledge, attitudes, or behavior of the group changed after exposure to a classroom curriculum or job training program. Time series designs are an extension of the before and after design that uses multiple measures of the outcome variables before an intervention begins and continues to take multiple measures after intervention is in place. If a change in the trend (direction or level) in the outcome occurs at, or shortly after the time of the intervention, the significance of the observed change is tested statistically. Time series measures may be based on larger groups or units that include but are not restricted to program participants. For example, crime rates for neighborhoods in which most or all youth participate in a delinquency prevention program might be used to assess reductions in illegal activity. Evaluation of a series of dropout prevention activities offered across the school year could examine the percentages of entering classes that graduate over a period of years. Time series designs should be considered when it is difficult to identify who receives program services or when the evaluation budget does not support collection of detailed data from program participants. Although new statistical techniques have strengthened the statistical power of these designs, it is still difficult to rule out the potential impact of non-program events using this approach.

The next two designs examine data at the individual level. Cross-sectional comparisons are based on surveys of groups of participants conducted after program completion. This design can be used to estimate correlations between outcomes and differences in the duration, type, and intensity of services received, yielding conclusions about plausible links between outcomes and services but no definitive conclusions about what caused what. Panel designs use repeated measures of the outcome variables for each individual. In this design, outcomes are measured for the same group of program participants, often
starting at the time they enter the program and continuing at intervals over time. For example, the evaluation of Health Planning and Promotion: Life Planning Education used pre-post data from participants to measure gains in understanding the best combinations of contraceptive methods and the consequences of early childbearing.(15) This design allows the characteristics of individual participants to be used in the analysis to identify different patterns of change associated with individual characteristics of participants and control for other events to which they were exposed.

Considerations/Limitations. Several limitations to non-experimental designs should be noted. First, the cross-sectional and panel designs provide only a segment of "dose-response curve," that is, only estimates of the differences in impact related to differences in the services received. These designs cannot estimate the full impact of the program compared to no service at all, unless estimates can be based on other information on the risks of the target population. Second, the designs that track participants over time (before and after, panel, and time series) cannot control for the effects of developmental changes that would have occurred without services, or for the effects of other events outside the program's influence. Third, the extent to which the results can be assumed to apply to other groups or other settings is limited, because this design provides no information for assessing the extent to which participants were selected into the program on the basis of factors which themselves influence outcomes.

Practical Issues. Non-experimental designs have considerable practical advantages because they are relatively easy and inexpensive to conduct. Individual data for cross-sectional or panel analysis are often collected routinely by the program at the end (and sometimes beginning) of program participation. When relying on program records, the evaluator needs to review the available data against the logic model to be sure that adequate information on key variables is already included, or to begin collecting additional data items if needed.

When individual program records are not available, aggregate statistics may be obtained from the program or from other community agencies with information on the outcomes among groups of participants. For example, crime rates, average promotion rates, and rates of births to teen mothers can be collected from existing records. The primary problem encountered in using such statistics for assessing impacts is that they may not be available for the specific population or geographic area targeted by the program. Often these routinely collected statistics are based on the general population or geographic areas served by the agency (e.g., the police precinct or the clinic catchment area). The rates of negative outcomes for the entire set of cases included may well differ from rates for the targeted group of vulnerable children and youth; this risk is greater for larger rather than smaller statistical areas.

A more expensive form of data collection for non-experimental evaluations is a survey of participants some time after the end of the program. These surveys can provide much needed information on longer-term outcomes such as rates of employment or earnings or high school graduation. As in any survey research, the quality of the results is determined by response rate rather than overall sample size, and by careful attention to the validity and reliability of the questionnaire items.

Example. The Youth Training Scheme (YTS) in Great Britain provides, through local agents, two years of vocational and on-the-job training for out-of-school and unemployed youth ages 16 and 17. The local agents are businesses or community organizations that receive government funds to design a training program, recruit and supervise youth, and provide at least 13 weeks of on-the-job training per year. Non-experimental evaluation of YTS was based on a follow-up survey of 63,000 former participants.(16) In addition to monitoring client satisfaction and job related outcomes, the survey was used in non-experimental comparisons of differences in outcomes related to differences among participants: job market outcomes were compared for graduates versus program dropouts and across youth who entered the program with different levels of motivation and past school achievement. Results indicate that
program graduates had better labor market outcomes than those who did not complete the program. Similarly, earning qualifications in the program (an interim outcome measure) was positively correlated with later labor market success (the longer term outcome). Non-experimental comparisons were also used to identify differences in outcomes related to characteristics of the participants or the training experience. The field of employment and type of local agent providing the training were significant predictors of labor market outcomes. Similarly, labor market outcomes were better for youth who began the program with higher levels of motivation and past school achievement. These findings are suggestive but not definitive. Because of the non-experimental design, participating youth might have been more likely to become employed than other youth even in the absence of the program.

**CHOOSING AMONG THE IMPACT DESIGNS**
Choice of an impact evaluation design begins by identifying the design that both offers the strongest capacity for isolating the independent causal effects of the program and is feasible given the structure of the program. The "decision tree" shown in Exhibit B illustrates a process for identifying which alternatives are feasible.

If the program will be provided to a limited number of youth who can be identified in advance and randomly selected for participation, then an experimental design should be considered. If the program will be provided to a limited number of youth, but the decision about who receives services is determined by organizational or geographic considerations (or other nonrandom selection rules), then quasi-experimental design variations should be considered.

The most difficult design challenges occur when the program is intended to serve all members of the target population. If the new program is implemented fully and rapidly, no youth will be available for a comparison group. Often, however, new full-coverage programs-for example, new health services—are intended for an entire population but not implemented in every community in the country, and certainly not at the same time. If some communities or groups are not included in the initial implementation, it may be possible to select as comparison sites communities that have not implemented the program and use a quasi-experimental design. This may not solve the problem of comparability sufficiently to allow such a design, however, if the communities where it was implemented have characteristics that are systematically different from those where it was not.

When non-experimental designs are necessary, the following can help guide the choice of design. If a program is implemented at different levels across sites but uniformly within sites, a cross-sectional design is suitable. If a target population is exposed to different levels of the program within a community, a panel study design is better to follow a sample of individuals, and record both outcomes and the amount of the program or intervention each individual received and when it occurred. If defining who is served by the program is difficult or the program is uniformly applied in all communities, then a time-series design is appropriate. Before-and-after designs without control groups are often used, but are subject to a number of threats to validity, including maturation and secular changes (discussed above).

**Performance Monitoring**

*Key Elements.* Performance monitoring is used to provide information on: 1) key aspects of how a system or program is operating; 2) whether, and to what extent, pre-specified program objectives are being attained (e.g., numbers of youth served compared to target goals, reductions in school dropouts compared to target goals); and 3) identification of failures to produce program outputs, for use in managing or redesigning program operations. Performance indicators can also be developed to 4) monitor service quality by collecting data on the satisfaction of those served, and 5) report on program efficiency, effectiveness, and productivity by assessing the relationship between the resources used
(program inputs) and the output and outcome indicators.

If conducted frequently enough and in a timely way, performance monitoring can provide managers with regular feedback that will allow them to identify problems, take timely action, and subsequently assess whether their actions have led to the improvements sought. Performance measures can also stimulate communication about program goals, progress, obstacles, and results among program staff and managers, the public, and other stakeholders. They focus attention on the specific outcomes desired and better ways to achieve them, and can promote credibility by highlighting the accomplishments and value of the program.

Performance monitoring involves identification and collection of specific data on program outputs, outcomes, and accomplishments. Although they


may measure subjective factors such as client satisfaction, the data are numeric, consisting of frequency counts, statistical averages, ratios, or percentages. Output measures reflect internal activities: the amount of work done within the program or organization. Outcome measures (immediate and longer term) reflect progress towards program goals. Often the same measurements (e.g., number/percent of youth who stopped or reduced substance abuse) may be used for performance monitoring and impact evaluation. However, unlike impact evaluation, performance monitoring does not make any rigorous effort to determine whether these were caused by program efforts or by other external events.
Design Variations. When programs are operating in a number of communities, the sites are likely to vary in mission, structure, the nature and extent of project implementation, primary clients/targets, and timeliness. They may offer somewhat different sets of services, or have identified somewhat different goals. In such situations, it is advisable to construct a "core" set of performance measures to be used by all, and to supplement these with "local" performance indicators that reflect differences. For example, some youth programs will collect detailed data on youth school performance, including grades, attendance, and disciplinary actions, while others will simply have data on promotion to the next grade or whether the youth is still enrolled or has dropped out. A multi-school performance monitoring system might require data on promotion and enrollment for all schools, and specify more detailed or specialized indicators on attendance or disciplinary actions for one or a subset of schools to use in their own performance monitoring.

Considerations/Limitations. In selecting performance indicators, evaluators and service providers need to consider:

*The relevance of potential measures to the mission/objective of the local program or national initiative.* Do process indicators reflect program strategies/activities identified in mission statements? Do outcome indicators cover objectives identified in mission statements? Do indicators capture the priorities at the community level?

*The comprehensiveness of the set of measures.* Does the set of performance measures cover inputs, outputs, and service quality as well as outcomes and include relevant items of customer feedback?

*The program's control over the factor being measured.* Does the program have influence/control over the outputs or outcomes measured by the indicator? If the program has only limited influence over the outputs or outcomes being measured, the indicator may not fairly reflect program performance.

*The validity of the measure.* Do the proposed indicators reflect the range of outcomes the program hopes to affect? Are the data free from obvious reporting bias?

*The reliability and accuracy of the measure.* Can indicators be operationally defined in a straightforward manner so that supporting data can be collected consistently over time, across data gatherers, and across communities? Do existing data sources meet these criteria?

*The feasibility of collecting the data.* How much effort and money is required to generate each measure? Should a particularly costly measure be retained because it is perceived as critically important?

Practical Issues. The set of performance indicators should be simple, limited to a few key indicators of priority outcomes. Too many indicators burden the data collection and analysis and make it less likely that managers will understand and use reported information. At the same time, the set of indicators should be constructed to reflect the informational needs of stakeholders at all levels -community members, agency directors, and national funders.

Regular measurement, ideally quarterly, is important so that the system provides the information in time to make shifts in program operations and to capture changes over time. However, pressures for timely reporting should not be allowed to sacrifice data quality. For the performance monitoring to take place
in a reliable and timely way, the evaluation should include adequate support and plans for training and technical assistance for data collection. Routine quality control procedures should be established to check on data entry accuracy and missing information. At the point of analysis, procedures for verifying trends should be in place, particularly if the results are unexpected.

The costs of performance monitoring are modest relative to impact evaluations, but still vary widely depending on the data used. Most performance indicator data come from records maintained by service providers. The added expense involves regularly collecting and analyzing these records, as well as preparing and disseminating reports to those concerned. This is typically a part-time work assignment for a supervisor within the agency. The expense will be greater if client satisfaction surveys are used to measure outcomes. An outside survey organization may be required for a large-scale survey of past clients; alternatively, a self-administered exit questionnaire can be given to clients at the end of services. In either case, the assistance of professional researchers is needed in preparing data sets, analyses, and reports.

Example. The Asociacion Salud con Prevencion (ASCP) in Colombia, South America, a non-governmental organization which provides primary prevention services which promote adolescent reproductive health, monitors outputs with data on the number of professionals trained, the number of youth given educational services, the number of workshops held, the number of condoms distributed, and the number of medical and counseling sessions provided. The results demonstrate that the program is providing promised services, but does not give an indication of the impact in terms of either immediate outcomes such as use of birth control or longer-term outcomes (which include reduced risk of out-of-wedlock births or early childbearing).

Process Analysis

Key Element. The key element in process analysis is a systematic, focused plan for collecting data to: (1) determine whatever the program model is being implemented as specified and, if not, how operations differ from those initially planned; (2) identify unintended consequences and unanticipated outcomes; and (3) understand the program from the perspectives of staff, participants, and the community.

Design Variations. The systematic procedures used to collect data for process evaluation often include case studies, focus groups, and ethnography.

Case studies involve the detailed analysis of selected program sites or clients to determine how the program is operating, what barriers to program implementation have been encountered, what strategies are the most successful, and what resources and skills are necessary. The answers to these questions are useful in providing guidance to policymakers and program planners interested in identifying key program elements and in generating hypotheses about program impact that can be tested in impact analyses. Case studies are sometimes used to test competing hypotheses about differences in the impact of services. This strategy is used to assess which approach is most successful in attaining goals shared by all when competing models have emerged in different locations. This requires purposely selecting sites to represent variations in elements or types of programs, careful analysis of potential causal models, and the collection of qualitative data to elaborate the causal links at each site.

Clients or sites chosen for case studies should represent wide variation in settings, program models, and clients. Identification of sample members within sites, interview topics, and key data elements begins with the logic

model as a guide. In a case study, qualitative data, collected using semi-structured interviews and observations of program operations, are often supplemented and verified by quantitative data on program operations and performance collected from records and reports.

Case studies may use several different approaches for collecting qualitative data for program evaluation. The most frequently used are semi-structured interviews, focus groups, and researcher observations while on-site. Semi-structured interviews allow for the discovery of unanticipated factors associated with program interpretation and outcomes. Protocols for semi-structured interviews contain specific questions about particular issues or program practices. The "semi" aspect of these discussion guides refers to the fact that a respondent may give as long, detailed, and complex a response as he or she desires to the question—whatever conveys the full reality of the program's experience with the issue at hand. If some issues have typical categories associated with them, the protocols will usually contain probes to make sure the researcher learns about each category of interest.

In case studies, observations at program sites provide an important method of validating information from interviews. In this case, the observations will often be guided by structured or semi-structured protocols designed to ensure that key items reported in interviews are verified and that consistent procedures for rating program performance are used across time and across sites.

Focus groups seek to understand attitudes through a series of group discussions guided by one researcher acting as a facilitator, with another researcher present to take detailed notes. Five or six general questions are selected to guide open-ended discussions lasting about an hour and a half. The goals of the discussions may vary from achieving group consensus to emphasizing points of divergence among participants. Discussions are tape-recorded, but the primary record is the detailed notes taken by the researcher who acts as recorder. Less detailed notes may also be taken publicly, on a flip-chart for all to see, to try to achieve consensus or give group members the chance to add anything they think is important. Soon after a particular focus group, the recorder and facilitator summarize in writing the main points that emerged in response to each of the general questions. When all focus groups are completed, the researchers develop a combined summary, noting group differences and suggesting hypotheses about those differences.

Ethnography relies almost exclusively on observation and unstructured interviews to study:

- Organizational and programmatic processes occurring at a program site;
- The community context in which the program is taking place;
- The relationship between program activities and other activities in the community;
- Causal processes as the participants view them; and
- Modes of decision-making.

Ethnography does not begin with the logic model. Its intent is to understand the program from the perspective of staff, participants, and others in the community. Ethnographers observe program operations as unobtrusively as possible, sometimes in the role of participant observer, and keep detailed field notes that are transcribed and coded to identify emerging themes and trends. The critical research goal is to provide data on the subjective experience of those in the program situation and to use this information to understand if the program goals are being achieved and, if so, how.

Ethnography uses procedures that are deliberately flexible. As a result, ethnography is helpful in
gathering information on unintended consequences and unanticipated outcomes. These unexpected observations may lead to an entirely new concept of program delivery. In a recent project examining service integration programs for at-risk youth, observations helped clarify that service integration needed to go beyond formal links and on-paper agreements, and provided insights into how informal processes bonded services together in their efforts to make a difference for high-risk youth in the community.(17) Observations from ethnographic studies are perhaps the hardest type of qualitative information to analyze, since they generate volumes of information, much of which may not be directly related to evaluation goals and may not be comparable across sites.

Practical Issues. Collecting qualitative data requires skilled researchers who are experienced with the techniques being used. To analyze these data, careful notes must be taken to ensure that responses are correctly recorded and to aid in interpreting them. In methods based on interviews, interviewers must be trained to understand the intent of each question, the possible variety of answers that respondents might give, and ways to probe to ensure that full information about the issues under investigation is obtained.

Analysis of qualitative data requires an in-depth understanding of programs, respondents and responses, and especially the context in which they are evaluated. Ultimately, the analyst makes judgments regarding the relative importance or significance of various responses. This requires an unbiased assessment of whether responses support or refute hypotheses about the way the program works and the effects it has.

One way to handle qualitative data is to treat one's interview and observational notes as text, and to conduct a textual analysis using specialized computer software that can search for the presence of specific themes or content. Qualitative software is available to facilitate the location and retrieval of information from massive textual files. This kind of software is expensive to use because huge amounts of text must be entered into a computer. Further, either the exact words one wants to search for must appear in the text, or the text marked for the presence of any theme or topic that the researcher wants to retrieve. Often researchers can achieve equal or better results with carefully constructed interview or data collection guides or structured focus groups, and systematically recording of responses or coding of data encountered in the field.

Example. Case studies of two pilot projects were used for the evaluation of mentoring in the juvenile justice system conducted by Public/Private Ventures. The program was designed to match 100 mentors to at-risk youth. Mentors were trained to meet with youth one-on-one before and after the youth's release from juvenile detention facilities, with the goal of establishing an attachment to an adult role model. Data were collected from mentor logs, program records, court records, structured interviews with mentors and youth before and after program participation, staff interviews, focus groups with mentors, youth and service agency staff, and in-depth interviews with mentor-youth pairs. The qualitative analysis examined the characteristics of successful matches, issues in program implementation, the style and content of mentoring interactions, and program staffing. Although it does not offer evidence on outcomes, the evaluation provides extremely useful information on the process of implementing a mentoring program and guidance for program development and replication.

Cost Studies

Key Elements. Cost studies are used to assess investments in programs by collecting information on: 1) direct program expenditures; 2) the costs of staff and resources provided by other agencies or diverted from other uses; 3) costs for purchased services; and 4) the value of donated time and materials. Costs for the first two items usually include expenditures for staff salaries; fringe benefits; special training costs (if any); travel; facilities; and supplies and equipment that have to be purchased. The value of
donated resources, which can be substantial, generally has to be estimated and requires careful
documentation of the donation. Cost analyses indicate that donations are a major cost item in many
youth programs. For example, the Cities in Schools (18) evaluation indicated that donations are between
74 percent and 90 percent of the total direct program costs, and that the wide variation among cities in
the types of donations received made the inclusion of these costs essential to an understanding of the
resources required to sustain. program operation.

The typical approach to cost studies is to calculate total program costs and then an average cost per
client, calculated by dividing the total by either the total number of clients served, or the total number of
clients who meet some standardized definition of success. This type of cost calculation can be linked to
results of an experimental or quasi-experimental impact evaluation to estimate costs per successful
client. It can also be used with performance indicators to assess the cost or cost-efficiency of achieving
program goals.

A second approach to cost estimation calculates the cost per unit of service. For example, the cost per
hour of classroom instruction or the cost per hour of counseling. This type of cost calculation is then
used in impact evaluations (including non-experimental evaluations) to look at the costs of different
outcomes. This type of cost analysis is difficult in multi-faceted, comprehensive programs in which the
level and type of service are highly variable and may involve a number of service providers. It is also
difficult in programs in which defining exposure to services is difficult. Where possible, it is preferable
to distinguish between fixed costs (e.g., rent or the director's salary) and variable costs (e.g., the costs of
special events or the hourly costs of the recreation director). The variable costs can then be used to
estimate the marginal cost of adding additional clients to the number receiving a specific unit of service.

Design Variations. Cost studies can be undertaken to describe the program costs and link these to the
level of outcomes achieved. In this application, the costs are compared to the level and type of outcomes
documented in performance monitoring outcomes. Decisions on whether the outcomes justify the costs
are based on opinions about the value of the outcomes (not monetized) and the likelihood that the
outcomes are attributable to the program.

Cost-effectiveness analysis is used to compare the costs of different approaches to providing some
standard level of service or desired level of outcome. This approach is most useful when multiple
programs are using different models to provide a service. The requirements are that the characteristics of
target populations served, the program goals, and the output or outcome measures be identical. For
example, cost-effectiveness studies could compare the relative effectiveness of residential and
nonresidential treatment for drug-abusing youth, provided that the youth served were similar in age and
drug use problems, and that the same measures of treatment success were used.

Cost-benefit studies provide estimates of the dollar benefits returned for each dollar spent on the
program—the key question from a policy perspective, but one that is not easily answered. This type of
evaluation has rigorous requirements for: 1) an estimate of program costs, either per client or per unit of
service; 2) estimates of the value of the benefits; and 3) comparative data on program impact—an estimate
of outcomes with and without the program. The first item should be obtainable from program financial
records, supplemented as needed by estimates of the cost of donated or reallocated resources. The
second can be obtained from an experimental or quasi-experimental evaluation of program impact or
another strategy for estimating the difference between what happened and what would have happened
without the program.

The primary barrier to conducting cost-benefit analysis of service programs designed to change behavior
stems from the third item: placing dollar values on benefits. Many benefits are of intrinsic value (e.g.,
reductions in family dysfunction and conflict) but quantifying that value is difficult.
Monetization of benefits to individuals requires assumptions about three matters, all of which are frequently controversial. First, the dollar value of the benefit may depend on personal values, for example, what residents are willing to pay for a crime-free neighborhood. Second, a dollar of benefit today is worth more than a dollar benefit realized next year. Thus, the benefits need to be time discounted, but by how much is a difficult question. Third, the beneficiaries need to be identified. Societal values become important when the beneficiaries differ in standing and perceived merit. For example, a high school equivalency degree for a violent youthful offender may result in the same gains in lifetime earnings for the offender as a violence victim would realize from physical therapy for the injury. Are they to be treated the same? To circumvent such difficult questions, the analyst may conduct a sensitivity analysis to reach conclusions based on explicit assumptions of value. For example, the neighborhood crime prevention program may be deemed cost-effective if "residents are willing to pay at least $100 per month for 10 percent lower rates of burglary" or "if the discount rate is less than 6 percent" or "if the offender's earnings are worth 50 percent of the victim's earnings."

Beyond benefits to individuals, the total value of benefits includes the social costs averted. These are the savings to the public that result from avoiding negative outcomes. These values must be based on studies that estimate the social costs of negative outcomes such as the costs of crime or drug abuse.(19)

These estimates are difficult to derive and are often based on tenuous assumptions. To compensate for problems in the reliability of estimates, cost-benefit calculations normally use a range of benefits to place an upper and a lower bound on the probable returns to investments in the program. A more significant problem is that monetary values based on public costs for the negative outcomes among the general population may be poor estimates of the value of benefits among the program's target population. For example, national estimates of the costs of drug abuse may not apply to reductions in amphetamine abuse among low-income adolescents in a single city. This problem needs to be acknowledged and value estimates revised to the extent possible to reflect savings for the program's participants. Other public benefits reflecting gains, not costs averted, are widely acknowledged, but rarely find their way into cost-benefit studies because there is no public consensus on their importance. Examples include improvements in the quality of life or the environment.

Considerations and Limitations. Documentation of gains to prevention programs is exceptionally difficult and requires estimating negative outcomes that did not occur. As described above, the most robust estimates of program impacts of this kind are based on experimental evaluations or quasi-experimental evaluations, which are difficult and expensive to conduct. When the program has total population coverage, it is possible to interpret differences between the observed trend and predicted trend in an outcome indicator over time to program impacts and estimate the monetary value of the benefits. This strategy was used to estimate the value of drug prevention efforts in the United States. National survey estimates of drug use in 1979 were used to estimate expected drug prevalence during the 1980s and early 1990s; the differences between these estimates and drug use prevalence rates based on national surveys during these years were attributed to federal investments in drug prevention programs.(20)

Practical Issues. Developing a conceptual framework that reflects all the issues in cost-benefit valuation, and then devoting the resources necessary for estimating the range of benefits, can require as much research time and expertise as determining whether the program had any impact. However, research dollars are always limited and evaluating program impact is usually the top priority, since valuing benefits is irrelevant if there is no program impact. A number of studies of the value of preventing negative outcomes among children and youth have been initiated recently. These can be expected to give program evaluators substantial help in estimating the value of reductions in youth problems for use in cost-benefit studies in the future.
Example. An evaluation of 13 delinquency prevention programs in Los Angeles County estimated cost effectiveness as a function of the delinquency risk of the population of youth served, costs, and success rate. This study compared cost to benefit ratios of alternative programs designed with a common goal and outcome measure—preventing subsequent arrest. Because the risks of delinquency varied among the youth served by different programs, estimates of the risk of delinquency was derived from existing research and used to classify the youth served by the program into four risk categories. Program costs were estimated by taking the total budgets from all sources divided by the number of clients. Costs of public expenditures for delinquency (costs to the community and justice system) were estimated from the proportion of the justice system budget (from the County budget) devoted to juvenile cases, divided by the number of juvenile cases at various stages of processing (from annual reports of the Los Angeles Probation Department, the California Youth Authority, and the U.S. Department of Justice).

The public costs averted were calculated by dividing the budget by the number of arrests of youth following program participation and calculating the savings as the difference between the two. The benefits of reductions in expected future arrests were estimated based on the probability of subsequent arrests reported in studies of criminal careers times the estimated public savings per arrest averted. Savings to victims were based on estimates of the costs of damage and loss for each type of juvenile offense from earlier research, adjusted for inflation. These costs per offense were applied to the expected lifetime arrests in the absence of the program and benefits were estimated as the difference between these costs and the absence of costs associated with no further arrests or victimization (estimating that for each arrest, there are four to five offenses that do not result in arrest). Thus, estimated program benefits were the sum of the public costs averted and the savings to victims.

The results were used to estimate the cost differential (costs divided by the value of benefits) to programs with different rates of success (measures as arrests prevented), controlling for the risk of offending of the juvenile population served. The findings were used to estimate the success rate required to show a positive rate of return given the delinquency risk of the population served for programs with different cost differentials. This estimate can be used in monitoring the performance of a wide variety of delinquency prevention programs.

Identifying Potential Evaluation Problems

A number of challenging problems face those who would apply research methods to the evaluation of human services programs. We summarize these, based on experience in reviewing and evaluating programs for vulnerable children and youth, to guide development of realistic evaluation plans. (21)

Defining Program Participation. Programs may be open-ended, lacking both formal intake procedures and policies for determining when the program is "completed." An evaluation can only yield interpretable results if participation is explicitly defined and uniformly measured. In the case of programs for vulnerable youth, for example, counselors may be contacted for several chats, followed some weeks later by an appointment, followed by intermittent participation in some, possibly not all, services offered. Youth may stop attending and then resume. Limiting participation in the evaluation to those who attend regularly is not an appropriate solution because dropping from consideration the youth who are most difficult to engage produces biased results. Often identifying who "participated" and for how long requires multiple categories to adequately reflect the variations in type, duration, and intensity of participation among the youth served. In addition, participants should be followed from the point of first contact and all major program activity documented. Evaluators also need to decide whether others who potentially benefit from the program--such as parents, boyfriends/girlfriends, or siblings--are defined as program participants. If so, their participation in program activities should also be tracked. If not, plans need to be made on how to count the gains made by these indirect program beneficiaries in evaluating program impact.
Evaluating the Relationship between Participation and Outcomes. Many programs emphasize individualized services tailored to need. In the youth services area, youth with the highest levels of risk are offered the greatest number or most intensive level of services. Obviously, assignment to treatment in this case is not random, and the multi-problem youth may never achieve the same level of positive outcomes as youth who began with fewer problems. For example, studies of the School-Based Health Centers in the U.S. show that frequent clinic users were at greater risk for alcohol and substance use, sexual activity, and poor family and peer relationships (22). Thus, comparing their outcomes to those for nonusers or those who used the clinic less frequently would be inappropriate. Similarly, comparisons between different programs must consider any differences in type and level of risk exhibited by participants. For this reason, data on the risks and needs of participants should be collected at intake for use in analysis and a pre-post design used when possible.

Defining the Unit of Analysis. Deciding on the appropriate unit of analysis can be difficult, particularly in evaluating comprehensive programs. Programs may target entire neighborhoods, classrooms, or families for change - sometimes planning activities directly for different groups, and sometimes planning carryover effects. Measurement at multiple levels is appropriate as long as each level is clearly defined. For example, crime reduction can be assessed by comparing neighborhood rates of calls for police services, household victimization rates, or youth delinquency surveys. Economic gains can be measured by changes in the area unemployment rate, average household or family income, or individual earnings. The selection should be closely linked to program goals and activities.

Evaluations of services integration programs, including most that use a case management approach, will face additional challenges in: 1) tracking the services received by participants; 2) developing common agreements among agencies on program goals and required components; 3) documenting service delivery by multiple agencies; 4) measuring effects of the service delivery system; and 5) differentiating services integration from service comprehensiveness. Each is discussed briefly below.

Tracking the services received by participants. Services integration usually involves referring participants to other agencies for needed assistance. A critical, and often difficult, problem is determining which services were actually received. Clients may or may not contact agencies to which they are referred, may or may not be accepted for services, and may or may not participate in services, if accepted. Documenting the chain of participation is essential to determine the extent to which services integration is being achieved, but is time consuming and often resisted by programs who see making the referral as the extent of their responsibility. Because staff turnover in service agencies is frequently high, preparing written agreements on data access and sharing is strongly recommended. In the absence of adequate agency documentation, information on service utilization can be collected in follow-up interviews with clients.

Developing common agreements among agencies on program goals and required components. The agencies collaborating in a services integration effort may differ in their vision of the program's goals, key strategies, and how youth needs will be evaluated and problems addressed. Evaluations tend to highlight these differences, which can constitute a barrier in gaining consensus on what is being evaluated. This is particularly true when multiple agencies recruit clients and/or case management services are not centralized. Time should be allocated for face-to-face meetings to get agreement on whom evaluators will count in selecting measures of program outcomes, and how service provision is expected to achieve program goals.

Documenting service delivery by multiple agencies. When many agencies coordinate and combine their resources to meet the needs of clients, one of the most difficult problems is assembling information on who received what types and amounts of service. Agencies have different methods of identifying clients. In the area of vulnerable children and youth, some use family identification numbers, others identify...
individual children served. Some group service records by family or child; others maintain records by contact, which introduces multiple records for single clients which then have to be checked to remove duplication. Agencies such as schools or juvenile courts can face legal or professional barriers to sharing client-based information with other agencies or evaluators. A systematic system for collecting the data needed to compile a complete picture of program participation must be developed early in the planning process and, as noted above, supported by written agreements and ongoing technical assistance and staff training in record-keeping procedures.

Measuring effects of the service delivery system. A primary goal of services integration is to change agency operations and increase effectiveness. These outcomes need to be measured at the agency, not individual, level. Evaluations of services integration need to document changes in agency procedures, increased participation in collaborative planning and service delivery, and decreases in barriers to interagency cooperation and client service associated with policies, and procedures. Referral patterns should show more diversity in planning. At the individual level, clients should report fewer unmet service needs, shorter waiting periods for service, and increased satisfaction with the response to their needs. Other evidence of integration includes increased staff knowledge and familiarity with the resources of other agencies and community groups.

Differentiating services integration from service comprehensiveness. Services integration is intended to provide not only faster, more appropriate services, but also services that would not otherwise be available to certain clients. The referral process educates clients on the options and assistance potentially available. Improved interagency planning and coordination reduces the barriers to obtaining additional services. All this makes the task of differentiating services integration from service comprehensiveness very difficult. Evaluation and program staff need to develop clear expectations on the extent to which the ease of obtaining services and the appropriateness of the service package can be distinguished from the extent to which the program is providing comprehensive services to meet the full range of client needs.

Conclusions
Strong pressure to demonstrate program impacts dictates making evaluation activities a required and intrinsic part of program activities from the start. At the very least, evaluation activities should include performance monitoring. The collection and analysis of data on program progress and process builds the capacity for self-evaluation and contributes to good program management and efforts to obtain support for program continuation—for example, when the funding is serving as "seed" money for a program that is intended, if successful, to continue under local sponsorship. Performance monitoring can be extended to non-experimental evaluation with additional analysis of program records and/or client surveys. These evaluation activities may be conducted either by program staff with research training or by an independent evaluator. In either case, training and technical assistance to support program evaluation efforts will be needed to maintain data quality and assist in appropriate analysis and use of the findings.

There are several strong arguments for evaluation designs that go further in documenting program impact. Only experimental or quasi-experimental designs provide convincing evidence that program funds are well invested, and that the program is making a real difference to the well-being of the population served. These evaluations need to be conducted by experienced researchers and supported by adequate budgets. A good strategy may be implementing small-scale programs to test alternative models of service delivery in settings that will allow a stronger impact evaluation design than is possible in a large scale, national program. Often program evaluation should proceed in stages. The first year of program operations can be devoted to process studies and performance monitoring, the information from which can serve as a basis for more extensive evaluation efforts once operations are running smoothly.
Finally, planning to obtain support for the evaluation at every level—community, program staff, agency leadership and funder—should be extensive. Each of these has a stake in the results. Each should have a voice in planning. And each should perceive clear benefits from the results. Only in this way will the results be acknowledged as valid and actually used for program improvement.

Notes


Appendix C

TEEN COURTS: A FOCUS ON RESEARCH
Teen Courts: A Focus on Research

Jeffrey A. Butts and Janeen Buck

Growing from a handful of programs in the 1960’s, the number of teen courts (or youth courts) now operating in the United States has been estimated to be as high as 675. Communities across the Nation continue to demand better information and assistance with which to start or enhance their own teen courts. This Bulletin helps to address that demand by providing information about the characteristics of established teen courts and the operational and managerial challenges they face. It also summarizes the evaluation literature on teen courts.

Background

Teen courts are spreading rapidly across the United States. Many people view them as a cost-effective alternative to traditional juvenile court for some young offenders. Until recently, relatively little information has been available about how teen courts operate or how they affect youthful offenders. This Bulletin presents the results of a national survey of teen courts. The findings suggest that most teen courts are relatively small and were established very recently. The findings also suggest that the most established teen court programs (i.e., programs reporting longevity in operations and/or little financial uncertainty) may be those that are housed within or closely affiliated with the traditional juvenile justice system.

The survey indicates that teen courts enjoy broad community support. Their popularity appears to stem from favorable media coverage and the high levels of satisfaction reported by parents, teachers, and youth involved in teen court programs, rather than from evaluation research showing that teen courts have beneficial effects on offenders. Little research has been conducted on outcomes for teen court defendants, although some studies offer encouraging results. Recent studies have found that teen court participation may be associated with low recidivism rates, improved youth attitudes toward authority, and increased knowledge of the justice system among youth. More research is required before claims about teen court effectiveness can be substantiated.

The Teen Court Concept

Teen courts are generally used for younger juveniles (ages 10 to 15), those with no prior arrest records, and those charged with less serious law violations (e.g., shoplifting, vandalism, and disorderly conduct). Typically, young offenders are offered teen court as a voluntary alternative in lieu of more formal handling by the traditional juvenile justice system (see figure 1). Teen courts differ from
other juvenile justice programs because young people rather than adults determine the disposition, given a broad array of sentencing options made available by adults overseeing the program. Teen court defendants may go through an intake process, a preliminary review of charges, a court hearing, and sentencing, as in a regular juvenile court. In a teen court, however, other young people are responsible for much of the process. Charges may be presented to the court by a 15-year-old “prosecutor.” Defendants may be represented by a 16-year-old “defense attorney.” Other youth may serve as jurors, court clerks, and bailiffs. In some teen courts, a youth “judge” (or panel of youth judges) may choose the best disposition or sanction for each case. In a few teen courts, youth even determine whether the facts in a case have been proven by the prosecutor (similar to a finding of guilt).

Adults are also involved in teen courts. They often administer the programs, and they are usually responsible for essential functions such as budgeting, planning, and personnel. In many programs, adults supervise the courtroom activities, and they often coordinate the community service placements where youth work to fulfill the terms of their dispositions. In some programs, adults act as the judges while teens serve as attorneys and jurors. The key to all teen court programs, however, is the significant role youth play in the deliberation of charges and the imposition of sanctions on young offenders.

Proponents of teen court argue that the process takes advantage of one of the most powerful forces in the life of an adolescent— the desire for peer approval and the reaction to peer pressure. According to this argument, youth respond better to prosocial peers than to adult authority figures. Thus, teen courts are seen as a potentially effective alternative to traditional juvenile courts staffed with paid professionals such as lawyers, judges, and probation officers. Teen court advocates also point out that the benefits extend beyond defendants. Teen courts may benefit the volunteer youth attorneys and judges, who probably learn more about the legal system than they ever could in a classroom. The presence of a teen court may also encourage the entire community to take a more active role in responding to juvenile crime. Teen courts offer at least four potential benefits:

- **Accountability.** Teen courts may help to ensure that young offenders are held accountable for their illegal behavior, even when their offenses are relatively minor and would not likely result in sanctions from the traditional juvenile justice system.

- **Timeliness.** An effective teen court can move young offenders from arrest to sanctions within a matter of days.
rather than the months that may pass with traditional juvenile courts. This rapid response may increase the positive impact of court sanctions, regardless of their severity.

- **Cost savings.** Teen courts usually depend heavily on youth and adult volunteers. If managed properly, they may handle a substantial number of offenders at relatively little cost to the community. The average annual cost for operating a teen court is $32,822 (National Youth Court Center, unpublished data).

- **Community cohesion.** A well-structured and expansive teen court program may affect the entire community by increasing public appreciation of the legal system, enhancing community-court relationships, encouraging greater respect for the law among youth, and promoting volunteerism among both adults and youth.

Researchers are beginning to report instances in which these potential benefits have been realized in some communities, but evaluation research on teen courts is still in the early stages. It is too soon to tell whether the positive results reported by some communities can be replicated reliably in other communities. Regardless of the limited evidence, however, teen courts are increasingly in use across the United States. This Bulletin describes the variety of teen courts and summarizes what researchers know about the effects of teen court programs.

### National Survey

As part of the Office of Juvenile Justice and Delinquency Prevention’s (OJJDP’s) Evaluation of Teen Courts Project, The Urban Institute recently conducted a national survey of teen courts and youth courts. With assistance from the National Youth Court Center (NYCC), which is housed at the American Probation and Parole Association and supported by funds from OJJDP, project researchers obtained addresses, telephone numbers, and personal contacts for all U.S. teen courts believed to exist as of the end of 1998, and they mailed questionnaires to nearly 500 programs. A handful of these programs had gone out of business by the time researchers tried to contact them. Of the remaining programs, 335 (more than 70 percent) completed and returned the survey. The responses documented the range of teen court programs used by jurisdictions across the country, the characteristics of their clients, the sanctions they imposed, the courtroom models they used, the extent of community support they received, and the challenges they faced.

### Program Characteristics

Recent growth in the number of teen court programs nationwide was reflected in the brief tenure of the programs responding to the national survey. Of all the programs that responded, 13 percent had been in operation less than 1 year and 42 percent had been in operation for only 1 to 3 years. More than two-thirds (67 percent) of all teen courts had been in existence for less than 5 years (see figure 2).

Many teen courts that responded to the survey were closely affiliated with the traditional justice system (see figure 3). Courts, law enforcement agencies, juvenile probation offices, or prosecutors’ offices operated slightly more than half (52 percent) of the programs responding to the survey. More than one-third (37 percent) of the programs were affiliated with the courts and 12 percent with law enforcement. Private agencies operated one-quarter (25 percent) of the teen court programs.

Most teen court and youth court programs were relatively small (see figure 4). More than half (59 percent) of the programs responding to the survey handled 100 or fewer cases annually. Just 13 percent of the programs handled more than 300 cases per year.

Very few programs relied on private funding to meet their operational costs (see figure 5). More than half (59 percent) of the teen courts received no private funding; 16 percent of the programs received up to one-fifth of their funding from private sources, and 11 percent received between one-fifth and one-half from private sources.

### Client Characteristics

Teen courts usually handle relatively young offenders with no prior arrests. Survey respondents reported that, on average, 24 percent of their cases involved youth under age 14 and 66 percent involved youth under age 16. More than one-third (39 percent) of the teen courts accepted only first-time offenders and another 48 percent reported that they “rarely” accepted youth with prior arrest records. Nearly all programs (98 percent) reported that they “never” or “rarely” accepted youth with prior felony arrests. Most programs (91 percent) also indicated that they “never” or “rarely”...
accepted youth who previously had been referred to a juvenile court.

To assess the nature of those cases typically handled in teen court, the survey asked each program to review a list of offenses and to indicate whether the program received such cases “very often,” “often,” “rarely,” or “never.” The offenses most likely to be received “often” or “very often” were theft (93 percent), minor assault (66 percent), disorderly conduct (62 percent), possession or use of alcohol (60 percent), and vandalism (59 percent) (see figure 6).

Sanctions
The principal goal of teen court is to hold young offenders accountable for their behavior. In a system of graduated sanctions, there is a consequence for every offense. Every youth who has admitted guilt or who is found guilty in teen court receives some form of sanction. In many communities, teen court sanctions do more than punish the offender. Sanctions encourage young offenders to repair at least part of the damage they have caused to the community or to specific victims. Offenders are often ordered to pay restitution or perform community service. Some teen courts require offenders to write formal apologies to their victims; others require offenders to serve on a subsequent teen court jury. Many courts use other innovative dispositions, such as requiring offenders to attend classes designed to improve their decisionmaking skills, enhance their awareness of victims, and deter them from future theft.

Survey respondents were asked to assess a list of typical sanctions and indicate how frequently the program used each one (i.e., “very often,” “often,” “rarely,” or “never”). Community service was the most commonly used sanction (see figure 7). Nearly all (99 percent) of responding teen courts reported using community service “often” or “very often.” Other frequently used sanctions included victim apology letters (86 percent), written essays (79 percent), teen court jury duty (74 percent), drug/alcohol classes (60 percent), and restitution (34 percent).

Courtroom Models
NYCC divides the courtroom approaches used by teen courts into four types (described in table 1): adult judge, youth judge, peer jury, and youth tribunal (National Youth Court Center, 2000). Findings
Figure 6: Offenses Handled in Teen Court

<table>
<thead>
<tr>
<th>Offense</th>
<th>Percentage of Teen Courts Reporting They Handle Selected Offenses “Often” or “Very Often”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theft (including shoplifting)</td>
<td>53%</td>
</tr>
<tr>
<td>Minor assault</td>
<td>66%</td>
</tr>
<tr>
<td>Disorderly conduct</td>
<td>62%</td>
</tr>
<tr>
<td>Alcohol possession or use</td>
<td>60%</td>
</tr>
<tr>
<td>Vandalism</td>
<td>59%</td>
</tr>
<tr>
<td>Marijuana possession or use</td>
<td>52%</td>
</tr>
<tr>
<td>School disciplinary problems</td>
<td>33%</td>
</tr>
<tr>
<td>Traffic violation</td>
<td>29%</td>
</tr>
<tr>
<td>Truancy</td>
<td>22%</td>
</tr>
<tr>
<td>Weapon possession or use</td>
<td>11%</td>
</tr>
</tbody>
</table>


from the national survey suggested that the adult judge model was the most popular. Nearly half (47 percent) of the responding courts used only the adult judge model. When the number of cases handled by adult judges in programs using a mix of courtroom models was added, the adult judge model accounted for more than half (60 percent) of all teen court cases.

The next most prevalent courtroom model was the peer jury, which accounted for 22 percent of all teen court cases. More than one in four (26 percent) teen court programs used this model for at least part of their caseloads. The youth judge and tribunal models were the least used, with each accounting for just 7 percent of all cases.

The use of courtroom models varied somewhat according to the agency sponsoring the program (see table 2). The adult judge model was the most popular among teen courts operated by local courts and probation agencies (58 percent) and those hosted by schools, private agencies, and other not-for-profit organizations (48 percent). There was no dominant model, however, among programs operated by law enforcement agencies or prosecutors. In fact, more than one-third (34 percent) of those programs used mixed models (i.e., a combination of two or more courtroom models).

Differences by courtroom model. The characteristics of teen courts were notably different when the analysis controlled for courtroom model (see table 3). For example, programs using the youth judge model were among the newest teen court programs. Fewer than one-fifth (19 percent) of these programs had been in operation for 5 years or more, compared with 31 percent of adult judge programs, 35 percent of programs using peer juries, and 34 percent of programs using the youth tribunal model. Most (58 percent) youth judge programs had been in operation for less than 2 years at the time of the survey.

Youth judge programs were also the smallest programs in terms of their annual caseloads. Only 14 percent of programs using the youth judge model reported more than 100 cases per year, compared with 40 percent of programs using the adult judge model and 38 percent of programs using peer juries.

Programs using the peer jury model were the least likely to depend on private funding. Nearly four-fifths (78 percent) of peer jury programs received no private funding and only 13 percent received more than one-third of their funding from private sources. For most other courtroom models, nearly half of the programs responding to the survey reported receiving some private funding (i.e., 45 percent of adult judge programs, 47 percent of youth
Table 1: Characteristics of Four Courtroom Models Used by Teen Courts

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Adult Judge</th>
<th>Youth Judge</th>
<th>Peer Jury</th>
<th>Youth Tribunal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge</td>
<td>Adult</td>
<td>Youth</td>
<td>Adult (limited role)</td>
<td>Youth (often 3)</td>
</tr>
<tr>
<td>Youth attorneys</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Role of the youth jury, if any</td>
<td>Recommends disposition</td>
<td>Recommends disposition</td>
<td>Questions defendant, recommends disposition</td>
<td>No jury</td>
</tr>
<tr>
<td>Percentage of teen courts using this model for all cases</td>
<td>47%</td>
<td>9%</td>
<td>12%</td>
<td>10%</td>
</tr>
<tr>
<td>Percentage of teen courts using this model for at least some cases</td>
<td>64%</td>
<td>13%</td>
<td>26%</td>
<td>12%</td>
</tr>
<tr>
<td>Percentage of teen court cases handled using this model</td>
<td>60%</td>
<td>7%</td>
<td>22%</td>
<td>7%</td>
</tr>
</tbody>
</table>

Note: In the national survey, the combination of the adult judge, youth judge, peer jury, and youth tribunal models accounted for 96 percent of all cases handled by responding programs. The remaining 4 percent were handled with other models, often variations of the more established models (e.g., youth tribunal with no prosecutor or defense attorney). The four courtroom models were first described by the National Youth Court Center, American Probation and Parole Association.


Community Support
The success of an individual teen court may depend on how well it is supported by various segments of the community. Teen court advocates have observed that it is essential for teen courts to be accepted by the larger
Table 2: Percentage of Teen Courts Using Each Courtroom Model, by Sponsoring Agency

<table>
<thead>
<tr>
<th>Courtroom Model</th>
<th>Administrative Host</th>
<th>Total (n=330)</th>
<th>Court/Probation Agency (n=121)</th>
<th>Police/Prosecutor Agency (n=50)</th>
<th>School/Private Agency/Other (n=159)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult judge</td>
<td></td>
<td>47%</td>
<td>58%</td>
<td>16%</td>
<td>48%</td>
</tr>
<tr>
<td>Youth judge</td>
<td></td>
<td>9%</td>
<td>4%</td>
<td>12%</td>
<td>13%</td>
</tr>
<tr>
<td>Peer jury</td>
<td></td>
<td>12%</td>
<td>18%</td>
<td>14%</td>
<td>6%</td>
</tr>
<tr>
<td>Youth tribunal</td>
<td></td>
<td>10%</td>
<td>3%</td>
<td>24%</td>
<td>10%</td>
</tr>
<tr>
<td>Mixed models</td>
<td></td>
<td>22%</td>
<td>17%</td>
<td>34%</td>
<td>23%</td>
</tr>
</tbody>
</table>

Notes: Percentages may not equal 100 due to rounding; n=number of respondents.

justice system in their local area (National Youth Court Center, 2000). To examine teen court program directors’ perceptions of community support for their programs, the survey asked each program to consider several prominent community groups and indicate whether each was “very supportive,” “moderately supportive,” “mildly supportive,” or “not at all supportive” (see figure 8). Judges were seen as the greatest supporters of teen court programs. More than 9 in 10 teen courts rated their local judges as “very supportive” (71 percent) or “moderately supportive” (21 percent). Other groups considered “very supportive” or “moderately supportive” of teen courts included law enforcement (87 percent), court intake and probation workers (86 percent), teachers and other school officials (86 percent), and prosecutors (84 percent). In general, teen courts perceived all of the named groups to be supportive. Even the groups ranking lowest on the list (elected officials and the business community) were considered by a majority of teen courts as either very or moderately supportive (78 and 67 percent, respectively).

Problems

As small, community-based programs, teen courts face a range of challenges and obstacles. To identify the type of problems facing teen courts, the survey asked each program to review a list of typical operational problems that might cause difficulties for teen courts. Each court was asked to indicate whether it had experienced the issue as a “serious” problem, a “minor” problem, something in between, or not a problem at all (see figure 9).

Not surprisingly, the operational problem reported most often by teen courts was funding (see figure 9). Forty percent of the programs reported “some problems” (25 percent) or “serious problems” (15 percent) with funding uncertainties. Only 38 percent of the programs reported that funding uncertainties caused no problems.

Other problems that presented significant challenges for teen courts included retaining youth volunteers (i.e., attorneys, judges, and jurors) and maintaining an adequate flow of referrals. More than one-fifth (21 percent) of the programs reported having “some” problems or “serious” problems keeping teen volunteers. Nearly one-third (29 percent) reported having “some” or “serious” problems with maintaining sufficient case referrals.

Several other issues were described as presenting “some” or “serious” problems for teen courts. These issues included cases in which too much time elapsed between a youth’s arrest and his or her referral to teen court (19 percent), difficulties in coordinating the efforts of teen courts with other agencies in the community (16 percent), and problems recruiting youth volunteers (19 percent) and adult volunteers (20 percent).

Differences by program characteristics.

The extent to which teen courts reported having problems in meeting specific challenges varied according to other program characteristics. Some differences were statistically significant. For example, teen courts operated by schools or private agencies were significantly more likely to report problems with funding uncertainties (see table 4). Among programs operated by private agencies and schools, 79 percent reported at least some problems...
### Table 3: Selected Characteristics, by Courtroom Models Used by Teen Courts

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Courtroom Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total (n=332)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Years in operation</strong></td>
<td></td>
</tr>
<tr>
<td>Less than 2</td>
<td>31%</td>
</tr>
<tr>
<td>2 to 4</td>
<td>37%</td>
</tr>
<tr>
<td>5 or more</td>
<td>31%</td>
</tr>
<tr>
<td><strong>Total annual caseload (cases/year)</strong></td>
<td></td>
</tr>
<tr>
<td>50 or fewer</td>
<td>34%</td>
</tr>
<tr>
<td>51 to 100</td>
<td>25%</td>
</tr>
<tr>
<td>More than 100</td>
<td>42%</td>
</tr>
<tr>
<td><strong>Sponsoring agency</strong></td>
<td></td>
</tr>
<tr>
<td>Court/probation agency</td>
<td>37%</td>
</tr>
<tr>
<td>Police/prosecutor</td>
<td>15%</td>
</tr>
<tr>
<td>School/private agency/other</td>
<td>48%</td>
</tr>
<tr>
<td><strong>Private funding sources</strong></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>59%</td>
</tr>
<tr>
<td>Less than 1/3 of budget</td>
<td>21%</td>
</tr>
<tr>
<td>More than 1/3 of budget</td>
<td>20%</td>
</tr>
<tr>
<td><strong>Authority to hold trials</strong></td>
<td></td>
</tr>
<tr>
<td>No—youth must admit to charges</td>
<td>87%</td>
</tr>
<tr>
<td>Yes—able to hold full trials</td>
<td>13%</td>
</tr>
<tr>
<td><strong>Paid program director</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>20%</td>
</tr>
<tr>
<td>Yes</td>
<td>80%</td>
</tr>
<tr>
<td><strong>Operation during summer</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>27%</td>
</tr>
<tr>
<td>Yes</td>
<td>73%</td>
</tr>
<tr>
<td><strong>Youth with prior arrests accepted</strong></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>39%</td>
</tr>
<tr>
<td>Rarely</td>
<td>49%</td>
</tr>
<tr>
<td>Often or very often</td>
<td>13%</td>
</tr>
<tr>
<td><strong>Youth with prior juvenile court referrals accepted</strong></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>50%</td>
</tr>
<tr>
<td>Rarely</td>
<td>41%</td>
</tr>
<tr>
<td>Often or very often</td>
<td>9%</td>
</tr>
</tbody>
</table>

**Notes:** Percentages may not equal 100 due to rounding; n = number of respondents.

**Source:** The Urban Institute. 1998. National survey of youth courts and teen courts.

with funding, compared with 44 percent of programs operated by courts and 49 percent operated by law enforcement or prosecutors. Teen courts operated by schools or private agencies were also significantly more likely than programs run by courts, law enforcement, or prosecutors to report problems with a lack of judicial support (38 percent) and difficulties coordinating with other agencies (63 percent).

Smaller programs were somewhat more likely than larger programs to report problems with a lack of judicial support and with a lack of clear program goals. More than one-quarter (28 percent) of teen courts that handled fewer than 50 cases per year reported having problems with goal clarity, compared with 15 percent of programs that handled more than 100 cases each year.
The length of time that teen courts had been in operation was also associated with problems regarding the clarity of their goals. Programs less than 2 years old were significantly more likely than programs in operation for 5 or more years to report having problems with goal clarity and with issues surrounding legal liability. Finally, programs that relied heavily on private funds (often those operated by private agencies) were significantly more likely than those that did not rely heavily on such funds to report a lack of judicial support, coordination difficulties, a lack of adult volunteers, and problems with retaining youth volunteers. Programs that depended on private funding were also significantly more likely to report problems with heavy caseloads. Among programs that received more than one-third of their funding from private sources, 35 percent reported problems stemming from too many referrals, compared with 16 percent of programs that received no private funding.

**Evaluation Research**

Despite broad and growing interest in teen courts, only a few studies have attempted to measure their effect on youth, and even the best of these studies have not yet produced the sort of evaluation data necessary to deem a program effective. Juvenile justice officials and practitioners generally praise teen courts, but these claims remain largely unsubstantiated. The Evaluation of Teen Courts Project conducted a comprehensive review of evaluation studies (published and unpublished) conducted in the past 20 years. These studies examined teen and youth court programs in States including California, Florida, Kentucky, Maryland, New York, North Carolina, and Texas. All of the studies were limited in scope and methodology, but together they offered insight to an essential question for State and local officials, “Do teen courts work?”

**Recidivism**

Recidivism would seem to be an obvious focus for evaluation studies of teen courts, but only a handful of evaluations have measured postprogram recidivism (see table 5). Most studies have relied on court records and official police data to detect recidivism. Few studies have attempted to collect personal data from teen court defendants. Only Swink's (1998) study of a teen court program in Onondaga County, NY, measured self-reported recidivism, and it relied on parents to report the illegal activities of their children.

Of the few studies that measured official recidivism, some found very low rates of reoffending among former youth court defendants. Several researchers found rates of postprogram recidivism that ranged from 3 to 8 percent within 6 to 12 months of appearance in teen court (Butler-Mejia, 1998; McNeece et al., 1996; SRA Associates, 1995). A few studies found recidivism rates in excess of 20 or 30 percent. One Texas study, for example, found that 24 percent of former youth court defendants participated in teen court programs in States including California, Florida, Kentucky, Maryland, New York, North Carolina, and Texas. All of the studies were limited in scope and methodology, but together they offered insight to an essential question for State and local officials, “Do teen courts work?”

**Figure 9: Extent of Problems Reported by Teen Courts**

<table>
<thead>
<tr>
<th></th>
<th>Percentage of U.S. Teen Courts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Serious problems</td>
</tr>
<tr>
<td>Funding uncertainties</td>
<td>15</td>
</tr>
<tr>
<td>Keeping teen volunteers</td>
<td>4</td>
</tr>
<tr>
<td>Not enough referrals</td>
<td>8</td>
</tr>
<tr>
<td>Delays between offense and referral</td>
<td>4</td>
</tr>
<tr>
<td>Difficult coordination with other agencies</td>
<td>2</td>
</tr>
<tr>
<td>Lacking teen volunteers</td>
<td>3</td>
</tr>
<tr>
<td>Lacking adult volunteers</td>
<td>3</td>
</tr>
<tr>
<td>Youth who deny charges after referral</td>
<td>6</td>
</tr>
<tr>
<td>Politics of juvenile crime</td>
<td>2</td>
</tr>
<tr>
<td>Lack of judicial support</td>
<td>1</td>
</tr>
<tr>
<td>Confidentiality issues</td>
<td>4</td>
</tr>
<tr>
<td>Lack of clarity regarding goals</td>
<td>7</td>
</tr>
<tr>
<td>Legal liability issues</td>
<td>5</td>
</tr>
<tr>
<td>Too many referrals</td>
<td>6</td>
</tr>
</tbody>
</table>

*Note: Percentages may not equal 100 due to rounding.*

### Table 4: Operational Problems, by Characteristics of Teen Court Programs

<table>
<thead>
<tr>
<th>Extent of Problem</th>
<th>Total (n=335)</th>
<th>Court DA (n=121)</th>
<th>Police, School (n=51)</th>
<th>Private (n=153)</th>
<th>Annual Caseload (number of cases)</th>
<th>Years in Operation</th>
<th>Percentage of Private Funding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Fewer Than 50 (n=109)</td>
<td>50 to 100 (n=80)</td>
<td>More Than 100 (n=135)</td>
<td>Less Than 2 (n=105)</td>
<td>2 to 4 (n=125)</td>
<td>5 or More (n=104)</td>
</tr>
<tr>
<td>Funding uncertainties</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not a problem</td>
<td>38%</td>
<td>56%</td>
<td>51%</td>
<td>21%</td>
<td>43%</td>
<td>30%</td>
<td>39%</td>
</tr>
<tr>
<td>Minor problem</td>
<td>23%</td>
<td>19%</td>
<td>20%</td>
<td>26%</td>
<td>16%</td>
<td>28%</td>
<td>22%</td>
</tr>
<tr>
<td>Definite problem*</td>
<td>40%</td>
<td>25%</td>
<td>29%</td>
<td>53%</td>
<td>41%</td>
<td>42%</td>
<td>39%</td>
</tr>
<tr>
<td>(\chi^2=40.75; p&lt;0.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of judicial support</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not a problem</td>
<td>73%</td>
<td>88%</td>
<td>76%</td>
<td>62%</td>
<td>67%</td>
<td>79%</td>
<td>76%</td>
</tr>
<tr>
<td>Minor problem</td>
<td>15%</td>
<td>9%</td>
<td>14%</td>
<td>18%</td>
<td>15%</td>
<td>11%</td>
<td>16%</td>
</tr>
<tr>
<td>Definite problem</td>
<td>12%</td>
<td>3%</td>
<td>10%</td>
<td>20%</td>
<td>18%</td>
<td>10%</td>
<td>8%</td>
</tr>
<tr>
<td>(\chi^2=24.95; p&lt;0.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legal liability issues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not a problem</td>
<td>79%</td>
<td>86%</td>
<td>80%</td>
<td>72%</td>
<td>82%</td>
<td>78%</td>
<td>74%</td>
</tr>
<tr>
<td>Minor problem</td>
<td>16%</td>
<td>11%</td>
<td>14%</td>
<td>20%</td>
<td>12%</td>
<td>15%</td>
<td>20%</td>
</tr>
<tr>
<td>Definite problem</td>
<td>6%</td>
<td>3%</td>
<td>6%</td>
<td>8%</td>
<td>6%</td>
<td>8%</td>
<td>5%</td>
</tr>
<tr>
<td>(\chi^2=14.58; p&lt;0.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of clear goals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not a problem</td>
<td>77%</td>
<td>76%</td>
<td>86%</td>
<td>75%</td>
<td>72%</td>
<td>71%</td>
<td>85%</td>
</tr>
<tr>
<td>Minor problem</td>
<td>8%</td>
<td>16%</td>
<td>8%</td>
<td>17%</td>
<td>17%</td>
<td>23%</td>
<td>10%</td>
</tr>
<tr>
<td>Definite problem</td>
<td>15%</td>
<td>8%</td>
<td>6%</td>
<td>8%</td>
<td>11%</td>
<td>6%</td>
<td>5%</td>
</tr>
<tr>
<td>(\chi^2=0.53; p&lt;0.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulties coordinating with other agencies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not a problem</td>
<td>48%</td>
<td>57%</td>
<td>60%</td>
<td>37%</td>
<td>48%</td>
<td>47%</td>
<td>49%</td>
</tr>
<tr>
<td>Minor problem</td>
<td>36%</td>
<td>29%</td>
<td>32%</td>
<td>44%</td>
<td>34%</td>
<td>39%</td>
<td>38%</td>
</tr>
<tr>
<td>Definite problem</td>
<td>16%</td>
<td>14%</td>
<td>8%</td>
<td>19%</td>
<td>18%</td>
<td>14%</td>
<td>13%</td>
</tr>
<tr>
<td>(\chi^2=2.15; p&lt;0.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lacking adult volunteers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not a problem</td>
<td>56%</td>
<td>63%</td>
<td>64%</td>
<td>48%</td>
<td>59%</td>
<td>49%</td>
<td>56%</td>
</tr>
<tr>
<td>Minor problem</td>
<td>24%</td>
<td>22%</td>
<td>14%</td>
<td>29%</td>
<td>17%</td>
<td>32%</td>
<td>26%</td>
</tr>
<tr>
<td>Definite problem</td>
<td>20%</td>
<td>15%</td>
<td>22%</td>
<td>23%</td>
<td>24%</td>
<td>19%</td>
<td>18%</td>
</tr>
<tr>
<td>(r=0.13; p&lt;0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Only three published studies (Hissong, 1991; North Carolina Administrative Office of the Courts, 1995; Seyfrit, Reichel, and Stutts, 1987) have used reasonably appropriate comparison groups to measure the possible effects of teen courts on recidivism (see table 5). Hissong’s evaluation of an Arlington, TX, teen court compared recidivism among teen court defendants with a group of non-teen-court participants matched on sex, race, age, and offense. The analysis suggested that teen court participants were significantly less likely to reoffend than the comparison group (24 percent versus 36 percent). Several important elements of the study, however, were poorly documented. The definition of recidivism used in the analysis (presumably rearrest) is unclear. The duration of the followup period is not described (subjects may have had different periods of risk), and there is a range of unexplored potential differences between the treatment group and the comparison group.

The North Carolina study used a comparison group that consisted of 97 cases diverted by police during a 6-month period prior to implementation of the teen court in Cumberland, NC. Researchers hypothesized that these youth would have been referred to teen court had the program been in existence. Teen court and comparison group cases were matched using several factors, including demographic characteristics and offense type, and researchers tracked the recidivism of both groups. The study failed to find statistically significant differences in the recidivism of the two groups. After 7 months, 20 percent of teen court participants had reoffended, compared with just 9 percent of the comparison group. The study also found little difference between the two groups in average time before a new offense (4.1 months for teen court offenders versus 4.6 months for the comparison group). Youth who successfully completed the teen court program were less likely to reoffend than were youth who began but failed to complete the program (11 percent compared with 42 percent), but this finding may reflect the greater tendency of low-risk youth to complete the program.

Seyfrit and her colleagues (1987) tracked recidivism outcomes for 52 youth referred to a Columbia County, GA, teen court program.
Table 4: Operational Problems, by Characteristics of Teen Court Programs (continued)

<table>
<thead>
<tr>
<th>Extent of Problem</th>
<th>Total (n=335)</th>
<th>Sponsoring Agency</th>
<th>Annual Caseload (number of cases)</th>
<th>Years in Operation</th>
<th>Percentage of Private Funding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Court DA (n=121)</td>
<td>Fewer Than 50 (n=109)</td>
<td>Less Than 2 (n=105)</td>
<td>None (n=193)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50 to 100 (n=80)</td>
<td>2 to 4 (n=125)</td>
<td>1-33% (n=70)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>More Than 100 (n=135)</td>
<td>5 or More (n=104)</td>
<td>More Than 33% (n=64)</td>
</tr>
<tr>
<td>Lacking teen volunteers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not a problem</td>
<td>51%</td>
<td>58%</td>
<td>43%</td>
<td>48%</td>
<td>54%</td>
</tr>
<tr>
<td>Minor problem</td>
<td>30</td>
<td>25</td>
<td>29</td>
<td>32</td>
<td>21</td>
</tr>
<tr>
<td>Definite problem</td>
<td>20</td>
<td>17</td>
<td>28</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>Keeping teen volunteers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not a problem</td>
<td>42%</td>
<td>44%</td>
<td>45%</td>
<td>39%</td>
<td>45%</td>
</tr>
<tr>
<td>Minor problem</td>
<td>37</td>
<td>41</td>
<td>29</td>
<td>37</td>
<td>31</td>
</tr>
<tr>
<td>Definite problem</td>
<td>21</td>
<td>15</td>
<td>26</td>
<td>25</td>
<td>24</td>
</tr>
<tr>
<td>Too many referrals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not a problem</td>
<td>79%</td>
<td>85%</td>
<td>76%</td>
<td>75%</td>
<td>87%</td>
</tr>
<tr>
<td>Minor problem</td>
<td>14</td>
<td>8</td>
<td>14</td>
<td>19</td>
<td>9</td>
</tr>
<tr>
<td>Definite problem</td>
<td>7</td>
<td>7</td>
<td>10</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Not enough referrals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not a problem</td>
<td>44%</td>
<td>45%</td>
<td>48%</td>
<td>43%</td>
<td>28%</td>
</tr>
<tr>
<td>Minor problem</td>
<td>27</td>
<td>30</td>
<td>27</td>
<td>26</td>
<td>29</td>
</tr>
<tr>
<td>Definite problem</td>
<td>29</td>
<td>26</td>
<td>25</td>
<td>32</td>
<td>43</td>
</tr>
</tbody>
</table>

Notes: Percentages may not equal 100 due to rounding. DA=District Attorney; n =number of respondents. Chi-square ($\chi^2$) measures the extent to which the values of one variable are systematically different across the categories of one or more variables. Probability ("p") measures the likelihood that a statistical relationship is due to chance. Typically, a relationship between two variables is considered statistically significant when the probability is less than 0.05. The correlation coefficient ("r") indicates the strength of association between two variables and ranges from –1.0 (strong inverse relationship) to +1.0 (strong positive relationship).

*Includes responses of "some problems" and "serious problems."


during an 18-month period in the early 1980's. They also collected data for a comparison group of 50 youth matched on demographics and offenses. The study found little difference between the two groups. Although 12 percent of the comparison group recidivated during the followup period, the same was true for 10 percent of the teen court defendants. Like the North Carolina study, the Seyfrit study was unable to control statistically for different periods of opportunity to reoffend. The followup periods ranged from 6 to 18 months, which reduced the researchers' ability to infer any real differences in the recidivism of the two groups.

**Other Outcomes**

Several studies have suggested that teen courts may have effects on youth other than reduced recidivism. These potential benefits include client satisfaction with the teen court experience (Colydas and McLeod, 1997; McLeod, 1999; Reichel and Seyfrit, 1984; Swink, 1998; Wells, Minor, and Fox, 1998), enhanced perceptions of procedural justice (Butler-Mejia, 1998), improved attitudes toward authority (LoGalbo, 1998; Wells, Minor, and Fox, 1998), and greater knowledge of the legal system (LoGalbo, 1998; Wells, Minor, and Fox, 1998).1

For example, McLeod's (1999) survey of former teen court participants found that at least 90 percent of youth referred to the Colonie (NY) Youth Court during 1997 and 1998 believed that the experience increased their understanding of the legal system, helped them improve their behavior, and helped them become more responsible. Nearly all survey respondents (95 percent) reported that going through teen court caused them to "make more thoughtful decisions." Nearly three in five (58 percent) reported better communication with their parents, and half (50 percent) reported improved grades in school. However, the study's very low response rate (24 percent of youth surveyed) raised the possibility that the youth responding to the followup survey may have been the most compliant and prosocial youth in the sample.

LoGalbo's (1998) evaluation of the Sarasota County, FL, teen court program also found

(continued on page 14)
Table 5: Findings of Studies on Recidivism Among Former Teen Court Participants

<table>
<thead>
<tr>
<th>Studies With Comparison Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>El Dorado County Superior Court, 1999</strong></td>
</tr>
<tr>
<td><strong>Measures:</strong> Uncertain</td>
</tr>
<tr>
<td><strong>Data Sources:</strong> Official records</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Key Finding:</strong> Measurable, but not significant, difference in favor of teen courts</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>North Carolina Administrative Office of the Courts, 1995</strong></td>
</tr>
<tr>
<td><strong>Measures:</strong> New court referral</td>
</tr>
<tr>
<td><strong>Data Sources:</strong> Official records</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Key Finding:</strong> Measurable, but not significant, difference in favor of comparison group</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Hissong, 1991</strong></td>
</tr>
<tr>
<td><strong>Measures:</strong> Uncertain</td>
</tr>
<tr>
<td><strong>Data Sources:</strong> Official records</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Key Finding:</strong> Significant difference in favor of teen courts</td>
</tr>
<tr>
<td><strong>Seyfrit, Reichel, and Stutts, 1987</strong></td>
</tr>
<tr>
<td><strong>Measures:</strong> Uncertain</td>
</tr>
<tr>
<td><strong>Data Sources:</strong> Official records</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Key Finding:</strong> Measurable, but not significant, difference in favor of teen courts</td>
</tr>
<tr>
<td>Table 5: Findings of Studies on Recidivism Among Former Teen Court Participants (continued)</td>
</tr>
<tr>
<td>---------------------------------</td>
</tr>
<tr>
<td><strong>Post-Hoc Studies (No Comparison Groups)</strong></td>
</tr>
<tr>
<td><strong>Harrison, Maupin, and Mays, 2000</strong></td>
</tr>
<tr>
<td><strong>Measures</strong>: Subsequent referral to juvenile probation</td>
</tr>
<tr>
<td><strong>Data Sources</strong>: Official records</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Minor et al., 1999</strong></td>
</tr>
<tr>
<td><strong>Measures</strong>: New court appearance</td>
</tr>
<tr>
<td><strong>Data Sources</strong>: Official records</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Butler-Mejia, 1998</strong></td>
</tr>
<tr>
<td><strong>Measures</strong>: Rearrest</td>
</tr>
<tr>
<td><strong>Data Sources</strong>: Official records</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>LoGalbo, 1998</strong></td>
</tr>
<tr>
<td><strong>Measures</strong>: Rearrest</td>
</tr>
<tr>
<td><strong>Data Sources</strong>: Official records</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Swink, 1998</strong></td>
</tr>
<tr>
<td><strong>Measures</strong>: New police contact</td>
</tr>
<tr>
<td><strong>Data Sources</strong>: Official records, questionnaires</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Table 5: Findings of Studies on Recidivism Among Former Teen Court Participants (continued)

<table>
<thead>
<tr>
<th>Study Sources</th>
<th>Methods/Findings</th>
</tr>
</thead>
</table>
| Wells, Minor, and Fox, 1998             | Methods/Findings: New court referral  
  Data Sources: Official records  
  Monitored subsequent court referrals for 55 teen court defendants handled by 18 Kentucky programs from 1994 to 1997.  
  Thirty-two percent of the teen court defendants recidivated (subsequent court contact for a new offense).  
  First-time offenders were less likely to recidivate than those with prior offenses. Successful completion of teen court sanctions was less likely for youth with prior offenses, but sanction completion was not correlated with recidivism.  
  Cautions: No comparison group. Significant subject attrition (88 percent of defendants from initial point of data collection) precludes meaningful analysis. |
| McNeece et al., 1996                   | Methods/Findings: Uncertain  
  Data Sources: Official records  
  Monitored caseload, sanctions, and client recidivism associated with Hernando County, FL, teen court during 1995 and 1996.  
  Researchers describe an analysis of official records that showed 8 percent of teen court youth processed since 1992 recidivated.  
  Cautions: No comparison group. Recidivism was not defined. Sample was not described. Followup period was not specified. |
| SRA Associates, 1995                   | Methods/Findings: New intake referral  
  Data Sources: Official records  
  Documented the number of cases heard, nature of sanctions imposed, and proportion of clients that recidivated after participation in a Santa Rosa, CA, teen court program.  
  Contacts with juvenile intake were tracked for defendants appearing in teen court between January 1993 and June 1994 (n=238).  
  Three percent of teen court defendants were again referred to juvenile intake following their appearance in teen court.  
  Cautions: No comparison group. Followup period not defined. Cases likely had varying exposure time for recidivism. Recidivism may be underreported because no data were collected from neighboring jurisdictions. |
| Rothstein, 1987                        | Methods/Findings: Uncertain  
  Data Sources: Police agency descriptions  
  Reported recidivism (presumably rearrest) for 87 youth referred to Odessa, TX, teen court in 1985 for misdemeanor drug and alcohol offenses.  
  Zero percent recidivism reported among teen court defendants during the 12-month followup period.  
  Cautions: No comparison group. Cases may have had varying lengths of exposure to recidivism risk. Recidivism results were based on claims made by the local police agency and not primary data collection by researchers. No discussion of data collection methods. Limited description of youth sample and selection methods. |

Source: The Urban Institute, Evaluation of Teen Courts Project.

(continued from page 11)

that teen court positively affected defendant attitudes toward authority and understanding of the legal process. LoGalbo surveyed 111 youth immediately after their initial interview with teen court staff and again upon completion of the program. The survey asked participants about their knowledge of Florida laws and the justice system, their attitudes toward nine authority figures (e.g., police officer, judge, parent, teacher), their attitudes toward teen court and toward themselves, and their perception of the fairness of teen court procedures. The study found teen court participation was associated with increased self-esteem and positive attitudes toward select authority figures (e.g., judges).
analysis also suggested that recidivism was less likely among defendants with improved attitudes toward authority figures.

Strong client satisfaction was also reported by researchers in Kentucky. Exit interviews conducted by Wells and colleagues (1998) revealed high levels of satisfaction among 123 teen court participants, with 84 percent indicating that their sentences were fair. Several positive features of the teen court experience were cited by the Kentucky subjects, including “educational advantages” (37 percent) and the actual sentences youth received (21 percent). Teens also consistently indicated that the opportunity to serve as a teen court juror was an important, positive aspect of the teen court process.

Conclusion

State and local jurisdictions across the country are embracing teen court as an alternative to the traditional juvenile justice system for their youngest and least serious offenders. Many jurisdictions report that teen court increases young offenders’ respect for the justice system and reduces recidivism by holding delinquent youth accountable for what is often their first offense. Moreover, a teen court may be able to act more quickly and more efficiently than a traditional juvenile court. Researchers are beginning to accumulate a body of findings on the effectiveness of teen courts, but more detailed information is needed for future practice and policy development.

The information discussed in this Bulletin is part of the Evaluation of Teen Courts Project, OJJDP’s response to the need for more detailed research about teen courts. The project, which is being conducted for OJJDP by researchers at The Urban Institute’s Justice Policy Center, is the first national, multisite evaluation of teen courts and youth courts. Four jurisdictions are participating in the study—Anchorage, AK; Tempe, AZ; Rockville, MD; and Independence, MO. The teen courts in these communities were selected for the study to maximize the number of courtroom models represented, the mix of geographic locations, and the overall quality and length of service of each program. The project features a quasi-experimental design with data in each jurisdiction being collected on a group of teen court participants and a comparison group of youth handled using traditional juvenile court procedures.

The Evaluation of Teen Courts Project is designed to address some of the key issues facing policymakers and practitioners as they consider investing more heavily in teen court programs in their own jurisdictions. The study will provide answers to the following questions:

- What do teen courts actually do with young offenders?
- What type of sentences are typically imposed on youth, and do the youth comply?
- Are youth and parents satisfied with their experiences in teen court?
- Do young offenders referred to teen courts have lower rates of recidivism than those handled in the traditional juvenile justice system?
- Do juveniles show improved attitudes toward law enforcement and the courts and improved relations with peers and family, and do they have a better understanding of the consequences of their illegal behavior?
- Do these outcomes vary across teen court models and across subsets of offenders?
- Have the most experienced teen courts learned any lessons that can be shared with other jurisdictions?
- What community-level factors contribute to the success of teen courts?

Findings from the entire Evaluation of Teen Courts Project will be available in 2001. Policymakers and practitioners will be able to draw on the study’s findings as they consider whether teen courts and youth courts should play a more prominent role in each jurisdiction’s system for responding to youthful offenders.

For Further Information

For more information about The Urban Institute, the Justice Policy Center, or the Evaluation of Teen Courts Project, see www.urban.org.

For more information about the National Youth Court Center, see www.youthcourt.net.

For more information about the Office of Juvenile Justice and Delinquency Prevention, see www.ojjdp.ncjrs.org.

References


This Bulletin was developed as part of the Evaluation of Teen Courts Project conducted by The Urban Institute in Washington, DC, and supported by funds provided to The Urban Institute by the Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice, under grant number 98-JN-FX-0003.

Points of view or opinions expressed in this document are those of the authors and do not necessarily represent the official position or policies of OJJDP or the U.S. Department of Justice nor of The Urban Institute, its trustees, or its funders.

Acknowledgments

Jeffrey Butts, Ph.D., Senior Research Associate, and Janeen Buck, Research Associate, are with the Evaluation of Teen Courts Project of The Urban Institute’s Justice Policy Center. Dionne Davis helped collect the data presented in this Bulletin. The Urban Institute’s Justice Policy Center is directed by Adele Harrell.

Both OJJDP and The Urban Institute gratefully acknowledge the efforts of the many teen courts and youth courts that contributed data for this study and the generous support of Tracy Godwin and the National Youth Court Center (www.youthcourt.net). Their participation made this Bulletin possible.
Update on Teen Court Recidivism Findings
### POST-HOC STUDIES (NO COMPARISON GROUPS)

<table>
<thead>
<tr>
<th>Year</th>
<th>Author(s)</th>
<th>Title</th>
<th>MEASURES</th>
<th>METHODS / FINDINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>Garrison, A.H.</td>
<td>“An evaluation of a Delaware teen court.” Juvenile and Family Court Journal, 52:11-21.</td>
<td>Subsequent re-arrest (a violation of Attorney General’s probation)</td>
<td>Tracked post-program recidivism for a sample (N=71) of youth referred to Kent County (DE) teen court from January 1999 to December 2000. 31% of teen court defendants were referred for new charges within the twelve months following their teen court appearance. Most recidivism (about 11%) occurred within the first three months following teen court. Defendants who received community service were less likely to recidivate within the first three months following court than defendants who did not receive this sanction. Caution: No comparison group.</td>
</tr>
<tr>
<td>2001</td>
<td>Harrison, P., J. R. Maupin, &amp; G. L. Mays</td>
<td>“Teen Court: An examination of processes and outcomes.” Crime and Delinquency, 47:243-264.</td>
<td>Subsequent referral to juvenile probation</td>
<td>Tracked post-program recidivism for a sample (N=478) of youth referred to Dona Ana County (NM) teen court from 1994 to 1998. 25% of teen court defendants were referred for new charges between teen court and 18th birthday. 32% of youth who did not complete teen court re-offended compared to 23% of program completers (statistically significant; p &lt; .001). Certain sanctions (community service and jury), as well as gender and age were associated with lower recidivism rates (findings were statistically significant). Caution: No comparison group. Recidivism offenses are reported in aggregate totals and cannot be attributed to individual youth. Methods do not control for subjects’ varying lengths of exposure to recidivism risk.</td>
</tr>
<tr>
<td>2001</td>
<td>LoGalbo, A.P. and C.M. Callahan</td>
<td>“An evaluation of teen court as a juvenile crime diversion program.” Juvenile and Family Court Journal, 52:1-11.</td>
<td>Re-arrest</td>
<td>Tracked post-program arrests among youth referred to Sarasota County (Florida) teen court between 1997 and 1998 (N=111). 13% of teen court defendants were re-arrested during 5-month follow-up. Improved attitudes toward self and authority figures (i.e., judge) were associated with lower incidence of recidivism among teen court youth. Caution: No comparison group for recidivism outcomes. Insufficient analysis of possible effects of sample attrition.</td>
</tr>
</tbody>
</table>
Appendix D
EVALUATION WEB RESOURCES
WEB RESOURCES RELATED TO EVALUATION

BJA Evaluation Website: A Comprehensive Evaluation Resource

“The Bureau of Justice Evaluation Website is designed to provide State Administrative Agency staff, criminal justice planners, researchers and evaluators, as well as local practitioners with a variety of resources for evaluating criminal justice programs. The website is maintained by the Justice Research and Statistics Association.”

This website by far provides the most useful and comprehensive information related to evaluation. It not only describes a comprehensive overview of evaluation issues, techniques and terms, it also references an impressive list of online resources available. The website also provides a lengthy list of links to online “How-To Guides” to evaluation.

URL:  http://www.bja.evaluationwebsite.org/

Evaluation Design

The GAO has constructed a thorough – albeit rather dry – description of all the relevant aspects of designing an evaluation. If you are planning on being heavily involved in this task, this could be a very useful resource.

http://www.gao.gov/policy/10_1_4.htm

Juvenile Justice Evaluation Center Online: Topic-Oriented Evaluation Information

“JJEC Online is a tool designed to assist juvenile justice practitioners, policymakers, and state agency administrators with the assessment and evaluation of programs and initiatives. JJEC Online is divided into four sections - JJEC Information, State Information, Juvenile Justice Program Evaluation, and Evaluation Resources - to provide professionals in the field with readily accessible evaluation assistance.”

In the Juvenile Justice Program Evaluation section, evaluation information is provided for 22 different juvenile justice topics with discussion of relevant process and outcome measures, evaluation issues, examples of evaluations, and publications specifically related to each particular topic.

URL:  http://www.jrsa.org/ijec/
National Center for Juvenile Justice: Measuring Program Performance

NCJJ is available to provide information related to program designs and knowledge of outcomes for particular programs. They will also provide practitioners and researchers with the technical assistance necessary to measure program performance.

URL: http://www.ncjfcj.unr.edu/homepage/ncjj/ncjj2/services/servicesProgplanning.htm

National Criminal Justice Reference Service: An Archive of Evaluations and Evaluation Reports

The NCRS website may also prove helpful in conducting evaluations by providing (1) examples of past evaluations for a large range of criminal justice topics and (2) by offering several reports related to evaluation issues.

URL: http://www.ncjrs.org/

Some examples of such reports provided by NCRS include:


Abstract: This report provides a framework for understanding the basic components of evaluation research and for giving program managers a preliminary sense of what to expect from an evaluation. The discussion notes that an evaluation can address the developmental issues of program start-up, program monitoring and modification, and the culmination of program activities in particular outcomes. It also provides program managers with suggestions when to use evaluation, when not to use it, and ways to modify it to address program needs. Individual sections explain the nature of evaluations; issues involved in conducting an evaluation internally or hiring a consultant; and the types of evaluations, including planning evaluation, process evaluation, evaluability evaluation, outcome evaluation, and efficiency evaluation.

Abstract: This document aims to provide grantees and contractors with protocols for training, technical assistance, and evaluation that represent basic tenets and standards for best practices in these areas. The protocols present guidelines, forms, and models that aim to ensure consistency and quality in the provision of training and technical assistance, as well as in the evaluation of the impact and outcomes of program initiatives.


Abstract: In an ongoing effort to provide resources and technical assistance to State and local service providers, this workbook is designed to assist grantees under Title V delinquency prevention programs in evaluating program effectiveness. The workbook provides information and resource aids on program planning, conducting evaluations, tracking programs, describing activities, monitoring risk factor data, performing analyses, and measuring outcomes and performance indicators. It is not a prescriptive evaluation tool but rather a guide that includes instructions and data-gathering forms that enable communities to assess their delinquency prevention program outcomes. The overall evaluation design used is a pre/post test design.

Standards for Evaluation: Utility, Feasibility, Proprietary, and Accuracy

Mary Ramlow of the Evaluation Center at Western Michigan University drafted a set of standards for program evaluation in order to “ensure that an evaluation will serve the information needs of intended users; will be realistic, prudent, diplomatic, and frugal; will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results; and will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.” These standards may be helpful to consider when initiating a program evaluation.


Joint Committee Standards on Evaluation: Program Evaluation

This document provides a step-by-step description of how to devise and execute an evaluation of a program.
Amoeba Web: Data Collection to Interpretation of Results

This website provides an overview of a whole host of research issues and topics ranging from data collection to research design to interpretation of results. Although the site discusses information related to quantitative research, it does provide more information related to qualitative research on the whole.

Electronic Textbook StatSoft

This website offers a description of statistical concepts and analytical techniques. The site provides statistical background at levels ranging from “elementary concepts” all the way to sophisticated, up-to-date statistical techniques. The site provides excellent support for researchers by offering a statistical glossary as well as an interactive Statistical Advisor that, based on your answers to a set of standardized questions, suggests what statistical methods would best suit your research purposes and then directs you to information related to that type of research.

Performance Measurement

This document walks you through how and why the government became so interested in performance measurement with the enactment of GPRA and how to go about designing and implementing this approach to evaluation.


http://www.ed.uiuc.edu/sped/tri/evalwkshp.htm
Appendix E
URBAN INSTITUTE PARENT AND YOUTH QUESTIONNAIRE 1
This survey is:
• for the Urban Institute only, we are not part of youth court
• confidential, your answers will not be shown to anyone
• for research only, it will have no effect on your case

Today’s ________ / ________ / ________
Date (month) (day) (year)

Your Birthday ________ / ________ / ________
(month) (day) (year)

What to do on the next page
For each question, please mark the answer that comes the closest, from “strongly disagree” to “strongly agree.”

If, for example:
you strongly agree
you disagree, but not strongly

YQ1-ID#________
<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most police officers try to do a good job.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can talk to my parent(s) about almost anything.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most of my friends drink (alcohol).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My friends get into trouble more than I do.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My teachers are proud of me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Getting good grades in school is important to me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Going through youth court is easy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My parent(s) are proud of me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I will probably be arrested again someday.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My friends think youth court is a joke.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Getting into a good college is important to me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth court is a waste of my time.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most kids shoplift from stores at least once.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My parent(s) take the time to really listen to me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth court makes you think about your future.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It’s worth getting into a little trouble to have fun.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We fight a lot in my family.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The police are usually fair to people like me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most of my friends are honest.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My parent(s) always know where I am.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

After answering every question, please place your survey inside the envelope.

Seal the envelope and drop it in the box provided.

Thank you for your participation in this study.
This survey:
• is for the Urban Institute, we are not part of youth court
• is confidential, your answers will not be shown to anyone
• will have no effect on how the youth court deals with your son or daughter

Your Child’s Birthday
(month) (day) (year)
(the child in youth court)

Your Child’s Sex: ☐ Male ☐ Female

What to do on the next page

For each question, please mark the answer that comes the closest, from “strongly disagree” to “strongly agree.”

If, for example:

you strongly agree ☐ ☐ ☐ ☐

you disagree, but not strongly ☐ ☒ ☐ ☐ ☐

PQ1-ID#_______
Please answer every item as it relates to your child who has been referred to youth court.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am proud of my child.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth court will go easy on my child.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My child doesn’t care much about school.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Getting good grades is important to my child.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth court will probably be a waste of our time.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My child will probably get in trouble again.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Going to a good college is important to my child.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth court will make my child straighten up.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sometime it's hard for me to talk to my child.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My child’s friends get into trouble a lot.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think most of my child’s friends drink (alcohol).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I don't know much about how youth court works.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My child will probably benefit from youth court.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My child won’t listen to me anymore.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most of my child’s friends have shoplifted.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My child thinks it is fun to get into trouble.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I have to yell at my child a lot at home.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The police are usually fair to people like me.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I usually know where my child is at any time.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most of my child’s friends seem to be honest.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Finally, please answer the following questions.

Your Sex:  
☐ Male  ☐ Under 30  ☐ Less than high school  
☐ Female  ☐ 30 - 39  ☐ High school/GED  
☐ 40 - 49  ☐ Some college  
☐ 50 - 59  ☐ College degree  
☐ 60 or older  ☐ Graduate degree

After answering every question, place your survey inside the envelope.

Seal the envelope and drop it into the box provided.

Thank you.
Appendix F
URBAN INSTITUTE PARENT AND YOUTH QUESTIONNAIRE 2
This survey is:
• for the Urban Institute only, we are not part of youth court
• confidential, your answers will not be shown to anyone
• for research only, it will have no effect on your case

Today’s _________ / _________ / ________
Date (month) (day) (year)

Your Birthday _________ / _________ / ________
(month) (day) (year)

What to do on the next page
For each question, please mark the answer that comes the closest, from “strongly disagree” to “strongly agree.”

If, for example:
you strongly agree

You disagree, but not strongly
**Please answer every item by marking one circle for each statement.**

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coming to youth court was a waste of time.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The punishment I got was not very tough.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think I was treated fairly by the youth court.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth court people don’t know what they’re doing.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What I did (my offense) wasn’t really that serious.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth court people were not really interested in me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I’m glad I came here rather than a real court.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I didn’t get to talk enough in youth court.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The kids working in youth court were o.k.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I faked being sorry when I was in youth court.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth court was more interesting than I expected.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I’ve done a lot of bad things nobody knows about.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It was unfair that I had to come to youth court.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth court people care about my rights.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I was treated about the same as other kids here.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I was treated worse than the other kids here.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It was my fault that I had to come here.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I wish they would have explained youth court better.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kids who work in youth court are mostly geeks.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth court made me want to know more about law.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Finally, please answer the following:

- **Your Age Now:**
  - ☐ Under 10
  - ☐ 10 years old
  - ☐ 11 years old
  - ☐ 12 years old
  - ☐ 13 years old
  - ☐ 14 years old
  - ☐ 15 years old
  - ☐ 16 years old
  - ☐ 17 years old
  - ☐ Over 17

- **Your Grade in School (as of October 2000):**
  - ☐ 6th
  - ☐ 7th
  - ☐ 8th
  - ☐ 9th
  - ☐ 10th
  - ☐ 11th
  - ☐ 12th
  - ☐ Not in school

- **Your Sex:**
  - ☐ Male
  - ☐ Female

Thank you.
Place your completed survey inside the envelope.
Seal the envelope and drop it in the box provided.
This survey:
• is for the Urban Institute, we are not part of Youth Court
• is confidential, your answers will not be shown to anyone
• will have no effect on how the Youth Court deals with your son or daughter

Your Child’s Birthday _________ / _________ / _________
(month) (day) (year)
(the child in youth court)

Your Child’s Sex: ☐ Male ☐ Female

What to do on the next page

For each question, please mark the answer that comes the closest, from “strongly disagree” to “strongly agree.”

If, for example:
you strongly agree ☐ ☐ ☐ ☐
you disagree, but not strongly ☐ ☐ ☐ ☐

PQ2-ID#________
Please answer every item as it relates to your child who has been referred to youth court.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coming to youth court was a waste of my time.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Youth court was pretty easy on my child.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I think my child was treated fairly.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Youth court people don’t know what they’re doing.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>What my child did wasn’t really that serious.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The youth court was not really interested in us.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I’m glad we came here and not a real court.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>My child did not get enough time to talk in court.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The kids working in youth court seem o.k.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>My child did not seem to take this seriously.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>They explained everything to me before court.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I wished they would have scared my child more.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>It was unfair to make my child come here.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>No one here really cared about my child’s rights.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I wish I could have talked more in youth court.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I think the youth court sees my child as a bad kid.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Youth court will help my child stay out of trouble.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>My child was treated about the same as other kids.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>My child will get better grades in school after this.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I am grateful to youth court for trying to help us.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Please answer the following questions.

Your Current Work Situation:
- ☐ Working full-time
- ☐ Working part-time
- ☐ Looking for work
- ☐ Staying home for now
- ☐ None of these

Your Current Living Arrangement:
- ☐ Renting house/apartment
- ☐ Own house/apartment
- ☐ Living with friends/relatives
- ☐ None of these

Does Your Family Have:
- Microwave oven ☐ No ☐ Yes
- TV ☐ No ☐ Yes
- Cable or satellite TV ☐ No ☐ Yes
- Computer ☐ No ☐ Yes
- Access to the internet ☐ No ☐ Yes
- Cellular phone ☐ No ☐ Yes

Place your survey inside the envelope.
Seal the envelope and drop it into the box provided.
Thank you.
This survey is:
• for the Urban Institute only, we are not part of youth court
• confidential, your answers will not be shown to anyone
• for research only, it will have no effect on your case

For each question, please mark the answer that comes the closest, from “strongly disagree” to “strongly agree.”

If, for example:
- you strongly agree
- you disagree, but not strongly

YQ3-ID#_______
Please answer every item by marking one circle for each statement.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Getting good grades in school is important to me.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Going through youth court is easy.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My teachers are proud of me.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My parent(s) are proud of me.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Getting into a good college is important to me.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth court is a waste of my time.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I will probably be arrested again someday.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My friends think youth court is a joke.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most of my friends drink (alcohol).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My friends get into trouble more than I do.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most police officers try to do a good job.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can talk to my parent(s) about almost anything.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My parent(s) take the time to really listen to me.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It’s worth getting into a little trouble to have fun.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most kids shoplift from stores at least once.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth court makes you think about your future.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most of my friends are honest.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We fight a lot in my family.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The police are usually fair to people like me.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My parent(s) always know where I am.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I wish I had gone to regular court, not youth court.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth courts are too hard on kids.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Being in youth court makes you a better person.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I will never get into serious trouble again.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think I will have a good job someday.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I’ve done a lot of bad things no one knows about.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most of my friends have smoked marijuana.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>You can learn a lot about the law in youth court.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most of my friends steal at least a little.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The police in my town do not like kids my age.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My parent(s) usually don’t care what I think.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My behavior is no worse than other kids my age.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Place your completed survey inside the envelope.
Seal the envelope.
Send us the envelope in the U.S. Mail.

Thanks for your help with this study and good luck with all your future plans.

Project Director
The Evaluation of Teen Courts (ETC) Project’s Protection of Human Subjects and Data Security Plan (Administrative Data) November 1999
Hiring and Working With an Evaluator
Hiring and Working With an Evaluator
Hiring and Working With an Evaluator

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

This briefing was prepared by Stan Orchowsky, Ph.D., JRSA's Research Director, with assistance from Taj Carson, Ph.D., former JJEC Project Manager, and Merideth Trahan, JJEC Project Manager. Editing was provided by Nancy Michel, JRSA's Director of Publications. Eric Peterson, our Grant Manager at the Office of Juvenile Justice and Delinquency Prevention, provided valuable support for which we are extremely grateful.

September 2001

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

Introduction ........................................................................................................ 1
Who is an Evaluator? ......................................................................................... 1
What Can an Evaluator Do For You? .............................................................. 2
Advantages of Hiring an Evaluator ................................................................. 3
Disadvantages of Hiring an Evaluator ............................................................. 4
Selecting an Evaluator ..................................................................................... 5
Developing an Evaluation Plan ......................................................................... 7
Specifying Evaluation Products ...................................................................... 8
Maximizing Collaboration ................................................................................ 9
Other Evaluation Resources ........................................................................... 10
Conclusion ....................................................................................................... 12
Introduction

Program evaluation is an essential tool for the management of juvenile justice programs. Evaluation can help program managers and staff members understand how their daily activities are related to specific goals and objectives, and how these goals and objectives relate to the juvenile justice problem or need that the program hopes to address. While most program managers are capable of conducting some evaluation activities, managers may want to employ a qualified evaluator to assist them with the process of measuring outcomes or designing an evaluation study that will determine whether they are achieving the outcomes they hope for. The purpose of this briefing is to provide information to juvenile justice program managers about how to go about hiring an evaluator. The briefing discusses how a qualified evaluator can assist a program manager in assessing her/his program's performance, what characteristics to look for when selecting a competent evaluator, and how to go about finding such a person.

Who Is an Evaluator?

An evaluator is someone who has received formal training in research and/or evaluation and has experience in conducting evaluations of programs. Unfortunately, there is no easy way to identify a qualified program evaluator. For example, there are no licensing or certification requirements for program evaluators. Although many evaluators are members of professional evaluation organizations, such as the American Evaluation Association or the American Educational Research Association, membership in these organizations does not imply qualifications. There are very few university degree programs in program evaluation, and program evaluators may have backgrounds in the social sciences, such as psychology, sociology, criminal justice, public administration, or education. In a later section we discuss some of the qualifications to look for when selecting a competent evaluator.
What Can an Evaluator Do For You?

A good evaluator is part facilitator, part researcher, and part program specialist. A good evaluator can help your program with some or all of the following:

- Develop a logic model; that is, provide a written description of how the activities and components of your program relate to each other and to the goals and objectives you are trying to accomplish.
- Develop measures to determine whether your program is meeting its goals and objectives.
- Develop an evaluation design to determine whether your program is having its intended impacts.
- Design data collection forms and procedures, and databases to capture and record data collected.
- Analyze data and present results and conclusions from the findings.
- Provide recommendations to the program regarding ways to improve service delivery.

Program managers and staff will be able to perform some of these tasks themselves. Even with these, however, the program will benefit from the experience and expertise of a qualified evaluator. Think of the evaluation process as being like the process of building one’s own home. If you have a great deal of time and expertise, you may be able to build your own home with no help from experts. If you have a thorough understanding of home building and expertise in all but the most difficult areas, you might do most of the work yourself, only hiring an expert to help with the most difficult aspects or those that require specialized ability or equipment, such as grading, pouring a foundation, or plumbing and electrical wiring. Finally, if you have little or no expertise (but lots of money), you may hire someone to build your home from top to bottom. Even in this last case, however, the home builder will still consult with you on a number of factors, ranging from the basic layout of the house to what color the rooms should be painted. In this same way, programs can hire evaluators to provide as much or as little help as they think they might need to successfully carry out evaluation activities.
Advantages of Hiring an Evaluator

There are, of course, advantages and disadvantages to hiring an evaluator. Some of the advantages of hiring an evaluator to help with some or all of your evaluation activities include:

• **Specialized Knowledge and Ability.** The primary advantage of hiring an evaluator is the same that comes with hiring any expert – specialized knowledge and ability. Evaluators understand how to document program operations and processes, how to measure program outcomes, and how to collect and analyze data to determine program effectiveness. Program managers and staff members usually do not have expertise in these areas.

• **Objectivity.** The hallmark of good program evaluation is objectivity; that is, the ability to look at information about the program and form unbiased conclusions about whether the program is achieving what it wants to achieve. Most program managers and staff believe in the effectiveness of their programs or they would not be working in them. Thus there is the potential that they will look at evaluation findings in a biased way (even if this is an unconscious bias). A good program evaluator will point out both the positives and negatives in the program's operations.

• **Credibility.** Precisely because program evaluators are more knowledgeable and objective, their conclusions and recommendations tend to carry more weight than those that might be produced by the program managers and staff. This credibility may be important to funders, for example, when making decisions about whether to continue to fund the program.

• **Perspective.** In part because they are objective, evaluators may come to the program with fresh views about program activities and relationships between program components. In addition, a good evaluator brings to the table a different way of thinking about program effectiveness, one grounded in empiricism (that is, understanding program operations and outcomes through the collection and analysis of data, both quantitative and qualitative).
Disadvantages of Hiring an Evaluator

There also may be some disadvantages to employing an evaluator to assist your program with its evaluation efforts. Some of these disadvantages include:

• **Cost.** Specialized knowledge and expertise do not come cheap, and evaluation is usually no exception. It is not simply that evaluators are highly paid professionals. Conducting a carefully controlled evaluation study to determine program effectiveness can be an expensive and time-consuming process.

• **Time.** Although careful evaluation takes time regardless of who is conducting the evaluation, it may take additional time for someone who is not familiar with your program to gain some knowledge of its structure and functions. The evaluator will need to review documents and conduct interviews to gain this knowledge. Not only will this take extra time on the part of the evaluator, it will also require staff resources to locate and make copies of program documents, sit through interviews with the evaluator, and so on.

• **Lack of Expertise.** Just as hiring a builder who is unqualified will result in delays, mistakes and perhaps a poor product, so too will hiring an unqualified evaluator produce potentially damaging results. An evaluator who is not qualified may alienate staff, intrude upon clients and staff-client relationships, misunderstand the program and its functioning, and draw conclusions that are incorrect or inappropriate. If such an evaluator’s work is released to funders and the public, the program may suffer damage to its reputation and may find its funding jeopardized.

Not all of these advantages and disadvantages will apply in every case. Program managers must weigh all of these factors carefully when they decide to hire a program evaluator.
Selecting an Evaluator

There are some basic qualifications you can look for when determining how to hire an evaluator. What follows are some considerations for selecting an evaluator.

- **Formal Education.** As noted previously, very few individuals have formal education in program evaluation. Most evaluators do have formal training in research methods, however, usually in a social science discipline. Graduate-level training should provide this knowledge; for example, someone with a Ph.D. in criminology should have the research knowledge required to conduct evaluations.

- **Experience.** While evaluation as a process has a great deal in common with conducting research, there are also many differences between the two. A qualified evaluator must not only have research skills, but must also have specific experience in working with programs. While experience working in a juvenile justice program is not a requirement for an evaluator, experience working with juvenile justice programs is. The juvenile justice system is a complex one, and familiarity with the system and with juvenile justice programs is essential for communication, collaboration, and appropriate interpretation of evaluation findings.

- **Evaluation Philosophy.** Much has been written about how evaluations should be conducted, and different evaluators view the evaluation process differently. For example, some evaluators may consider themselves to be “experts” and view their role as one of an outsider who reviews program materials, interviews managers and staff, and then makes recommendations for “fixing” the program. Others see themselves more as researchers than evaluators and may avoid providing feedback to program managers and staff until after the evaluation is completed so as not to “contaminate” the evaluation. Neither of these are particularly productive evaluation philosophies for working with juvenile justice programs. Instead, you should be looking for an evaluator who believes that the evaluation process is a collaborative one between the evaluator and program managers and staff. In this philosophy, program managers and staff are seen as the experts, and evaluators work closely with them throughout the process of documenting program activities, developing performance measures,
interpreting evaluation findings, and making recommendations for program improvement. The goal of such evaluations is to improve the program, not to declare the program a success or failure. More formal names for this evaluation philosophy include “participatory evaluation,” “utilization-focused evaluation,” and “empowerment evaluation.”

• Communication Skills. Evaluators must be able to communicate with a wide variety of individuals who have a vested interest in the results of their work. Program staff and managers, funding agency representatives, legislators, city council members, and even program clients are some of the “stakeholders” to whom evaluators may be called upon to present their evaluation findings, conclusions, and recommendations. Evaluators should not only be personable and engaging, but should be able to clearly present findings and conclusions both orally and in written form.

The process for hiring an evaluator is similar to that for hiring any employee. Carefully review the evaluator’s resume to determine if she/he has experience conducting evaluations of programs similar to yours. Be sure that references include directors of programs that the evaluator has worked with in the past, and ask those individuals about their experiences with the evaluator, including how well the evaluator worked collaboratively with the program managers and staff. Interview the evaluator and determine if this is a person with whom you would be comfortable working. Ask for samples of the evaluator’s work, including evaluation reports. Review the materials to be sure they are written clearly, without a great deal of jargon, and in a way that would be understandable to you and to those with whom you would like to share the evaluation findings.
Developing an Evaluation Plan

When you have selected an evaluator, you must specify in writing what the evaluator will do. Early in the process both you and the evaluator should agree on the questions to be addressed by the evaluation, the tasks that need to be performed to address those questions, who will be responsible for these tasks, and when they will be accomplished. For example, one of the questions you might want answered is whether the juveniles in your program have developed more positive attitudes toward authority figures, such as parents, teachers, and probation officers, as a result of your program. You and the evaluator will need to agree on the tasks that need to be performed in order to answer this question— for example, identifying or designing an instrument to measure attitudes toward authority figures, administering the instrument to juveniles at the beginning and end of their program participation, scoring the measurement instrument and entering the scores in a database, and analyzing the data and presenting the findings. These may all be responsibilities of the evaluator, or you may wish to save some money by having staff members, for example, administer and score the measure and enter the resulting scores in a computerized database. In any case, spelling out the responsibilities of the evaluator will avoid confusion and duplication of effort and ensure that all tasks are completed in a timely fashion.
Specifying Evaluation Products

In addition to an evaluation plan, you should be sure that you have agreed with the evaluator on what the products of the evaluation will be. Evaluators should produce a formal report at the end of the evaluation explaining what was done and what was found. However, you may also wish to ensure that other products are produced, such as an executive summary of the evaluation report or a briefing to your county commissioners regarding the evaluation findings. It is important that you and the evaluator agree on when the evaluator’s involvement in the process ends. For example, does the evaluator’s role end when the final report is submitted to you, or will the evaluator participate in the process of disseminating the findings, including presentations to various stakeholder groups and being available to answer their questions? Other considerations include whether the evaluator will provide formal progress reports during the course of the evaluation, and whether the final products will include a database, data collection forms, and other products related to data collection.
Maximizing Collaboration

A good evaluator will seek to work collaboratively with you and your staff. However, there is always the potential for conflict between the evaluator and the program staff. Staff members may feel defensive and be reluctant to answer the many questions the evaluator may have. An evaluation often means more work for staff members, who must sit for interviews, provide access to files and notes, and collect data for the evaluation. These are considerations that must be addressed at the outset of the evaluation process so that the evaluator and staff can work collaboratively.

The evaluator and program manager must explain the purpose of the evaluation to staff and assure them that the evaluator is not there to examine the job performance of individual staff members. The evaluator should explain that she/he is committed to working together with program managers and staff to improve the program and the services it provides to youth. Addressing these issues early in the process will greatly improve the ability of the evaluator and program staff to work together effectively.

The evaluator and program manager and staff should work collaboratively in implementing all phases of the evaluation plan. Specifically, they should work together to identify program goals and objectives, link activities to goals and objectives, develop performance measures, determine what data to collect, and interpret the findings of the data analysis. Program staff should also have input on any recommendations resulting from the evaluation.

A close working relationship between the evaluator and program staff will reap many benefits for both. The evaluator who works closely with program managers and staff will have a much clearer sense of how the program functions and will be in a much better position to provide useful feedback. Program managers and staff will benefit from the fresh perspective that an evaluator can provide on their daily activities and how these relate to what their program is trying to accomplish. The result will be an evaluation that is relevant and useful, and one that has the endorsement of the program’s staff and managers.
Other Evaluation Resources

Up to this point, we have been assuming that programs may want to hire an evaluator to conduct evaluation activities with their programs. However, a number of resources may be available to help you with evaluation activities, and these resources may be inexpensive or cost-free. These resources may also be able to provide you with suggestions about where to find a qualified evaluator. The following are some places where evaluation assistance may be available:

• **Federal Agencies.** The Office of Juvenile Justice and Delinquency Prevention (OJJDP), the National Institute of Justice (NIJ), and other Federal agencies may fund program evaluation activities at the local level. Moreover, OJJDP contracts with a number of organizations and individuals to provide assistance to states and localities that wish to evaluate their juvenile justice initiatives. JRSA’s Juvenile Justice Evaluation Center project is an example of one such initiative.

• **State Agencies.** The agency in your state that administers funds from OJJDP and other Federal agencies may have resources available to assist you with evaluation. Many of these agencies have program evaluators on staff, as do other state agencies, such as Departments of Correction and Juvenile Justice Services, along with other social service agencies.

• **Local Agencies.** Depending on the size of your locality, there may be funds and/or evaluation expertise available from the city or county government. County social and juvenile service agencies, as well as mental health service agencies, may have evaluators on staff who can provide assistance to your program.

• **Colleges and Universities.** Colleges and universities can be valuable resources for finding individuals who can provide assistance with evaluating your program. Although few universities have formal program evaluation departments, knowledgeable faculty members may be found in departments of criminology, education, psychology, social work, and public policy. These departments may also have graduate students...
available who while they may not be qualified evaluators, may have
enough knowledge to assist you with particular evaluation tasks, such
as developing measurement instruments or automated databases.

- Professional Organizations. As noted previously, many evaluators are
members of professional organizations, and these organizations may
be able to provide information on how to contact qualified evaluators.
Those organizations include the American Evaluation Association, the
American Educational Research Association, the American Society of
Criminology, the Academy of Criminal Justice Sciences, the American

Resources for Hiring and Working with An Evaluator

- Justice Research and Statistics Association
  http://www.jrsa.org
- Office of Juvenile Justice and Delinquency Prevention
  http://www.ojjdp.ncjrs.org
- American Evaluation Association
  http://www.eval.org
- American Educational Research Association
  http://www.aera.ucsb.edu
- American Society of Criminology
  http://www.asc41.com
- American Psychological Association
  http://www.apa.org
- American Sociological Association
  http://www.asanet.org
Conclusion

There are many benefits to hiring an evaluator to help programs collect and analyze data about their effectiveness. Juvenile justice program managers who do wish to hire an evaluator should look for someone who has experience working with similar programs and who believes in a collaborative approach to evaluation. By clearly specifying in advance the tasks the evaluator will perform, program managers can be assured of getting the assistance they need in developing performance measures and instruments, analyzing data, and presenting findings to key constituent groups.
Program Evaluation Briefing Series

#1 Juvenile Justice Program Evaluation: An Overview

#2 Hiring and Working With an Evaluator