PRIMARY PREVENTION AND EVALUATION RESOURCE KIT

Volume 2:

Evaluating Prevention Strategies

PENNSYLVANIA COALITION AGAINST RAPE

This resource was supported with funding from the Pennsylvania Department of Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Pennsylvania Department of Public Health.

Recommended citation:

Townsend, S. (2014). *Primary prevention and evaluation resource kit* (Vol. 2). Enola, PA: Pennsylvania Coalition Against Rape

TABLE OF CONTENTS

OVERVIEW

For many years, the staff of sexual violence prevention programs have worked to get a foot in the door at schools, community organizations and faith communities. They have convinced educators and community leaders that sexual violence happens in their communities and that youth and adults need to learn about it. While the work is ongoing, **great strides have been made in dispelling myths and shifting the blame away from survivors**. Now the field is being asked to do more. Based on research that consistently shows that **changing attitudes does not change behaviors**, the emphasis is now on changing behaviors through building skills, altering social norms and other strategies to shift the cultural foundations of sexual violence.

> Primary prevention is about getting to the root of the problem and changing our culture to one that promotes safety, equality and respect.

At the same time, **preventionists throughout the nation are increasingly being asked to show evidence for the outcomes of their programs**. Evaluation is important on many levels. Most importantly, evaluation helps to identify and sustain what is working and examine and change what is not. It also strengthens funding proposals, opens the doors to do programs in new settings and helps preventionists build credibility within the community.

However, **the growing need for evidence-based programming is not necessarily paralleled by a comparable increase in funding**. Therefore, **the task of evaluating prevention programs is largely falling to preventionists**, some of whom are well versed in evaluation and others who are only starting to learn about evaluation methods.

While current and future programming may demand greater effort and skill in the area of evaluation, **the long-term rewards promise to outweigh the short-term costs**. Evaluation will enable preventionists to build upon strengths and improve programs. Through evaluation, sexual violence preventionists will be able to show what they have always trusted: that their programs do make a positive difference.

> Evaluation can be used to build upon strengths and to leverage support for programming.

This resource kit is intended to support preventionists in building upon what they are already doing to evaluate their programs. It is divided into four volumes:

VOLUME 1:
Introduction to Primary Prevention
Choosing Primary Prevention Strategies

VOLUME 2:
Introduction to Program Evaluation
Basic Steps for Evaluating Your Programs

VOLUME 3:
Analyzing Quantitative Evaluation Data
Interpreting and Using Findings

VOLUME 4:
Analyzing Qualitative Evaluation Data
Interpreting and Using Findings

In the **first volume**, we defined primary prevention and discussed some of the challenges of it. Then we looked at a framework for choosing primary prevention strategies.

This **second volume** provides an introduction to program evaluation. It walks the reader through six steps for evaluating programs and offers suggestions for when and how to hire an outside evaluator. A supplemental volume includes measures you may find useful when evaluating your prevention programs.

The **third volume trains** the user on how to analyze quantitative evaluation data (i.e., numbers) using frequencies, percentages, averages, tests of change over time and tests for group differences. It will show the reader how to use Microsoft Excel to analyze data and includes step-by-step instructions on how to use the software.

The **fourth volume** trains the user on how to analyze qualitative evaluation data, such as notes from interviews and focus groups. It shows the reader how to use some simple techniques that provide a systematic approach to finding themes in the data.

There are many approaches to evaluation, each with its own advantages.

This resource kit is <u>not</u> intended to be a blueprint for programs to follow. There are many approaches to evaluation, each of which has its own advantages. It is also important to match the evaluation strategy to your agency and community, choosing an approach that matches your needs, strengths, challenges, resources, history, personalities and priorities. The possibilities for primary prevention and program evaluation are limited only by our own imaginations.

INTRODUCTION TO PROGRAM EVALUATION

- Why Evaluate Primary Prevention Efforts?
- What Are the Different Types of Evaluation?
- What Are the Challenges of Evaluating Primary Prevention?
- What Have Previous Evaluations of Sexual Violence Prevention Programs Found?
- How Can Program Staff Use Research and Evaluations?

WHY EVALUATE PRIMARY PREVENTION EFFORTS?

There are four common reasons to evaluate a program:

Evaluation provides a basis for making informed decisions.

Evaluation can help program staff make informed decisions about continuing or modifying a program. Evaluations can be used to identify programs that show promise or that demonstrate clear success in areas that are a priority. These programs would likely be continued. Evaluations can also identify programs that are not showing sufficient impact or that are having effects in areas that are not as important to the mission of the agency. These programs may need to be modified or discontinued. It is also important to consider unintended negative effects. Evaluations can reveal whether the program is inadvertently having effects that are not desirable and changes can be made to correct those effects. By **providing a basis for informed decisions,** evaluation protects programs from making capricious decisions. It also provides checks and balances so that other factors (political climate, personal preferences, etc.) do not lead to arbitrary decisions about the continuation or cancellation of a program.

Evaluation can protect a program from outside criticism.

Evaluation can also help protect a program from outside criticism. There are many reasons that people may not support sexual violence prevention programming, especially in the schools. A common fear is that the program will "plant ideas." This fear can be ameliorated by surveys that show how familiar students already are with sexual violence. For example, some people fear that talking about drug-facilitated sexual assaults will introduce students to the idea and lead them to try something that they otherwise would not have thought of. However, surveying students before prevention programs start often shows that students already have heard about drug-facilitated sexual assaults.

Evaluation provides insight into how a program is working.

Evaluation can also provide insight into how or why a program is working or not. Understanding the mechanisms by which programs work provides a foundation that staff can build on when developing new programs. Rather than reinventing the wheel or trying things haphazardly, staff can take elements and strategies that are most successful and use them in other programs. Likewise, precious time and resources will not be wasted on strategies that have minimal impact.

Evaluation shows that we are holding ourselves accountable.

Finally, evaluation is a mechanism for accountability. Public funds are limited. Citizens are entitled to know that their tax dollars are being used wisely. Private funders similarly want to know that their dollars are having a positive impact.

WHAT ARE THE DIFFERENT TYPES OF EVALUATION?

Often when people think about evaluation they have in mind something like the *Consumer Reports*. They want to rate programs so that they know which ones work and which ones are best. Or they may think of evaluation like a report card: How well did this program do as measured on some supposedly objective scale? However, evaluation is actually a much broader concept. **Different types of evaluations answer different questions.** It is important that program staff identify what type of evaluation will best answer their questions.

Different types of evaluations answer different questions.

There are five main types of evaluation (Rossi, Freeman, & Lipsey, 1999):

TYPE OF EVALUATION	PURPOSE	USES	COMMON METHODS
Needs Assessment	Identify and prioritize needs	Decisions about how to allocate resources; whether, where and when to start new programs	■ Interviews ■ Focus Groups ■ Surveys ■ Existing data ■ Observations
Program Theory	Clarify underlying theory about why and how program works	Improve how program is conceptualized; identify intermediate and long-term effects	■ Interviews ■ Document review ■ Logic model
Process and Performance	Describe how a program is operating	Identify implementation problems; fidelity checks; assessing satisfaction	■ Routine data collection ■ Satisfaction surveys ■ Fidelity checks
Program Impact	Determine if program has intended effects and how strong effects are	Decisions about continuation or expansion of program; modifying program	■ Pre-test / Post-test comparisons ■ Community surveys ■ Interviews ■ Focus Groups
Efficiency	Compare program costs to outcomes	Large scale policy and funding allocation decisions	■ Cost-benefit analysis ■ Impact analysis

1. *Assessing the need for a program*

Identifying, comparing and prioritizing needs can help when making **decisions about how to allocate resources and whether, where and when to start new programs.**

Needs assessments are often done through:

- **Interviews, focus groups and surveys** with community leaders, members of the intended audience (e.g., adolescents if it will be a school-based program), others who may be affected by the program (e.g., parents), others who deal with sexual violence from a different perspective (e.g., law enforcement, medical personnel, mental health professionals) and professionals in related fields (e.g., school personnel, other social services providers).
- **Data from existing sources** such as police and court records, crime reports, school disciplinary reports, hotline records, client usage reports and news reports.
- **Observations** of physical environments (e.g., graffiti in the community), social behaviors in public settings (e.g., of teachers to assess gender equity in classrooms) and social messages (e.g., advertising and other media).

> Needs assessments can be used to determine how to allocate resources across different strategies.

When conducting a needs assessment, it is important to include all of the relevant people who may have insight into or be affected by the program. A common mistake is to get input from other professionals but leave out the people who will be participating in or otherwise affected by the program. Youth are commonly left out of assessments and program development. This is a grave error. Finding out the experiences and views of the people who will be participating in or benefitting from the program is a valuable source of information about how the intended program may be received, how well needs are currently being met, what needs are not being met and strategies that may be more or less effective.

2. *Assessing program theory*

Assessing program theory is used to **articulate and clarify the underlying logic about why and how the program should work**.

> Clarifying the theory of how what you do leads to change can help identify what kinds of outcomes you might realistically expect. It is also a way of identifying gaps in your programming.

This type of evaluation can be useful in itself because a program that is based on a weak or faulty theory has little chance of achieving intended results. Assessment of program theory **can help to improve how the program is conceptualized. It can also identify the kinds of effects that you might expect and that you will want to measure.** This is especially important when the process of prevention is expected to be a long-term endeavor. In the case of sexual violence, we cannot expect to see the rates of sexual violence drop in the immediate future. If that is our only measure of success we may set ourselves up for appearing like we have failed by measuring rates of violence in the short-term. Or if we start a program and delay its evaluation for many years, we may be disappointed to learn that the resources we have been devoting to the effort have been for naught. However, if we understand clearly the theory of the program including the incremental changes and chain reactions that we expect to see, then we can measure those changes and assess our progress every step of the way.

Assessing program theory usually involves describing program goals and objectives and the chain reaction that leads from the activities of the program to the intended outcomes. This is often done through:

- **Interviews** with program staff and program participants
- **Review of program documents** including mission statements, written objectives, curricula, program materials and prior evaluation results
- Using the information collected through interviews and documents to create a **logic model or change model** which is a picture that illustrates the chain reaction between activities and outcomes

When assessing program theory, it is important to repeatedly refine the model. It is often helpful to talk with program staff, create a first draft, go back to program staff for feedback, revise the model, go back to program staff again and continue this process until there is a consensus that the model accurately captures the program. Often the model will become more complex as underlying, unspoken assumptions become evident.

3. *Monitoring program process and performance*

Monitoring program process and performance **describes how a program is operating**.

This type of evaluation may help **identify problems in how the program is being implemented and making adjustments along the way**. It can be used to assess whether the program is delivered in the way it is intended, if it is reaching the targeted audience, how satisfied participants are with the program and what is going well or not going well.

> Monitoring how a strategy is operating can help make it run more smoothly.

Monitoring is often done through:

- **Routine data collection** including the number of programs done, frequency of programs, number of participants, components of the program used, etc.
- **Program satisfaction surveys** that are given to program participants; in schools these often include surveys of both students and teachers. It is important to note that satisfaction surveys are a method of monitoring process and performance. **They are <u>not</u> a method of assessing program impact.** Just because participants enjoyed the program does not mean that it is an effective program. This is especially true when thinking about primary prevention. The ultimate goal is to change behaviors in ways that prevent sexual violence. Participant satisfaction does not measure whether that goal has been achieved.
- **Fidelity checks** determine whether the program is being implemented like it is intended; with curricula this often involves observing the presenter to see whether they are covering the topics in the curriculum and how much time they are spending on each topic

Program monitoring is the type of evaluation that program staff tend to be the most familiar with because they are already doing it. Sometimes the monitoring process is limited to what is required by funders. However, it can be worthwhile to consider whether there is additional information that would be useful to the program.

4. *Assessing program impact*

Assessing program impact is done in order to **determine whether a program has the effects it is intended to have and how strong those effects are**. This is typically the type of evaluation that people think of when they hear "program evaluation."

Researchers most often assess program impact by randomly assigning people to receive the program or not receive it. They then test some outcome of interest (e.g., acceptance of rape myths, self-report of perpetrating behaviors, number of bystander interventions, etc.). Usually the test is given twice: once before the program is delivered and again afterward. The two groups are then compared. If the program is effective then the group that participated in the program should show significantly different results than the group that did not participate.

> Assessing program impact tells you how well you are achieving the objectives of the program.

However, this approach to evaluation (often called a randomized pre-post test or an experimental design) is generally not feasible for community-based sexual violence programs. Besides the time and resources it takes to test people who are not receiving the program, it is often awkward to ask that people who could potentially benefit from the program not receive it for purposes of evaluation. A common strategy used by researchers in this situation is to delay the intervention for one group. Thus, the evaluation is done and then the group that did not receive it has the chance to go through the program. However, this may be practically difficult for sexual violence prevention programs and the benefits may not outweigh the costs.

An alternative approach is to rely on **one group of participants, all of whom receive the intervention. In this case you would compare the Pre-test and Post-test and consider whether there were other factors besides the intervention that could have accounted for any changes.** Other factors might include school curricula, exposure to sexual violence issues through news and entertainment media, or violence prevention programs being run by other organizations such as law enforcement, youth services, or faith communities.

It is also necessary to re-think your evaluation strategy when you want to evaluate an intervention that is not a program or curriculum per se. For example, what if the intervention is a message campaign that involved posters, radio and television public service announcements? In this case there is no identifiable subgroup to assess. Instead, your target was to change something in the community at large starting with people who saw the campaign.

You can still assess program impact by **surveying people in the community**. It would be advisable to survey people prior to the start of the campaign to get a baseline assessment of where the community is in regard to the message. Then conduct a second survey after the campaign has happened to see if the community has changed in the ways you intended.

While you might want to ask about the extent of their exposure to the campaign, keep in mind that even people who did not see the campaign may still show changes if they are influenced by people who did see it. So you may want to focus on the comparison of before and after the campaign, not so much on the comparison between people who saw the campaign and those who did not.

In this case, you are not surveying the same people both times. Rather, you are taking a sample of the community before the campaign and comparing them to a sample of the community after the campaign. This is referred to a cross-sectional design. It does require surveying more people to increase the chances that the sample from before the campaign is similar to the sample after the campaign.

Regardless of what kind of intervention you are assessing or the exact evaluation strategy you use, **it is critical that the questions you ask or other measures you use directly relate to what the program is intended to do.** One of the most common mistakes in sexual violence prevention is to say that the program is designed to prevent sexual violence (which is a behavior) and then assess it by asking factual questions about sexual violence and what people think about it (which are knowledge and attitudes). There must be a clear correspondence between the program's goals and what you measure. The logic model can help with this process.

> One of the most common mistakes is to ask about knowledge and attitudes when what you really want to know about are behaviors.

It is also important to **consider how effects may change over time**. Many programs assess changes immediately after the program ends. The logistics of doing this are relatively easy. However, it does not answer the question of whether those changes are sustained over time. What do participants think a month later? A year later? You cannot assume that the changes will be maintained over time. These questions can be answered by collecting follow-up data to see if the effects are sustained over time.

5. *Measuring efficiency*

Evaluations of program efficiency or cost-benefit analyses are used to **compare program costs to the outcomes** in order to determine whether the program is worth continuing to invest in.

> Cost-benefit analyses compare a program's costs to its outcomes.

WHAT ARE THE CHALLENGES OF EVALUATING PRIMARY PREVENTION PROGRAMS?

Just as there are challenges when doing primary prevention, there are similar challenges when evaluating primary prevention programs. Four challenges are particularly worth keeping in mind:

1. **Program staff often face barriers to doing the type of prevention work they think will be most effective.** Preventionists often know that they need to do programs that have more sessions, address more of the root causes of sexual violence, include social activism and that build on prevention messages over time. All too often, however, their efforts are thwarted by others in the community who (often for very legitimate reasons) can't provide preventionists with the time or type of access that is needed. Therefore, they may have to evaluate a program that, from the outset, they do not think is optimal. This may require lowering the expectations of what outcomes can be achieved. This requires that what is measured matches those lower expectations.

2. **The reduction of sexual violence is a long-term outcome**. If you measure the rate of new sexual assault cases in the short-term it will look like your prevention program failed even if it is actually working. It is simply not feasible to expect the rate to decline within a fiscal year. In fact, by creating a more supportive climate for survivors you may see the rate of reporting increase, making it look like the problem is actually getting worse. Alternatively, if you focus only on theory and process evaluations you may never get around to assessing impact.

3. **Intermediate outcomes need to be identified so that steps along the way to the long-term goal can be evaluated.** However, it can be difficult to know precisely what those intermediate steps are. This is why program theory is so important. A clear, detailed change model that shows the chain reaction that leads from your activities to the prevention outcomes will point to intermediate outcomes. One of the most common gaps is between increasing knowledge and decreasing victim-blaming attitudes to changing behaviors. Knowledge and attitudes may be necessary precursors to behavioral change, but what comes in between? The CDC logic model indicates that there are community-level changes that create new norms about sexual violence (perceiving costs of sexual violence and benefits of prevention; integrating sexual prevention into goals and activities; reforms to eliminate social inequalities and to increase justice and accountability). It also indicates that there are individual changes that lead to individuals behaving in ways that are consistent with non-violence (perception of norms and awareness of social rewards and consequences).

4. **Evaluation tools tend to overlook community-level changes.** Most evaluations of sexual violence prevention programs rely on surveys that are administered to individuals who participated in an education program and that ask them about their own knowledge, attitudes, beliefs and behaviors. However, the changes that we are trying to bring about are social changes. How do we measure changes in social norms? Cultural values? Changes in systems? New tools are needed, but most local programs are not equipped to create them. Much more collaboration needs to be done between the field and researchers who have the resources to do this type of development on behalf of the field. The CDC is working to develop evaluation tools that correspond with their logic model for RPE programming. These should be very useful for programs when they are available.

These challenges are not impossible to resolve, but they do require close attention. Otherwise, we risk the appearance of failure when, in fact, our programs and interventions may be having important effects. This manual provides a process and resources for program evaluation that will help programs do evaluations that are mindful of these challenges.

WHAT HAVE PREVIOUS EVALUATIONS OF SEXUAL VIOLENCE PREVENTION PROGRAMS FOUND?

Using evaluations that have been done by researchers is one way of checking the **potential effectiveness** of your own programs. If you are doing a similar type of program that researchers have already found to be effective, then you are warranted in thinking it may be effective in your community as well. If you are using a program that was developed by someone else you may need to make modifications to it in order to meet the needs of your audience, to be culturally relevant, or to accommodate limitations you face. However, be aware that the more you modify a successful program the less confident you can be that it will have the same effects with your audience. This does not mean you should not modify it, merely that you need to be certain you do your own evaluation as well so that you can determine whether the program works with your modifications.

There are four main types of sexual violence prevention programs that have been evaluated for program impact in the research literature. An overview of the programs and the major findings follows.

TYPE OF PROGRAM	COMMON CONTENT	SHORT-TERM EFFECTS	LONG-TERM EFFECTS
Short Educational Programs	Myths, definitions, signs, resources, risk reduction (healthy relationships)	■ Increased knowledge ■ Decreased myth acceptance	■ None found
Theater Programs	Myths, definitions, signs, resources, risk reduction (healthy relationships)	■ Increased knowledge ■ Decreased myth acceptance	■ None found
Intensive Programs	Same as above over more time plus more skill-building, gender roles, analyzing media, gender equity, social activism	■ Increased knowledge ■ Decreased myth acceptance ■ Increased intent to intervene ■ Decreased rates of physical and sexual violence	■ Increased knowledge ■ Decreased myth acceptance ■ Increased intent to intervene
Bystander Education	Skill-building to confront sexist and violent peers	■ Increased positive attitudes toward bystander behavior ■ Increased confidence for acting as a bystander ■ Increased intent to act as a bystander ■ Increased reports of bystander interventions	■ Increased positive attitudes toward bystander behavior ■ Increased confidence for acting as a bystander ■ Increased intent to act as a bystander ■ Increased reports of bystander interventions

1. *Short Educational Programs*

This is perhaps the most common type of prevention program. It typically involves one to two sessions that are about one hour each. Most often they are done with mixed-gender groups, but sometimes the groups are single-gender. These programs focus on myths about sexual assault, legal definitions, signs of potential perpetrators and local resources for survivors. Some also include a general discussion about healthy relationships and/or tips for reducing risk (e.g., safe dating strategies). For young children, short programs tend to focus on defining types of touches, skills for getting away from risky situations and skills for seeking help.

Short educational programs can increase knowledge and decrease acceptance of rape myths.

The programs often try to be interactive by using didactic presentations, discussions and interactive exercises.

Most of the evaluations of this type of program have found that participants show:

- Increases in knowledge about sexual assault
- Decreases in acceptance of rape myths
- However, follow ups at later times show that the effects tend to weaken over time.
- Only a few studies have measured actual behaviors by asking participants about whether they have engaged in specific sexually aggressive behaviors or have been the victim of such behaviors. These studies show no reductions in sexual violence.

As one group of researchers noted following their study, **"It may simply be unrealistic to expect that long-held, deeply ingrained attitudes and beliefs will be changed in any lasting way as the result of a one- or two-hour program. The danger of such programs is that they can make us think that we are doing something, even if we are not"** (Frazier, Valtinson, & Candell, 1994, p. 156). This is a challenging statement, but an important one to consider. A more positive way of looking at the situation is that these educational programs are effective at increasing knowledge and changing attitudes. This is a necessary starting point. However, if we want to do primary prevention we must go further and do more intensive or different programs that target changing behaviors. Providing more intensive programs will require increased access to participants as well as funding and staffing resources in prevention departments of community-based sexual violence centers.

2. *Short Theater-Style Presentations*

These programs are similar to the short educational programs. They are usually done in single sessions with mixed-gender audiences. The topics they address are also similar with an emphasis on defining sexual violence, debunking myths, identifying signs of potential perpetrators and dangerous situations, sometimes acting out ways to avoid potential assaults, presenting ways to help survivors and sometimes presenting ways to confront potential perpetrators or to intervene in rape culture.

> Interactive theater-style presentations can also increase knowledge and decrease rape myth acceptance.

The major difference between these programs and short educational programs are that they use actors to present vignettes of scenarios related to sexual violence. Sometimes the action is stopped to obtain feedback from the audience on what they think is going on and to get suggestions for what the characters should do. In some programs audience members are invited to take on a character's role to experience how the scene might go differently if the characters were to follow the audience's suggestions. Some programs involve discussion afterwards.

Evaluations of theater-style programs have shown very similar results as the short educational programs:

- Short-term increases in knowledge
- Short-term decreases in rape myth acceptance including victim blaming
- Effects weaken over time
- No reductions in sexual violence perpetration or victimization have been reported.

3. *Intensive Educational Programs*

Intensive educational programs have many of the same goals as short programs. However, they are much longer, including at least six sessions with the same participants over time. The increased time allows for a number of differences:

- Each topic can be explored in more depth with more opportunities for participants to process their reactions to what they are learning.
- More interactive activities, including more skill-building exercises, can be used.
- More topics can be addressed. The additional topics are often focused on understanding and changing specific aspects of a rape culture such as exploring gender roles, analyzing media for how it objectifies women, identifying gender inequity and promoting social activism.
- There is ample time to complete an evaluation of the program without detracting from the program.

> Intensive programs that include skill-building and peer support for using those skills can change behaviors related to prevention.

Evaluations of two intensive programs are particularly worth noting, both as examples of content and the potential effects that programs might see when taking a more intensive approach.

Safe Dates is a 10-session high school curriculum that was designed to be taught by teachers. In the original research study that documented the effectiveness of Safe Dates, those teachers received 20 hours of training from the developers of the curriculum. The program also includes a theater production by peers, a school poster contest, services in the community for adolescents in abusive relationships and training given to community service providers.

- The initial evaluation of Safe Dates showed promising results (Foshee et al., 1998). There were significant increases in students' knowledge and changes in attitudes as well as 60% less sexual violence perpetration in the school that received the full intervention.
- However, at a one-year follow up the knowledge and attitude changes remained but the differences in perpetration rates had disappeared (Foshee et al., 2000).
- This evaluation supports intensive programs that have multiple components within the school and community as a way of decreasing sexual violence in the short-term.
- However, it may be that multi-year programming is needed in order to maintain the prevention effects over time.

Healthy Relationships (Men for Change, 2000) is a three-year program designed for grades 7 through 9. It is an interactive, activities-based curriculum. Each year includes approximately 20 lessons. In the first year students learn to recognize emotions that can lead to violence and basic communication skills that can be used to solve problems. In the second year students examine gender stereotypes with a specific emphasis on media images. Students use critical thinking skills to analyze power and control dynamics in popular culture and to connect those images to physical and sexual violence in relationships. In the third year students explore the connections between sexist attitudes and violence in relationships. The curriculum

reinforces values of safety, equality, respect, empathy, personal responsibility and personal empowerment. It also has a social action component.

- A three-year evaluation of the Healthy Relationships curriculum showed promising results in many areas despite the fact that the evaluation was based on a small selection of activities from the curriculum rather than the entire curriculum.
- Students who participated in the program: were more likely to use assertive rather than aggressive responses to conflicts; demonstrated increased knowledge about television violence, relationship violence and boundary setting; showed fewer beliefs in gender stereotypes; indicated that they were more likely to talk to abusers about their behavior, to talk with a teacher or guidance counselor about what to do if they know someone is being abused and to talk directly with the person being abused; were more confident in using the skills taught in the program, including identifying stereotypes in advertising, resisting peer pressure and recognizing power dynamics.
- In terms of actual behaviors, students who participated in the program were more likely to break up with a violent dating partner and reported significantly lower rates of physical violence, passive-aggressive tactics and psychological abuse in their relationships.
- This evaluation supports intensive programs that occur over multiple years, that build basic communication and assertiveness skills and that strengthen critical thinking about cultural messages and norms.

4. *Bystander Empowerment Programs*

Bystander empowerment programs train students to be empowered bystanders who confront sexist and abusive peers. Students show other students, through example and mentoring, that sexism and gender violence are not acceptable and will not be tolerated in the school culture or other setting where the program is implemented.

There is a strong theoretical basis for bystander programs, especially as a way of changing norms in a community (Banyard, Plante, & Moynihan, 2004). Drawing from work in the social sciences and community development bystander empowerment is theorized to increase the likelihood of people intervening in rape culture and high-risk situations by:

- Increasing their awareness of sexual violence and its impact on survivors
- Increasing their sense that they are responsible for solving the problem
- Strengthening their skills for intervening
- Developing their belief they can use those skills effectively
- Creating new social norms when interventions are made

A bystander empowerment program that is increasingly being used with college students is *Bringing in the Bystander*. This three-session program has been evaluated by researchers using a rigorous methodology (Banyard, Moynihan, & Plante, 2007).They found that up to two months after participating in the program, college students:

- Decreased their acceptance of rape myths
- Increased their knowledge of sexual violence
- Increased their beliefs about prosocial behaviors
- Strengthened their sense that they can effectively intervene
- Reported making more interventions

Another notable program is *Mentors in Violence Prevention* at Northeastern University. An evaluation of this program is currently underway (Katz, 1994).

Evaluations of sexual violence prevention programs found that common educational programs are good at increasing knowledge and decreasing acceptance of rape myths, including victim blaming. However, they have not been found to be effective at changing behaviors, other than possibly how people respond to a friend/family after an assault.

There is evidence that more intensive, multi-component programs may decrease perpetration. However, the effects on behavior tend to weaken over time so we need to look at multi-year programs.

Bystander empowerment programs, although relatively short, do show promise for changing knowledge, attitudes and behaviors that are focused on intervening in high risk situations. However, they are not designed to change other causes of sexual violence.

HOW CAN PROGRAM STAFF USE RESEARCH AND EVALUATIONS?

Existing research evaluations can be used in a number of ways by program staff.

- **Research can inform your decisions about continuing, changing and/or expanding your programs.** You may want to continue your short educational programs for other important reasons such as increasing awareness about services, building bridges in the community and strengthening community awareness about the need for prevention efforts. However, do not view them as primary prevention. The research has consistently found that they do not succeed at preventing sexual violence.
- **You can use research findings to advocate with others in the community (e.g., schools) for why more intensive programs are needed.** Schools especially may be very comfortable with the current arrangements, especially if it involves your staff coming in as guest speakers for a day or two once a year. It takes more commitment on their part to have you come in for multiple sessions and to create space in their curriculum for a multi-year program. However, the research evaluations can be used to build a strong argument for a multi-session, multi-year program.
- **You can also use this information to make connections between sexual violence prevention and other efforts schools are making to meet learning objectives.** In particular, curricula like the Healthy Relationships curriculum have numerous components that can be used to help meet academic learning standards. While sexual violence may not be the first issue that school administrators and teachers think of when planning their curricula, they may be receptive to arguments for how sexual violence prevention (especially programs that build skills) fit the standards they are required to meet. Some of the Pennsylvania academic standards (Pennsylvania Department of Education, 2013) to consider when talking with school personnel include:
(See next page)

PENNSYLVANIA ACADEMIC STANDARDS

STANDARD	GRADE 3	GRADE 6	GRADE 9	GRADE 12
10.2 Healthful Living	Identify media sources that influence health and safety	Explain the media's effect on health and safety issues	Analyze media health and safety messages and describe their impact on personal health and safety	Compare and contrast the positive and negative effects of the media on adult personal health and safety
	Identify the steps in a decision-making process	Describe and apply the steps of a decision-making process to health and safety issues	Analyze and apply a decision-making process to adolescent health and safety issues	Examine and apply a decision-making process to the development of short-term and long-term health goals
10.3 Safety and Injury Prevention	Recognize safe / unsafe practices in the home, school and community	Explain and apply safe practices in the home, school and community	Analyze the role of individual responsibility for safe practices and injury prevention in the home, school and community	Assess the personal and legal consequences of unsafe practices in the home, school, or community
11.2 Balancing Family, Work and Community Responsibility	Recognize conflict situations and identify strategies to avoid or resolve	Describe strategies to avoid or manage conflict and violence	Analyze and apply strategies to avoid or manage conflict and violence during adolescence	Analyze the impact of violence on the victim and surrounding community
	n/a	Contrast the solutions reached through the use of a simple decision making process that includes analyzing consequences of alternative solutions against snap decision making methods	Solve dilemmas using a practical reasoning approach	Justify solutions developed by using practical reasoning skills

PENNSYLVANIA ACADEMIC STANDARDS

STANDARD	GRADE 3	GRADE 6	GRADE 9	GRADE 12
11.2 Balancing Family, Work and Community Responsibility	Explain daily activities that fulfill family functions in meeting responsibilities	Compare and contrast how different cultures meet family responsibilities within differing configurations	Contrast past and present family functions and predict their probable impact on the future of the family	Assess the relationship of family functions to human developmental stages
	Identify how to resolve conflict using interpersonal communication skills	Describe positive and negative interactions within patterns of interpersonal communications	Justify the significance of interpersonal communication skills in the practical reasoning method of decision making	Evaluate the effectiveness of using interpersonal communication skills to resolve conflict
11.4 Child Development	Identify health and safety needs for children at each stage of child development	Identify ways to keep children healthy and safe at each stage of child development	Evaluate health and safety hazards relating to children at each stage of child development	Analyze current issues in health and safety affecting children at each stage of development

- These evaluations can also help in advocating for more funding and additional resources. Again, the evidence is compelling that one- to two-session programs are not achieving primary prevention goals. Therefore, sources that fund prevention efforts need to be supporting multi-session, multi-year programs. This information can be used when doing public policy advocacy for increases in state and federal prevention funds as well as when writing requests for proposals (RFPs).

Although limited, the existing research and data can be used to help rape crisis centers market their programs and to make the case for additional funds, programs and partnerships in prevention.

BASIC STEPS FOR EVALUATING YOUR PROGRAMS

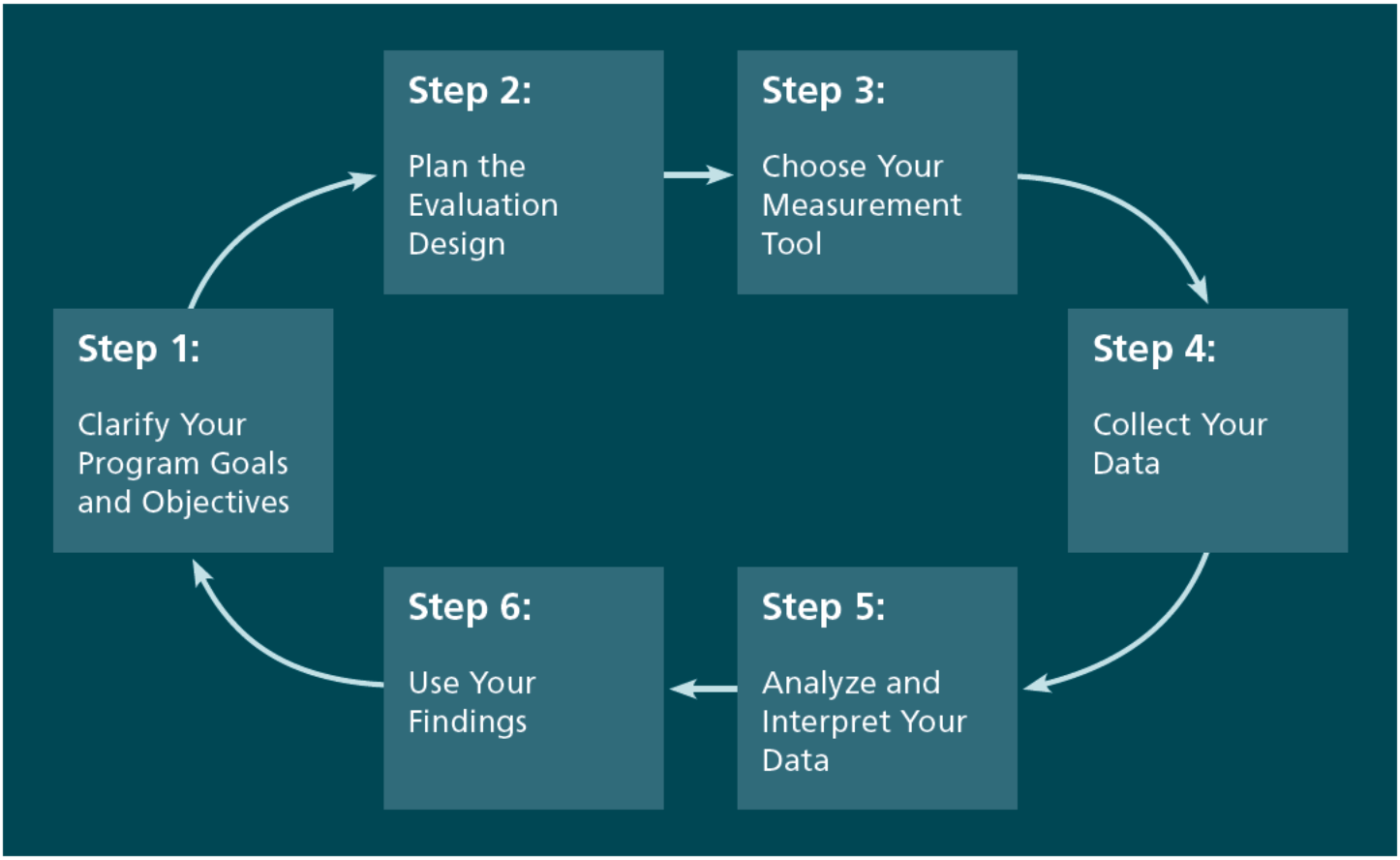
- Scope
- Step 1: Clarify Program Goals and Objectives
- Step 2: Plan Your Evaluation Design
- Step 3: Choose Your Measurement Tools
- Step 4: Collect Your Data
- Step 5: Analyze and Interpret Your Data
- Guidelines for Hiring an Outside Evaluator

SCOPE

There are many types of evaluations and techniques that can be used. It is best to have various evaluation techniques from which you can choose. Your staff are capable of learning numerous evaluation strategies and using them effectively.

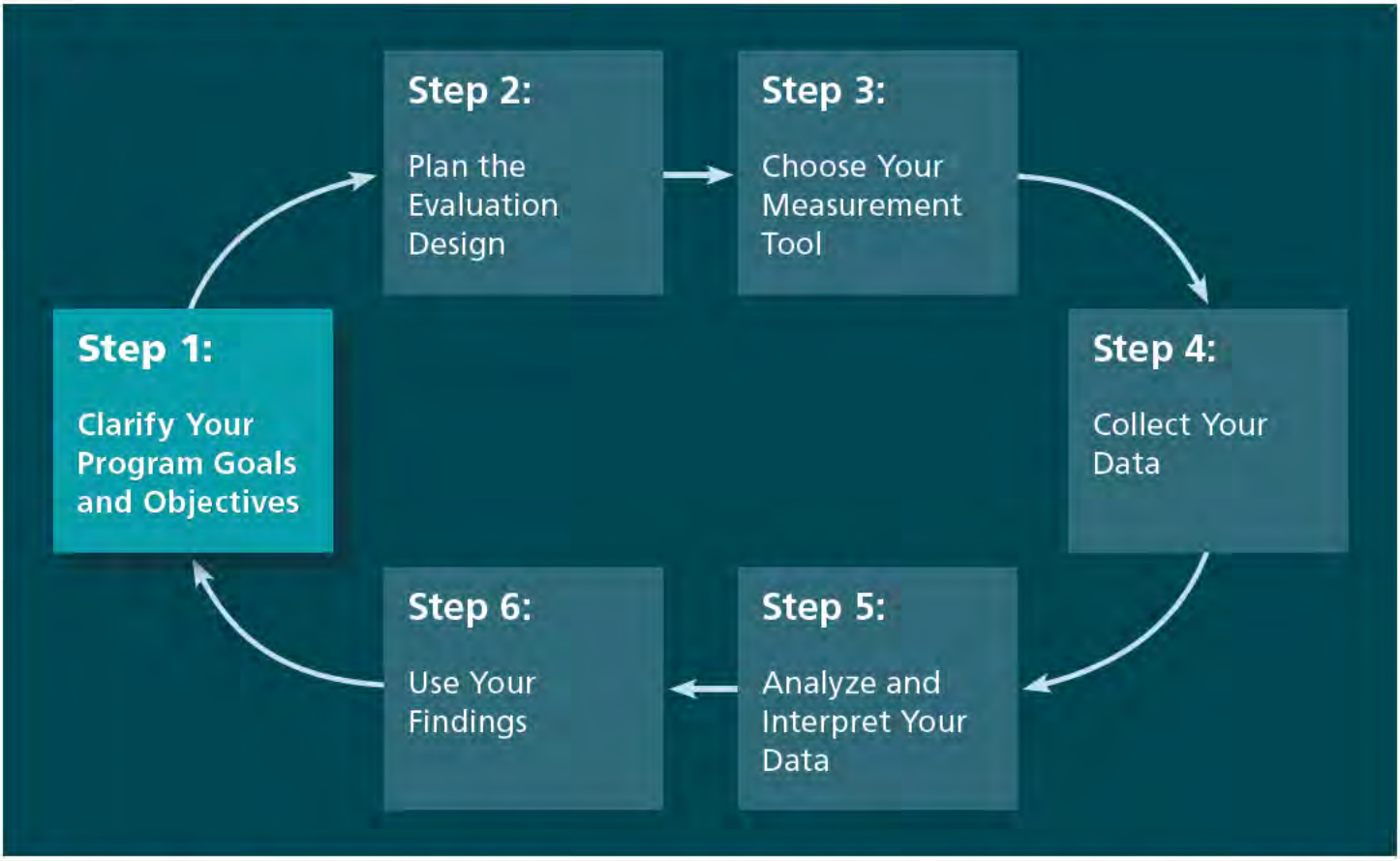
This section includes a detailed outline of the six basic steps of evaluating program impact. The focus is on program impact because this is the type of evaluation that staff usually want to learn how to do. This part of the manual will:

- Walk you through the basic steps of evaluating program impact.
- Introduce you to the key issues you need to consider at each step.
- Help you to identify where your program has the capacity to do evaluation on your own.
- Identify where you might need to hire outside help.



Note: Steps 5 and 6 are covered in Volumes 3 and 4 of this toolkit.

STEP 1: CLARIFY YOUR PROGRAM GOALS AND OBJECTIVES



The first step in any evaluation of program impact is to clarify the program's goals and objectives. This involves two tasks:

- Task 1: List or describe the changes that should occur as a result of your program.
- Task 2: Based on that theory, define the program's goals and objectives.

Task 1: List or Describe the Changes that Should Occur as a Result of Your Program

It is important that you clearly communicate the changes that you expect to see as a result of your program. Often we have an intuitive sense of what we expect, but we have not always articulated the specific changes. It is critical to do so when we are talking about long-term changes in behavior and social norms. **There are many changes that occur in the process of preventing sexual violence and we need to identify what those changes are.** This requires that we have a clear sense of how change should happen.

> Defining how you expect change to happen will bring clarity to your work and identify what you need to measure.

When articulating our theory of change we typically develop a graphic representation of the changes we expect to see. This picture illustrates what we think will happen as a consequence of the program. In other words, "If we do A, then B will happen. If B happens then C will happen, etc." until we get to our ultimate goal for the program.

There is no one right way to illustrate program theory. Some approaches specify certain boxes or columns that you should have and programs fit what they do into that pre-set structure. Excellent examples of this approach are found in the *W. K. Kellogg Foundation Logic Model Development Guide* (free download available at www.wkkf.org). The main advantage of a structured approach is that it helps people who are new to theory models get started and provides them with a clear sense of direction. The disadvantage is that the structures can sometimes restrict creative thought and may lead to important dynamics being overlooked.

Other approaches encourage starting with a blank sheet and developing the structure that best captures the particular program. This approach allows for an end product that is unique to the program and that may better capture the particular dynamics of the program and the way change occurs in that community. What follows are suggestions for taking the latter approach.

The main question we are trying to answer with a change model is, **How does change happen?** In other words, how does what we do in our prevention programs lead to the goal of preventing sexual violence? In order for our programs to prevent sexual violence, what needs to change in individuals? In social settings? In organizations and systems? In culture and values? With a change model we are trying to illustrate how one change leads to another.

How to Develop a Change Model

Change models are developed through an iterative process. It is important to allocate time for many revisions. It may be helpful to start with a small group that works most closely with the program and then present the draft to a wider range of staff and volunteers. Include people who know about the issues but who are not directly part of the program. You should expect to go through many revisions. It is important to continue revising until no significant changes are suggested. At that point you will know that you have the best model possible at that time.

> It is helpful to allow plenty of time for reflection, revisions and input from multiple perspectives.

Change models should be periodically revisited throughout the life of the program to make edits based on what you have learned, how the program has evolved and whether the results you are seeing from your evaluations support the assumptions you made about the connections between different steps in the change process.

The following are some questions that can help develop your change model. Your answers to these questions will help you to clarify what the different steps of the change process should look like and how they are connected.

Note: Although we read a finished change model from left to right, some people find it easier when they are developing the model to work from right to left. In other words, start with the end goal and work step-by-step backwards as you think about what would need to happen immediately prior to this step.

- What is the problem your program is trying to solve or the issues your program will address? Are there multiple problems that are interrelated?
- What needs to change in order for the ultimate goal to be reached? If there are multiple goals, identify which changes go with which goal(s).

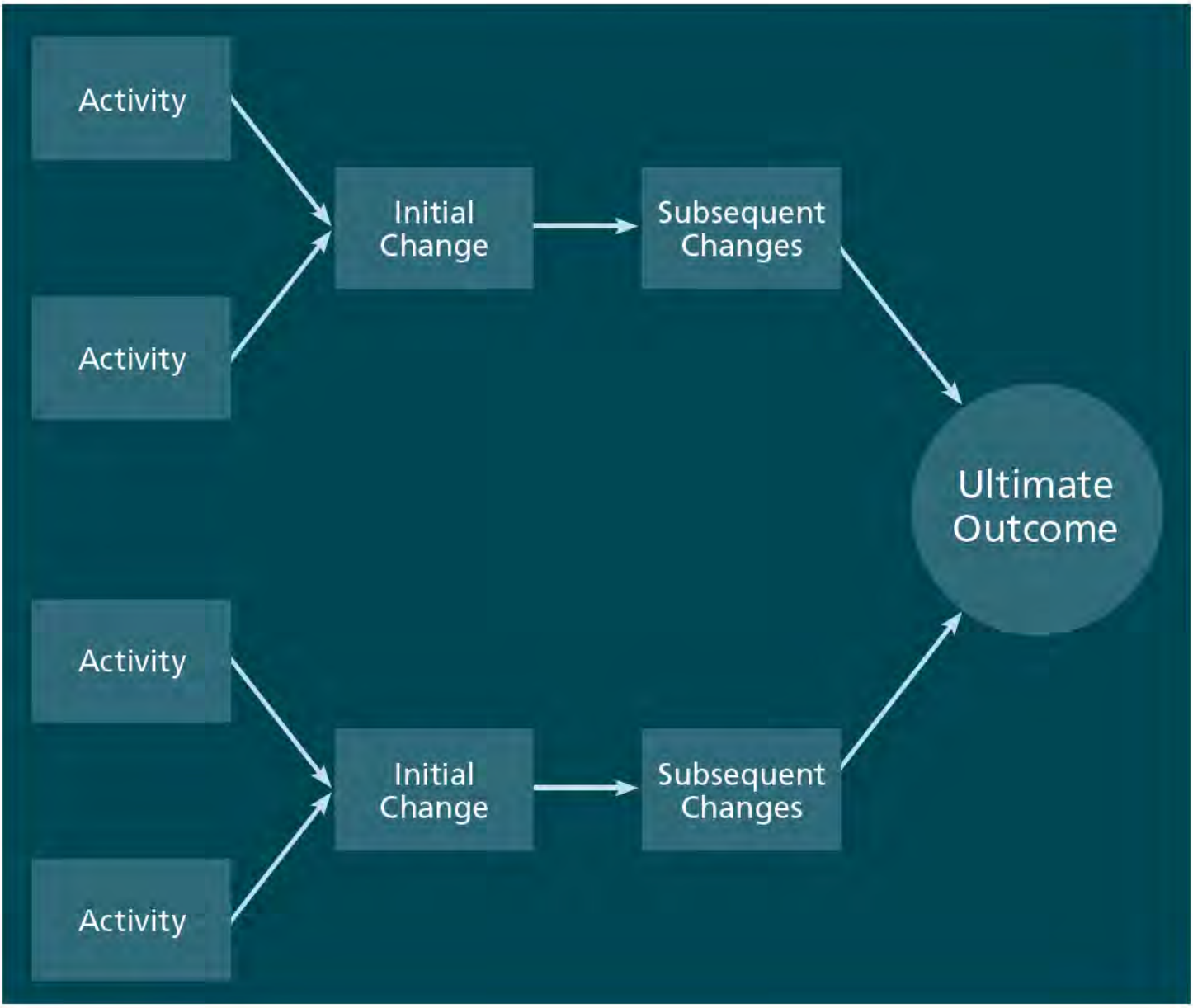
- Are there are different types of change needed, such as changes in awareness, attitudes, perceptions, knowledge, intents, skills and behaviors? Remember that because sexual violence is a behavior, primary prevention requires changes in skills and behaviors.
- For each change ask if there is something that needs to happen before it in order for the change to occur.
- What are potential barriers to change? Is there anything you need to do to address those barriers in order for your program to be successful? Significant barriers should be addressed in your model.
- What are potential supports to change? How might these supports help you accomplish your goals? Can you be successful without those supports? Supports that are necessary to the success of your program should be included in your model.
- As your model develops, do the connections between steps seem plausible? If not, then there are probably missing steps that need to be added.
- In addition to talking about the program, take some time to review documents related to it (e.g., brochures, procedure manuals, curriculum guides, etc.). What actions are reflected in these documents? What is the rationale for them?
- If the program is already running, observe it in action. What issues come up during the program? How are they responded to and what does the response say about the kinds of changes the staff are trying to bring about?
- If the program is already running, to what extent do you think you reach the audience? What helps or hinders your reaching them? Include factors that help in your model as necessary components for success. Include how you respond to factors that hinder success.

How to Draw a Change Model

Each arrow shows how one step leads to the next.

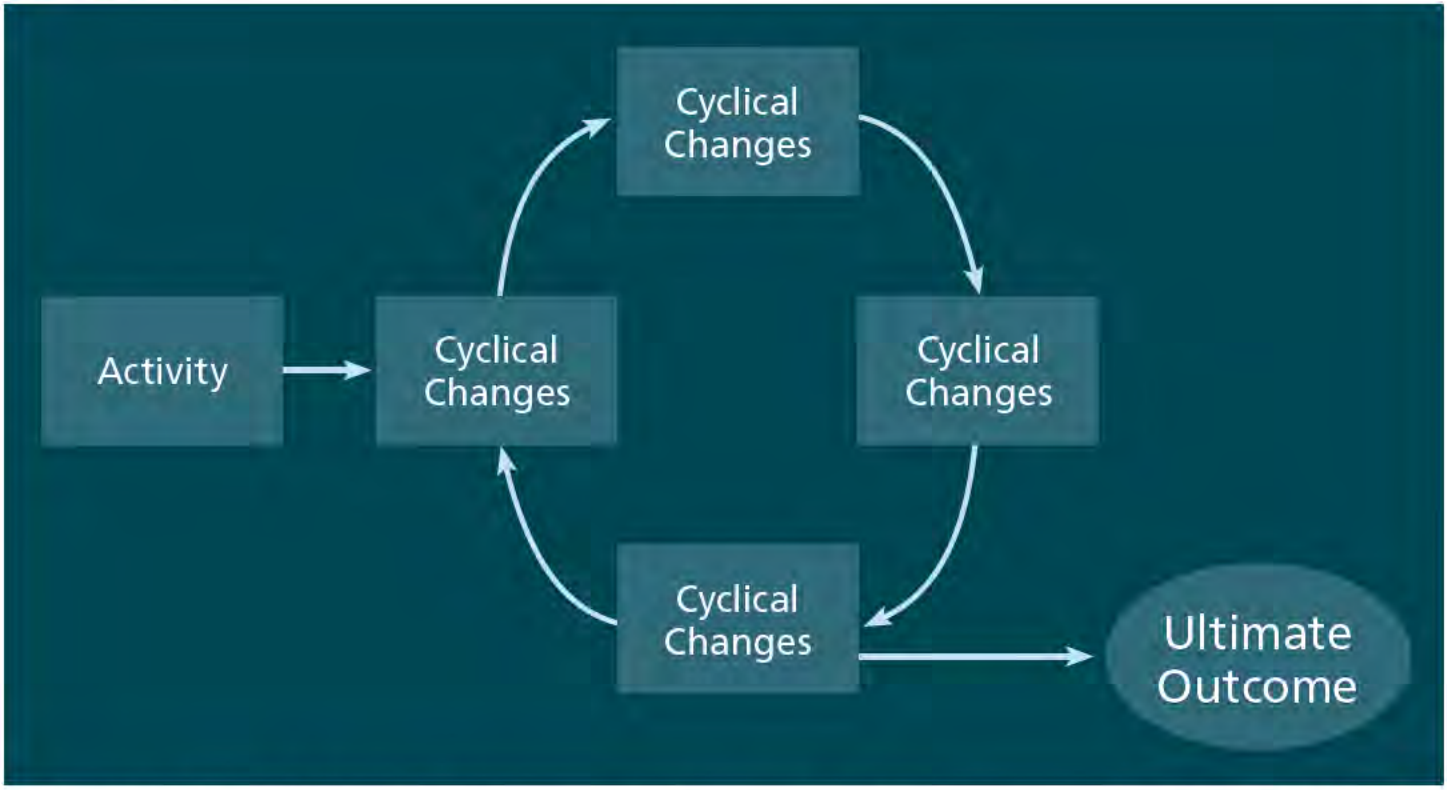
After discussing the questions above you can begin to draw your illustration of the change process. Each square or circle represents a step in the change process. Each arrow indicates that one step should lead to the next step(s). Arrows are single-headed if the change process only goes in one direction. They are double-headed if there is a back and forth or some type of iterative process. Typically, the graph is drawn so that it is read from left to right. On the left-hand side we usually see the activity. Then come the initial, direct changes that the activity should cause. Then come subsequent changes that result. This continues until we reach the far right side where we have the ultimate outcome we are looking for from the program. An example of how a change model might be structured is seen in Example 1.

Example 1:



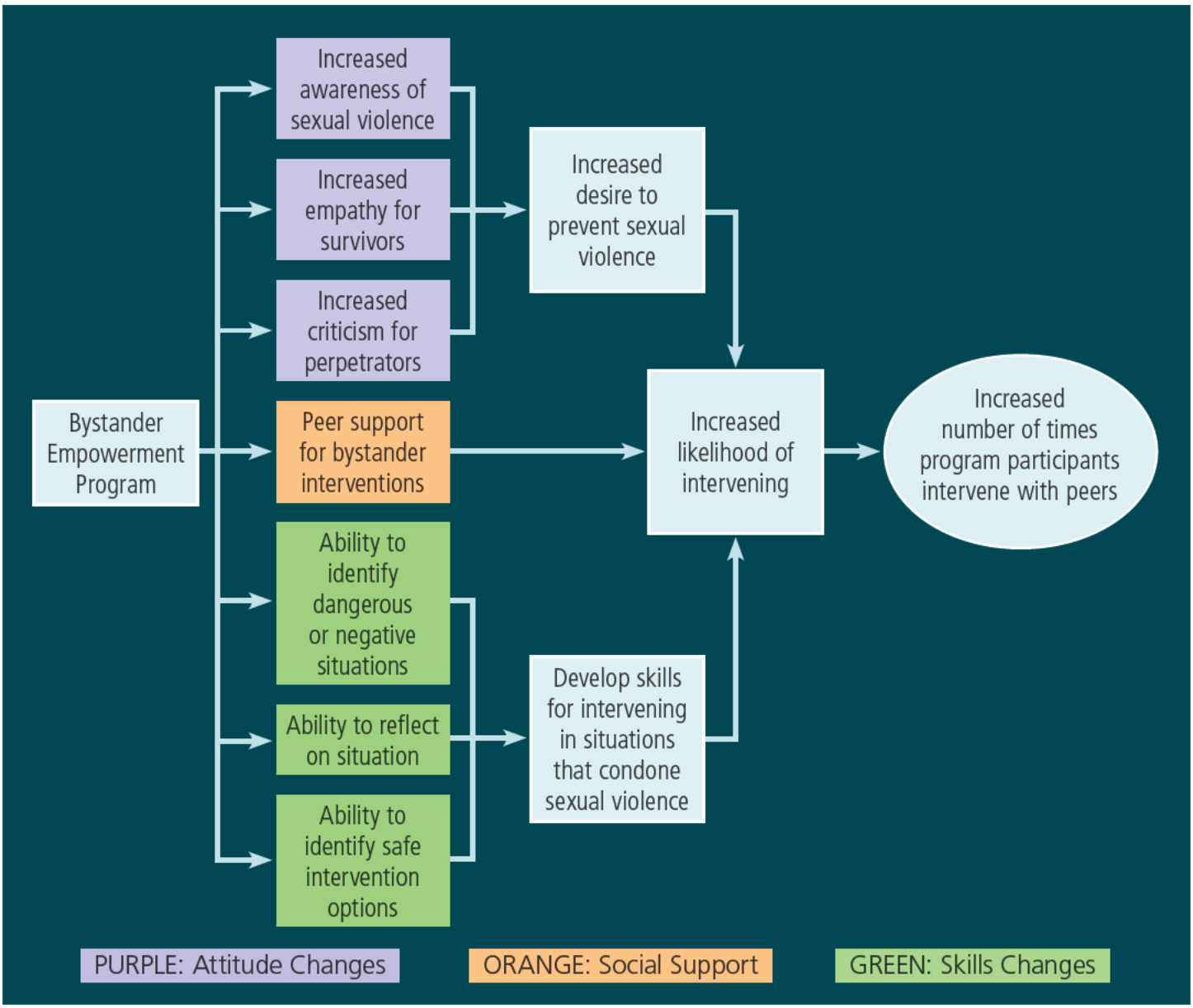
Although change models may look like a linear process, there is often an iterative nature to them as the program builds. You may find creative ways to reflect that back and forth process. Think about when it is logical to use double headed arrows. Also consider how you arrange the squares or circles—sometimes you will find that a different arrangement can capture a more complex dynamic such as shown in Example 2.

Example 2:



While you can fill in the boxes in whatever way makes sense to you, it is common that the far left boxes reflect the major activities of your prevention program, the middle boxes represent the intermediate changes and the far right box reflects the long-term goal.

The following is an example of a change model for a bystander empowerment program:



This program's change model says that the bystander empowerment program will increase participants' awareness of sexual violence, empathy for survivors and criticism of perpetrators. In turn, these three effects will combine to increase participants' desire to prevent sexual violence.

At the same time, the program will develop participants' skills for intervening in situations that condone sexual violence. Specifically, they will develop the ability to identify dangerous or negative situations, to reflect on the situation and to identify safe intervention options.

The interactive structure of the program, extensive use of peer feedback and norms of respect created in the group will also foster peer support for intervening.

The combination of increased desire, skill development and peer support will increase the likelihood of participants intervening that will then increase the number of times they intervene with their peers.

To capture the relationship-level and community-level impacts that occur the more times community members intervene, the model will need to be expanded.

Task 2: Define the Program's Goals and Objectives

Once you have a model of how your program creates change, you can define the specific goals and objectives. These goals and objectives will be much more specific and comprehensive than if you tried to write them without a clear understanding of how your program works.

What do we mean by goals and objectives? ***Goals* refer to the general effect** you want the program to have. They are stated in broad terms. ***Objectives* refer to the specific effects the program will have on the participants or community.** They are narrow statements of exactly who and what will change. The objectives will determine the data you will collect for your evaluation. Each goal must have at least one objective. These goals and objectives should correspond with your change model. For example, a program may identify one of their goals as:

Notice how these goals and objectives relate to the change model on Page 24.

GOAL 1: Participants will have greater skills for intervening in high-risk situations.

The objectives used to measure whether this goal has been achieved might be:

OBJECTIVE 1: Participants will distinguish between high-risk, moderate-risk and low-risk situations.

OBJECTIVE 2: Participants will determine how situations are consistent with or in contrast to their own values.

OBJECTIVE 3: Participants will identify multiple safe options for intervening and evaluate how comfortable they are with using each option.

A common mistake in program evaluation is having objectives that do not directly measure the goals. For example:

GOAL: Reduce sexual perpetration committed by high school boys.

OBJECTIVE 1: Boys who participate in the program will report significantly more knowledge of what sexual violence is than they did before the program.

OBJECTIVE 2: Boys who participate in the program will report significantly fewer victim-blaming attitudes than they did before the program.

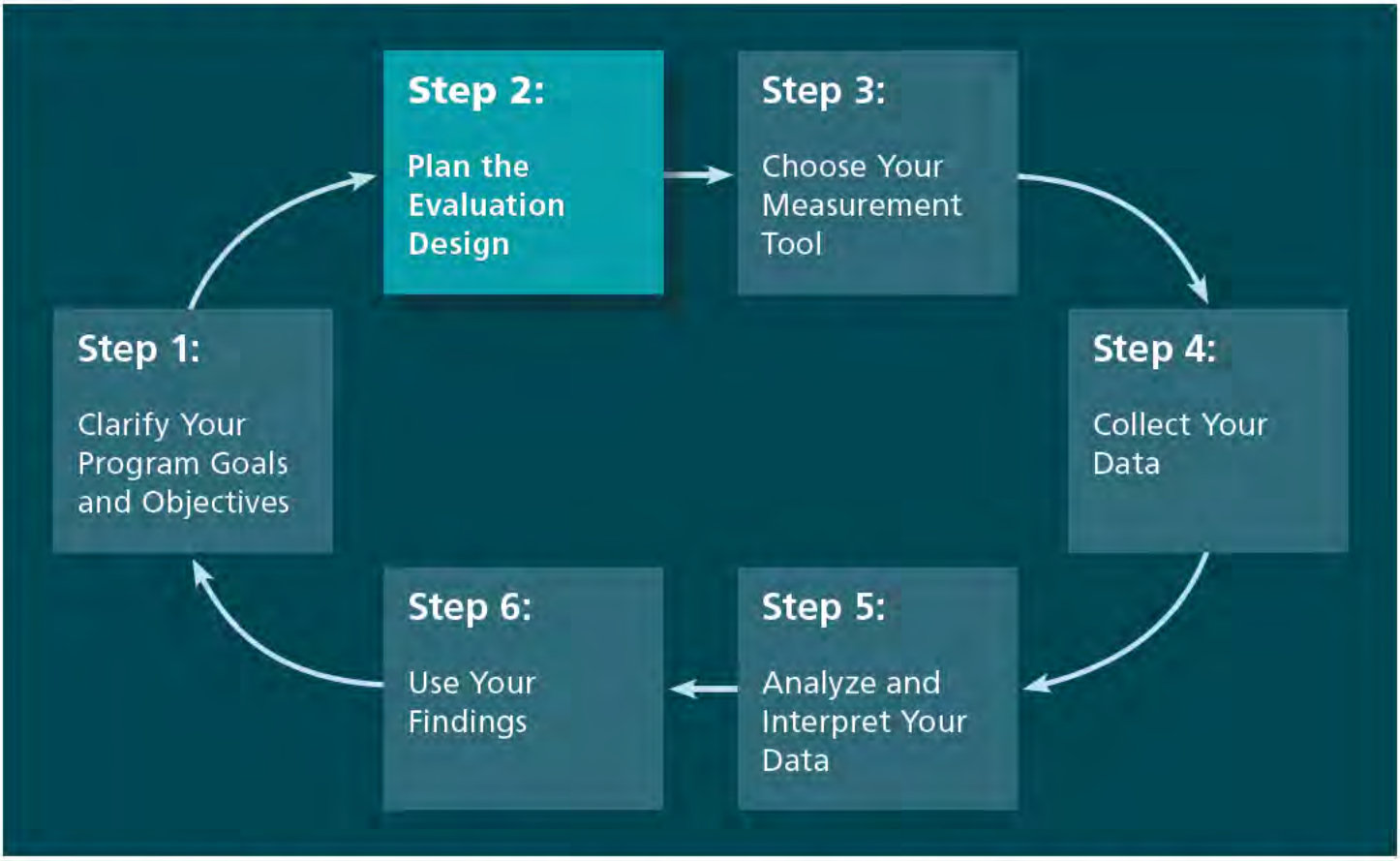
PROBLEM: The goal is to reduce actual acts of sexual violence but the objective is measuring the boys' knowledge and attitudes. Knowledge and attitudes do not change behavior. We all have probably had the experience of knowing we should or should not do something and even having positive attitudes, but not following through on those thoughts. (Think of your last New Year's resolutions!) Similarly, we sometimes do things that we said we never would do. (Anyone who is a parent has probably had this experience!)

SOLUTION: Change the objectives to be about actual acts committed within a specified time period. For example: *"Significantly fewer boys who participate in the program will report using verbal or physical coercion to get sex after participating than did before the program."*

Practical Tips: Here are some practical tips to keep in mind when clarifying your programs' goals and objectives (Campbell et al., 1998) :

- **Goals and objectives should focus on the most important parts of your program. It is easy to generate a long and overwhelming list.** By creating your change model first, you should have gained consensus among your staff and others with whom you work as to what information will be the most useful for improving your program and for making decisions about the program.
- **Make your goals and objectives easy to understand.** Avoid jargon. Your goals and objectives should be a useful tool for you when talking with the public about what you are doing.
- **Make sure that each goal and objective contains only one idea.** This will make your evaluation much more precise and will help you determine what aspects of the program are working well and which may require a different approach. (If you use each box in the middle of your change model as an objective, you should be fine).
- **Avoid borrowing goals and objectives from other programs.** Although most sexual violence prevention programs have the same ultimate mission, there can be details that don't transfer well from one program to another. It is worth taking the time to write your own goals and objectives. This way they will reflect <u>your</u> program's values, priorities, expectations and activities.
- **Be realistic about what can be achieved given the level of intensity of your program.** For example, if you are working in a school where are allowed to make a single, 40-minute presentation to an auditorium full of students, then it is not realistic to expect that you will see significant changes in the incidence of sexual assaults as a result of that presentation. However, it may be realistic to expect that students will gain some basic knowledge about sexual violence or that they will learn about support services that are available through the rape crisis program.

STEP 2: PLAN THE EVALUATION DESIGN



There are multiple designs that are used when doing program evaluation. The design you choose is based on:

- The type of evaluation you are doing (e.g., program impact, program performance, needs assessment, etc.)
- The resources you have available for the evaluation (e.g., skills, time and funding)
- How you weigh the advantages and limitations of the different designs

This manual presents the most common design used for assessing program impact by community-based programs: the Pre-Post Design. The advantages and limitations of this design will be discussed with some thoughts presented about alternative approaches.

Planning your evaluation design involves three tasks:

- Task 1: Understand the design
- Task 2: Determine the timing of your evaluation
- Task 3: Determine who will participate in the evaluation

Task 1: Understand the Pre-Post Design

The evaluation design that is recommended is what is commonly called a Pre-Post Design (with Follow Up, if possible). The design looks like this:

- **Pre-test:** The Pre-test is a way of measuring people's knowledge, attitudes, intents and behaviors before you do your prevention program. This is sometimes called a *baseline. The Pre-test can be done in many ways, including surveys, interviews, focus groups, or observations.*
- **Prevention Program:** This is your intervention. It may be a skill-building curriculum, social marketing campaign, change in a system, or other strategy that you are using to prevent sexual violence.
- **Post-test:** This is the same measure you used at the Pre-test. You give the test a second time shortly after the prevention program is completed. By comparing the results of the Pre-test and Post-test, you can see whether or not there have been changes.
- **Follow-up:** This is a third time of assessing the people you wanted to affect with your program. Most often it uses the same measure as the Pre-test and Post-test. The difference is that it is done at a much later time, usually a minimum of four weeks and a maximum of one year after the intervention is completed. The follow-up assessment lets you see whether or not the changes you saw at the Post-test are sustained over time. The challenge with using a Follow-up is that you will need to have access to participants at a later time.

The advantage of the Pre-Post Design with Follow-up is that it lets you see how program participants change over time. If you see substantial changes from the Pre-test to Post-test and those changes are in the direction you intended, then it is likely because your program was effective. Further, if you see no drop from Post-test to Follow-up then you can conclude that the changes were sustained over time.

> This design lets you see how participants change over time.

The assumption underlying this evaluation design is that **nothing else caused the changes.** This is why researchers often have a *control group.* This is a group that does not get the prevention program but is given the same tests. If the group that goes through the prevention program shows substantial changes but the control group does not, then that supports the idea that it was the program and not something else that caused the change.

Using a control group requires additional resources. It can also be difficult for a community-based program to convince the public to withhold the program from a group (such as from one school) but still invest in testing that group. For this reason, the use of a control group is not often used by community-based programs.

In order to be more confident that the changes you see are in fact due to your program and not something else, it is important to take note of other events that might be influencing the outcomes. For example, it is worthwhile to keep a log of:

- Major news stories about sexual violence.
- Major events in the community that could also influence the outcomes of your program (e.g., Take Back the Night rallies, Clothesline Project displays, etc.).
- Major changes in personnel in the settings where you do the program (e.g., changes in school administrators, counselors, or teachers who may be addressing sexual violence outside of your program).
- Major changes in community-based systems that respond to sexual violence (e.g., law enforcement, prosecution and social services).
- New laws about sexual violence.
- Other programs, campaigns, or interventions that your agency is running simultaneous to the program you are evaluating.

To strengthen your evaluation, take note of any events outside of your program that might influence the outcomes.

You may not be able to determine how much of the changes you see are due to your prevention program and how much are due to these other events. But if you are aware of other possible influences then you can make a reasonable judgment about how to interpret your results.

Task 2: Determine the Timing of Your Evaluation

This is a fairly straightforward task. However, there are some things to keep in mind:

- **Your Pre-test needs to happen before the program begins, but not too far ahead of time.** For example, if you do the Pre-test in September but the intervention doesn't start until December, there could be important events that occur in between those two times that impact your results. To make it easy, many programs do the Pre-test right at the start of the program. However, this can take precious time away from your program and it can also affect the experience people have of the program. There is no best time to do a Pre-test, but it is worth doing it a few days to a week before the program starts. Otherwise, do it immediately before the program begins.
- **The Post-test needs to happen after the program ends, but not too soon or too late.** If you do the Post-test the same day that the program ends, then people may give answers that they remember by rote but have not really internalized. You are also more likely to get the answer that they think you want to hear. Whenever possible, it is best to delay the Post-test for a short time, perhaps a day to a week. This will be a better assessment of what the participants actually retained when they left the classroom or other setting in which the program was delivered.
- **The timing of the follow-up should be determined by the timeframe you are interested in as well as what is feasible.** If possible, do multiple follow-up assessments, for example at four weeks, 12 weeks and one year. This will let you see how quickly (if at all) the effects are wearing off. Although a follow-up is ideal, it is often not feasible for community-based prevention programs.

Task 3: Determine Who Will Participate in the Evaluation

You do **not** need to collect survey or other data from every participant in your program.

It is easy to think that the more surveys you administer, the better. This is not true. As long as you have collected data from a group that fairly represents your audience, you can draw reasonable conclusions. Collecting more data may simply mean more work for your staff.

If you choose to collect from only a sample of participants, keep the following ideas in mind:

- **Getting a good sample starts by understanding the larger group.** To use an obvious example, if your program is intended to reach both high school boys and high school girls then your sample needs to fairly represent both genders. If you do presentations to all ninth grade students and there are 50% girls, then a sample that consisted of 80% girls would underrepresent the boys. When working in school settings some factors to consider in determining whether your sample is adequate are: gender, age/grade, academic performance, ethnicity, students in special education classrooms or receiving academic support services, socioeconomic status, language and literacy. The question is: How well does your sample represent the students who receive your program?
- **The default way to sample is randomly.** Of course it would be awkward to only survey some students in the same classroom. So it is more common to randomly select classes or to randomly select schools (if the schools you work in have similar demographics). Usually random sampling will result in adequate representation. The number of people you have in your sample will depend in part on the type of analyses you plan on doing. It is useful, therefore, to plan your analyses ahead of time, working with a consultant if necessary. In general, if you are testing changes in attitude or knowledge a sample size of 50–100 people is probably sufficient. If you are testing changes in behavior you will want a larger sample.
- **Randomly sampling people does not mean that you cannot be strategic as well.** If there is a particular group that is very small then a random sample might not include them. For example, if you randomly select half of the classrooms that you do your program in, but there is only one special education classroom, then there is a good chance that the special education classroom might not be included. Knowing how well the program works for them could be very important. In this case you can do a random selection of the mainstreamed classrooms and also survey the special education classroom.

Evaluating Other Kinds of Programs and Efforts

Although the Pre-Post Design with Follow-up was described here in regard to educational programs, it can also be used to evaluate other types of community efforts. The following are a few examples that may help in thinking about your own evaluations.

Social Marketing Campaigns

Whether you are doing a PSA campaign or some other type of social marketing, you can use a similar evaluation design. You could develop a community survey to evaluate the campaign's objectives and administer it to people (A) before the campaign starts, (B) immediately after the campaign ends and (C) at a later follow-up date. The biggest difference here is who you assess. You have two main options.

Option #1 is to survey the same people at all three times. This ensures that any changes you see are really changes in those people and that you didn't get different results simply because you surveyed different people. The disadvantage to this strategy is that it is very difficult to keep the same group of people over time. It takes resources on your part to track people over time. Some people will drop out because they lose interest or for other reasons. When people drop out that makes it more difficult to know if any changes you see are true changes in the group or simply due to the fact that certain people didn't take the next survey. For this reason it is important to record as much relevant information as possible about the participants so that you can determine whether people who drop out are somehow different from the people who complete all of the surveys.

Option #2 is to survey different people at each time. This may be a much easier approach in terms of your resources. However, the disadvantage to this strategy is that it's harder to know if any changes you see are true changes in the community or due to differences between the groups. There are ways of dealing with this, for example by making sure that each group has the same demographic make-up and recording information about any factors that you reasonably think could influence the outcomes. This approach requires a larger sample size.

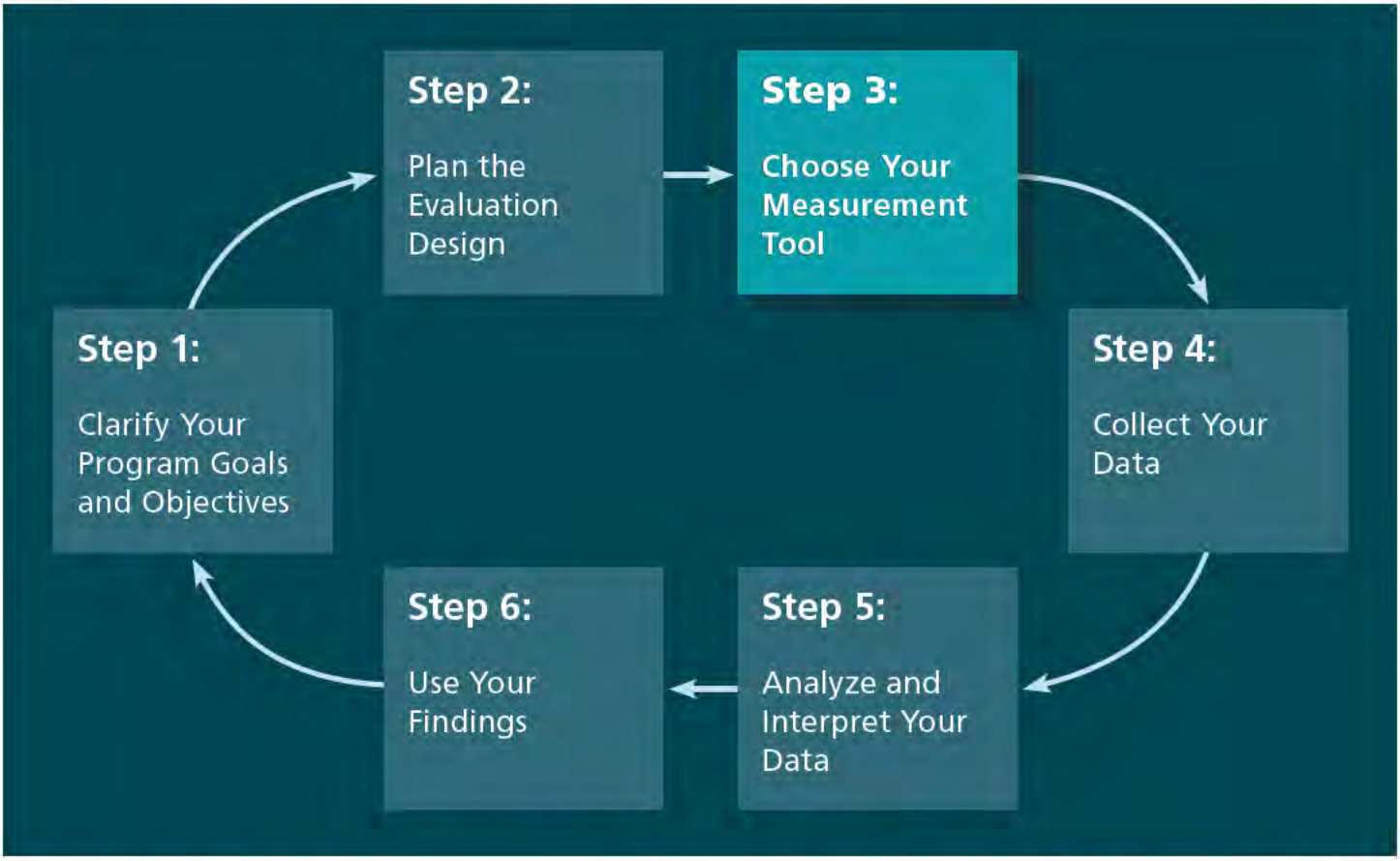
Systems Changes

Your prevention effort may involve changing the way systems work, such as school disciplinary procedures, law enforcement or medical responses, or social services. These efforts can also be evaluated using a similar design. The major difference is that your objectives are likely to be focused on actions taken within the system. In this type of evaluation your data often comes from system records. You may be interested in who takes certain actions, how long they take, the results of the actions, etc. The evaluation design is still basically the same: (A) review documents before the change takes place, (B) review them again after the change has been implemented for a time long enough to expect to see change and (C) review them a third (or more) time.

There are a few things to keep in mind when using records as your source of information:

- Records are rarely complete. You will need to have a plan for how to deal with missing information (e.g., ignore it, substitute an average answer, estimate it based on other information, etc.)
- Old records may not have the information that you need to evaluate your objectives. In the course of planning the system change you will ensure that new records will have the required information, for example by creating new forms, but you may not have comparable information to compare with if you are using existing records for your Pre-test.
- Systems are a complex web of interactions. Changes made in one part of the system can lead to changes in many other parts of the system. Therefore, it is important to have a good understanding of the entire system you are working in so that you can be reasonably sure that any outcomes you see are a result of your program and not of some other change in the system.

STEP 3: CHOOSE YOUR MEASUREMENT TOOLS



Once you have figured out the design for your evaluation, you need to select the actual tools you will use to measure effectiveness. There are two main tasks in this step:

- *Task 1: Select the* type of measure you want to use
- *Task 2: Select the* specific measure you will use and modify it as needed OR create your own measure

Task 1: Select the Type of Measure You Want to Use

There are many types of measures, each with their own advantages and disadvantages. The most important factor determining which type of measure you use is **what type of measure will give you the richest and most useful information.** Different questions may be best answered using different measures (see table on next page). You will also want to consider issues of feasibility, including:

- the time it takes to use a particular measure
- whether your staff has the requisite skills to use it
- how receptive participants will be to the measure

There are four main types of measures you can use to measure the effects of your program:

	GOOD FOR ASSESSING	ADVANTAGES	DISADVANTAGES
Surveys	■ Knowledge ■ Attitudes ■ Intentions ■ Behaviors	■ A quick and inexpensive way to get information from a large number of people ■ It's easy to be consistent in how you administer the surveys ■ Analyzing surveys is relatively straightforward	■ Writing a good survey is harder than many people realize ■ It's easy to get flooded with surveys and for inputting data to take longer than expected ■ Behaviors are self-reported
Focus Groups	■ Attitudes ■ Opinions ■ Interpretations	■ Let you get more in-depth information ■ Discussion among a diverse group of people can lead to insights that you would not get from individuals ■ Relatively low-cost and low-time investment	■ Results will be influenced by group dynamics; requires skill in group facilitation ■ How to interpret the group discussions is not always self-evident
Interviews	■ Attitudes ■ Opinions ■ Interpretations ■ Motives ■ Experiences	■ Let you get in-depth information ■ Participants may disclose information and details that they would not write about on a survey or talk about in a focus group	■ Time intensive ■ Being consistent across interviews is challenging ■ Requires good interviewing skills ■ How to interpret the interviews is not always self-evident
Observations	■ Behaviors ■ Environments	■ Record actual behaviors versus self-reports ■ Gives insight into interactions between individuals and their physical and social settings	■ Need to have clear definitions of what you are looking for ■ Requires good observation skills and the opportunity to witness the behaviors ■ Difficult to be consistent across observations

In selecting the type of measure you want to use, keep in mind the following:

- **The type of measure you use must match the goals and objectives of your program.** In some cases this leads to more than one option. For example, three of these methods are good for assessing attitudes.
- **When you have more than one option, the type of measure you use will depend on weighing** the advantages and disadvantages as well as determining the skills, time and other resources you have available.
- **The most common types of measures used by community-based programs are surveys.** However, they are not the only option nor are they always the best option. Carefully consider the kind of evaluation questions you want to answer, the resources you have available and your audience. Then decide if surveys are the best option for your evaluation.
- **Focus groups are too often overlooked as a useful and very feasible approach.** Focus groups are small group discussions that get in-depth information on specific topics. Although they are facilitated, the goal is to get participants talking with one another. The facilitator's role is primarily to get conversation going and to keep it going. You may want to consider using focus groups as either a replacement for or a supplement to surveys if you are finding that surveys aren't giving you the rich details you want or you are left with too many unanswered questions.
- **You may want to use different types of measures to answer different evaluation questions.** For example, you may want to use surveys to answer questions about changes in participants' knowledge, attitudes and behaviors. This could be augmented with focus groups to explore how different parts of the program impacted people in different ways and what their experience of the program was.
- **Samples** of measures are found in the Supplemental Volume of this toolkit.

Task 2: Select the Specific Measure You Want to Use

Make sure your measure fits your program and audience.

Once you have selected the type of measure, you need to choose the specific tools you will use. Because surveys are the most common type of measure used and there are so many available, here are tips for selecting surveys:

- **The specific survey you use must match the goals and objectives of your program.** If they don't, then you will be evaluating something, just not the thing you want to evaluate. Each question that you include in your survey must relate to your outcome goals and objectives. If a question doesn't relate, then don't ask it.
- **If your goals are about changing behaviors, then assess behaviors.** If your goals are about changing attitudes, then assess attitudes. If your goals are about increasing knowledge, then assess knowledge. If your goals are about changing intents, then assess intents. This may sound obvious, but often there is a mismatch between our goals and what we ask on surveys.
- **Make sure that the survey is a good fit with the people who will be using it.** Think about reading level, language and cultural sensitivity. Consider attention spans and survey length. Make it appear interesting — visual layout can go a long way toward maintaining interest, especially with teenagers.
- **Writing surveys that give you consistent answers and measure what you want is harder than it seems.** For this reason, you may want to start with a survey that has been written and tested by researchers. You may need to modify it to fit your program goals, objectives and audience. But starting with a survey that has already gone through a process of testing and revision will get you started on the right track. Be cautious in making changes. Measures that have gone through careful development have often been determined to be what researchers call *valid and reliable measures. Altering the questions or range of answers can change the validity and/or reliability. However, changes may be made to ensure that the survey fits your program goals and objectives or that it is understandable to your audience*. There is no reason to reinvent the wheel. You may find a survey that fits your program very well. **But this only works if the survey matches your goals and objectives.**
- **If time permits, include a couple of open-ended questions.** These are questions that the respondents answer in their own words, as opposed to multiple choice or true/false questions. The advantage to open-ended questions is that people can give their own answers and don't have to make their thoughts or experiences fit the narrow multiple choice categories. The disadvantage is that most people will write short, uninformative answers. You can improve the quality of the answers you get by using open-ended questions sparingly and only for questions that you think people will be interested in writing about.
- **Examples** of surveys are found in the Supplemental Volume of this toolkit.

Make sure questions yield clear and honest answers.

If you are modifying a measure or writing your own from scratch, keep in mind the following Guidelines (Campbell et al., 1998):

- **Try to give people a range of responses from which to choose.** It's better to use a scale such as *strongly disagree, disagree, feel neutral, agree, strongly agree* than it is to use "yes/no" answers. Using scales like these is less limiting and lets you measure change better. For example, if a person changes their belief about a specific question a little bit but all that they have are "yes" and "no" for options, they will probably answer the same way on both surveys. But if they have more choices then you might find that they went from "strongly agree" to "agree."
- **Make sure that you are only asking one thing in each question.** For example, imagine people are asked to say how much they agree with this statement:

  "Advertising and music portray women as sex objects."

  What if someone thinks that only one of these portray women as sex objects? How are they supposed to answer the question? This problem is easily solved by breaking the statement into two different questions. Any time you use the word "and," double check that you are truly only asking one question.
- **Make sure that the answers you give for people to choose from are mutually exclusive.** To illustrate this, think about the question:

  "How much do your teachers talk about sexual harassment?"
  They talk a lot about it
  They talk enough about it
  They talk a little about it
  They don't talk at all about it

  Talking "enough" is not exclusive of the other answers. A student could think that teachers only talk a little bit about sexual violence, but that is enough because they don't think it needs to be talked about more.
- **Avoid loaded questions that may bias people's answers.** For example, if you start a question with *"Do you agree that…"* you are implying that people should agree at least a little bit with what you say. Instead, you can ask *"How do you feel about…"* or *"What do you think about…"*
- **Give people permission to give unacceptable answers.** When asking about controversial issues or behaviors it is important to give people permission to give their honest answer even if it is not socially acceptable. For example, most people if asked whether it is okay to coerce someone to have sex will say "no" because they know it's not acceptable to do so. However, they may actually approve of using coercion in some circumstances or they may think that certain types of coercion are okay. Therefore, you need to phrase the question in a way that indicates that you know that people do use coercion sometimes. For example, *"How often have you…"* or *"Under what circumstances would you…"* make it easier for people to admit to unacceptable behaviors, intents, or opinions. You still need to give "never" as one of the possible answers.

Decide if responses will be anonymous or confidential.

If you are using a survey you need to decide whether the answers will be anonymous or confidential. **Anonymous measures** are ones that do not contain any information that can identify the person who gave the answers. In the case of surveys this means that there are no names or other unique identifiers written on the survey. You may ask for demographic information. The bottom line of anonymous data is that there is no way you can match the answers to an individual.

Surveys that ask about satisfaction with the program should always be completed anonymously**.** There is no reason to have identifying information on these surveys and you will get more honest feedback on what participants liked and did not like about the program if you make them anonymous.

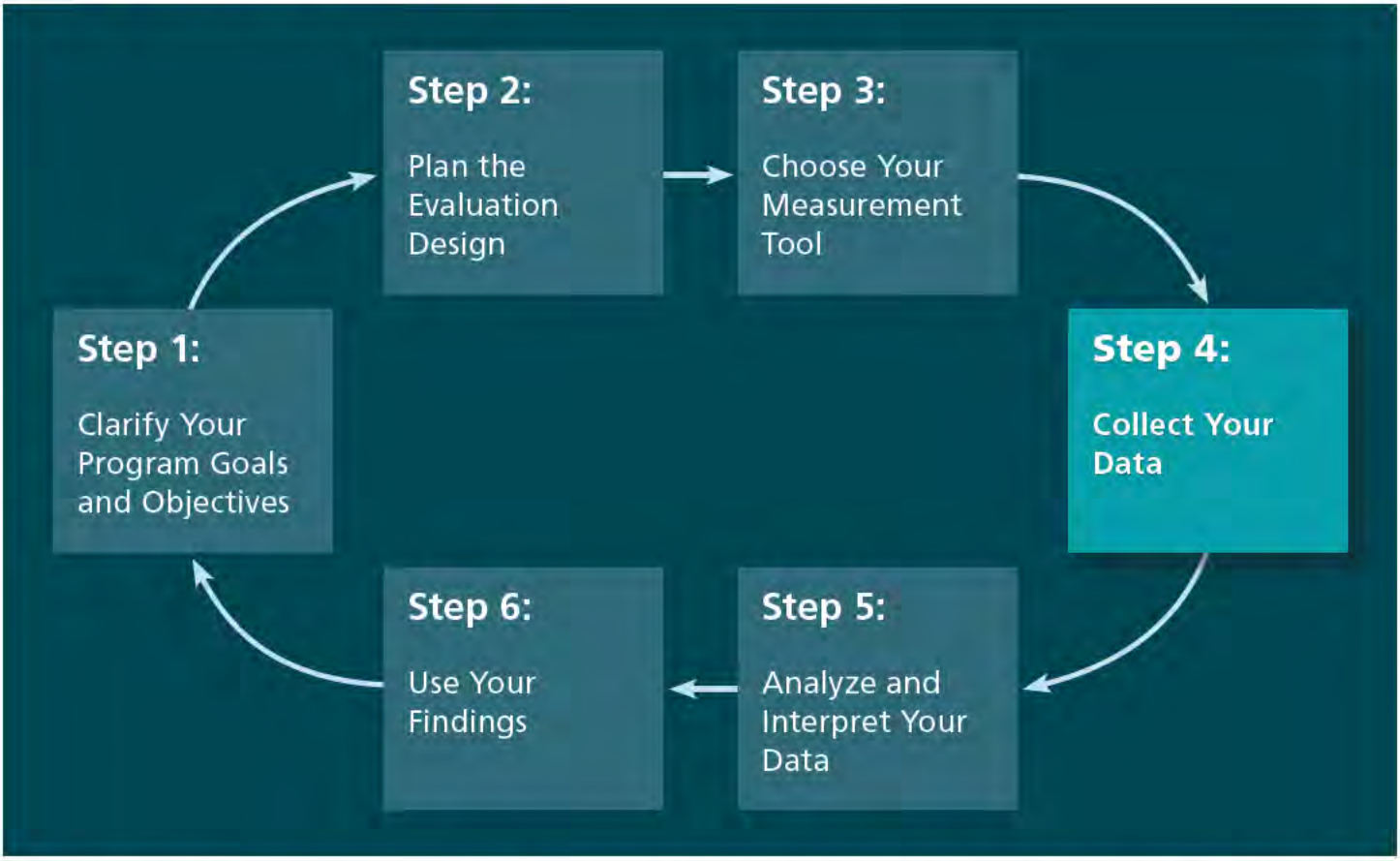
Anonymity in focus groups and interviews is slightly different because you typically have some knowledge of who the people are due to how they were invited to participate. However, there may be times when an outside person (e.g., a teacher or school administrator) has chosen the participants and you do not have any personal knowledge of them. In either case, you can provide anonymity for their answers by using pseudonyms in the notes you take and when you write up your results in reports.

Confidential measures are ones in which participants' names are not used, but there is some kind of ID number used instead. Why would you need to use an ID number?

- **When doing a pre-post survey, you must use ID numbers so that you can match the surveys.** Part of the analysis that is done on pre-post surveys requires that we know which surveys go together. However, it does not require that you know who actually completed the surveys. A common way to assign ID numbers is to use the last four digits of a person's phone number. Or you can make up a number using questions that will give a unique combination of answers when strung together (e.g., # of siblings, # of pets, street address). In either case you can reassure participants that you have no way of connecting their ID number to them.
- **ID numbers may be used for necessary administrative purposes.** For example, if you are compensating participants for participating you may need to be able to verify that they completed the survey before mailing them the check. In this case, in truth you could connect their answers to them. Therefore, you need to explain that the list of ID numbers and matching participants will be kept separate from the surveys and that no one other than your staff will have the list.

Whether your measures are anonymous or confidential, you need to explain to the people participating in your evaluation which they are and what that means. Most people, especially teenagers, do not realize that there is a difference between the two terms. **Teenagers will be especially concerned about whether their parents/guardians or teachers will find out their answers.** They can come up with some remarkable ideas about the ways program staff and teachers can figure out which survey belongs to which person. So you will need to be prepared to explain anonymity and confidentiality and to get them to trust you. Even with adults these reassurances are important, especially if you are asking behavioral questions.

STEP 4: COLLECT YOUR DATA



Before you collect data you should think carefully about exactly what you are going to do. **The goal is to ensure that all of the information is collected in a similar manner.** This is true whether you are using surveys, focus groups, interviews, or observations. You want to eliminate any variations that could influence your findings. The following are some basic guidelines to consider when collecting your data.

Surveys

Be clear in your instructions and anticipate likely questions.

While surveys seem to be the simplest approach to evaluation, there are still ways you can accidentally introduce unintended variations that can affect your findings. These variations often happen when you have multiple people who are collecting the data or when you are collecting data in multiple settings. While you may need to show some flexibility, especially when working with schools, you cannot be so flexible that you compromise the evaluation. School personnel understand the need for demonstrating that programs are effective and for using standardized procedures, so a brief explanation of your evaluation procedures should go a long way toward obtaining cooperation. When planning how to collect your survey data, consider the following:

- **Ensure that you have enough time for the surveys.** If participants hurry through the questions then they will not give thoughtful answers. They may also take the survey less seriously, leading to flippant answers. The need for sufficient time is another reason it may be good to do the surveys at a different time than the program itself (e.g., the day before and the day after).

- **Script out the instructions.** It is important that all participants be given the same instructions and explanations. Some points to include when giving instructions are (Campbell et al., 1998):
  - Let the group know what they are being asked to do (i.e., complete a survey).
  - Tell them why they are being asked to fill out a survey and how their responses will be used (e.g., to find out what parts of the program are working well and to improve it).
  - Explain how the surveys are anonymous OR confidential.
  - Tell them how to complete the survey. Be as specific as possible (e.g., circle your answers on the page).
  - Reassure them that this is not a test and that they will not be graded or judged for their answers. Emphasize that you are interested in hearing what they think. Encourage them to be honest.
  - Tell them how long you think it will take them to finish the survey.
  - Encourage them to ask questions if they don't understand a question or if they are unsure of the instructions.
- **Script out answers to common questions.** In addition to the instructions, it can be useful to anticipate some of the common questions participants might ask and to script out the answers. This list can be revised over time. Note: As mentioned earlier, youth are especially prone to worrying about whether or not you can figure out which survey belongs to which person. For example, if you are using the last four digits of phone numbers as ID numbers it is common that someone will say, "But the school knows what all of our phone numbers are." It's also common to have a student say, "But the teachers know what our handwriting looks like." Think about how you want to alleviate these concerns and make sure that similar answers are always given.

Focus Groups

Foster natural conversation and interactions among participants.

If you run more than one focus group there will be variations between them due to the different group dynamics. This is fine. However, the basic procedures you follow from one group to the next should be the same even if the conversations themselves unfold in different ways. Keep the following guidelines in mind when running focus groups:

- **Who participates in your groups will greatly influence the results.** It is generally recommended that each group represents one "audience" for your program. For example, if you are interested in the experiences of both students and teachers you would want to run two separate groups. Within the group, however, you want a diversity of views. For example, if you are running a group with teachers you might want to make sure that the group includes new and experienced teachers, teachers who instruct core subjects as well as those who teach electives and others who have a role with students such as coaching staff.
- **Keep groups small.** Typically groups range from 6–12 people. Fewer than six can make people feel too scrutinized. More than 12 makes it difficult for everyone to participate.
- **Choose facilitators wisely.** You want to use facilitators with whom participants will be comfortable, especially when working with youth. If you have a mixed gender group then it is

best to have both a male and female facilitator. Even if only one person facilitates while the other takes notes, the mere presence of both male and female facilitators can put people at ease.

- **Have a designated notetaker.** The person(s) facilitating the discussion should not be burdened with simultaneously taking notes. They should be able to focus completely on the people participating in the discussion.
- **The facilitator should have a guide to work from during the group session.** The guide should include introductions, a few general questions to get discussion going on the major topics you are interested in and a few follow-up questions to make sure that you have gotten the information you want. The guide should be flexible and short (3–6 focal questions). It is not a rigid agenda. Discussion should be allowed to flow naturally with participants responding to and feeding off of one another's comments. The guide is mostly a way to get discussion started, to jump start discussion when there are extended lags and to ensure that by the end of the focus group the essential evaluation questions have been answered. By being flexible and following the natural flow of the discussion, you will find out information that you never even thought to ask about, which is one of the advantages of focus groups over surveys.
- **Focus groups usually last 60–90 minutes.** You want ample time for discussion, but not so long that it becomes burdensome for participants.
- **Make sure the facilities are comfortable.** Participants should be able to see each other; a circle format is usually best for promoting discussion. Make sure that chairs, temperature and lighting are as comfortable as possible. Having snacks available can help people relax.
- **Check equipment ahead of time.** Focus groups are usually taped, either with videotape or audiotape. This is done because the discussions often jump around a lot and can move at a quick pace. It is difficult to rely on notes alone. If you do any taping, the equipment should be unobtrusive (although participants must be informed that a tape is being made), which includes testing it ahead of time so there is no awkward and time-wasting dealing with equipment. All sound checks should be done prior to the participants entering the room.

Interviews

Interviews are an excellent way of getting rich, detailed information from people. The key is to get people talking. Like in advocacy and counseling, you want to be **more focused on listening to what they are saying and less focused on what you are going to ask.** While well-crafted questions and following up for more details make for a successful interview, don't be so focused on what you want to ask that you forget to listen.

> "Get an informant on a topic of interest and get out of the way."
>
> — H. Bernard Russell

There are a number of considerations to keep in mind when doing interviews.

- **People who are willing to be interviewed have their own reasons for participating and they are not necessarily the same as your reasons.** It is often helpful if you can think about why people would want to participate in your evaluation interviews. You need to figure out what the major reasons are someone would want to be interviewed and then make sure that their needs are met during the course of the interview. When interviewing survivors the driving need is often for the survivor to tell her or his story. Some researchers have found that it's helpful to start out by asking, "Tell me your story" or "Tell me why you wanted to do this interview." This way the survivor can put out there what she or he wants to say and have it be heard in an empathic way. Some of what is said will be relevant to your purposes, some of it will not. But you can then proceed with the interview and get what you need out of it and the person being interviewed has also gotten what she or he needs out of it.
- **The interview should be focused but not rigid.** The most common approach to interviewing is what researchers call a semi-structured interview. In this approach you have some main questions that you want to ask (a good number is 4–10 questions). These questions are the same for all interviews. For each question you might ask some additional questions to clarify the answer or to get more details about specific parts of the answer. These additional questions, usually called probes, are going to depend in part on what the individual person says and how articulate they are in their answers. There may be variation between interviews in how you probe the answers. Although the main questions are usually asked in the same order, this approach allows for variation. For example, if you're on Question 3 and the interviewee spontaneously starts talking about something related to Question 6, you will want to go with the flow and jump ahead rather than asking them to hold onto those thoughts so you can come back to them later. The order you ask the questions doesn't matter as much as getting rich, meaningful answers. That richness depends in part on establishing rapport, helping the person feel comfortable and having the interview feel natural.
- **Use open-ended questions.** The point of interviews is to obtain a rich, detailed understanding of people's experiences. In the case of program evaluation, it is to understand the needs they see in their community (needs assessment), their experiences of participating in a program (process evaluation), or the impact the program has had on them (impact evaluation). In order to get this rich, detailed understanding, you need to ask questions that are sufficiently broad enough to capture many different experiences, including ones that you would never think to ask about directly, yet sufficiently narrow enough that people understand what you're asking.

- **You need to know what constitutes an "adequate" answer so that you can probe more when needed.** Although we are not looking for "right" answers, we do have a sense of how much depth we are looking for with our questions. There is nothing more frustrating than doing an interview and then later realizing that you should have asked for more details. Think ahead of time about what you mean by each main question and what would make for a full answer. Use that to think of how you can probe for more details. Remember the basics: who, what, when, where and why.
- **Always end the interview on a positive note.** Concluding questions should be slanted toward positive experiences. Depending on what your evaluation is about, these may be questions about what they liked about the prevention program, the strengths they see in their community for responding to sexual violence, hopes that they have, etc.
- **Be prepared for disclosures.** Although what we are talking about here are interviews for the sake of program evaluation, you may still have people disclose their experiences with sexual violence. Be supportive. If the person is distraught then you may have to stop the formal interview and go into crisis intervention mode. PCAR's statewide hotline connects callers to the nearest rape crisis center: **1-888-772-PCAR**. However, if they are not in crisis then you want to provide them empathy and information without having it derail the interview.
- **Use the interview as an opportunity for education.** Regardless of what the interview is about or whether or not a disclosure occurs, this is an opportunity to provide information and resources. Have a short handout available to give to all participants at the end of the interview that provides some basic, supportive information about sexual violence and that includes a hotline number or other pertinent resources.
- A list of interview questions that you might find helpful is found in the Supplemental Volume of this toolkit.

Observations

There are many ways to do observations. The approach that may be most useful for program evaluation is the use of **structured observations**. This method may be especially helpful when conducting needs assessments. It can also be used to assess program impact. Sometimes the same type of observation can be used for both purposes. For example, documenting that youth frequently use coercion when interacting with each other can be used to justify the need for a program that aims to reduce the acceptance of and use of coercion. You can then use that same method of observing as a type of pre-post test to determine whether youth are, in fact, using coercion less frequently.

Some of the types of evaluation questions that can be explored through observations include:

- In what ways do youth in your community maintain power over their peers, especially in interactions between genders? How common are acts of verbal coercion, social manipulation (e.g., ostracizing, cliques, etc.) and physical manipulation?
- How do peers respond when they witness acts of power over another person? How do adults respond?
- How common are public acts of violence among youth? What forms of violence are most common?
- How do peers respond when they witness acts of violence? How do adults respond?

- How widespread are images of sexual objectification, exploitation and violence? Where are these images accepted? Where are they prohibited, explicitly or implicitly? Who generates them (e.g., images generated by media, graffiti, personal decorations, etc.)?
- How accessible is pornography to adults in your community? To youth?
- Where do youth gather in your community? When are they most often there?

Observations can also be used to evaluate skill-building curricula. For example, you might observe:

- Bystander strategies used during role plays
- Assertiveness skills used during role plays
- Assertiveness skills (and negative communication) used during group discussions
- Acts of respect (and disrespect) displayed during group discussions and exercises
- Materials and products created during curriculum activities

Structured Observations Involve Four Main Tasks.

Task 1: Identify Behaviors

The first and most important step in doing structured observations is to identify behaviors that are related to your evaluation question(s). You must have clear definitions for the behaviors that you will count as evidence. To do this, you will want to consider a few questions.

- **Are there relevant subcategories of behavior?** For example, if you are interested in coercion you might want to think about ways that people can coerce verbally, socially and physically.
- **In each category, what specific behaviors will count?** You need to do an exhaustive brainstorm to make sure you don't leave any important behaviors out. Keep in mind that "behavior" is being used here in a broad sense. Behaviors may involve interpersonal interactions, but they can also include environmental characteristics (e.g., what types of posters are on display, where are pornographic magazines located in the store, what type of graffiti do youth draw, etc.). Also, behaviors can include what people do not do. For example, when youth commit abusive acts and their peers ignore it, the ignoring is as much an act as speaking up on behalf of the victim.
- **In each category, what behaviors will not count?** It can be useful to think of the kinds of behaviors that an observer might think to count but that do not actually fit what you mean by that behavior. For example, if you want to assess how youth respond when their peers act abusively, does it not count if the response involves physical or verbal aggression? In other words, do you only want to count positive, assertive responses?
- **Whose behaviors will count?** In any setting there may be people whose behaviors you are interested in as well as people whose behaviors you do not want to include. Be clear about whose behaviors will count. For example, you may be interested only in youth behavior, or only in behavior between two people, or only in mixed-gender groups.

Task 2: Develop a Dictionary

Once you have defined the behaviors you are interested in, you need to **develop a dictionary.**

Define what counts and does not count in very concrete and specific terms.
This can be a simple table that lists the name of the behavior, what counts and what does not count. The people who will do the observations will use this as their guide. You want to keep your dictionary as short as possible, but also as clear as possible. The goal is to have a guide that results in consistent ratings. For example:

BEHAVIOR	WHAT COUNTS	WHAT DOES NOT COUNT
Physical Aggression	Kicking, hitting, shoving, pushing, grabbing, etc.	Physical acts that are done in self-defense and that do not escalate the situation

Task 3: Develop Codes

Decide how you will record the behaviors you see.

Once you have determined what behaviors you will be looking for, you need to **determine how** you will record them. There are two common strategies used:

- **Yes/No coding:** This approach is the simplest. You simply record whether or not you observed the behavior. At the end of your observations you can total up the behaviors and look for patterns. For example, did you observe more physical aggression at the school or at the mall? Did you see more boys or girls acting physically aggressive?
- **Scaled coding:** This approach allows you to record variations or degrees in the behavior. For example, you might have a scale that ranges from 1–4 where mild physical aggression is a "1" and extreme acts of physical aggression are a "4." This approach can let you not only count the number of incidents, but also to make distinctions based on severity. For example, a yes/no coding scheme may make it appear that girls and boys both act physically aggressive with about the same frequency. However, if boys are more severely aggressive that can be captured by a 1–4 scale if you see, for example, that the average rating you observed for boys was a "3" and the average rating you observed for girls was a "1." This additional precision is the advantage of using scaled coding. The disadvantage is that you need to have clear definitions not only of what counts as "physical aggression" but also of what counts as a 1, 2, 3, or 4 on the scale. So the upfront work and maintaining consistency can be more time-consuming.

Task 4: Observe

Finally, you need to conduct the observations. Here are some things to keep in mind to make your observations as successful as possible:

- **Scout out the setting.** Before you do your observations you want to have a good sense of the setting where you will be. This includes being familiar with the environment and the kinds of activities you can expect to be taking place there. There is a big difference between making observations in a setting with a lot of people and high levels of activity (where you will probably need to focus on only a small number of people or a small area because you won't be able to pay attention to everyone) and a setting with a small number of people and low levels of activity (where you might be able to include everyone in your observations).
- **Choose the best time to observe.** Scouting out the setting can also help you determine when the best time is to observe. You want to have the best chances of seeing the behaviors you are interested in, so some times may be better than others.
- **Be unobtrusive.** Whether or not the people you are observing know what you are doing, you need to try and blend in as much as possible. Think about how you fit naturally into the environment. Consider what to wear, how to record your observations (e.g., clipboard, small notebook, smart phone, tablet, mental count, etc.), where to be and how to act.

- **Make sure you have addressed any possible ethical issues.** While program evaluation is not subject to the same protections of human subjects that is required in research, you still want to make sure that you are engaging in ways that are ethical and that are acceptable in your community. For this reason, observations are usually done in public settings. If it is a private setting, you will need to be honest with people ahead of time about who you are and what you are doing. You will need to consider issues of privacy, confidentiality and anonymity. If you are observing minors you will need to consider whether additional steps need to be taken.

Special Topic: Evaluating Children's Programs

Many prevention programs choose to work with young children to develop skills for risk reduction and to establish a foundation for healthy relationships and respect for boundaries. Evaluating these programs poses some interesting challenges. At the same time, these challenges represent opportunities for creative approaches to evaluation.

Responding to survey questions—whether they are asked aloud or written—is a complex mental process (Borgers, deLeeuw, & Hox, 2000). Respondents have to:

1. Understand the question
2. Retrieve the relevant information (knowledge, attitudes, behaviors) from their memory and come up with their true answer
3. Match their true answer to the responses on the survey they are allowed to choose from
4. Decide whether to edit their answer to fit the situation or to fit what they think they are expected to say

When working through this complex process, children and adults may give an answer that appears reasonable, but it is actually superficial because they did not actually work through the entire mental process of answering the question (Borgers, Hox, & Sikkel, 2003). For example, they may estimate how frequently they have engaged in specific behaviors rather than actually counting the number of times or give answers based on general impressions rather than concrete situations. Because answering survey questions is more difficult for children, they are more likely to give answers that look satisfactory, but actually are filled with ambiguity.

The following table illustrates how evaluation methods need to be adapted to children's developmental levels (Borgers et al., 2003):

AGE	CHALLENGES	RECOMMENDED METHOD	TIPS FOR USING METHOD
Less than 4 years	■ Language and thought processes limited	■ No direct evaluation with children	■ Use caregivers and teachers as informants
4–8 years	■ Developing basic skills ■ Very literal ■ Very suggestible and reluctant to express own thoughts or feelings ■ Short attention span ■ Can't process questions that are not directly about themselves	■ Interviews ■ Small focus groups ■ Structured surveys where questions and answers are given aloud (children are not given anything written)	■ Keep questions very simple and direct ■ Incorporate play into the tasks ■ Use child's own words ■ Be vigilant about not influencing answers through words, body language, tone, etc.
8–11 years	■ Very literal ■ Difficulty processing negative statements (e.g., "not," "never," etc.) ■ Difficulty processing questions that are not directly about themselves	■ Interviews ■ Small focus groups ■ Surveys that are administered in small groups	■ Use visual stimuli when asking questions ■ Use response cards (e.g., index cards children choose from to show their answers) ■ Avoid negatively phrased questions ■ Keep questions very clear
11–15 years	■ Answers may be influenced by the context in which the questions are asked ■ Lack of motivation ■ Concerns about confidentiality	■ Written surveys ■ Computer-assisted surveys	■ Continue to use graphical and visual questions (e.g., frowning / smiling faces) ■ Watch out for ambiguous questions ■ Continue to watch out for literal interpretations
16 years and older	■ May have their own group norms ■ Answers may be influenced by the context in which the questions are asked	■ Written surveys ■ Computer-assisted surveys	■ May be treated like adults for surveying, but be aware of extra sensitivity to context

Incorporating the evaluation in the program itself can save time and make for a more interactive evaluation.

The need to **incorporate play and visual stimuli into measures with children 11 years and younger** provides us with unique opportunities to be creative. The following are a few ideas for carrying out evaluations with children. One thing you may notice is how the evaluations are often worked into the program itself. This has the added advantage of allowing more time for actual programming/education.

Visual Answers

A program that was doing a risk reduction and assertiveness building program with kindergartners and first graders wanted to know if students could identify situations that were high risk versus those that were safe. In their program, they talked about these as "stop" situations (high risk) and "go" situations (safe).

To evaluate the children's ability to distinguish between high-risk and safe situations, the program staff used a Pre-test/Post-test design.

Pre-test:

- At the start of the program, staff used puppets to act out three scenarios.
- They gave children a red stop sign with "stop" written on the sign and a green circle with "go" written on it.
- After each scenario was acted out, they asked the children to look at the signs lying flat on the floor in front of them and to choose whether a specific character should stop what they are doing or whether they can keep going.
- Then they told the children to pick up either the stop sign or the go sign, close their eyes and hold the sign in the air.
- The teacher then recorded the number of children who held up each sign (see below).

Post-test:

- At the end of the program, the staff again used puppets to act out three scenarios.
- The story lines were different, but they were comparable to the Pre-test scenarios.
- The "stop" and "go" procedure was repeated.

Data Analysis:

- After they returned to their offices, the staff took the data recorded by the teacher and recorded the children's answers as either being "correct" or "incorrect."
- They looked to see whether the percentage of correct answers increased from the Pre-test to the Post-test.

Using Findings:

- Because of the positive changes they saw, the staff decided to continue using the exercises and discussions for this part of their program.
- They shared the results with the school where they were working with the children to demonstrate that the program is working well and to gain support for continuing it.
- They also used the results when talking with other schools and community groups about bringing the program to their settings.
- Finally, the agency's development staff used the results to demonstrate their effectiveness when asking for funds to continue the program.

Behavioral Checklists

A program was doing a risk reduction and assertiveness building program with preschoolers. They were teaching the children using a "No–Go–Tell" format. They wanted to assess how much changed in children's skills for saying "no" assertively, going to find a safe adult and telling what happened.

To assess behavioral changes, the staff first generated a list of what it looks like when children effectively use the skills they were teaching. They turned that list into a checklist that could be used to record what children did during the program. They then used a Pre-test/Post-test design:

Pre-test:

- Before they showed children how to use the "No–Go–Tell" strategy, they introduced the basic concepts.
- They then sampled three children to show how they would say "no" in a specific situation.
- While the children demonstrated what they would do, a trained volunteer who was helping with the program used the checklist (see below) to record the skills that were and were not used.
- They used a similar approach when asking children who they would tell about certain kinds of situations and what they would say.
- They then proceeded with their education and skill building program. Throughout the program, they used what the children had done at the beginning as a springboard into discussion and practice.

Post-test:

- At the end of the program they had the same children repeat the exercises and recorded skills that were and were not used.

Data Analysis:

- After doing the program with multiple groups of children, the staff pooled the data.
- They compared the average number of skills used at the start of the program to the average number of skills used at the end of it.
- In addition to the overall averages, they looked at each behavior to see if some of them showed greater change than others.

Using the Findings:

- The staff found that while, overall, there were significant changes from the beginning of the program to the end, some skills changed more than others.
- While saying "no" showed a lot of changes in voice and eye contact, assertive body posture was still rarely used. Therefore, they made changes to their program to do more practice on this skill.
- The changes they wanted to make would require adding another session to the program, so they used the findings to advocate for more time with their community partners.

Interactive Assessment

Staff running a bystander empowerment program for middle school students knew from past experience that the students didn't take surveys seriously. They understood why, in light of all the written tests students take, they would be tired of paper-and-pencil assessments. So the staff wanted a more interactive approach to evaluation. They also wanted the evaluation to be a part of the program itself rather than something added on at the end.

They prepared questions ahead of time about different scenarios where a bystander intervention might be needed. For each scenario, they came up with four responses: (1) a direct and effective intervention, (2) an indirect and effective intervention, (3) a direct intervention that would actually make the situation worse or pose other problems and (4) ignoring the situation. Instead of simply using these four labels, they came up with responses that were specific to each scenario and had a volunteer who was artistic draw cartoons on newsprint to illustrate each response. For example:

SCENARIO: I see four boys I know from science class standing in a circle around a girl. They're making comments about how she "looks hot," "is really sexy," and how they wish they could "get our hands on you." The girl looks uncomfortable. I...

{PCAR: Can you draw a cartoon-like or stick-figure-type illustration of this scenario?}

RESPONSE 1: Tell the guys to leave her alone because they're acting like jerks.

{PCAR: Can you draw a cartoon-like or stick-figure-type illustration of this response?}

RESPONSE 2: Scare the guys off by calling out that a teacher is coming around the corner.

{PCAR: Can you draw a cartoon-like or stick-figure-type illustration of this response?}

RESPONSE 3: Challenge the guys to a fight.

{PCAR: Can you draw a cartoon-like or stick-figure-type illustration of this response?}

RESPONSE 4: Go on to class.

{PCAR: Can you draw a cartoon-like or stick-figure-type illustration of this response?}

They used a Pre-test/Post-test design, but did so with real-time, interactive assessments where they had students move around the room to show their answers:

Preparation:

- Knowing that students would be giving answers publicly, the staff spent the first session of the program doing interactive activities to build trust and rapport among the group and to set a tone of respectful communication.

Pre-test:

- Before they introduced any activities or concepts about bystander interventions, they had students gather in a circle in the center of the room.
- The stack of newsprints containing the scenarios were posted at the front of the room. The stack of newsprints with the responses were posted in each of the four corners of the room.
- The facilitator described the first scenario and had students quietly think about what they might do. While students thought, the facilitator went around and revealed each of the answers by unveiling the newsprints in the four corners of the room.
- When all four answers were revealed, the facilitator asked the students to move to the picture that best matched what they would do.
- Then the facilitator had students move back to the center of the room, revealed the next question and responses, and repeated the procedure for 4–5 scenarios. Participants indicated their responses in silence.
- While the activity was taking place, the teacher recorded how many students stood in each corner of the room.
- At the end of the assessment, the facilitators led a discussion with students about why they made the choices they did, what it was like to make their choices and to have their peers see their answers. This led into an in-depth discussion about peer influences on our actions and was the springboard into talking about bystander roles and the power of bystander interventions. The teacher took notes on the discussion.

Note:

- There is a possibility that students would be influenced by what their peers do in this exercise. That is why the facilitators took the time for the trust-building exercises during the prior session.
- Throughout the exercise, they repeatedly reminded students that there were no right or wrong answers and that we might do different things in different situations.
- The facilitators were comfortable with the possibility of peer influence during the exercise because, in reality, students will be in social situations when they have the opportunity to intervene as bystanders. Therefore, the facilitators thought their public choices during this activity were a better reflection of what they would really do, than an anonymous survey where they might be more likely to say what they would want to do.

Program:

- During the remainder of the five-session curriculum, the facilitators would use what came out of the activity as a springboard into further discussion and as a reference point for students to reflect on their choices. This helped to make the discussions more concrete and to go in greater depth.

Post-test:

- During the last session, the facilitators repeated the Pre-test assessment, using different but similar scenarios.
- Following the assessment, they talked about what it was like to do the activity and what students noticed in how their choices were similar to or different from the choices they made during the prior session. Again, notes were taken.

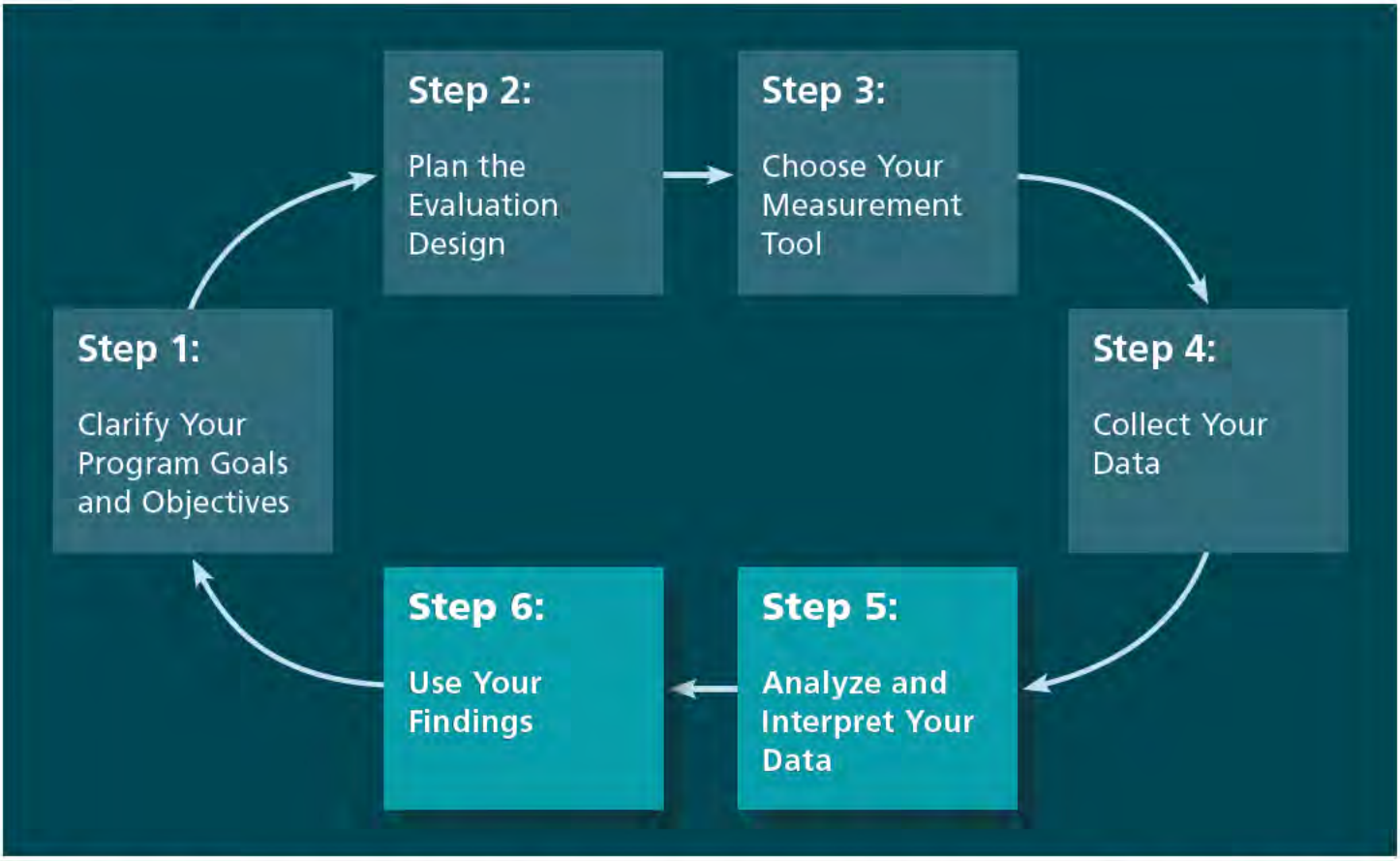
Data Analysis:

- Following the completion of the five-session curriculum, the program staff compared the counts for the actions students would take. They were looking for an increase in direct and effective and indirect and effective interventions. They were looking for a decrease in direct but problematic interventions and ignoring the situation.
- They also reviewed the notes on students' responses to see if there were any common themes that would help them better understand students' motivations and decision making.

Using the Findings:

- Program staff noticed that there were significant increases in effective interventions and significant decreases in problematic and no interventions. They shared this information with the school to continue fostering support for the program.
- They noticed that there were a lot more direct and effective actions chosen than indirect and effective interventions. In thinking about this, they realized that during the curriculum they spent a lot more time talking about direct interventions. They wanted to make sure that students also valued indirect interventions because sometimes those are the safest and the ones students are more likely to use because they involve less confrontation. So they made adjustments to the activities and discussions to make sure they gave equal time to both direct and indirect interventions.
- They also used the findings, along with testimonials from the middle school teachers, to approach a school in the neighboring district about adopting the program. They found that illustrating the changes with a simple graph got people interested, but it was the quotes from the students and teacher testimonials that sold them on wanting to bring the program to their school.

STEPS 5 & 6: ANALYZE, INTERPRET AND USE YOUR DATA



Once you have collected your evaluation data, you need to analyze and interpret it. **The type of analysis you use will depend on a number of factors**, including:

- What evaluation questions you want to answer
- Whether your data are quantitative (either are numbers or can be converted to numbers) or qualitative (open-ended answers in which participants give the answer in their own words)
- The skills your staff have in data analysis
- Outside help you can get from volunteers, interns, or consultants
- How quickly you need results

There are many sophisticated approaches to data analysis that someone with a background in data analysis can do. You may want to consider contracting with an outside consultant for this step. However, there are also simple analyses that you can do using widely available software such as Microsoft Excel. A step-by-step tutorial in data analysis is outside the scope of this manual. However, Volume Three describes how to:

- set up spreadsheets
- determine the appropriate method(s) of data analysis
- conduct four common methods of analysis (frequencies, percentages means and *t*-tests to compare pre-post data)
- handle missing or unclear data
- interpret and summarize results

As a preview to those detailed instructions, keep the following suggestions in mind about how to manage, analyze and interpret your data.

Quantitative Data Management

- **Make sure that every survey has an ID number.** Even if it is a satisfaction survey on which participants did not make up a unique identifier you still need to number them. If questions or problems arise in the data entry process you can easily find the survey in question.
- **Quantitative data need to be organized and stored in a spreadsheet.** Typically this means that each participant's survey or each observation will be a row in the spreadsheet and each question will be a column.
- **For analysis purposes, every answer needs to be a number.** This is called coding the data. You can easily convert many answers to numbers. For example:

Strongly Disagree =	1
Disagree =	2
Agree =	3
Strongly Agree =	4

Never =	1
Rarely =	2
Sometimes =	3
Often =	4
Always =	5

No/False =	0
Yes/True =	1

Male =	1
Female =	2

- Codes need to be logical and consistent. When possible, higher numbers should indicate "more" of something. For example, in the above coding for the agreement scale higher numbers mean participants agree more. You also need to make sure that higher numbers have the same meaning. For example, suppose you had participants say how much they agreed with these two statements:

  If a woman doesn't physically fight back, you can't call it rape.

  No matter what a woman does, rape is never acceptable.

  People who say they "Strongly Agree" with the first statement are accepting rape myths. In contrast, people who say they "Strongly Agree" with the second statement are rejecting rape myths. In this case you need to do what we call "reverse coding." First determine what you want higher numbers to mean. If you want high scores to reflect rejection of rape myths then for the first question you need to reverse the numbers so that someone who says they "Strongly Agree" that a woman has to fight back will get a "1" for that answer instead of a "4." This way when you average all of the questions about rape myths the average will be based on consistent scores. If you need to reverse code and fail to do so then your numbers will "cancel each other out."

- **Before you run any analyses, screen your data.** Check to make sure that there are no invalid numbers. Invalid numbers usually represent typos. For example, if the scale is from 1 to 4 and you have a 33 in there, then probably the real answer was a 3 and the person entering the data simply hit the key twice. This is why you need ID numbers—You cannot assume the typo. You must find that survey and double check the answer before you make a correction.

Quantitative Data Analysis

- **Frequencies:** Calculate frequencies when you want to know the **number of times** people gave a certain answer.
- **Percentages:** Calculate percentages when you want to know the **proportion of times** people gave a certain answer. Percentages are especially useful when you want to compare answers between groups. For example, what if you want to compare how many girls accepted a specific rape myth versus how many boys accepted that same myth? If you have the same number of girls and boys you can simply use the frequencies. However, usually we don't have the same number. In this case, you can compare the percentage of girls who accepted the myth with the percentage of boys who accepted it. Note: If the difference is very small (e.g., 43% of girls versus 40% of boys), we must be very cautious in conclusions we draw because this difference may be "due to chance" and may not represent a "real" difference.
- **Means (averages):** Calculate means when you want to know the average for a set of questions. For example, if you ask 10 questions about rape myth acceptance you might want to get a sense of how much, overall, people accept rape myths. In this case, you can take all 10 questions and calculate the average score. This will greatly simplify the data and provide you with a single score that summarizes people's answers.
- **T-tests and ANOVAs:** Calculate t-tests and ANOVAs when you want to compare groups or compare Pre-tests, Post-tests and follow-ups. These types of analysis let you determine whether the difference you see between two groups is what we call "statistically significant." If it is significant then we can infer that the difference (for example, between the Pre-test and Post-test scores) is a "real" difference and it is not simply "due to chance."
- **Regressions:** There are a wide range of statistical techniques that are based on what statisticians call regressions. In general, these techniques let you determine how well a variety of factors predict some outcome. For example, imagine that you are evaluating a bystander empowerment program. One of the questions you asked was, *"When you hear sexist comments, how likely are to you to say that you disagree with the comments?"* There are many factors that could influence someone's answer: how much they accept rape myths, their attitudes toward women, their general level of self-esteem, their gender, their age, etc. Regression analyses can let us test how much those factors influence the likelihood of voicing disagreement with sexist comments.
- **Think about how you will analyze your data when you first choose your measures.** There is nothing more frustrating than going to analyze your data and suddenly realizing that you don't have the information you need to answer the evaluation questions you're interested in.

Data Interpretation

Interpreting your data requires that you summarize and synthesize the results in meaningful ways. A long list of numbers will be overwhelming to you and to anyone with whom you share your results.

- **Think about the purpose of your evaluation.** Decide what information is most relevant. Do not present every detail of the results. Instead, highlight the most important findings.
- **Summarize the findings in a succinct way.** Think in terms of bullet points and take-away messages.
- **When possible, use graphs, charts, tables and diagrams.** Visual summaries are often easier for you and others to understand.
- **If you have negative findings (i.e., something didn't work), think about the positive lessons you can learn from them.** It is valuable to know what objectives you are not achieving so that you can plan for how to improve the program.
- **Don't be modest about your successes.**

Step-by-step instructions on how to analyze quantitative data are found in Volume Three of this toolkit.

GUIDELINES FOR HIRING AN OUTSIDE EVALUATOR

If you skipped over the five steps of program evaluation because you think that it is not possible for your program to do its own evaluation, stop here and go back and read the preceding pages. While hiring an outside evaluator is sometimes a wise and efficient strategy, it is important that you understand the steps of evaluation because **some steps of program evaluation an outside evaluator cannot do for you**. They may be able to help you do them but they cannot do them alone. **Only you know:**

> There are some parts of an evaluation that only you can do.

- What your program goals are
- How you define success
- What outcomes are most important to your program
- What questions you want answered by an evaluation
- What values your program holds and want to see reflected in the evaluation
- How you want to use the evaluation findings

Furthermore, **you probably have more insight than an outside evaluator** does on issues such as:

- What kinds of evaluations people in your community will be most accepting of and even enthusiastic about doing
- Political pitfalls of evaluation in your community
- What has worked and failed in the past when you have done evaluations
- What resources your program has available for evaluation
- What will be of interest to funders or others with whom you might share evaluation findings

Before hiring an outside evaluator it will be useful for you to consider carefully what specific parts of the evaluation you need help with and what parts you can do on your own. You may find that you can do much of the work, thereby cutting down on the costs of the evaluation contract.

Finding an Evaluator

There are numerous ways you can find an evaluator. Some places to start include:

- **Talk with other programs or with PCAR** for names of evaluators that they have used and recommend
- **Go to eval.org and look at their listings of evaluators under the tab "Find an Evaluator."** Note: Most evaluators who list at this site are evaluation firms. Independent consultants tend to drum up business more by word of mouth. So if you don't see a person's name listed there, don't worry about it.
- **Contact the Division of Violence Prevention at the CDC** for names of evaluators that they recommend
- **Contact the faculty of nearby colleges or universities** to inquire about their interest or if they know anyone they could recommend. Likely departments to contact include: psychology, social work, sociology, criminal justice, public health, or nursing.

Hiring an Evaluator

In evaluating sexual violence prevention programs, **it is especially important to work with an evaluator who** either has some knowledge of sexual violence, of prevention in general, and/or of sexual violence prevention in particular — or who is willing to learn and able to get up to speed quickly on the specific issues that the sexual violence prevention/rape crisis movement faces.

There are unique issues when evaluating sexual violence prevention programs. Not all evaluators will be well suited for this work.

Some questions that may help you in hiring an outside evaluator include:

- **Is the evaluator a member of the American Evaluation Association (AEA)?** Members of AEA subscribe to professional principles and standards of practice. They may also have access to evaluation resources through the AEA network.
- **What background or training does the evaluator have in program evaluation?** Program evaluation is not the same as research, although many of the methods overlap. It is generally best to work with someone who has some type of training or experience specifically in program evaluation. Many people who have graduate degrees in social work, psychology, public health, applied sociology, nursing, or criminal justice have been trained in program evaluation. However, do not assume that just because they have a degree in one of these areas that they have the kind of training you need. Ask them specifically about their training in program evaluation.
- **What methods does the evaluator use?** Not all evaluators use both quantitative and qualitative methods. If you know that you would like to have information from interviews or focus groups, you need to make sure that the evaluator is willing to do qualitative research and analysis. If you have specific things you want to learn from a survey, you need to make sure that the evaluator can do the necessary statistical analyses. You don't need to know what those analyses are; that's the evaluator's job. But you do need to know that if you say, "We want to know _____" that the evaluator can do the necessary analyses to find the answer. If they say they can't answer that question but they could do analyses to answer a different question then you may want to find a different evaluator.
- **What experience does the evaluator have with sexual violence prevention or rape crisis work?** Some familiarity with sexual violence issues is important. You want an evaluator who has at least a basic understand of myths and facts about sexual violence and the kind of work that is done by similar programs. An understanding of the larger political climate (e.g., funding structures, history of the movement, etc.) may also be useful.
- **What kind of relationship does the evaluator want with you?** There is no one right answer to this question. However, there will be answers that fit with the kind of relationship you are looking for and answers that don't fit. How much collaboration do each of you want — working jointly on all aspects of the evaluation? working jointly on planning the evaluation but then the evaluator works more independently once you've agreed on the plan? the evaluator taking the lead and running major steps/products by you for approval?

- **How much time does the evaluator need to complete the project?** Be sure that you establish a clear timeline for the project and that it is spelled out in the contract. Deadlines may need to be negotiated as situations arise. However, your needs should take priority in establishing the deadline. An evaluation that comes in late is no good if it means that you can't use it for a grant proposal, board retreat, etc.
- **Does the evaluator have any samples from prior evaluations?** Feel free to ask for copies of evaluation reports, evaluation summaries, etc. This is a good way of ensuring that the evaluator can communicate with you in a clear, jargon-free way. Keep in mind that the evaluator may have limits on what can be shared due to confidentiality agreements with previous clients. However, she should be able to provide you with some type of sample. If an evaluator gives you an article from an academic journal as a sample of evaluation work or provides you with a list of professional publications, this may be an indicator that she is not accustomed to translating results into a clear, jargon-free format.

Responsibilities of an Evaluator

If you do decide to hire an outside evaluator/consultant to help you, it is important to be aware of what you can expect from a professional evaluator. According to the American Evaluation Association, an evaluator should abide by five professional principles (American Evaluation Association, 2004):

Systematic Inquiry: Evaluators should conduct systematic, data-based evaluations. This includes:

- Exploring strengths and shortcomings of different approaches that might be used in an evaluation
- Communicating their methods and approaches in a way that you can understand
- Being open to questions and critiques you may have of the methods and approaches

Competence: Evaluators should provide competent services. This includes:

- Possessing the appropriate education, abilities, skills and experience for the job
- Ensuring that the evaluation is done in a way that is appropriate for the cultural context, including considering gender, race, ethnicity, religion, socio-economics, or other factors that may be relevant to the evaluation
- Turning down evaluation opportunities if they do not have the necessary training or skills for the particular project

Integrity/Honesty: Evaluators should display honesty and integrity in their own behavior and work to ensure the honesty and integrity of the evaluation. This includes:

- Negotiating honestly with clients about costs, tasks to be undertaken, limitations of the evaluation and how the data may be used
- Disclosing any potential conflicts of interests
- Informing you in a timely fashion of any changes that need to be made to the evaluation plan and the likely impact of those changes
- Being open about their own interests and values concerning the evaluation
- Taking all possible action to correct any use of the evaluation that is misleading

Respect for People: Evaluators should respect the security, dignity and self-worth of respondents, program participants, clients and other evaluation stakeholders. This includes:

- Abiding by professional standards and regulations about protecting participants from potential risks and ensuring informed consent
- Maximizing the benefits and minimizing unnecessary harms that may result from negative findings
- Fostering social equity in the evaluation

Responsibilities for General and Public Welfare: Evaluators should take into account the diversity of general and public interests and values that may be related to the evaluation. This includes:

- Including relevant perspectives from a full range of stakeholders
- Considering the broader implications and potential side effects of the evaluation
- Allowing you access to all evaluation information in ways that will not compromise confidentiality
- Presenting the results to you in a way that is clear and understandable

If you hire an outside evaluator who fails to meet these responsibilities, you should voice your concerns to them. You are the client and it is the evaluator's obligation to do everything in their power to ensure that you are being given the service that you want. **Although they may be an expert on evaluation, remember that you are the expert on sexual violence prevention and you are the expert on your community. An evaluator should work with you, drawing in as much of your expertise as possible.**

REFERENCES

American Evaluation Association (2004). *Guiding Principles for Evaluators.* Washington, D. C.: American Evaluation Association.

Banyard, V. L., Plante, E. G., & Moynihan, M. M. (2004). Bystander education: Bringing a broader community perspective to sexual violence prevention. *Journal of Community Psychology, 32,* 61–79.

Banyard, V. L., Moynihan, M. M., & Plante, E. G. (2007). Sexual violence prevention through bystander education: An experimental evaluation. *Journal of Community Psychology, 35,* 463–481.

Borgers, N., deLeeuw, E., & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin de Methodologie Sociologique, 66,* 60–75.

Borgers, N., Hox, J., & Sikkel, D. (2003). Response quality in survey research with children and adolescents: The effect of labeled response options and vague quantifiers. *International Journal of Public Opinion Research, 15,* 83–93.

Campbell, R., Davidson, W. S., Ahrens, C. Aponte, G., Dorey, H., Grubstein, L., Naegeli, M., & Wasco, S. (1998). *Introduction to Evaluation Training and Practice for Sexual Assault Prevention.* Okemos, MI: Michigan Public Health Institute.

Foshee, V. A., Bauman, K. E., Arriaga, X. B., Helms, R. W., Koch, G. G., & Linder, G. F. (1998). An evaluation of Safe Dates, an adolescent dating violence prevention program. *American Journal of Public Health, 88,* 45–50.

Foshee, V. A., Bauman, K. E., Greene, W. F., Koch, G. G., Linder, G. F., & MacDougal, J. E. (2000). The Safe Dates program: 1-year follow-up results. *American Journal of Public Health, 90,* 1619–1621.

Frazier, P., Valtinson, G., & Candell, S. (1994). Evaluation of a coeducational interactive rape prevention program. *Journal of Counseling and Development, 73,* 153–158.

Katz, J. (1994). Mentors in Violence Prevention (MVP) Trainer's Guide. Northeastern University's Center for the Study of Sport in Society. Boston: Northeastern University.

Men for Change (2000). *Healthy Relationships Curriculum.* Halifax: Men for Change.

Pennsylvania Department of Education (2013). *Standards aligned system.* Downloaded from www.pdesas.org.

Rossi, P. H., Freeman, H. E., Lipsey, M. W. (1999). *Evaluation: A Systemic Approach.* Thousand Oaks: Sage Publications.

PENNSYLVANIA COALITION AGAINST RAPE
125 North Enola Drive • Enola, PA 17025
717-728-9740 • 800-692-7445 • TTY 877-585-1091 • pcar.org