Why Evaluate?

A 1996 literature review conducted by the RAND Corporation revealed that while hundreds of arts programs for at-risk youth exist, very few rigorous evaluations of such programs have been conducted. Instead, arts agencies have relied on anecdotal evidence of program success to leverage the resources needed to support their arts programs for at-risk youth. The YouthARTS project was designed, in part, to provide the "hard evidence" of program effectiveness needed to get the attention of funding agencies and policy makers and to raise general awareness about the role that the arts can play in promoting healthy youth development.

One of the YouthARTS project's primary goals was to conduct a rigorous evaluation of the impact of arts programs on juvenile delinquency and related behavior problems. To accomplish this ambitious task, program staff at the three YouthARTS sites collaborated with Caliber Associates, a management consulting firm under contract with the Federal Office of Juvenile Justice and Delinquency Prevention, as well as local data collectors (such as Portland State University) and other local program partners. Through these collaborative arrangements, each YouthARTS site gathered data to support the national evaluation of program effects on the knowledge, attitudes, and behaviors of youth at risk for juvenile delinquency and crime.

The rigorous evaluation of YouthARTS program outcomes in Atlanta, San Antonio, and Portland has shown that arts programs really can have an impact on youth. Not only can such programs enhance young peoples' attitudes about themselves and their futures, they also can increase academic achievement and decrease delinquent behavior. Equally important, the evaluation provided YouthARTS sites with valuable information about program implementation and service delivery—the feedback needed to refine their program activities and maximize their success.

This chapter is designed to help you evaluate the effectiveness of your arts program for at-risk youth. It was written, in large part, by Rebecca Schaffer from Caliber Associates.

Toward the end of the chapter is a short section on how to measure improvements in art skills. It describes the data collection methods and instruments that the YouthARTS programs used to assess changes in art knowledge among participating youth.

This chapter covers:

- [ ] Why evaluate?

- [ ] Conducting your own process and outcome evaluation

- [ ] The benefits and challenges of a well-planned outcome evaluation

- [ ] A step-by-step approach for evaluating your arts program's outcomes

- [ ] Where to go for evaluation assistance

- [ ] How to measure improvements in art skills

- [ ] Best practices from the field

- [ ] Other resources

The appendices contain evaluation resources discussed in this chapter, including the data collection instruments and data collection guide used in the YouthARTS evaluation. Copies of the final YouthARTS evaluation report will be available through the National Criminal Justice Reference Service (NCJRS).

Conducting Your Own Process and Outcome Evaluation

"Evaluation is a formal process for collecting, analyzing, and interpreting information about a program's implementation and effectiveness. You collect this kind of information every day. To have a true picture of how well your program is doing, however, you need to follow procedures that are systematic, objective, and unbiased."
—Abt Associates, consulting firm

Close your eyes and imagine yourself conducting an evaluation of your art program's outcomes. What do you see?

Hopefully, you see yourself collecting information about program implementation and outcomes that you need to enhance your program and justify its funding—information that will help you to assess your program operations, determine the extent to which your goals are being met, and pinpoint the factors that facilitate or impede your program's success. You also see yourself using evaluation activities to initiate or enhance strong collaborative relationships with program stakeholders throughout your community, increasing their interest in and understanding of your program and the evaluation process.

If this *wasn't* what you saw when you closed your eyes (and if you're reaching for the extra-strength pain reliever), you're not alone. Many program managers view evaluation as a boring obligation that uses up valuable resources without providing much in return. But, however much we'd like to see arts programs receive ample funding without having to justify every move, funding sources demand accountability;

they want to know how their dollars are being used and what impact they are having. So, where does that leave you? Knowing that you need to conduct an evaluation and knowing *how* to do so are not the same thing. You need information about how to plan and implement an effective evaluation. This portion of the handbook aims to provide just that. It presents the benefits and challenges of program evaluation, lays out specific steps to take when evaluating your program, and provides a list of additional resources that can help you along the way. While this chapter is not intended to turn you into a professional program evaluator, it will help you to think about, understand, and appreciate what it takes to develop and implement an evaluation that will help you to answer the question, "Does my program work?"

Benefits and Challenges of a Well-Planned Outcome Evaluation

'While this chapter is not intended to turn you into a professional program evaluator, it will help you to think about, understand, and appreciate what it takes to develop and implement an evaluation that will help you to answer the question, "Does my program work?"'

A well-planned evaluation has the following characteristics:
- From beginning to end, it involves communication and collaboration among the key program partners, first to develop realistic expectations for what the program can achieve during the evaluation time period, and then to design and implement the evaluation.
- It is based on a program planning model that lays out the relationships among the targeted problem(s), program activities, and intended immediate, intermediate, and long-term effects (or outcomes).
- It addresses clearly articulated evaluation questions.
- It includes an upfront assessment of evaluation resources, including the feasibility of collecting the necessary data.
- It provides information about program implementation and operations, and program outcomes. It includes both process evaluation and outcome evaluation components. The process evaluation component collects information about program implementation and service delivery, which is needed to monitor and refine program activities. The outcome evaluation component collects the information needed to determine the program's effects on participating youth.
- It considers contextual factors that may affect the evaluation results.

While conducting an evaluation with these characteristics can be quite costly and challenging, it usually produces numerous benefits that, in the long run, outweigh the costs.

The benefits of a well-planned outcome evaluation

> *"An evaluation can be an important tool in improving the quality of a prevention program if it is integrated into the fabric of the program rather than added on after the fact."*
> — Lana Muraskin, evaluation consultant

A well-planned evaluation addresses clearly articulated evaluation questions.

The benefits of a well-planned evaluation include:

- **Program clarification**. Too often, program administrators develop and operate programs with vague assumptions about the types of outcomes the program will have and how it will accomplish them. A well-planned evaluation requires you to clarify your assumptions about the links between your target population, program activities, and expected immediate, intermediate, and long-term outcomes. This clarification process should help you and your key program stakeholders to keep your "eyes on the prize" throughout the program's duration by focusing your attention on the link between your program activities and its intended outcomes.
- **Program monitoring**. Tracking the number and type of activities you offer, the number and type of participants involved, and your activity-related expenses can help you monitor how close you are to achieving your service goals.
- **Program justification**. Promising results from a well-planned evaluation can be used to justify program expenditures, maintain the commitment of existing funding sources, and leverage additional resources from the community.
- **Program improvement**. The information that you collect will help you determine which program operation strategies are most effective and identify areas where improvement is needed.
- **Addition of knowledge to the field**. Information on program outcomes and "best practices" can be shared with your peers, other communities, government agencies, and other audiences in order to help promote effective practices and programs, as well as useful evaluation methods.

126

In general, any well-planned evaluation of a delinquency prevention or intervention program does the entire field a favor. As the competition for crime reduction funding grows, so does the need to provide "hard" evidence that prevention and alternative intervention strategies help reduce crime. Without this evidence, these strategies are likely to be the first crime-reduction strategies cut from national, state, and local budgets.

The challenges of evaluation

> *"Despite their value, evaluations are not always welcomed. Because they carry risks and use scarce resources, and because staff may be unsure how to conduct them, evaluations are often a low priority for programs . . . By understanding the potential difficulties before designing an evaluation, however, it is possible to avoid some of those risks or to minimize their effects."*
> —Lana Muraskin, evaluation consultant

Few arts-based prevention and intervention programs have undertaken well-planned outcome evaluations, in part because the challenges to planning a successful evaluation can seem overwhelming and often appear to outweigh the benefits. These challenges tend to fall into the four categories described below.

Competition for limited resources. The cost of evaluation varied based on the evaluation questions you are asking. A rigorous evaluation can be quite costly but can provide extremely valuable information. The sidebar on page 136, "Making Do With What You Have," looks at the cost of conducting evaluations.

Fear of "negative" evaluation results. Not all evaluation findings indicate success. When conducting an evaluation, you will need to be prepared to deal with "negative" evaluation findings. If the results conflict with your first-hand knowledge of the program, you may want to re-examine the design and implementation of the evaluation for flaws that could have affected the evaluation results. For example, an evaluation of the DARE drug prevention program, which targets fifth and sixth graders, indicated that the program had not made significant changes in drug use among its target population. As discussed by researchers at the Urban Institute, "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 . . . " Evaluation questions must address outcomes that the program is likely to affect and that are likely to occur within the time frame of the study.

'Evaluation questions must address outcomes that the program is likely to affect and that are likely to occur within the time frame of the study.'

127

If, however, you decide that the evaluation design and evaluation questions were appropriate and the negative evaluation results are accurate, you may need to modify your program to improve its effectiveness or, in extreme cases, reallocate resources to more promising prevention and intervention efforts. Programs that are committed to identifying and overcoming their own shortcomings are most likely to succeed in the long run.

Methodological paralysis. Formal evaluation methods are foreign to most program personnel and, at times, appear to be too complex for the layperson. Don't let the fear of the unknown paralyze you. Keep in mind that help is available. In addition to this guide, hundreds of resources are out there to assist you in completing a meaningful evaluation, including information and resource clearinghouses, publications, on-line help, and outside consultants. A list of useful resources is provided later on in this chapter. Also keep in mind that your evaluation doesn't have to be perfect. In fact, few (if any) evaluations are. Even professional evaluators run into problems along the way. Despite inevitable setbacks, most well-planned evaluations are able to meet the information needs of program stakeholders.

Difficulties of collaboration. Evaluating prevention and intervention programs requires the collaboration of some of the busiest and more over-extended individuals in your community. Program partners and outside consultants must devote the time needed to design and implement an evaluation that will produce accurate and useful results.

While these challenges require careful consideration *before* you embark on your evaluation, they should not deter you from beginning a well-planned evaluation. Once you have completed the first three steps described in the next section of this chapter, you will be able to determine what a well-planned evaluation can do for you *and* how feasible it would be for you to conduct one. With this information, you then will be ready to decide whether to evaluate your program and what level of resources you will need to allocate to ensure a successful evaluation.

'Programs that are committed to identifying and overcoming their own shortcomings are most likely to succeed in the long run.'

A Step-by-Step Approach for Evaluating Your Arts Program's Outcomes

Follow these steps to conduct a well-planned evaluation of your arts-based program:

1. Prepare your planning model.
2. Develop your evaluation questions.
3. Assess whether an evaluation is feasible.
4. Plan your evaluation.
5. Identify data collection instruments.
6. Collect data.
7. Analyze data.
8. Present findings and disseminate lessons learned.
9. Use evaluation findings.
10. Think about conducting a follow-up.

The remainder of this section describes each of these steps in detail.

Step 1: Prepare your planning model

The first step in developing your program evaluation is to clarify your assumptions about the specific changes the program is intended to cause within the target population, how it will achieve these changes, and how you will know when these changes have occurred. The planning model will help you to accomplish this step.

You may recall from the Program Planning chapter that, in general, a planning model shows the causal links among a targeted problem, the program activities designed to address the problem, and the immediate, intermediate, and long-term outcomes achieved by the program. Comprehensive planning models also show the resources that are allocated to support the program activities—such as staff and collaborative relationships—and the environmental factors that may affect program implementation and outcomes, such as other prevention programs or socio-economic changes in the community.

'A planning model shows the causal links among a targeted problem, the program activities designed to address the problem, and the immediate, intermediate, and long-term outcomes achieved by the program.'
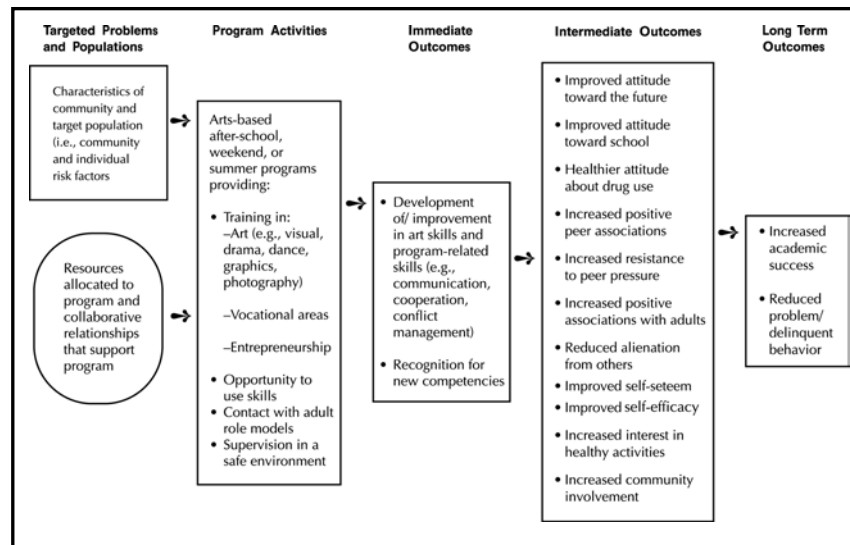
'Including immediate and intermediate outcomes in your planning model is particularly important, because long-term outcomes may not be detectable until after your evaluation is completed. If you were to measure only long-term outcomes, you might assume that the program has had no impact on the youth, while, in fact, it has been making incremental progress toward its goals.'

At minimum, developing a planning model such as the YouthARTS planning model in **Table 1** (above; repeated from page 27) involves describing the following components:

- **Targeted problems and populations**. What specific problems (low school achievement? high juvenile crime rates?) is your program designed to address among which populations? The clearer your definition of the targeted problems and populations, the easier it will be to determine if your program addresses them. As described in the Program Planning chapter, you should prepare a problem statement that defines this component, as well as those that follow. (Each YouthARTS site wrote a problem statement; see pages 20.)

- **Program activities**. What types of activities does your program provide for which youth? When and with what intensity/duration are they provided? Who provides these services? Where are they provided? Are there any referral services or follow-up activities for program participants? (See the Program Planning chapter for detailed information on developing program activities.)

- **Expected outcomes**. What do you expect your program to accomplish in the long run? What smaller (or interim) changes will lead up to these long-term outcomes? Your planning model should include descriptions of your expected immediate, intermediate, and long-term outcomes and show that each type of outcome leads to the next. Including immediate and intermediate outcomes is particularly important, because long-term outcomes, or impacts, may not be expected or detectable until long after your evaluation is completed. If you were to measure only long-term outcomes, you might conclude that the program has had no impact on the youth, while, in fact, it has been making incremental progress toward its long-term goals.

While you won't include all of the program information in the graphic depiction of your planning model, you should document it and update it as needed. Keeping a comprehensive record of program implementation and operations (as well as noting changes in your target problem/population or expected outcomes) is a critical component of program monitoring and evaluation.

Once you have completed your planning model, you should review it to make sure that your assumptions about your program's expected outcomes are realistic, given that outside factors such as socioeconomic conditions and family relations are likely to influence youths' attitudes and behaviors. Do all of your program activities seem to lead to your expected outcomes? Does anything seem superfluous? Is anything missing?

The process of developing and/or reviewing the planning model provides an ideal opportunity for all program partners, including youth from the community, to share their assumptions about and perceptions of the program and to ensure that their expectations are realistic given the scope of the program and the nature of the targeted problem. It can help the stakeholders to reach a consensus about the program's immediate, intermediate, and long-term goals and how it expects to achieve them. Partner "buy-in" to the planning model will help ensure that everyone agrees on the program's "measures" of success. You can convene all program partners to develop the planning model, or you can develop a draft model and then distribute it for review by other partners.

Once you have completed, reviewed, revised, and achieved stakeholder consensus on your planning model, you are ready to "operationalize" it. That is, you can begin identifying potential measures, or indicators, of each planning model component and

Many factors must be considered in order to develop realistic expectations for an arts-based delinquency prevention or intervention program. These factors include the characteristics of the target population, the types of risk factors addressed by the program, the length and intensity of the program, attendance and participation rates, and a range of outside factors—such as gang activity in the youths' community—that might influence the participants' attitudes and behaviors.

131

determine how the relevant data can be collected. For example, in the YouthARTS planning model, decreased juvenile delinquency is an expected long-term outcome. An indicator that the program has achieved this outcome is a decrease in the number of times program participants are referred to the juvenile courts. Data on this indicator could be gathered from juvenile court records. Another indicator of decreased delinquency is a decrease in self-reported delinquency, which could be measured through youth surveys or interviews. **Table 2**, a portion of an evaluation data map, provides a list of the YouthARTS planning model components and their respective indicators and data sources. Once you have begun identifying indicators and data sources, you can create a similar data map for your evaluation.

Table 2: Partial YouthARTS Evaluation Data Map

Research Question and Logic Model Component Indicators	Definition of Indicator	Data Source
What are the program attendance and completion rates?		
Program attendance	# of absences from program	Program attendance records maintained by artist/counselor
To what extent do youth benefit from participation in program activities?		
Improved art skills	Exhibits increased art knowledge and skill	Art Knowledge Quiz Participant Skill Assessment
Improved teamwork	Behaves in a cooperative way in a group	Participant Skill Assessment
Improved work skills	Works on a task from beginning to end	Participant Skill Assessment
Improved participation	Actively participates in class/sessions	Participant Skill Assessment
Improved communication skills	Communicates effectively with adults Communicates effectively with peers	Participant Skill Assessment Participant Skill Assessment
Production of art	Completed assigned projects/produced art	Participant Skill Assessment
Recognition of effort	Received public recognition for effort	Participant Skill Assessment
To what extent do participant and comparison groups show changes in risk factors at the completion of the program?		
School attitude and value	How often are school value statements true	YOCS (6-9)
Peer influence/Decision making	Frequency of participation in harmful activities with friends	YOCS (21-25)
Self-esteem	Agreement with statements about pride, failure, confidence, etc.	YOCS (54-58)
Referrals to Juvenile Court	Total number of referrals before program period Number of referrals during program period	Referral Form Exit Form
Academic GPA (by semester)	GPA in core academic courses per term	Academic Data Form
School attendance (by semester)	Number of excused absences in semester Number of unexcused absences	Academic Data Form

Step 2: Develop your evaluation questions

When you look at your completed planning model, you should be able to identify numerous questions that an evaluation could answer. For example, an evaluation could determine whether the program actually addresses the identified problem or

need, whether it actually served members of the target population, whether it was implemented as planned, or whether it achieved its expected outcomes. To keep the scope of your evaluation manageable, you will need to rate your evaluation questions in order of priority based on your information needs, the needs and requirements of your funders and other audiences, and your time constraints. One way to do this is to imagine that you will present your evaluation results to an important audience. Ask yourself the following questions:

What are the three most important points you would like to be able to make about your program? That is, on which parts of your planning model would you like to focus? Think about the people and organizations to whom you plan to present your evaluation results. Different audiences will be interested in different parts of your model. For example, representatives from arts agencies will be more interested in whether you can show that your program increased participants' art knowledge and creativity than will representatives from juvenile justice organizations. Figuring out who your audience includes will help you to prioritize your questions and, thus, keep the scope of your evaluation manageable.

When do you need to make this presentation? You will likely want to use findings from your evaluation in proposals for new or continuation grants, progress reports to existing funders, and similar fundraising and marketing efforts. Keeping a calendar of key dates by which findings are needed will help you to develop realistic evaluation questions. For example, if findings are needed for a continuation grant at the end of the program's first year, you will need to focus some of your efforts on gathering information about program implementation and about the immediate outcomes of your program. You will not be able to measure its long-term or overall effectiveness within the program's first year. Specifying a time frame will help you to determine which evaluation questions are realistic to answer.

Once you have thought through these questions, you should be ready to develop your evaluation questions. Remember that your questions should test some aspect of your planning model and be clear, specific, and realistic.

Step 3: Assess whether an evaluation is feasible

"Evaluability assessment is a systematic procedure for deciding whether program evaluation is justified, feasible, and likely to provide useful information."
—Adele Harrell, evaluation consultant

Assessing early on whether it is feasible for you to conduct an evaluation at all can save you a lot of time and energy and help ensure meaningful evaluation results. This process, sometimes called evaluability assessment, involves answering the following questions:

Do sufficient resources exist to support the evaluation? Before beginning your evaluation, you will need to ensure that you have adequate resources to support the entire evaluation effort, including funding, computer equipment and software, staff time (and commitment), and possibly an outside consultant.

How feasible is it for you to access existing data sources? Let's say that one of your main evaluation questions is whether your program affects academic performance. Your planning model and data map include grade point average as an indicator for

134

this outcome and school records as the primary data source for this indicator. How feasible is it for you to access these school records? Are they complete? Are they in a comprehensible format? How long will it take for you to obtain them? Who will you need to contact to do so? In order to answer these questions, you will need to contact the agencies and organizations from which you intend to collect data. After explaining your program and what the data collection effort will entail, you will need to obtain from them a written agreement that they will provide the requested data in the specified time period. You also will need to obtain sample records to assess their completeness, whether they are comprehensible, and how long it takes for the organization to provide them.

How feasible is it to collect *new* data? You also will need to assess the feasibility of collecting information from any new data sources, such as interviews with program staff, youth, and other stakeholders. You will need to determine how willing and how able key program stakeholders are to participate in and support your planned data collection activities. For example, you will need to find out if the parents of the youth you plan to include in the study are likely to give their children permission to participate in the study and how difficult it will be to maintain contact with and collect data from the youth during the evaluation period. Moreover, you should identify any existing data collection instruments that have been used in similar evaluation efforts (see below); using or adapting existing instruments (instead of developing new ones) will save you both time and effort.

Do similar evaluation efforts exist? Try to learn about similar evaluation efforts. You may be able to obtain permission to use or adapt data collection instruments that have been used for similar evaluations. Moreover, many final evaluation reports and publications provide important "lessons learned" about evaluating prevention programs, which can help you to replicate promising approaches and avoid common pitfalls. Similarly, contacting programs or outside consultants who have conducted similar evaluations and are willing to discuss their evaluation experiences may prove to be a very valuable endeavor.

After finding answers to these questions, you should be able to decide whether it is possible for you to conduct an evaluation that will answer your evaluation questions and meet the information needs of your intended evaluation audience. You also should be able to estimate the level of effort required to gather your evaluation data and determine whether you will need outside assistance.

'You will need to determine how willing and how able key program stakeholders are to participate in and support your planned data collection activities.'

135

If you decide that it is not possible for you to conduct a well-planned outcome evaluation that will answer all of your evaluation questions, don't despair. The section "Making Do With What You Have," (below) can help you to determine the types of evaluation activities you can successfully complete given the amount of money available for you to spend on evaluation.

(While the remaining evaluation steps described in this chapter are designed for a thorough outcome evaluation, most contain information that you can use to conduct less extensive evaluation efforts. For example, Step 5 provides tips for developing surveys—such as simple program satisfaction, or feedback, surveys. So, *read on*!)

At minimum, determine the number of youth served, the services provided or products developed, and information about the characteristics of participating youth (for example, age, sex, race, and juvenile court status). Take the time to find out how satisfied participants were with the program. And don't forget to keep complete attendance records.

Making Do With What You Have

KRA Corporation, a management consulting firm, recently developed a very useful evaluation manual—The Program Manager's Guide to Evaluation—under a contract with the Administration on Children, Youth and Families (see Other Resources, page 174). The manual provides general cost information for different types of evaluation activities. We modified the following excerpts from the manual to address specific issues that might be of concern to you.

The cost of conducting your evaluation will depend on a variety of factors, including which aspects of your program you decide to evaluate; the size of your program (that is, the number of staff members, youth, components, and services); the number and type of outcomes you want to assess; who conducts the evaluation (for example, program staff, a local university, or an independent consultant); and your agency's available evaluation-related resources. Costs also vary based on economic differences in communities and geographic locations.

Sometimes funders will establish a specific amount of grant money to be set aside for an evaluation. The amount usually ranges from 15 to 20 percent of the total funds allocated for the program. If the amount of money earmarked for an evaluation is not specified by the agency, you may want to talk to other program managers in your community who have conducted evaluations. They may be able to tell you how much their evaluations cost and whether they were satisfied with what they got for their money.

Although a dollar amount cannot be specified, it is possible to describe the kinds of information you can obtain from evaluations at different cost levels:

Lowest cost evaluation activities: If you spend only a minimal amount of money, you will be able to determine the number of participants served, services provided or products developed, and information about the characteristics of participating youth (for example, age, sex, race, and juvenile court status). You also may be able to find out how satisfied youth were with the program. This information may be useful for progress reports, continuation grant applications, or publicity efforts. It also can help you to determine whether you are reaching your desired youth population. Finally, it provides the foundation for a more extensive evaluation. **At minimum, all programs should collect these types of information—especially complete attendance records—on an ongoing basis.**

Low-moderate cost evaluation activities: If you increase your evaluation budget slightly, you will also be able to determine whether your participants' knowledge, attitudes, or behaviors have changed over the course of the program using a pre- and post-program assessment. (See the discussion on outcome evaluation under Step 4 to learn more about this type of assessment.) In addition, you will be able to collect in-depth information about the *process* of implementing your program. As discussed under Step 4, process evaluations provide valuable information needed to ensure that the target population is being reached, that the provision of key services is running smoothly, and that any program weaknesses are addressed—three key accountability issues of concern to program funders.

Moderate-high cost evaluation activities: Adding more money to your evaluation budget will allow you to use control or comparison groups to determine whether short-term changes in participants' knowledge, attitudes, and behaviors were caused by your program—that is, whether the changes were *outcomes* of your art program. (See the discussion on outcome evaluation design under step 4 for information about control and comparison groups and their role in outcome evaluations.) You also may be able to determine whether modifications to your program activities have affected program outcomes. (Let's say, for example, that you reduce the number of artists in a classroom, and, subsequently, the youths' scores on an art knowledge test drop

'Process evaluations provide valuable information needed to ensure that the target population is being reached, that the provision of key services is running smoothly, and that any program weaknesses are addressed.'

137

substantially. If nothing else has changed in the program or in the youths' lives, you may conclude that the reduction in artists negatively impacted the youths' ability to acquire and demonstrate new art knowledge.)

Highest cost evaluation activities: At the highest cost level, you will be able to obtain all of the information available at the other cost levels *and* determine your program's lasting outcomes and impacts—that is, the effects that your program is expected to have on program participants after they have left the program. This type of evaluation is particularly costly because it requires tracking—or maintaining contact with—program participants (and possibly control or comparison youth) after they have left the program. While expensive, this type of evaluation is important because it determines whether the changes in knowledge, attitudes, or behaviors that your participants experienced initially were maintained over time.

As illustrated in these descriptions, the more money you are able to invest in an evaluation, the more useful the information that you will obtain about your program's effectiveness will be, and the more useful these results will be in helping you refine and justify your program.

Step 4: Plan your evaluation

At this point, you've already accomplished much of the leg work required to conduct a sound outcome evaluation. You know the types of questions you would like to answer, the types of data you'll need to collect to answer them, and the sources of these data. You've established your desired evaluation time frame and assessed the resources available to support your evaluation effort. The next step is to plan your evaluation effort. Ideally, this step involves preparing a detailed written document that can be circulated to and reviewed by the key players involved in the evaluation. Reviewing evaluation plans can lead program partners to provide additional information about the program and their expectations for the evaluation, which can help guide the evaluation in the right direction.

A comprehensive written evaluation plan includes the following components:
- background and purpose of the evaluation
- outcome evaluation design
- process evaluation plan
- data collection strategy (data map and data collection instruments)
- data analysis plan
- draft outline for the final evaluation report
- timeline.

If it is not possible for you to produce a detailed written plan, you should at least try to address each of these components in outline format and discuss them with your program partners. Let's examine each component in more detail.

Background and purpose: This first section of a written evaluation plan provides important contextual information. It presents a brief program description, the program planning model, the evaluation questions, and an explanation of how the evaluation results will be used.

> *"Choosing a strong evaluation design is important, because your findings may be invalid if someone can easily find another explanation for outcomes you attribute to your program. A good design will increase confidence that clients are changing for the better and that the program itself is producing these results."*
> — Abt Associates, consulting firm

Outcome evaluation design: An evaluation design specifies when, from whom, and about whom you will collect outcome evaluation data. It determines how you will measure changes in program participants and how you will prove that these changes resulted from your program.

The ideal outcome evaluation design is an **experimental design**, which involves collecting data from youth randomly assigned to treatment groups (youth from the target population who receive program services) and control groups (youth from the target population who do not receive program services). Random assignment of youth to the two groups–maybe by flipping a coin–ensures that the groups are comparable at the start of the evaluation and, consequently, that any differences between the two groups' outcomes at the end of the evaluation period can be attributed to the program.

To develop an experimental design, take the following steps:
- Select a large pool of youth from your target population.
- Randomly assign the youth to treatment and to control groups.
- Invite the youth in the treatment group to participate in the program and invite the youth in the control group to participate in a study. If possible, provide incentives such as cash, gift certificates, or time off probation for control group members to complete the required surveys, interviews, and so forth.

When done correctly, random assignment usually ensures comparability between treatment and control groups. However, because participation in the two groups is most often voluntary and some youth will choose *not* to participate in the study, you will need to ensure that the *final* treatment and control groups are comparable. You will need to gather demographic, attitudinal, and behavioral information from the youth who have agreed to participate in each group at the beginning of the program period in order to determine whether the two groups are similar on key characteristics—such as sex, age, and level of court involvement—that may influence program outcomes. If you find that the two groups are very different on key characteristics, you may decide to start over or adjust your data analysis plan to take these differences into account.

While experimental designs provide the strongest evidence of a program's effects, they are not always feasible or desirable for several reasons. First, program staff or other program partners may feel that randomly assigning potential program participants to treatment and control groups is unethical because it deprives control group members who could benefit from the program from receiving its services. They may decide that admitting the neediest or most interested candidates to the program is more important than achieving the most rigorous evaluation design. Second, the pool of program candidates may be too small to divide into treatment and control groups. Third, using a control group requires considerable effort. The process of randomly assigning youth to the treatment and control groups requires careful planning, and maintaining contact with (and collecting data from) control group members during the evaluation period may require considerable time and effort, even if you have elected to use incentives.

The next best thing to an experimental design is a **quasi-experimental design**. This type of design involves first selecting a treatment group and then selecting a comparison group of youth from the target population who are as similar as possible

'The process of randomly assigning youth to the treatment and control groups requires careful planning, and maintaining contact with (and collecting data from) control group members during the evaluation period may require considerable time and effort, even if you have elected to use incentives.'

to the youth in the treatment group on important characteristics (such as age, race, grade level, delinquent behavior) but who have not and will not participate in the program before or during the evaluation period. Using a quasi-experimental design involves the following steps:

- Select a treatment group.
- Select a comparison group of youth who are comparable to (that is, who match) the youth in the treatment group on the characteristics that you believe are most likely to affect program outcomes. (For example, if you believe that age will have a large effect on program outcomes and half of the youth in your treatment group are 16 years old and half are 12 years old, you will need to make sure that your comparison group reflects a similar split between youth ages 16 and 12 years old.) Matching youth on key characteristics can be quite tricky, particularly if your target population and treatment group are diverse and you have identified a number of characteristics that may affect program outcomes. In fact, you may decide that you will need outside help to complete this step.
- Invite the youth in the treatment group to participate in the program and invite the youth in the comparison group to participate in a study. If possible, provide incentives such as cash, gift certificates, or time off probation for comparison group members to complete the required surveys, interviews, and so forth.

Unlike the experimental design, the quasi-experimental design does not involve random assignment; thus, it is less certain that you will begin the evaluation with comparable groups. Despite this potential drawback, this type of design is the best alternative to the experimental design and, when the final comparison and treatment groups are carefully matched on key demographic characteristics, can produce strong evaluation findings.

In order to demonstrate change, you will need to collect outcome evaluation data on program participants and control or comparison youth at the beginning and end of the program period. Collecting pre- and post-program data will allow you to assess any changes that have occurred in program participants' knowledge, attitudes, and behaviors over the course of the program. Comparing these changes to the changes in the control or comparison group members' knowledge, attitudes, and behaviors will allow you to determine if the program contributed to these changes.

A Non-Experimental Design: Assessing Short-Term Changes

While using an experimental or quasi-experimental design is recommended for most program evaluations, it is not always necessary. Let's say, for example, that you want to determine the immediate effects of a three-day conflict resolution training program on participants' knowledge of conflict resolution techniques. By testing the program participants at the beginning and end of the training program, you can easily determine whether their knowledge of these techniques has changed during the program. Moreover, because it is unlikely that any outside factors would have caused this particular change over such a short time period, you can be fairly confident, without using a control/comparison group, that any changes resulted from your program. **Thus, pre- and post-program assessments of program participants** are best used to assess short-term changes when few alternative explanations for your findings exist.

'Pre- and post-program assessments of program participants are best used to assess short-term changes when few alternative explanations for your findings exist.'

Process evaluation plan: Well-planned outcome evaluations also include process evaluation activities that answer questions about how the program was intended to operate and how it actually operates on a daily basis. They provide valuable information about factors that facilitate and impede program implementation, promising program strategies and areas that need improvement, as well as the contextual information needed to interpret changes in participants' knowledge, attitudes, and behaviors. Process evaluation topics fall into five main categories:

- **background information**: history and purpose of program, target population and community characteristics, underlying program theories (planning model), and future plans for program
- **organization, staffing, and interagency collaboration**: program administration; hiring, training, and roles and responsibilities of program staff; and collaborative arrangements with other agencies
- **program access**: methods used to recruit program participants from target population and factors that decrease and increase access to program

- **program activities and services**: regular program activities, special activities (such as field trips), and referral services; intensity and duration; attendance and participation rates; and changes in program activities
- **budget and costs** (optional): funding sources and expenditures

Your evaluation plan should document the topics about which you plan to collect data. At minimum, you should be able to present the planning model; describe the program activities, target population, and intensity and duration of program activities; and provide attendance and participation rates. This information is needed to provide a context for the outcome evaluation findings.

Data collection strategy: This section of the evaluation plan describes how you will collect the data needed to answer the evaluation questions, using a **data map** and **data collection instruments**. As discussed in Step 2 and Step 3, a data map shows how you plan to answer your evaluation questions. Specifically, it is a table that links the planning model and evaluation questions to the indicators and data sources. (See **Table 2** on page 132.) You should describe all of the data sources included in your data map—surveys, intake forms, school records, interviews, and so forth. If you are developing or adapting program-specific data collection forms or surveys, you should describe their contents and include copies of these instruments in an appendix.

Data collection plan: This plan describes the "who, what, when, and where" of data collection. That is, it tells *who* will be responsible for collecting data from the sources included in the data map and describes any training the data collectors will receive to prepare them for this task. It also describes *how* and *when* the data collection instruments will be administered to the appropriate subjects. For example, you might ask participants to complete written surveys at the beginning of the first art session, read written surveys aloud to the comparison group in a classroom after school on the first day of the program, and interview program staff at the end of the program using an interview guide. The plan should contain explicit survey administration instructions and describe strategies for overcoming potential difficulties, such as language barriers. (Step 6 describes these topics in more detail.)

Analysis plans: These plans describe how the collected data will be analyzed and how these analyses will be used to answer the evaluation questions. You can describe the methods that you intend to use to analyze your data in text and/or include them in a column of your data map. You also should describe any anticipated constraints on

Even if you decide not to conduct an outcome evaluation, you might still conduct a process evaluation. Process evaluations can provide information needed to ensure that the target population is being reached, that key services are being provided, and that any program weaknesses are addressed. Process evaluations also can help assure program funders that the program is operating as planned and that the administrator and staff are holding themselves accountable for the success of the program.

143

your analyses. For example, small program sizes limit the extent to which small changes in participant outcomes can be assessed—that is, only large changes can be identified. (Step 7 describes the data analysis methods that you are likely to use.)

Draft outline for final evaluation report: It's a good idea to include a draft outline for your future evaluation product in your evaluation plan, whether it will be an evaluation report, briefing, article, or other type of written or oral presentation. Laying out what you plan to say about your program and evaluation effort will help you to "stick to the point." A sample evaluation report outline appears in **Table 3**.

Table 3: Evaluation Report Outline

EXECUTIVE SUMMARY

I.   INTRODUCTION

    1. Program Description and Logic Model
    2. Evaluation Purpose and Overview
    3. Report Layout

II. EVALUATION METHODOLOGY

    1. Evaluation Questions
    2. Evaluation Designs
    3. Data Collection Strategy (Methods, Instruments, Plan)
    4. Data Analysis Plan
    5. Evaluation Constraints

III. PROCESS EVALUATION FINDINGS

IV. OUTCOME EVALUATION FINDINGS

V.   CONCLUSIONS AND RECOMMENDATIONS

APPENDIX

Timeline: Timelines are critical components of an evaluation plan. Not only do they help you to keep track of when different tasks must be accomplished, they also help you to assess the level of time and effort that will be required from program staff (and possibly an outside evaluator) at different points throughout the evaluation period. Your timeline should include each of the evaluation steps discussed in this chapter, as well as the specific tasks that will occur within these steps, including conducting meetings, distributing draft items—for example, planning models or data maps—to program stakeholders for review, and developing data bases. A portion of a sample timeline format is provided in **Table 4**.

Table 4: Partial Evaluation Timeline (Sample)

TASK	TIME PERIOD
1. Prepare and Operationalize Logic Model	
1.1　Conduct kick-off meeting with stakeholders	January 3
1.2　Revise logic model based on stakeholder input	January 4-9
1.3　Contact data sources and prepare data map	January 4-17
1.4　Distribute revised logic model and data map	January 18

Step 5: Identify data collection instruments

When selecting or preparing your data collection instruments, you will need to consider which type of instrument best suits your needs:

- **Written surveys or questionnaires** are often used to gather large amounts of information from many people (for example, to assess treatment and comparison group members' knowledge, attitudes, and behaviors; treatment group members' thoughts about the program; program staff's perceptions of program implementation and operations, staff training, and program outcomes; treatment group family members' perceptions of program outcomes; and other program stakeholders' thoughts about the program).

- **Interviews (by phone or in person)** are often used to gather qualitative information about program implementation and operations and program outcomes from program participants, staff, and other partners on an individual basis.
- **Focus group interviews** can be used to gather qualitative information about program implementations and operations and program outcomes from small groups of youth, program staff, or other partners.
- **Observation forms or checklists** can be used to record information about social interactions or group processes gathered by observing program sessions, classrooms, or treatment and comparison group members' homes.
- **Program implementation/operations reporting forms** are used by program staff to document information about program implementation activities and daily program operations, such as duration of activities or attendance levels.
- **Extraction forms** are used to gather data from existing records, such as court histories and school records.

Evaluators frequently use several types of data collection instruments depending on the kinds of data that they need to collect, the availability of data sources, and the characteristics of the subjects from whom data are being collected (for example, oral interviews or focus groups would be necessary to collect information from participants who have difficulty reading).

Fortunately, numerous instruments exist that you can use or adapt to collect your data. We recommend that you use existing instruments, if possible, because it will save you the considerable time and effort needed to develop new data collection instruments and usually will save you the trouble of proving that your new instruments are reliable—that is, that they actually collect the data that you need. We have included in **Appendix 24** some of the data collection instruments used in the YouthARTS outcome evaluation—art knowledge surveys, a participant skills assessment form, a youth attitude and behavior survey, an academic data form, and a court referral/exit form—as well as several other sets of process and outcome evaluation instruments, which you can use or adapt to meet your specific data collection needs. In the "Other Resources" section of this chapter is a list of resources that you can use to locate additional existing instruments, and appropriate resources to consult if you should decide to develop your own instruments. Focus group questions are included in Appendix 25.

Appendix 24:

Data collection instruments

Appendix 25:

Focus group questions

'Fortunately, numerous instruments exist that you can use or adapt to collect your data. We recommend that you use existing instruments, if possible, because it will save you considerable time and effort.'

Tips for Survey or Test Questions

Here are general points to keep in mind when you are developing surveys or tests that will be administered to program stakeholders.

- Use clear, simple language that all respondents will be able to comprehend.
- Make sure the questions ask what you want to know.
- Each question should ask about only one thing.
- Avoid generalizations; each question should be specific.
- Do not use a leading question (that is, one that suggests that there is a preferred response).
- Make sure respondents understand what you are asking.
- Make sure respondents are familiar with the topic you are asking about.
- Identify whether the respondent should mark one choice or all choices that apply.
- Response choices should be comprehensive (include "not applicable" if necessary). and exclusive (choices should not overlap).

Regardless of whether you are using or adapting existing instruments or developing brand new ones, you will need to test them to ensure that they work with your respondents *before* you use them to evaluate your program. A *pilot test* for a youth survey, for example, involves administering the survey to a group of youth from your target population and then examining their responses and interviewing them to determine if the survey was easy to complete and if their responses to the survey questions were accurate. If any of the survey responses surprise you, you may want to ask the youth if they misunderstood the particular questions or if some other factor influenced their responses. You will need to revise questions that seemed to mislead the youth, produced little variation in response, or produced results that differed substantially from those expected. If the survey takes much longer than expected to complete or is too difficult for the youth, you may need to shorten it or consider replacing it with an oral survey or interview.

'If you fail to administer
pre- and post-program
surveys to respondents in
the appropriate time period,
you risk losing valuable
information about the program's
impact on the items measured.'

Step 6: Collect data

Your data collection plan (within your overall evaluation plan) should specify the "who, what, when, and where" of data collection. It is critical that you follow this plan to administer the data collection instruments to the appropriate respondents in a systematic fashion during the designated time period. If you fail to administer pre- and post-program surveys to respondents in the appropriate time period, you risk losing valuable information about the program's impact on the items measured. For example, if pre-program art knowledge surveys are administered several weeks into the program, the youth already may have gained new art knowledge that you will have missed measuring.

Before collecting your program evaluation data, you should complete the following steps:

Develop a data collection timeline based on your overall evaluation timeline. You should collaborate with all of the individuals who will be involved in the data collection process to develop this time line, ensuring that they will be willing and able to administer the appropriate instruments at the appropriate time.

Clarify the roles and responsibilities of all individuals who will be involved in this process. If more than one individual—artists, probation officers, outside evaluators—will be involved in administering surveys, conducting interviews or focus groups, or observing program activities, you should develop detailed protocols and/or provide detailed training to ensure that everyone is doing these things the same way. Differences in survey administration procedures—for example, reading the survey questions aloud instead of telling the respondents to read the survey questions silently to themselves—may cause differences in survey responses, which may decrease the strength of the evaluation findings.

Obtain permission from parents to gather information from their children. You will need to prepare a *written* informed consent form to be signed by the parents (or other legal guardians) of all youth included in the study. This form should describe the purpose of the study and the types of information that are being collected, promise confidentiality, and ask for a parent's signature. It should be written in a language that each parent understands. A sample consent form is included in **Appendix 7**.

Develop procedures to ensure confidentiality. You will be collecting data of a sensitive nature; therefore, it is very important to promise respondents that the information they

provide will be confidential (that is, no one but the person collecting the data will know how they responded to the survey or interview questions). One way to ensure confidentiality is to assign each respondent a unique number and place that number on the appropriate data collection instruments at the beginning of the study. This process will enable the data collector to match up the pre- and post-program surveys and other data collection instruments for each youth without using their names. The list that links respondent names and numbers should be destroyed only after the evaluation efforts, including follow-up evaluations, have been completed.

Detailed instructions for completing these steps and administering data collection instruments are included in the YouthARTS data collection implementation guide, which can be found in **Appendix 24**.

Evaluation Data Types

Evaluation data fall into two categories:

Quantitative data include pieces of information that can be expressed in numerical terms, counted, or compared on a scale. Examples include reading test scores, the number of people who responded positively to an interview question, the number of female program participants, and the average age of participating youth.

Qualitative data include pieces of information that are difficult to measure, count, or express in numerical terms. Examples include people's perceptions about the fairness of a program requirement, descriptions of program activities, and descriptions of problems that participating youth encountered. Qualitative data often provide the context needed to interpret quantitative findings.

In the following example, this sentence provides a quantitative finding: "By the end of the program period, approximately 25 percent of program participants had stopped attending program activities." This sentence provides the qualitative data needed to interpret that finding: "Program staff believe that this drop in attendance was a direct result of the new discipline policies mandated by the program manager."

Try not to underestimate the number of staff and the amount of time required to collect data. For example, if you decide to administer pre- and post-program surveys to participating youth and your program serves more than one group of youth at a time, you will need at least two people to administer the surveys at the appropriate times, review the surveys for completeness, and ensure that the pre- and post-program surveys are clearly labeled.

Step 7: Analyze data

You do not have to be a statistician to analyze quantitative data. However, you do need to be familiar with some basic mathematics (such as calculating averages and percentages). This section is designed to walk you through some of the basic methods that you will need to use to analyze your outcome data. It is *not* designed to teach you all of the ins and outs of statistical analysis—for that, you will need to refer to a statistics textbook or enroll in a statistics course at your local university. (If, after reading the following information, you need further assistance or would like to move beyond the methods presented here, please refer to the data analysis resources in the "Other Resources" section of this chapter.)

Let's say that you have collected demographic data and art knowledge test scores for five treatment group members (Will, Sally, Vanessa, Peter, and Jessica) and five control group members (Steve, Rob, Gina, Rachael, and Danielle). **Table 5** presents these data.

'You do not have to be a statistician to analyze quantitative data. However, you do need to be familiar with some basic mathematics (such as calculating averages and percentages).'

Table 5: Sample Evaluation Data

| Group Members | Demographic Data | | | Art Knowledge Test Scores | |
	Age	Sex	Race	Pre-Program Test Score	Post-Program Test Score
TREATMENT GROUP					
Will	13	Male	African American	95%	95%
Sally	12	Female	White (not Hispanic)	85%	90%
Vanessa	12	Female	Hispanic	75%	95%
Peter	9	Male	White (not Hispanic)	85%	80%
Jessica	13	Female	Asian American	80%	90%
CONTROL GROUP					
Steve	12	Male	White (not Hispanic)	90%	95%
Rob	11	Male	African American	85%	90%
Gina	12	Female	Hispanic	95%	90%
Rachael	12	Female	White (not Hispanic)	75%	80%
Danielle	11	Female	African American	80%	85%

Two types of analyses can be conducted on these quantitative data to evaluate your program's outcomes: **descriptive analyses** and **comparative analyses**. Let's look at each of those in detail.

Descriptive analyses

Descriptive analyses, which can be used to *summarize* and then *compare* the characteristics of two or more groups, include frequencies, averages and ranges, and cross-tabulations. Calculate **frequencies** to determine the number and percentage of individuals with a certain characteristic. For example, you can determine the number or percentage of treatment group members in each age group, race, and sex.

In this case, one of five treatment group members (20 percent) is 9 years old; two out of five (40 percent) are 12 years old; and two out of five (40 percent) are 13 years old. Similarly, three out of five treatment group members (60 percent) are female; and two out of five (40 percent) are male. Finally, one out of five treatment group members (20 percent) is African American; two out of five (40 percent) are white, not of Hispanic origin; one out of five (20 percent) is Hispanic; and one out of five (20 percent) is Asian American.

You could also use frequencies to describe the results that the treatment group youth achieved on their art knowledge tests. For example, four out of five members of the treatment group (or 80 percent) received a pre-program test score of 80 percent or higher. Similarly, five out of five treatment group youth (100 percent) received a post-program test score of 80 percent or higher. All of these frequencies could also be calculated for the comparison group members.

Calculating **averages** and presenting **ranges**—the highest and lowest points—are also useful methods to summarize information for selected groups of youth. For example, you can calculate the average age of youth in the treatment and comparison groups and present the age range for each group. You may also decide to calculate the average pre-program test score or the average post-program test score for each group.

The average age of youth in the treatment group is calculated by adding the five ages (13 + 12 + 12 + 9 + 13 = 59) and then dividing the total by the number of youth in the group (59/5 = 11.8). Thus, the average age of youth in the treatment group is 11.8. Since the lowest age is 9 and the highest is 13, the age range is 9 to 13. Using the same methods, you can determine that the average age of the control group members is 11.6 and that the range for this group is 11 to 12.

The average pre-program test score for the treatment group is calculated by adding the five pre-program scores (95 [percent] + 85 + 75 + 85 + 80 = 420) and then dividing the total by the number of youth in the group (420/5 = 84 percent). Thus, the average pre-program score achieved by youth in the treatment group is 84 percent, and the range is 75 to 95 percent. Using the same method, you can determine that the average pre-program score achieved by the control group members is 85, and the range is 75 to 95 percent.

Cross-tabulation is used to determine the number or percentage of individuals with two selected characteristics—such as age and group status (that is, treatment group or control group). To compare the ages of the youth in the treatment group to the ages of the youth in the control group, you will need to create a table in which each age is assigned a column and each group status is assigned a row. (See **Table 6**.)

Table 6: Sample Cross-tabulation

Group Status	Age				
	9	10	11	12	13
Treatment Group	1	1	0	1	2
Control Group	0	0	2	3	0

While the *average* ages of the treatment and control groups are nearly identical (11.4 and 11.6, respectively), this cross-tabulation shows that the individual ages of the youth in the two groups show more variation. While the treatment group members are spread out across the entire age range (9 to 13), the control group members are concentrated within a smaller range (11 to 12).

Note: Usually, the characteristic that you are most interested in appears across the top of the table, and the other characteristic appears along the left side of the table. In our example in **Table 6**, we were interested in the age of the youth in the two groups, so age appeared across the top of the table. If we were more interested in determining the group status of youth in different age groups, we probably would have put group status across the top of the table and age down the left side.

Comparative Analyses

The second type of analyses to be used on quantitative data, comparative analyses, can be used to assess changes in the knowledge, attitudes, and behaviors of both treatment and control group members—that is, they can be used to assess program outcomes. This discussion focuses on two types of comparative analyses: (1) calculating and comparing change scores and (2) calculating and comparing the proportion of youth who show improvement in their scores.

Calculating and comparing change scores: Change scores are calculated by subtracting pre-program scores from post-program scores in order to measure the size (and determine the direction) of changes between pre- and post-program knowledge, attitudes, and behaviors (that is, program outcomes). For example, by calculating average change scores on art knowledge tests, you can determine the level of the program's effect on participants' level of art knowledge, if any. By comparing the average change scores of the treatment group to those of the control/comparison group, you can determine whether any changes noted were caused by the program. (If the treatment group demonstrates a larger change in the desired direction than does the control/comparison group, you will have evidence that your program works.)

In this case, calculating the treatment group's average art knowledge test change score involves two steps:

First, calculate each treatment group member's change score by subtracting his/her pre-program test score from his/her post-program test score. (See **Table 7**.) Note that some of the change scores may be negative numbers, as is Peter's (80 - 85 = -5 percentage points).

Table 7: Sample Change Scores Calculation

Treatment Group Member	Post-Program Test Score		Pre-Program Test Score		Change Score
Will	95	-	95	=	0
Sally	90	-	85	=	5
Vanessa	95	-	75	=	20
Peter	80	-	85	=	-5
Jessica	90	-	80	=	10
Total	**450**	**-**	**420**	**=**	**30**

153

Second, add the five youths' change scores (0 + 5 + 20 + -5 + 10 = 30), and then divide the total by the number of youth (30/5 = 6). Thus, the average change score for this group is 6 percentage points.

By completing the same two steps for the control group, you would find that its average change score is 3. Thus, the treatment group achieved a higher change score (6) than the control group did (3), and you can reasonably conclude that the art program achieved its expected outcome of increasing program participants' knowledge about the arts.

Calculating and comparing the proportion of youth who show improvement: Simply by comparing pre- and post-program scores for treatment and control groups and calculating the proportion of youth in each group who show improvement between the beginning and end of the program, you can determine whether your program has had an effect on the selected outcome measure. For example, if 50 percent of the youth in the treatment group obtain post-program art knowledge scores that are larger than their pre-program scores, and only 25 percent of the youth in the control group do so, you can conclude that your program has had a positive effect on art knowledge.

In this case, three out of five (60 percent) of the treatment group members showed improvement in their art knowledge test scores (that is, their post-program test scores were higher than their pre-program test scores); one out of five (20 percent) stayed the same; and one out of five (20 percent) had a declining score. (See **Table 8**.)

Table 8: Sample Calculation of Proportion of Youth Showing Improvement

Group Members	Pre-Program Test Score	Post-Program Test Score	Score Improved	Score Stayed the Same	Score Declined
Will	95	95		X	
Sally	85	90	X		
Vanessa	75	95	X		
Peter	85	80			X
Jessica	80	90	X		

At the same time, four out of five (or 80 percent) of the control group members showed improvement in their art knowledge test scores, and one out of five (20 percent) had a declining score. Thus, the percentage of control group members who showed improved scores was greater than the percentage of treatment group members who did so. Based on these results, you *cannot* conclude that the art program achieved its expected outcome of increasing program participants' knowledge about the arts.

Note: When calculating and comparing the proportion of youth who show improvement on a survey or test, such as an art knowledge test, you need to determine at what point differences between pre- and post-program scores are meaningful (that is, at what point they actually show improvement). For example, if a youth receives a 95 on a pre-program survey and a 96 or a 94 on the post-program survey, you may decide that a one-point difference isn't really meaningful—that is, it doesn't *really* show an improvement or decline in art knowledge. You may decide to focus only on differences of five or more points (or the number of points that would constitute a change in an academic letter grade). Thus, if three youth improved by only three points, and two youth improved by ten or more points, the proportion of youth showing (meaningful) improvement in art knowledge would be two out of five, or 40 percent.

Clearly, the two comparative analyses of art knowledge test scores yielded quite different results. The calculation and comparison of average change scores showed that the treatment group members achieved higher average change score than the control group did, indicating that the program *did* achieve its intended outcome of increasing art knowledge. However, the calculation and comparison of the proportion of youth who showed improvement showed that the percentage of youth who showed improvement was greater for the control group than for the treatment group, indicating that the program did *not* achieve its intended outcome of increasing art knowledge. Although they indicate opposite conclusions, both findings are accurate.

The reason that the findings of the two analyses differ is that one of the youth in the treatment group increased her test score by 20 percentage points over the course of the program and another increased her score by 10 percentage points, which dramatically increased the average change score for the treatment group. Because the other eight youth across both groups showed much smaller changes (usually 5 percentage points), these two youth are considered outliers (that is, extreme cases that differ substantially from the rest of the group). Such outliers are often removed from calculations of average change scores to ensure that the results reflect the level of change demonstrated by the majority of the group. (The modification to the data set

You should use the comparative analytical method that is most meaningful to you and your evaluation audience. If you are more interested in the level of change achieved by the group as a whole, you should calculate and compare average change scores for the treatment and comparison groups. If you are more interested in the proportion of youth in each group that showed any improvement, you should calculate and compare the proportion of youth who show improvement.

155

and the individual change scores for the outliers are discussed when the final results are presented.) To avoid this issue, the national evaluation of YouthARTS usually chose to calculate and compare the proportions of youth who showed improvement in various outcome areas, because these analyses are not affected by outliers.

The data analysis resources in the "Other Resources" section of this chapter describe the methods used to determine whether the changes that you have identified are *statistically significant* (that is, whether they were caused by your program rather than by chance). While conducting tests of statistical significance adds another level of rigor to your evaluation (and the more rigorous your evaluation, the more credible its findings), it is not necessary; common sense should tell you if the improvements you have identified are meaningful.

While numerous methods exist to analyze *qualitative* data—including some of the information gathered through interviews, focus groups, or artist journals—we recommend that you create simple tables that summarize different types of information for different respondent groups. You can then refer to these tables when interpreting your quantitative data and preparing your final evaluation product. (See **Table 9** for an example.)

Table 9: Sample Qualitative Data Table

Interview Questions	Program Manager	Artists	Youth	Probation Officers
What could we do differently to improve the project?	-involve the probation officer in selecting artist -link arts programming directly to probation programs	-small group work -flexible curriculum -consistency among artists and probation officers	-leave it the way it is -more field trips -classes an hour longer -more practice with the camera -bigger projects and more of them	-more chances for kids to continue on with art -select an art medium that energizes probation officers -involve the probation officers in selection of artists

Step 8: Present findings and disseminate lessons learned

You are now ready to share what you have learned from your program evaluation effort by presenting your findings in a written report, an executive summary, an oral briefing, or another type of presentation. In general, your presentation should describe the program and its planning model, the purpose and methodology of the evaluation, and the process and outcome evaluation findings. If you choose to write an evaluation report, you should include the data collection instruments and supporting documents in an appendix. (See the sample report outline in Table 3.) The type of presentation, its format, and its level of specificity should reflect the needs and preferences of its intended audience. If you are preparing the report for a government agency that is funding your program, you may want to contact that agency to find out if it has any specific requirements or preferences for evaluation products. Many government agencies expect evaluation reports to contain executive summaries, which highlight key findings, conclusions, and recommendations and help facilitate decision-making. You also may want to disseminate your findings and any lessons that you have learned to wider audiences through press releases, newspaper articles, the Internet, or other media.

Step 9: Use evaluation findings

Once the final evaluation reports and/or other products have been completed and distributed to program stakeholders and other key audiences, you may feel as though you can relax a bit. No doubt, you deserve a rest after all that you've accomplished. However, it should be a brief rest, because your work is not done. No evaluation effort is truly complete until its findings have been used.

In general, evaluation findings can be used to:
- fine-tune, expand, or curtail a program
- make management and administrative changes
- influence policy decisions
- add to existing knowledge about juvenile delinquency prevention
- undertake a new evaluation effort

No matter how informative an evaluation is, its worth lies in the extent to which the program and/or the field are able to use the information to improve existing programs, create new programs, replicate promising approaches, and/or conduct new research that will guide future programming efforts.

Before you disseminate any evaluation product or related informational document, make sure to ask all program stakeholders to review it for accuracy, clarity, and tone.

Step 10: Think about conducting follow-up

While basic pre- and post-program data will provide you with valuable information about the program's immediate and, perhaps, intermediate outcomes, you will need to collect follow-up data in order to assess your program's long-term effects on program participants. Evidence of long-term, positive program outcomes is a persuasive argument for continued or increased program funding. Despite the potential benefit of follow-up evaluation activities, few studies include these activities because they often require too great an effort. Maintaining contact with and collecting data from treatment and control/comparison group members after the initial evaluation period can be very difficult. In fact, if you decide to conduct a follow-up, you can expect the size of both your treatment and comparison groups to shrink substantially because some youth will choose not to participate and others will have moved to another location or transferred to a different school.

In order to decide whether to conduct a follow-up, you should consider the following questions:

- How useful would positive follow-up results be to your program?
- What would you expect the long-term outcomes of your program to be, given what you now know about your program and the target population?
- How difficult will it be for you or your data collector to track—contact and collect data from—members of the treatment and comparison groups? Do you think you could find enough of the youth to make it a worthwhile effort?
- How committed are program staff and others involved in the evaluation process to the follow-up evaluation effort?

If your answers to these questions are encouraging enough for you to conduct a follow-up evaluation, you then will need to decide what data to collect. While it is a good idea to re-administer some of the same instruments that you used during your original evaluation (to assess trends over time), you may want to include only those questions that focus on topics of particular interest. You also may choose to administer new instruments, such as interview and focus group protocols, that gather more qualitative information about program outcomes or focus on potential program outcomes that were not addressed in the original evaluation.

Where to Go for Evaluation Assistance

If you don't already have questions or concerns, you probably will once you begin planning your evaluation. Fortunately, numerous resources are available to help you plan and implement a program evaluation. This section briefly covers three types of resources: clearinghouses and resource centers, printed evaluation resources—guides, books, and forms—and evaluation consultants and technical assistance providers.

Choosing an Evaluation Consultant

At minimum, you will need to make sure that your evaluation consultant meets the following criteria:

- is knowledgeable about juvenile delinquency prevention and intervention programs, and possibly about arts-based approaches to prevention and intervention
- is interested in your evaluation questions
- is willing and able to commit to your evaluation time frame (which may change based on their advice)
- is able to communicate clearly both orally and on paper
- is experienced in conducting and managing a comprehensive program evaluation, collecting and analyzing evaluation data from sources similar to those you have identified in your data map, and producing user-friendly reports
- is committed to collaborating with you and investing the time needed not only to assist with the evaluation but also to enhance your knowledge about and skills in conducting your own evaluation

159

Evaluation consultants and technical assistance providers

If you decide that you need assistance to conduct your evaluation, consider the following potential sources of technical assistance:

- **federal, state, and local government agencies** such as the U.S. Office of Juvenile Justice and Delinquency Prevention, the U.S. Department of Education, the Center for Substance Abuse Prevention of the U.S. Department of Health and Human Services, state health and human service agencies, and the offices of state juvenile justice specialists (individuals responsible for coordinating juvenile justice grants at the state level)
- **local universities or colleges**, particularly their public policy, social work, criminology/criminal justice, education, sociology, and statistics departments
- **research firms and management consulting companies** such as Caliber Associates, Abt Associates, the Urban Institute, RAND Corporation, Research Triangle Institute, and Developmental Research and Programs
- **private foundations, professional organizations, and other institutions** such as the Annie E. Casey Foundation, the Ford Foundation, Pew Charitable Trusts, the Academy for Educational Development, the National Council on Crime and Delinquency, the National Resource Center for Youth Services, the President's Crime Prevention Council, and the American Evaluation Association
- **regional consortia of arts, education, and/or human and social service organizations** such as the New England Foundation for the Arts and the Northwest Regional Educational Laboratory

The options are limitless if you are willing to explore a bit. (Simply conducting an Internet search using search terms such as "management consulting company," "evaluation," and a geographic area will locate the Web sites of hundreds of companies from which to choose a qualified assistant.) You might even be able to identify an organization (or individual researcher) that would be willing to provide assistance in exchange for the opportunity to collect and possibly publish data on your program. (See **Appendix 26** for consultant contact information.)

In any case, refer to *The Program Manager's Guide to Evaluation*, by KRA Corporation, for detailed information about selecting and managing an evaluation consultant (see the "Other Resources" section of this chapter).

Developing Instruments
to Measure Improvements
in Art Knowledge

One of the immediate program outcomes that all three YouthARTS sites wanted to measure was the improvement of participants' knowledge of the arts. Because the content and format of the arts instruction differed across the three programs, it was not possible to develop one standardized instrument to collect data at all three sites. Instead, each program developed and administered its own *written* "art knowledge survey/test," focusing on the art disciplines taught during their art sessions. In addition, the "skills assessment form," which asked artist instructors at all three sites to rate each youth's performance in various outcome areas, obtained artists' perceptions of the youths' art knowledge and skills at the beginning and end of the program period. Finally, the programs in Portland and San Antonio used other data collection methods, such as interviews and focus groups with the artists and participating youth, to collect information on art knowledge outcomes.

This section first describes the process of developing and administering the art knowledge surveys and highlights key survey results at each YouthARTS program. It then describes the other methods used by Youth Arts Public Art and Urban smARTS to gather additional information about art knowledge.

Art knowledge surveys

At the outset of the program, **Art-at-Work** administrators developed one 24-item multiple-choice survey designed to test knowledge about the specific art disciplines that the youth would study: pottery, silk-screening, drawing/painting, sculpture, photography, and printmaking. The survey was then administered to the treatment youth at the beginning and end of the program period. The analysis of survey data showed that the youths' art knowledge did not change significantly during the program period. After discussing this finding, the program administrators realized that two factors may explain the lack of change. First, the survey, which was developed at the outset of the program, was not revised to reflect changes that the artists had made to their curricula after the program had started. Thus, it is possible that the youth were

tested on topics and concepts that were not actually taught during that program period. Second, it is possible that the youth knew the material but experienced difficulties completing the written survey. The art knowledge survey is included in **Appendix 24**.

Urban smARTS administrators developed an eight-item multiple-choice art quiz to test treatment youths' knowledge of the three disciplines taught during the program period: dance, visual arts, and drama. The pre-program survey was first administered several weeks after the program began and then again at the end of the program. The survey results showed little change in the number of youth who passed the quiz over the course of the program period. A probable explanation for the lack of change is that a very high percentage of youth received a passing score on the pre-program survey, leaving little room for improvement over the program period. One possible explanation for the high pre-program scores is that by the time the survey was administered to the youth (several weeks into the program), the participants had already learned some of the terms and concepts covered by the quiz. Had the survey been administered at the very start of the program, the pre-program scores might have been lower, leaving more room for change. A second possible explanation for the high pre-program scores is that the youth entered the program with more knowledge of the arts than was anticipated by program staff. If this was true, program staff should raise program expectations and introduce the youth to more complex or difficult art concepts.

Instead of developing one art knowledge survey for all three **Youth Arts Public Art** projects, each artist developed his/her own quiz that covered the relevant art discipline—photography and poetry, videography, or theater. The program encountered a number of challenges in administering the surveys at the beginning and end of the project periods. Of the 23 program participants included in the national evaluation, only five youth (in the poetry and photography project) completed both pre-program and post-program quizzes. Thus, it was only possible to assess changes in art knowledge for those five youth. The survey results showed substantial improvement in their knowledge about poetry and photography—four of the five youth improved from a failing grade to a passing grade over the course of the program.

All three programs are currently revising their pre- and post-program art knowledge surveys/tests and are committing the time and resources needed to ensure that, in the future, the pre- and post-program surveys are administered in a timely manner.

Appendix 24:
Data collection instruments

162

Other methods used to assess art knowledge

In addition to the written art quiz, **Urban smARTS** conducted focus groups with artists, teacher liaisons, case workers, and youth to determine their perceptions of the program and the art produced by the youth. Similarly, **Youth Arts Public Art** conducted focus groups with juvenile justice counselors, artists, and youth to find out how satisfied they were with the Youth Arts Public Art projects, what they liked best, what they didn't like, and what they felt could be done differently to improve the project. The court counselors were also asked how they felt about the artwork produced, and how they liked working with the kids on the art projects. Youth were asked what new art skills they learned and what other art skills they would still like to learn. Caliber Associates, as a part of the national evaluation, conducted focus groups at all three sites. Focus group questions are included in **Appendix 25**.

Best Practices from the Field

This section highlights several evaluation methodologies that other arts organizations have used to measure their effects on participating youth.

The Co-Arts Assessment Plan. Between 1991 and 1996, Harvard Project Zero—a research group at the Harvard Graduate School of Education—conducted a two-phase research project, known as Project Co-Arts, to develop and test a self-assessment model for organizations implementing arts-based youth programs. The resulting model, known as the *Co-Arts Assessment Plan*, provides a framework that community art centers and other educational institutions can use to document and assess the educational effectiveness of their arts-based youth programs.

In the first phase of the project, Co-Arts researchers visited, surveyed, and interviewed (by phone) hundreds of community art centers across the country. They then developed a framework to help administrators make thoughtful decisions as they attempt to offer quality education, often on a shoestring budget. The resulting *Co-Arts Assessment Plan* has two objectives: (1) to guide educators in an ongoing process of self-examination through "assessment forums," and (2) to document the process with an "organizational process folio," which may include materials such as tape-recorded interviews, correspondence with parents, memos from staff members, and youth enrollment figures for individual classes.

Appendix 25:
Focus group questions

Evaluation is time–consuming work. Be sure to assign an adequate number of well-informed personnel to administer pre-program art knowledge surveys to the youth *before* the program begins and again at the end of the program. Also, make sure that the youth are taught all of the topics and concepts that appear on the surveys. Finally, keep a portfolio of the youths' work; improvement in their artwork is proof in itself of increased art knowledge.

163

Co-Arts used the methods contained in the *Co-Arts Assessment Plan* to gather the data needed to write thirty sketches and six detailed portraits of educationally effective community art centers. In developing these products, Co-Arts used "interpretive description portraiture," a process of developing a literary narrative based on anecdotal evidence through which unifying themes (and emergent themes for hypothesis-testing) are identified.

Selected Co-Arts Findings

Co-Arts identified the following unifying themes concerning the educational effectiveness of community arts programs:
- power of art to transform and/or articulate personal identities
- cultivation of strong relationships among center constituents (teachers, students, parents, and staff)
- knowledge of and attention to the interests and needs of the communities served
- provision of enduring oases (safe havens) for students and families.
- attention to own process of development and transformation

Co-Arts also identified the following distinguishing characteristics of effective artist instructors:
- careful attention to process through ongoing reflection
- interest in learning from their mistakes (that is, identifying areas for improvement)

These Co-Arts findings parallel some of the key lessons learned through the YouthARTS project.

In the second phase of the project, Co-Arts researchers worked with selected community art centers around the country to implement and test the assessment plan and determine how organizational process folios could best be incorporated into program management. They also maintained a clearinghouse for resources and information regarding the inspirational field of out-of-school, community-based arts education. The clearinghouse produced a database with information about more than 500 community art centers in the United States, files of materials from more than 300 of these centers, and a library of relevant books and articles.

The *Co-Arts Assessment Plan* is presented in the following volumes, published by Harvard Project Zero, Cambridge, MA:

The Wheel in Motion: The Co-Arts Assessment Plan from Theory to Practice (with the accompanying *Co-Arts Assessment Tool Kit*), by J. Davis, B. Solomon, M. Eppel, and W. Dameshek (1996; $30).

Safe Havens: Portraits of Educational Effectiveness in Community Arts Centers that Focus on Education in Economically Disadvantaged Communities, by J. Davis, E. Soep, S. Maira, N. Remba, and D. Putnoi (1993; $30).

The Co-Arts Assessment Handbook, by J. Davis (1993; $30).

Another Safe Haven: Portraits of Boulevard Arts Center Then and Now, by J. Davis, M. Eppel, M. Galazzi, P. Gonzalez-Pose, S. Maira, and B. Solomon (1996; $20).

For more information on these and other related resources, call Jessica Davis at Harvard Project Zero, (617)495-4342, or see the Project Co-Arts Web page on the Project Zero Web site, http://128.103.182.32/Left/PZInfo/Research/Restxt/Coarts.htm.

Manchester Craftsmen's Guild. The Manchester Craftsmen's Guild, in Pittsburgh, which operates several arts programs for youth in inner-city neighborhoods and public schools, has used the *Co-Arts Assessment Plan* to guide their organizational self-assessment. The Guild measures students' artistic performance based on information collected from student journals, which the students can use to assess and reflect on their own program involvement and artwork, and student projects that result in exhibitions and portfolios, which include a written personal statement focusing on aesthetic development and technical inquiry. Students also participate in individual and group critiques to acquire communication and critical-thinking skills. In addition to these self-assessment activities, the Guild has contracted with outside evaluators to assess program effectiveness, particularly its effectiveness in increasing the number of students who continue with their education beyond the high-school level. For more information about the Guild or its evaluation efforts, contact Joshua Green, director of educational programs, (412)322-1772. Additional information is available at the Guild's Web site, http://artsnet.heinz.cmu.edu/mcg/pages/Youth.html.

Children of the Future. Children of the Future is a daily arts and public safety program that serves youth ages five to twelve at eight inner-city recreation centers in Columbus, Ohio. This nationally recognized program describes itself as "an unconventional

crime prevention program that uses the arts to create a safe neighborhood haven." Artists work with children after school and during the summer to provide them with avenues for expression, constructive communication, and conflict resolution skills development. Program activities such as role playing, theater games, writing, and visual arts projects are designed to help the youth address the risk factors present in their homes and communities. Since its inception in 1995, nearly 100 program artists have served more than 6,200 participating youth.

Children of the Future is an AmeriCorps project administered by the Greater Columbus Arts Council, in partnership with the City of Columbus' Departments of Recreation and Parks, Public Safety, and the Columbus Metropolitan Housing Authority. Evaluation has been an important component of this program. For a number of years the Greater Columbus Arts Council hired professional evaluators to gather anecdotal information about program effectiveness by surveying the children's parents and conducting focus groups with both the artists and the children. In 1997, the arts council initiated a study that used a quasi-experimental design to identify any links between regular program attendance and changes in school-related behaviors over the course of the school year. Assisted by the Columbus Public Schools, the arts council completed the study and reported the following findings:

"Children in the Children of the Future program, as contrasted with a comparison group of similar children, demonstrated significant change over two nine-week grading periods in areas that are important in the school and classroom environment. They showed a significant, positive gain in their overall attitude towards school. Specifically, they expressed a higher level of motivation to work hard in school. They exhibited increases in the 'ability to use school time effectively, to persist in and concentrate on instructional tasks, to seek and use feedback, and to evaluate one's own work.' Their grades in art improved significantly in contrast to the comparison group. More of the participants got higher grades in art from the beginning to the end of the study and fewer got lower grades than children in the comparison group. Participants in the program exhibited a significantly greater gain in positive attitudes towards art over the two nine-week periods than did the comparison children. Specifically, they reported an increase in pride of family members in the art products they produced and in their skills in art. Participants, in contrast to the comparison group, reported significantly increased activities related to art including the areas of visual art, theater, music, dance, and writing. The participants exhibited a significantly improved overall attitude towards art. All of these changes would be expected to positively impact the overall education experience of Children of the Future participants."

166

For more information about Children of the Future and/or the evaluation effort, contact Timothy Katz, program director, Greater Columbus Arts Council, (614)224-2606. Additional information is also available on the Children of the Future Web page located on the Greater Columbus Arts Council Web site, http://www.gcac.org/cof.htm.

Mill St. Loft. The Mill St. Loft in Poughkeepsie, New York, uses a combination of methods to evaluate its arts-based job-training programs—including pre- and post-program tests, youth questionnaires, teacher surveys, and staff-written anecdotal evaluations. Rating scales are used to assess the youths' job- and life-skill development. Portfolio assessments are used to assess changes in art skills. Youth develop comprehensive portfolios that contain resumes, photographs of artwork, and writing samples, and program staff are trained in portfolio development and assessment. School records are used for baseline assessments. Together, these methods are used to conduct both formative and summative program evaluations on an ongoing basis. (Formative evaluations are process-oriented assessments of new programs and services that enable staff to identify and address areas for improvement during the program's early stages. Summative evaluations are outcome evaluations that focus on the program's overall effectiveness.) For further information call Carole Wolfe, executive director, (914)471-7477.

Tucson-Pima Arts Council. In order to evaluate its arts-based job-training programs, the Tucson-Pima Arts Council in Arizona uses pre- and post-program tests to measure changes in academic achievement and attitudes. In addition, information from youth and artist journals is used to assess the youths' self-image, ability to work within a team, and attitudes. Art knowledge surveys are used to assess the development of new art skills over the course of the programs. The arts council is trying a new approach to attitudinal testing. Following a model developed elsewhere, youth are given color markers to use in marking their responses to the questions on the survey. They are told that red is for good/best, blue for okay, and green for don't like/worst. The theory is that people respond to color differently than they respond to words and that using color to mark answers instead of using a pencil is a way to use the arts, make answering the survey more fun. Dian Magie, executive director, can be reached at (520)624-0595.

Youth development and the arts in nonschool hours

Between 1987 and 1997, a team of more than 15 researchers, led by Milbrey W. McLaughin and Shirley Brice Heath, conducted an extensive study of organizations judged by local youth to provide effective and desirable learning environments outside of school. The study focused on 120 community-based organizations

providing a wide range of youth programs in 34 urban and rural geographic locations from Massachusetts to Hawaii. Approximately 30,000 youth passed through these sites during the study period.

Six major data sources (within the comparative framework of ethnology) were used:
- interviews with policy makers, social service workers, juvenile justice officials, and adult community organization leaders
- audio-recordings and field notes produced at the program sites during program activities
- youth logs covering daily activities, transportation opportunities, media engagement, and activities linked to literacy and the arts
- sociodemographic statistics related to economic and education changes
- interviews that local youth conducted with other community members
- the National Education Longitudinal Survey

Three-hundred youth were tracked across the 10-year study period. Using this follow-up information, 60 case studies focusing on their learning ecologies were developed.

Originally, the study paid no particular attention to arts-based community organizations. However, as the research progressed and interim findings became available, noteworthy patterns among youth involved in arts programs emerged. To examine these patterns more closely, a separate two-year analysis of the data collected from arts-based community organizations was conducted.

The results of this separate analysis are described in an article prepared by two of the study's researchers—Shirley Brice Heath and Elisabeth Soep of Stanford University—for future publication. This article—*Youth Development and the Arts in Nonschool Hours*—focuses on the effects that arts programs have on youth who are "placed at high risk through circumstances in their communities, schools, and families." The following paragraphs are excerpts from the article:

" . . . the arts, by virtue of their very nature, carried a particular power for learning achievement both in the arts themselves and in closely related competencies upon which successful performance and knowledge in the arts depends. . . . Outcomes reveal that involvement in arts-based youth organizations led to an intensity of certain characteristics among the young participants including motivation, persistence, critical analysis, and planning. Young people at art sites were more likely to win an academic honor than youth from a national sample of students across the U.S. as

measured by the National Educational Longitudinal Survey. They were also more likely to say that they plan to continue education after high school and to be recognized for community service and school attendance."

The article continues: "Arguments to discount these findings might assume that since these young people elect to participate in youth organizations they probably boast a remarkable talent and enjoy benefits not available to other youngsters. Quite the contrary. Using a 'risk index' of eight factors—such as violence in school and neighborhood, domestic instability, and economic deprivation—young people at youth organizations emerged as having a higher risk index than students in the national sample."

The authors conclude that "close examination of how the arts work at the level of everyday interactions in effective youth organizations reveals that the arts promote cognitive, linguistic, socio-relational, and managerial capacities. These achievements are mediated through risks of imagination and interaction, rules that guide but always change, and demands that create identities based in resourcefulness and accomplishment. All artists—especially the young—must be willing to make a leap of commitment. This step involves risks of greater variety than those required to go out for basketball or work on a neighborhood teen board."

Involvement in the arts and success in secondary school
In the article "Involvement in the Arts and Success in Secondary School," James S. Catteral describes the relationships between student involvement in the arts and academic achievement. Based on a longitudinal study of 25,000 students in the eighth to tenth grades, the study showed that "academic grades, standardized test scores, measured reading levels and attitudes concerning commitment to community were all higher for students maintaining high levels of activity in music, chorus, drama, and the visual arts." Theories for why the arts make a difference are not proposed. However, the analysis does show that students involved in the arts "are doing better in school than those who are not—for whatever constellation of reasons." For a copy of this article, contact Americans for the Arts at (202)371-2830.

The arts and public safety impact study
In *The Arts and Public Safety Impact Study: an Examination of Best Practices* (Rand, 1998), Ann Stone, David McArthur, Sally Ann Law, and Joy Moini report on a partnership between local arts agencies in Los Angeles, Chicago, and New York City, and Americans for the Arts. The goal of this study is to demonstrate that arts programs

can contribute in quantifiable and positive ways to solving social problems such as crime and violence. Contact Randy Cohen at Americans for the Arts for further information, (202)371-2830.

Other Resources

Clearinghouses and resource centers

Hundreds of clearinghouses and resource centers exist to help program administrators and service providers locate the materials needed to evaluate their programs. We've selected several that are relevant, comprehensive, and user-friendly. Although some of these clearinghouses and centers may seem to cover very specific subject matter, they all provide more general materials that can help you evaluate your arts program. For example, the National Clearinghouse on Child Abuse and Neglect Information can provide you with a copy of KRA Corporation's publication, *The Program Manager's Guide to Evaluation*, which is a very useful evaluation resource designed for program managers in a wide range of human and social service settings.

National Criminal Justice Reference Service (NCJRS)
P.O. Box 6000
Rockville, MD 20849-6000
Local phone: (301)519-5500
Toll-free phone: (800)851-3420
E-mail: look@ncjrs.org
Internet: http://www.ncjrs.org/

Juvenile Justice Clearinghouse
P.O. Box 6000
Rockville, MD 20849-6000
Phone: (800)638-8736
Fax: (301)519-5212
E-mail: askncjrs@ncjrs.org
Internet: http://www.ncjrs.org/ojjdp/html/pubs.html

National Crime Prevention Council On-Line Resource Center

1700 K Street, NW, Second Floor
Washington, DC 20006-3817
Phone: (202)466-6272
Fax: (202)296-1356
Internet: http://www.ncpc.org/

National Clearinghouse for Alcohol and Drug Information

P.O. Box 2345
Rockville, MD 20847-2345
Local phone: (301)468-2600
TDD: (301)230-2687
Toll-free phone: (800)729-6686
Fax: (301)468-6433
E-mail: info@health.org
Internet: http://www.health.org/

National Clearinghouse on Families and Youth

P.O. Box 13505
Silver Spring, MD 20911-3505
Phone: (301)608-8098
Fax: (301)608-8721

National Clearinghouse on Child Abuse and Neglect Information

P.O. Box 1182
Washington, DC 20013-1182
Local phone: (703)385-7565
Toll-free phone: (800)394-3366
Fax: (703)385-3206
Internet: http://www.calib.com/nccanch

Educational Resources Information Center (ERIC)
Clearinghouse on Assessment and Evaluation
The Catholic University of America
210 O'Boyle Hall
Washington, DC 20064-4035
Phone: (202)319-5120
Toll-free phone: (800) GO4-ERIC
E-mail: eric_ae@cua.edu
Internet: http://ericae2.educ.cua.edu

Harvard Family Research Project
38 Concord Avenue
Cambridge, MA 02138
Tel: (617)495-9108
Fax: (617)495-8594
E-mail: hfrp@hugse1.harvard.edu
Internet: http://hugse1.harvard.edu/~hfrp/

Innovation Network, Inc. (InnoNet)
E-mail: info@inetwork.org
Internet: http://www.inetwork.org/

American Evaluation Association
401 E. Jefferson Street, Suite 205
Rockville, MD 20850
Phone: (301)251-7700
Fax: (301)279-6749
E-mail: aea@phoenixpp.com
Internet: http://www.eval.org/

Appendices 27, 28:
Resources for evaluation topics

Printed evaluation resources (guides, books, journals, and forms)

This section is organized by evaluation topic. The first category of resources lists comprehensive evaluation resources, which provide detailed information across numerous topics. Subsequent categories include resources that provide detailed information about one or more evaluation topics. (Some publications are included in more than one category.) **Appendix 27** tells where to find these resources and, if possible, how much they cost. **Appendix 28** includes a more extensive list of evaluation resources.

Remember to check out **Appendix 24**, which contains data-collection implementation guides, sample work sheets, sample data maps, and other resources that have been used to evaluate YouthARTS and other prevention programs. **Appendix 29** includes a copy of the computerized data collection forms from the *Community Self-Evaluation Workbook,* prepared for OJJDP's Title V Delinquency Prevention Program (a national community-based delinquency prevention grants program). The *Workbook* is designed to guide OJJDP Title V grantees through the process of assessing their community needs, developing appropriate local delinquency prevention strategies, and evaluating their efforts. It contains numerous forms that can be adapted to meet your planning and data collection needs. The computerized workbook forms included in the appendix were created to meet the growing demand for the Workbook. The forms do not contain the detailed instructions contained in the Workbook, nor do they include the user's guide that was prepared to help users navigate the computerized forms. For copies of the Workbook or the user's guide, contact NCJTS, toll-free, (800)851-3420.

Comprehensive evaluation resources

Herman, J.L. (Ed.). *Program evaluation kit*, (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc. (1988).

This evaluation guide includes nine separate volumes that focus on different aspects of program evaluation methodology. The nine volumes include:

> Vol. 1, *Evaluator's handbook*, by J.L. Herman, L.L. Morris,
> and C.T. Fitz-Gibbon
> Vol. 2, *How to focus an evaluation*, by B.M. Stecher and W.A. Davis
> Vol. 3, *How to design a program evaluation*, by C.T. Fitz-Gibbon
> and L.L. Morris
> Vol. 4, *How to use qualitative methods in evaluation*, by M.Q. Patton
> Vol. 5, *How to assess program implementation*, by J.A. King, L.L. Morris,
> and C.T. Fitz-Gibbon
> Vol. 6, *How to measure attitudes*, by M.E. Henerson, L.L. Morris,
> and C.T. Fitz-Gibbon
> Vol. 7, *How to measure performance and use tests*, by L.L. Morris,
> C.T. Fitz-Gibbon, and E. Lindheim
> Vol. 8, *How to analyze data*, by C.T. Fitz-Gibbon and L.L. Morris
> Vol. 9, *How to communicate evaluation findings*, by L.L. Morris,
> C.T. Fitz-Gibbon, and M.E. Freeman

Appendix 24:
Data collection instruments
Appendix 29:
Community Self-Evaluation
Workbook

Rossi, P., and Freeman, H. *Evaluation: A systematic approach* (5th ed.). Newbury Park, CA: SAGE Publications, Inc. (1993).

Wholey, J., Hatry, H.P., and Newcomer, K.E. (Eds.). *The Handbook of Practical Program Evaluation.* San Francisco, CA: Jossey Bass (1995).

Developing planning models and data maps

Harrell, A.; Burt, M.; Hatry, H.; Rossman, S.; Roth, J.; and Sabol, W. *Evaluation Strategies for Human Services Programs.* Washington, DC: The Urban Institute (1996).

Hulett, S. *Program Planning and Evaluation: Using Logic Models in Arts Programs for At-Risk Youth.* Americans for the Arts Monograph Vol. 1, No. 6 (1997).

KRA Corporation. *The Program Manager's Guide to Evaluation: An Evaluation Handbook Series from the Administration on Children, Youth, and Families.* Washington, DC: Administration on Children, Youth and Families. (Undated).

Kumpfer, K.; Shur, G.; Ross, J.; Bunnell, K.; Librett, J.; and Millward, A. *Measurements in Prevention: A Manual on Selecting and Using Instruments to Evaluate Prevention Program* (DHHS Publication No. (SMA)93-2041). Rockville, MD: U.S. Department of Health and Human Services, Center for Substance Abuse prevention (1993).

National Institute on Drug Abuse. *How Good is Your Drug Abuse Treatment Program? A Guide to Evaluation* (NIH Publication No. 95-3609). Washington, DC: U.S. Government Printing Office (1995).

Wong-Rieger, D., and David, L. "Using program logic models to plan and evaluate education and prevention programs." In A. Love (Ed.), *Evaluation Methods Sourcebook II* (pp. 120-136). Ottawa, Ontario: Canadian Evaluation Society (1995).

Developing evaluation questions

Kumpfer, K.; Shur, G.; Ross, J.; Bunnell, K.; Librett, J. and Millward, A. *Measurements in Prevention: A Manual on Selecting and Using Instruments to Evaluate Prevention Program* (DHHS Publication No. (SMA)93-2041). Rockville, MD: U.S. Department of Health and Human Services, Center for Substance Abuse prevention (1993).

National Institute on Drug Abuse. *How Good is Your Drug Abuse Treatment Program? A Guide to Evaluation* (NIH Publication No. 95-3609). Washington, DC: U.S. Government Printing Office (1995).

Muraskin, L. *Understanding Evaluation: The Way to Better Prevention Programs* (Report No. ED/OESE-92-41). Rockville, MD: Westat, Inc. (ERIC Document Reproduction Services No. ED361604) (1993).

United States General Accounting Office, *Designing Evaluations*, (GAO/PEMD-101.4). (March 1991).

Conducting evaluability assessments

Rossi, P., and Freeman, H. *Evaluation: A Systematic Approach* (5th ed.). Newbury Park, CA: SAGE Publications, Inc. (1993).

Ruttman, L. *Planning Useful Evaluations: Evaluability Assessment.* Newbury Park, CA: SAGE Publications, Inc. (1980).

United States General Accounting Office, *Designing Evaluations*, (GAO/PEMD-101.4) (March 1991).

Selecting an evaluation design

Fitz-Gibbon, C.T., & Morris, L.L. *How to Design a Program Evaluation.* Thousand Oaks, CA: SAGE Publications, Inc. (1988).

Harrell, A.; Burt, M.; Hatry, H.; Rossman, S.; Roth, J.; and Sabol, W. *Evaluation Strategies for Human Services Programs.* Washington, DC: The Urban Institute (1996).

National Institute on Drug Abuse. *How Good is Your Drug Abuse Treatment Program? A Guide to Evaluation* (NIH Publication No. 95-3609). Washington, DC: U.S. Government Printing Office (1995).

Rossi, P., and Freeman, H. *Evaluation: A Systematic Approach* (5th ed.). Newbury Park, CA: SAGE Publications, Inc. (1993).

United States General Accounting Office, *Designing Evaluations*, (GAO/PEMD-101.4) (March 1991).

Wholey, J.; Hatry, H.P.; and Newcomer, K.E. (Eds.). *The Handbook of Practical Program Evaluation.* San Francisco, CA: Jossey Bass (1995).

Collecting and analyzing data
Fitz-Gibbon, C.T., and Morris, L.L. *How to Analyze Data.* Thousand Oaks, CA: SAGE Publications, Inc. (1988).

Henerson, M.E.; Morris, L.L; and Fitz-Gibbon, C.T. *How to Measure Attitudes.* Thousand Oaks, CA: SAGE Publications, Inc. (1988).

Morris, L.L.; Fitz-Gibbon, C.T.; and Lindheim, E. *How to Measure Performance and Use Tests.* Thousand Oaks, CA: SAGE Publications, Inc. (1988).

National Institute on Drug Abuse. *How Good is Your Drug Abuse Treatment Program? A Guide to Evaluation* (NIH Publication No. 95-3609). Washington, DC: U.S. Government Printing Office (1995).

Patton, M.Q. *How to Use Qualitative Methods in Evaluation.* Thousand Oaks, CA: SAGE Publications, Inc. (1988).

United States General Accounting Office. *Quantitative Data Analysis: An Introduction.* (PEMD-10.1.11) (June 1992).

What, Me Evaluate? A Basic Evaluation Guide for Citizen Crime Prevention Programs. Washington, DC: The National Crime Prevention Council (1986).

Wholey, J., Hatry, H.P., & Newcomer, K.E. (Eds.). *The Handbook of Practical Program Evaluation.* San Francisco, CA: Jossey Bass (1995).

Presenting and using evaluation findings
KRA Corporation. *The Program Manager's Guide to Evaluation: An Evaluation Handbook Series from the Administration on Children, Youth and Families.* Washington, DC: Administration on Children, Youth and Families. (Undated).

Morris, L.L.; Fitz-Gibbon, C.T.; and Freeman, M.E. *How to Communicate Evaluation Findings.* Thousand Oaks, CA: SAGE Publications, Inc. (1988).

National Institute on Drug Abuse. *How Good is Your Drug Abuse Treatment Program? A Guide to Evaluation* (NIH Publication No. 95-3609). Washington, DC: U.S. Government Printing Office (1995).

Rossi, P., and Freeman, H. *Evaluation: A systematic approach* (5th ed.). Newbury Park, CA: SAGE Publications, Inc. (1993).

Wholey, J.; Hatry, H.P.; and Newcomer, K.E. (Eds.). *The Handbook of Practical Program Evaluation.* San Francisco, CA: Jossey Bass (1995).