In **Preventing Violence in Relationships: Interventions Across the Life Span,** Edited by Paul A. Schewe, American Psychological Association, 2002.

10

EVALUATING PREVENTION PROGRAMS: CHALLENGES AND BENEFITS OF MEASURING OUTCOMES

PAUL A. SCHEWE AND LARRY W. BENNETT

The purpose of the preceding chapters has been to highlight promising directions for the prevention of violence in relationships. These chapters have provided specific techniques and general guidelines for developing successful prevention programs. Although one method of improving preventive services is to learn from the experiences of others, another way is to gather specific and critical feedback about the outcomes of interventions. This chapter provides basic information on why agencies should (and should not) evaluate their programs and how to design and conduct an evaluation that will provide maximally useful information.

WHY IS PROGRAM EVALUATION IMPORTANT?

There are a variety of reasons for evaluating programs. The motivation to evaluate a prevention program may originate from the community,

external funders, internal decision makers (i.e., administrators or boards of directors), staff, or participants in the program. Requests for evaluation usually hinge on accountability to a funding source, a board of directors, or an external service network. Funders and board members want to know whether the program is doing what it intended to do, whether it is having the desired effect, and what the effect costs. Requests for accountability may spark frustration with agency staff, who are struggling to provide badly needed services with scant resources. Funders and board members are sometimes viewed as detached from the work and interested only in the "bottom line." Staff may use this perception (often a misperception) to justify their fear or disinterest in evaluation.

Program staff themselves often drive evaluation efforts. Program staff may want to develop and improve the program and may view evaluation as one source of assistance. Knowing "what works" is important to program administrators and direct service providers alike. In some cases, the stimulus for program evaluation is to learn about unexpected effects of programs. Prevention educators are usually aware that certain program elements seem to be more effective than other program elements. Evaluation may shed light on these disparate experiences or even provide evidence that their perceptions are not well founded. Another motivation agencies may have for evaluating prevention programs is their belief that their program is unique and should be showcased. This motivation is a two-edged sword because agency staff may be less likely to accept findings that show no effect, or worse, show a negative effect.

Evaluation also has the potential to increase the effectiveness of the services evaluated, even beyond the "Hawthorn effect" of staff performing better just because they know they are being evaluated. Some practitioners embrace evaluation as a necessary aspect of practice. Admittedly, the mutual reinforcement of evaluation and practice is not the way most staff will view evaluation, at least to begin with. Initially, the link between evaluation and practice must usually be made by modeling, in which experienced and respected staff embrace evaluation. Unfortunately, the opposite effect, in which evaluation is seen as external interference, may also be modeled by experienced staff and administrators.

Linking evaluation to the allocation of funds or other resources is one of the reasons staff and agencies cite for not wanting to evaluate their programs. Service providers might fear that if their program does not show a positive impact in the evaluation, the program might get cut. The reality of these fears should be addressed prior to beginning any evaluation and are just one reason why, particularly in the relatively young field of interpersonal violence prevention, evaluations should be directed primarily at program development and improvement. In interpersonal violence prevention, the most appropriate research questions are usually developmental in nature: "What are the strengths and weaknesses of the program? Who does

this program serve best? What kind of an effect does this program have on participants?" rather than "Is this program effective in reducing child abuse, bullying, rape, dating violence, domestic violence, or elder abuse?"

Edleson and his colleagues at the Domestic Abuse Project in Minneapolis identified several motivations for evaluation that may lead to problems (Edleson & Frick, 1997). One of these motivations is using evaluation to postpone making a decision about a program or service that should be made using other criteria. For example, an administrator may use the results of an evaluation to justify eliminating or replacing an existing program. In essence, the decision to alter or eliminate the program shifted to the evaluator. A second problematic motivation to evaluate is public relations or marketing, in which evaluation factoids are used as sound bites and visual bits to entice funders, politicians, and the public to support a program. Along with public relations and marketing, evaluation data may be sought for other issues such as politics, turf, or competition. Although such phenomena are often unavoidable in agency practice, using evaluation to gain an edge is probably a misdirected effort. A mundane, but more important, reason to not evaluate a program is lack of resources such as money or staff and adequate support for evaluation.

For all the aforementioned reasons, staff and administrators may resist evaluation. Practitioner resistance to evaluation usually has roots in lack of familiarity about the mechanics of evaluation and in resentment of the allocation of time and resources to something perceived as not contributing to the prevention of violence. To expect practitioners to overcome their own resistance to evaluation is a mistake akin to blaming the victim. If lack of familiarity and resource concerns are where resistance is experienced, then agencies, funders, evaluators, and other relevant networks must take responsibility for cocreating the conditions that adequately support the evaluation process.

The bottom line in evaluation is accountability: to administrators and prevention educators; to communities and taxpayers; and most importantly, to the participants in the programs whose lives we are hoping to improve. An agency that does not critically examine its effects on those it seeks to help is probably not accountable to its community or its clients.

OUTCOME EVALUATION, PROCESS EVALUATION AND SATISFACTION SURVEYS

Many staff believe that they already evaluate their programs. Each week they record the number of students they presented to, indicate which lessons they taught and what materials they used, and review the satisfaction surveys they gathered at the end of every class to determine what the

students liked and did not like about the program. These are good examples of process evaluation and satisfaction surveys.

Process evaluations reveal the characteristics of clients that receive the services, the number of clients, and the type and content of the service that they received. An example of a thorough process evaluation report might look like this:

> On May 5th I gave a presentation to 25 11th-grade students at Washington school. 53% of the students were female, and 60% were White, 25% were Black, 10% were Hispanic, and 5% were Asian. Students come from a mostly middle- to upper-middle-class community. 45% of the students in the class indicated that they had already experienced violence in their dating relationships. For this presentation, we showed the 10-minute dating violence video, handed out a copy of the "power and control wheel," discussed power and control in friendships and dating relationships for approximately 15 minutes, spent 10 minutes discussing an incident involving a student bringing a gun to school last week, and ended the session by discussing local resources and handing out our brochure and satisfaction survey. 21 of the 25 students returned completed satisfaction surveys. After the class, one student asked to speak to me about a violent relationship, and after speaking with her for a few minutes, I was able to give her referral information for one of our counselors.

Satisfaction surveys are a systematic way of learning about students' perceptions about a program. Typical questions include "What did you like/dislike about the program? Would you recommend this program to a friend? How helpful was this program? As a result of this program, do you feel that you are better able to avoid violence in your relationships?" Results of satisfaction surveys give indications of how well the program was received, suggest aspects of the program that might be modified or expanded, and show whether the program is perceived as being helpful. Also, because the results of satisfaction surveys tend to be positive, they can be a useful tool in increasing service providers' morale and preventing caregiver fatigue.

Extreme caution must be used, however, to avoid confusing high marks on program helpfulness with positive outcomes. In an experiment by Schewe and O'Donohue (1993b), two independent prevention programs were presented to different groups of students. Although both programs received equally high marks on a satisfaction survey covering the credibility and helpfulness of the program, one of the programs was not at all effective in achieving desirable outcomes, and the other program was moderately effective in producing desired changes among program participants. If only the satisfaction survey were used in this study, both programs would have been shown as equally effective. Use of an outcome evaluation, however, revealed that students in one of the programs evidenced positive changes in knowledge, attitudes, and behavioral intentions, whereas students in the other group evidenced almost no change.

Formative evaluation is another type of process evaluation that consists of two kinds of assessment: needs assessment and evaluability assessment. The purpose of *needs assessment* is to aid in the design or expansion of services. In a needs assessment, an agency may survey a community to determine whether there is a need for an intervention, study other unmet needs of the target population, or conduct an audit of staff competence to deliver the intervention.

Evaluability assessment is often the first step toward designing and conducting an outcome evaluation of an existing program or intervention. The purpose of evaluability assessment is to determine whether an agency or institution is prepared for evaluation by focusing on how information is processed, how fully developed the program is, and whether goals are quantifiable. For example, the goal of many media campaigns is to raise awareness about violence. Although this might be a worthwhile goal, quantifying or measuring an increased awareness about violence is difficult. Regarding program development, if a prevention program is not developed to the point at which another agency or educator could replicate it, it is probably not ready to be evaluated. If a program exists only in the mind of its creator or implementor and changes significantly from week to week or setting to setting, the results of any evaluation will be useless because it will not be clear what was evaluated.

Outcome Evaluation

Outcome evaluations seek to answer the question, "Is the program meeting its goals and objectives?" For programs whose goals and objectives are not so clear, an alternate evaluation question might be, "How are students different after participating in the prevention program?"

Measuring Outcomes

Outcomes assessed for prevention programs often fall into the categories of knowledge, attitudes and beliefs, skills, and behavior or behavioral intentions. The category of behavioral intention is included here because those in the field of primary prevention might not be able to measure the presence or absence of the behavior that we are trying to prevent. For example, if the goal of a rape prevention program were to prevent men from raping, trying to measure the success of the program by tracking rape convictions would be nearly impossible (Schewe & O'Donohue, 1998). Moreover, it would be misleading, because the chance a man will be arrested for rape is only a fraction of the chance he will rape, and it is an even smaller fraction of the chance he will think about rape. This issue warrants further discussion.

Researchers evaluating the efficacy of treatments for convicted sex offenders have commonly used police reports of rearrest or reconviction to

track the progress of both treated and untreated sex offenders (Marshall, Jones, Ward, Johnston, & Barbaree, 1991). However, tracking incidence rates is not an effective way to measure the outcome of primary prevention interventions. First, this type of research takes years to complete and is extremely costly. Second, there is the problem of measuring incidence. Given a normal or even high-risk population, only a small percentage of men would be expected by chance to ever commit rape. Estimates using the Sexual Experiences Survey have suggested that about 9% of the male population commit acts that meet the legal definition of rape or attempted rape (Koss & Harvey, 1987). Police reports of incidents would be unacceptable because only approximately 10% of all rapes are reported to the police, and only 50% of these reports result in arrest (Missouri Division of Health, 1985). Of those 50% that are arrested, only 66% are prosecuted, and 85% of these result in conviction (M. Cavins, Cook County State's Attorney's Office, personal communication, June 1996). Thus, the probability of finding that a participant has committed rape by monitoring rape convictions is roughly 1 in 1,000.

Studies of domestic violence incidents have revealed a similar ratio of incidents to official detection. One careful study found that the ratio of victim-reported domestic violence to arrest was 35 to 1; that is, for every reported arrest, there were 35 assaultive actions (Dutton, Bodnarchuk, Kropp, Hart, & Ogloff, 1997).

Because of the difficulties in tracking the incidence of violence among study participants, it becomes necessary to use proximal outcome measures to evaluate the efficacy of primary prevention programs. Such proximal outcome measures include changes in knowledge, attitudes, and beliefs that have been associated with the ultimate outcome (behavior) that prevention educators are trying to influence (Schewe & O'Donohue, 1993a). Typically, violence potential has been measured by assessing attitudes and behaviors thought to be relevant to committing an act of interpersonal violence. For example, in the field of rape prevention it is common to assess acceptance of rape myths, acceptance of interpersonal violence, and the extent to which individuals believe that sexual relationships are adversarial in nature (Burt, 1980). Another approach has been to directly ask men to rate their likelihood of committing rape (Malamuth, 1989) or to ask them whether they have committed acts of sexual coercion in the past (Koss & Oros, 1982). Rape potential is then inferred; that is, rape-supportive attitudes indicate rape potential; self-reported future likelihood of raping indicates rape potential; or self-reported past history of raping indicates rape potential. In domestic or dating violence prevention, proximal measures might include attitudes toward women, beliefs about battered women, or beliefs about gender-based power.

Another proxy for measuring target behavior is to track changes in behavioral intention. For instance, if before the program 25% of the students indicate that they might commit rape if they could be assured of not getting caught on Malamuth's (1989) Attraction to Sexual Aggression scale, and afterward only 5% indicate that they might commit rape, then there is at least some indication of the program's success. How confident we are about this success will depend largely on our experimental design, to be discussed later.

In intimate partner violence, common proximal measures may include hostility, alcohol use, or belief in the situational acceptability of violence. It is critical that proximal measures be linked theoretically to the goal of the intervention. This link often engenders considerable debate. For example, some practitioners believe alcohol use is theoretically linked to dating violence, whereas others argue there is no established causal link between drugs and violence (Bennett & Fineran, 1998; Nicholson et al., 1998; O'Keefe, 1997).

Logic Models

To determine what measures to include in an outcome evaluation, prevention researchers regularly rely on *logic models*, or diagrams that outline the prevention program from its process or content to its multiple objectives, to its ultimate goal or goals. An example of a logic model is found in Table 10.1. Logic models are useful tools that can be used in several ways.

TABLE 10.1
Sample Logic Model

Process (what we do for whom)	Strategies	(Measurable) objectives	Impact
Provide a 1-hour rape prevention program to male and female eighth-grade students	Address rape myths	Decrease rape supportive attitudes	Reduce the incidence of rape
	Teach communication skills	Improve intergender communication	
	Portray the negative consequences of rape	Increase empathy for victims of rape	
	Instruct students about how they can help a rape victim	Increase empathy for victims of rape Decrease rape supportive attitudes	
	Provide information about local rape crisis centers	Increase knowledge local resources Increase participants' sense of safety	Increase the number of victims receiving assistance

Commonly, they are used by evaluators to identify the measurable objectives of an existing program. In this case, the logic model is usually built from left to right. The prevention educator starts by describing the process or the practical characteristics of the program. Next, the content or the specific strategies used in the intervention are noted. Then the objectives of these strategies are noted. At this point an effort is made to identify objectives that are specific enough to be measurable. To say that the objective of an exercise is to "increase knowledge about domestic violence" is not very helpful in developing an instrument for measuring that objective. "Increasing knowledge of the causes and consequences of domestic violence, being able to identify risk factors for an abusive relationship, and increasing knowledge of specific resources for victims of domestic violence" are examples of objectives that are much easier to measure. The idea is to provide enough detail about the program content and the intended objectives to develop a measurement strategy.

One danger is that, in some cases, the measurable objectives may have a weak relationship to the ultimate goal or impact of the prevention program. For instance, changing attitudes about a particular behavior has a relatively small influence on actually changing the behavior. Raising awareness or increasing knowledge in a particular area has an even smaller effect on behavior. In determining which objectives to measure, it makes sense to measure those most closely related to the behavior itself.

Another way to use the logic model is in program development. Here the model is generally completed from right to left. We start with the ultimate goal or impact that we hope to achieve. Next, we identify those subgoals or objectives that are empirically or theoretically related to our goal, along with ways to measure those objectives. Finally, we develop a program (strategies) designed to best achieve those objectives.

Building an Evaluation Instrument

After identification of the measurable objectives of a program, the next step is to identify or develop tools for measuring those objectives. Conducting a search of the scientific literature is one way to find existing outcome measures. Asking researchers in that field of practice is another way to learn about appropriate existing measures. The rigor required in selecting measurement tools is linked to how the evaluation will be used. If, for example, program staff want to gain recognition for their program by writing a journal article or conference presentation, they must use valid and reliable measures that can pass at least some level of scientific review. Evaluation measures can be designed locally as well, which is likely to occur and may be necessary if the evaluation results will be used only by staff and local or state funders. If staff must develop their own evaluation instruments to document the specific outcomes of their unique program, then we

recommend that they seek outside assistance from someone who has experience in developing outcome evaluation instruments. This assistance will ensure that measures are appropriate to the task at hand and will provide the information needed.

Other Considerations for Measuring Development

Keep Outcome Measures as Simple and Brief as Possible

When researchers develop outcome measures, they are pressured to be as comprehensive as possible so that every positive outcome can be measured and accounted for. Practical constraints such as the amount of classroom time available and the tolerance of the audience to complete long questionnaires necessitate briefer measures. Use of a logic model of the intervention helps focus the evaluation on the major components of the intervention and its intended outcomes. The evaluation should be focused on the primary outcomes; the evaluation does not need to measure every possible positive outcome.

Be Careful of Ceiling Effects

Teenagers and adults today are well informed and may score surprisingly well on knowledge and attitude questionnaires, regardless of high levels of violence in their relationships. A pilot test of the outcome measures should be administered to the intended audience to be sure that there is room for improvement as a result of the intervention.

Measure Behavior Whenever Possible

The relatively weak link between knowledge, attitudes, and actual behavior argues for the use of behavioral measures whenever possible. However, when collecting data regarding criminal behavior (e.g., assault, drug use), using hypothetical situations or asking questions about behavioral intent is one way to protect participants and may allay concerns of review boards.

Be Clear About Confidentiality Issues

Participants will not answer questions honestly unless they can trust the test giver and are aware of the procedures that will be taken to protect the confidentiality of their answers. Although it has never been tested in a court of law, obtaining a "certificate of confidentiality" is an additional step that researchers can take to protect the confidentiality of information from research participants.

EVALUATION DESIGN

After program staff have identified or developed their evaluation instruments, they still need to decide what type of evaluation design to use. Evaluation design decisions are often based on practical considerations such the amount of time and resources that are available and the availability of appropriate control groups, as well as the research questions that are driving the evaluation.

Single-Group, Posttesting Only

This simplest of evaluation designs is useful in two instances. One instance is when program staff gather information in the form of satisfaction surveys. The other instance is when the program is training to mastery. The quizzes and tests that are taken in school are an example of a single-group, posttesting only design. If the program objective is that "participants will be able to identify three characteristics of an abusive relationship," and after the program, 90% of the participants are able to identify three characteristics of an abusive relationship, then there is at least some evidence of the program's success. One problem with this design, especially in the area of interpersonal violence (IPV) prevention, is that a program is usually not trying to teach to mastery but rather dealing with a wide range of attitudes, beliefs, skills, and knowledge, and the aim is often to move people in the right direction rather than getting them to achieve some level of mastery. Another problem with this evaluation design arises when someone asks, "How would the participants have performed on the test if they hadn't had the program?" Perhaps participants already possessed that knowledge or those attitudes before the program. To state that the results of the evaluation are attributable to the program, a more sophisticated evaluation design is necessary.

Two Groups, Posttesting Only

One way to strengthen the argument that a prevention program was responsible for positive findings is to administer the evaluation instrument to a group of people who did not receive the intervention (a no-treatment control group). By using a no-treatment control group, researchers are able to demonstrate how participants in the program are different from a similar group of people who did not receive the intervention. Here, the difficulty is finding a group of people that do not differ significantly on important characteristics from the people who participated in the program. If the groups do differ, then one could argue that the observed differences between the two groups on an assessment instrument might be attributable to the differences in age, gender, race, education, and so forth. One way to increase the likelihood that the two groups are as similar as possible is to randomly assign individuals to the intervention group and to the no-treatment control group.

In this way, any extraneous variables that might affect performance on the outcome measure, such as level of education, should be equally distributed between the two groups. However, random assignment does not ensure that some important characteristic is not disproportionately assigned to one group, and hence differences between groups might continue to exist regardless of the intervention.

Another way to ensure equality between the two groups is to use a matched sample. In this method, pairs of individuals are matched on important characteristics; then one member of each pair is assigned to the treatment group and the other is assigned to the control group. The drawback is that we often do not know what all of the important extraneous variables might be, or there might be so many variables that matching becomes impractical or impossible.

One Group, Pre- and Posttesting

Another way to strengthen the argument that a prevention program was responsible for the positive findings is to administer the evaluation instrument to participants before and after the intervention. When little time exists between the pre- and the posttest, researchers can be more confident that the changes in participants' scores are a result of the program. However, if the program spans several weeks, one might argue that other life experiences might be responsible for the improvement in scores rather than the intervention. Here again, adding a no-treatment control group can increase the level of confidence that the positive outcomes can be attributed to the program.

Treatment and Control Groups, Pre- and Posttesting

The primary advantage of using pre- and posttesting with both treatment and control groups is that researchers can compare scores between the two groups preintervention to determine if any important differences exist that might influence their scores postintervention. If no differences exist preintervention, then any differences postintervention can be confidently attributed to the effects of the intervention. If differences do exist between group preintervention, statistical techniques can be used to help control the effects of these preexisting differences. So finally, there is an evaluation design that allows us to confidently attribute any significant differences between the treatment and no-treatment control groups after the intervention to the effectiveness of the intervention.

Although this level of evaluation design represents the state of the art in the evaluation of prevention programming, several problems exist with this design. One problem is the practical problem of finding a no-treatment control group that is willing and able to complete the assessment instruments at two points in time similar to the intervention group. School administra-

tors who are reluctant to give prevention educators access to students to deliver a service are sometimes more reluctant to allow class time for an evaluation of that service. Another problem with a treatment and no-treatment control group design is that when the study is finished, all that really can be said is that the intervention is better than nothing. This type of evaluation does not show which aspects of the program were responsible for the observed changes and does not allow the statement that this program is better than any other program (unless another program used identical evaluation instruments and similar evaluation procedures). Fortunately, there is an evaluation design that solves both of these problems.

Alternate Treatments, Pre- and Posttesting

The very best control group for an intervention is an alternate intervention (Horvath, 1988). By providing two interventions, evaluators avoid the practical (and ethical) problem of not providing services to a group of participants in the study. Two interventions also strengthen conclusions by allowing researchers to state that not only is Program A effective, but it is more effective than Program B. Then we can begin to examine the differences between programs A and B to determine which components of the program might be responsible for the positive changes. When selecting an alternate intervention, there are several options. The easiest option for a control group is to use an existing program, if there is one. A second option is to provide a program that is similar to the one being evaluated, minus one or two components. This is also known as a treatment dismantling design and allows assessment of the importance of selected components of the program. Another choice is to select a program that might be considered the state-of-the-art in a particular field of IPV prevention. Of course, there is the very real possibility that the alternate program will outperform the program being evaluated. In this case, staff should consider adopting the other program or, at the very least, modifying the current program to draw on the strengths of the alternate. If practical limitations prevent implementation of a similar prevention program, a prevention video or other easily administered IPV intervention could be used.

One potential problem is that both programs might result in equivalent changes on the selected outcome measures. In this case, the better intervention will maintain its effect several months after the intervention, thus creating the need for long-term follow-up assessments.

USING EVALUATION RESULTS

Obviously, the most important aspect of conducting an evaluation is using the results of that evaluation to fulfill its original purpose. When an evaluation study is done correctly, reporting the results is a relatively simple task because the entire study should have been designed with this purpose in mind. For instance, if the purpose of the evaluation is to share positive initial outcomes with potential funders, practitioners should have a good idea of what they want to say in the report before the evaluation begins. For any given service, the agency would probably like to report how many people the program serves; the types of needs that participants in the program generally have; the number of people that successfully complete the program; barriers to program completion; the benefits that participants report; and the gains in skills, knowledge, attitudes, behavior, or health and well-being among participants following the program. The evaluation study, then, is simply a tool to provide some science and objectivity to the evaluation and a means to provide some numbers for the measurable outcomes chosen for evaluation. The type of information presented, and the way that it is presented, will vary depending upon the audience and the purpose of the presentation. This section provides suggestions for the presentation of evaluation results to stakeholders and funders for the purpose of generating additional support for programs, to the community and other stakeholders for the purpose of improving public relations and generating public support for programs, and to service providers and program administrators for the purpose of modifying and improving services. Although there are several other potential audiences for the results of an evaluation, these three groups will be used here as guidelines that can be modified for other potential audiences.

Funders

Written reports and face-to-face presentations use numbers to support the story that the program wants to tell. Presenters should supplement numbers with testimonials and statements from participants, but sparingly. Funders may also want to see the ratio of program effects to program cost (e.g., the cost per unit of attitude change). Such cost–effect analysis may be more useful to funders and agency administrators than to practitioners or community constituents.

Community

Evaluation results may be reported in brochures, web pages, press releases, letters, and memos. Again, if the program has a story to tell, staff will want to use more narrative, using numbers only to highlight important points; using participant statements and testimonials more freely; and using charts, graphs, pictures, and bulleted lists to make the information easy to read and understand. Explanations about programs geared to community members should be simple and to the point, avoiding jargon and avoiding most of the process-oriented detail service providers enjoy.

Service Providers

Written reports, face-to-face presentations, and workgroups can best be used. Service providers and evaluators will need to get into the details of the measurement tools and the raw data. Service providers need to know exactly how and where participants improved and, more importantly, where they did not improve. Evaluation reports need to describe who improved (including age, ethnicity, gender, education, income, family make-up, and types of presenting problems) and who did not improve. In some cases, the results might indicate that the measurement tool was flawed or did not capture the constructs that the program intended to change. In other cases, results might suggest that the program needs to be adapted to accommodate different populations, emphasize different constructs (i.e., skills, knowledge, attitudes, behavior), or modify other aspects of the service delivery (i.e., time, location, language). Service providers want to know about outcome, but they are often equally interested in process evaluations.

SUMMARY

Conducting an outcome evaluation of a prevention program can be challenging and rewarding. Potential challenges include staff resistance to evaluation; practical difficulties regarding the time and effort required to conduct an evaluation; theoretical difficulties linking the specific intervention techniques to the desired goals and objectives; and technical difficulties related to developing outcome measures, experimental designs, and analyzing data. Potential rewards include a significantly improved intervention, documentation of positive outcomes, and a staff that is more motivated and more focused on the goals and objectives of the interventions. Various funding agencies, social science departments of local colleges and universities, state departments of health and human services, the Centers for Disease Control and Prevention, or the National Institutes of Health can assist in conducting an outcome evaluation.

REFERENCES

Bennett, L., & Fineran, S. (1998). Sexual and severe physical violence of high school students: Power beliefs, gender, and relationship. *American Journal of Orthopsychiatry, 68*, 645–652.

Burt, M. (1980). Cultural myths and supports for rape. *Journal of Personality and Social Psychology, 38*, 217–230.

Dutton, D. G., Bodnarchuk, M., Kropp, R., Hart, S. D., & Ogloff, J. R. P. (1997). Wife assault treatment and criminal recidivism: An 11-year follow-up. *International Journal of Offender Therapy and Comparative Criminology, 41*, 9–23

Edleson, J. L., & Frick, C. (1997). *Evaluating domestic violence programs*. Minneapolis, MN: Domestic Abuse Project.

Horvath, P. (1988). Placebos and common factors in two decades of psychotherapy research. Psychological Bulletin, *104*(2), 214–225.

Koss, M., & Harvey, M. (1987). *The rape victim: Clinical and community approaches to treatment*. Lexington, MA: Stephen Greene.

Koss, M., & Oros, C. (1982). Sexual experiences survey: A research instrument investigating sexual aggression and victimization. *Journal of Consulting and Clinical Psychology, 50*, 455–457.

Malamuth, N. M. (1989). The attraction to sexual aggression scale: Part one. *Journal of Sex Research, 26*, 26–49.

Marshall, W. L., Jones, R., Ward, T., Johnston, P., & Barbaree, H. E. (1991). Treatment outcome with sex offenders. *Clinical Psychology Review, 11*, 465–485.

Missouri Division of Health. (1985). *Myths and facts about sexual assault*. Jefferson City, MO: Department of Social Services.

Nicholson, M. E., Maney, D. W., Blair, K., Wamboldt, P. M., Mahoney, B. S., & Yuan, J. (1998). Trends in alcohol-related campus violence: Implications for prevention. *Journal of Alcohol and Drug Education, 43*, 34–52.

O'Keefe, M. (1997). Predictors of dating violence among high school students. *Journal of Interpersonal Violence, 12*, 546–568.

Schewe, P. A., & O'Donohue, W. T. (1993a). Rape prevention: Methodological problems and new directions. *Clinical Psychology Review, 13*, 667–682.

Schewe, P. A., & O'Donohue, W. T. (1993b). Sexual abuse prevention with high-risk males: The roles of victim empathy and rape myths. *Violence and Victims, 8*, 336–348.

Schewe, P. A., & O'Donohue, W. T. (1998). Psychometrics of the Rape Conformity Assessment and other measures: Implications for rape prevention. *Sexual Abuse, 10*(2), 97–112.