10

Evaluating Ecological Mental Health Interventions in Refugee Communities

Jon Hubbard and Kenneth E. Miller

As the preceding chapters have shown, considerable progress has been made in the design and implementation of ecological mental health projects with communities displaced by political violence. Innovative strategies have been developed to help refugee communities respond effectively to their own mental health needs, using methods that integrate local and Western knowledge, beliefs, and practices. The critical question now is whether these programs are achieving their goals of empowering communities and improving the mental health and psychosocial well-being of community members. That is, to what extent are ecological mental health interventions with refugees *effective*?

In order to answer this question, the editors of this volume asked the contributing authors to (1) describe the methods they used in evaluating their projects, (2) summarize their evaluation findings, and (3) discuss any challenges they encountered while carrying out their evaluations. For authors who were unable to evaluate their interventions, the editors asked that they discuss the obstacles they had encountered to conducting systematic evaluations. The emphasis on identifying challenges and ob-

stacles to conducting evaluations was intended to generate a discussion of commonly encountered evaluation roadblocks, and to begin exploring a range of possible solutions to those roadblocks.

In reviewing the evaluation sections of the preceding chapters, we found that two themes were readily apparent. First, it is considerably easier to carry out systematic evaluations of ecological interventions in the comparatively safe and stable environments of resettlement countries such as the United States. In contrast to projects implemented in or near zones of ongoing violent conflict, interventions in the United States and other industrialized nations do not have to contend with the recurrent threat of violence and forced relocation; they have greater access to evaluation resources (materials, computers, consultants); and they typically have greater control over participation in their interventions, which allows for the development of more rigorous evaluation designs. Conversely, ecological interventions in or near conflict zones must contend with precisely the opposite conditions: ongoing vulnerability to further acts of violence that may result in repeated experiences of displacement, a lack of evaluation resources, and minimal control over who participates (and who does not) from week to week in the project. From an evaluation standpoint, such conditions are clearly far from ideal. In fact, they appear to be sufficiently formidable as to discourage program staff from attempting to carry out systematic evaluations of their projects—the second theme evident in our review of the evaluation sections of the chapters in this book. The authors in this volume are not alone; indeed, a review of other published accounts of ecological interventions with refugees in areas of ongoing conflict reveals a similar pattern: innovative program designs, compelling implementation strategies, and minimal discussion of actual outcome data (e.g., de Jong, 2002).

We appreciate the magnitude of the obstacles that program staff working in or near zones of conflict commonly encounter when trying to evaluate their interventions, and mean no disrespect in underscoring the lack of sound evaluation data. Indeed, we have encountered these same obstacles in our own work. We are concerned, however, about inadvertently creating what psychologist Robin Dawes (1994) has termed a "house of cards"—a substantial body of anecdotal evidence with little empirical data to support it. In the absence of sound evaluation findings, we cannot know the extent to which our interventions are truly effective.

We have no way of knowing which aspects of our programs are working well, and which components need to be altered or discarded. Confidence in the effectiveness of our interventions becomes more a matter of faith than knowledge, based more on subjective impression than organized assessment. This was precisely the situation with the highly funded Project DARE ("Dare to keep kids off drugs"), a U.S.-based project designed to prevent drug abuse and promote healthy psychosocial development among youth. Faith-based confidence in the DARE program was extremely high, and the program was implemented, at a cost of millions of dollars, in schools throughout the country. Unfortunately, a systematic evaluation of the DARE program in the state of Illinois showed it to have only minimal impact on students' drug use immediately after the intervention and *no impact at all* at 1 and 2 years post-intervention. In addition, the program had no effect on enhancing children's social skills (Enett et al., 1994). As it turned out, confidence in the DARE program was indeed something of a house of cards, one that cost a great deal of time and money while yielding few, if any, beneficial results.

Clearly, evaluations are essential if we wish to have well-founded confidence that our interventions are achieving their intended goals. Further, evaluations can help us identify problems with the design or implementation of our programs that may be diminishing their effectiveness. Finally, well conducted evaluations can help us answer a range of other interesting questions, such as: Who are we reaching with our programs? Who are we failing to reach, and why? What unanticipated effects are our programs having in the community, both positive and negative?

The goals of this chapter are twofold: (1) to describe the rationale and methods of two key types of program evaluation: process evaluations and outcome evaluations; and (2) to suggest strategies for carrying out process and outcome evaluations in refugee settings that are in or near situations of ongoing conflict. We recognize that organizations working with refugees in developing countries typically have limited budgets, work under chaotic and often stressful conditions, and may lack staff members with expertise in program evaluation. However, we believe that informative evaluations can be conducted using modest resources, that evaluation designs can be tailored to the demands of highly challenging settings, and that sound evaluations can be carried out with

a minimum of evaluation expertise. Our hope is to both demystify the evaluation process, and to outline a set of evaluation strategies that can provide meaningful data yet still be implemented under the difficult circumstances in which ecological interventions with refugees are typically conducted.

PROCESS EVALUATIONS

Several years ago, one of us helped to evaluate the outcome (effectiveness) of an ecological mental health project for rural communities in a Latin American country that had been devastated by a widespread campaign of state-sponsored violent repression. For nearly 3 years, representatives of numerous villages had traveled to a central location in order to participate in week long trainings in the theory and methods of the mental health intervention. The expectation was that they would return to their villages and implement what they had learned, developing workshops designed to help community members, and children in particular, heal from the effects of the violence and subsequent displacement.

Unfortunately, it was impossible to evaluate the effectiveness of the participants' work in their own communities, for it turned out that no one had actually implemented the mental health activities they had learned in the trainings. This was quite a surprise for the staff of the organization that had provided the training, who had assumed that project trainees were actively putting into practice the knowledge and skills they had acquired. In a subsequent workshop held to explore the reasons for the trainees' failure to implement the mental health intervention in their home communities, the participants offered a list of significant obstacles they had encountered, including a lack of supervision and consultation by program staff, the lack of a written manual to which they could turn for descriptions of the various intervention techniques, and resistance from religiously conservative community members who regarded the intervention as subversive and had threatened to alert the army if the intervention was carried out. These and other data gathered during this *process evaluation* workshop provided invaluable information to the staff of the project, who were then better informed about the kinds of support

trainees needed, and the kinds of challenges they faced upon returning to their home communities.

Process evaluations are designed to answer several key questions:

- To what extent has an intervention been implemented as planned?
- What factors are causing an intervention to be implemented differently than planned (or to not be implemented at all)? How might those factors be addressed?
- To what extent is an intervention reaching the intended (target) population?
- Who within the target population is *not* being reached by the intervention, and what are the obstacles to participation for these individuals (families, groups, communities)?

Evaluations that address the first two questions are sometimes referred to as implementation or fidelity evaluations, while those addressing the latter two questions are sometimes called efficiency evaluations (Dalton, Elias, & Wandersman, 2001; Miller, 1999). In this chapter, we use the term *process evaluation* to refer collectively to any evaluation that addresses any or all of these questions.

It is an axiom of program evaluation that an intervention is unlikely be effective if it is not implemented appropriately. In the example offered earlier, the intervention was not implemented at all. More commonly, however, interventions are not effective because they have been implemented poorly rather than not at all. For example, one of our colleagues was involved in a well designed community intervention designed to increase safe sex behavior among gay men in a large urban area of the United States. At the completion of the multisession group intervention, the program was found to have had little effect on increasing the participants' safe sex behavior. Although it would have been easy to conclude that the project was poorly designed and therefore ineffective, a closer examination revealed something quite different. It turned out that one of the group leaders was quite uncomfortable talking explicitly about the kinds of sexual behaviors that put gay men at risk for the transmission of HIV, the virus that causes AIDS. The group leader's discomfort led to an

avoidance of the very sort of discussions that were to essential to the program's success. No wonder that the program showed few beneficial effects—it had not been implemented as designed.

This example illustrates an important point: Process evaluations can help us determine whether an intervention that is not effective suffers from a poor *program design*—in which case the intervention itself should be modified, or from problems of *program implementation*, in which case, obstacles to effective implementation should be addressed (Dalton et al., 2001). There is another important point to be drawn from this example and that of the project in Latin America: until we can be reasonably certain that a program has been well implemented, it makes little sense to evaluate the outcome or effectiveness of the program. *The effectiveness of any intervention depends in part on the quality of its implementation.*

Strategies for Evaluating the Implementation of Ecological Interventions

The first and most obvious question asked by a process evaluation is whether an intervention was actually conducted. Although it may seem odd to even ask such a basic question, our earlier example of the mental health project that was *not* implemented by trained paraprofessionals in their home communities underscores the importance of ensuring that people are actually conducting the intervention. In our experience, the risk of non-implementation increases when paraprofessional staff are trained in a central location and expected to implement what they have learned in their home communities without regular supervision and consultation from project staff. Mental health work, especially with survivors of extreme violence and forced displacement, is inherently complex and challenging. A brief but intense period of training cannot substitute for ongoing guidance and consultation. Of course, support and supervision are needed whether paraprofessional staff are working in communities distant from the central training site, or in the same community in which the training is offered.

Let us assume that trained community members are receiving adequate support from more experienced project staff members, and are actively involved in implementing an ecological mental health interven-

tion with other community members. The question now becomes one of assessing the fidelity of the project's implementation; that is, is the intervention being conducted as planned? If not, why not?

There are several ways of examining the issue of fidelity.

- **Co-facilitation.** Co-facilitation involves having trained community members co-lead the intervention with a more experienced staff member (who may also be a member of the target community), who can serve as a role model for the appropriate implementation of the intervention while also ensuring that activities are implemented as designed.

- **Observation.** Experienced staff members can observe directly the implementation of an intervention, noting areas of fidelity as well as obstacles to implementing the intervention as designed.

- **Post-session interviews with trained paraprofessionals.** Supervisory staff can regularly conduct interviews with trained paraprofessionals following each session of the intervention. In addition to ensuring that all of the planned activities were carried out (or to discovering why some activities were not implemented), regular interviews also provide time for providing paraprofessional staff with support and supervision.

- **Videotaping sessions or events.** Videotaping intervention sessions or specific intervention events provides an invaluable source of data regarding what actually happened during the implementation of project activities. For example, the process of videotaping the implementation by schoolteachers of the *Playing to Grow* intervention with Guatemalan refugee children in Mexican refugee camps provided the project staff with invaluable data regarding the kinds of activities that worked well and those that needed to be altered, and also helped staff identify areas in which the schoolteachers needed additional training (Miller & Billings, 1994;

Miller, Billings, & Farias, 1995). A major concern with videotaping is the extent to which participants feel comfortable being videotaped. In our experience, children are often more comfortable being videotaped than adults. It is also possible that cultural differences in what it means to allow oneself to be photographed or videotaped may shape a given community's openness to having the intervention (or parts of it) captured on video.

- **Focus groups with project participants.** Focus groups are a form of group interview, in which a small group of participants is asked to discuss a set of questions related to a particular topic (Dean, 1994; Krueger, 1994). Focus groups are ideal for conducting process evaluations, because it is relatively easy to ask participants in an intervention to talk about the activities that were conducted, and to comment on those activities they found most helpful and those they believe should be altered or dropped. Focus groups are efficient, as data are gathered from several people at once. Focus groups typically have about 8 to 10 people, and can run as long as participants are willing to continue discussing the questions at hand (we usually limit focus groups to a maximum of about 2 hours). The groups can take place in any community setting, such as a school room, a clinic, or a community member's home. *The most effective focus groups are those in which the facilitator is able to generate discussion among group members, rather than fostering dyadic interactions between each group member and the facilitator.* For those interested in learning more about focus groups, Richard Krueger (1994) and Debra Dean (1994) have written excellent guides.

- **Questionnaires.** Another approach to ensuring fidelity of implementation is to have paraprofessional staff (i.e., trained community members) complete brief questionnaires or checklists, indicating which of the planned activities were actually carried out, and which were not. Supervisory staff should then discuss with the paraprofessionals the reasons

that specific activities were either not carried out, or were implemented differently than planned. It is also useful, although time and labor intensive, to have *project participants* complete **participant satisfaction questionnaires** at the conclusion of each meeting, or alternatively, at the midpoint and conclusion of the intervention. The questionnaires should ask participants to indicate their level of satisfaction with the specific intervention activities, and may also cover other areas such as satisfaction with the performance of the paraprofessional staff, the quality and appropriateness of the intervention materials, and so forth. For participants with limited reading skills, project staff can read the questionnaire items aloud and participants can indicate their responses using such answer choices as an empty circle ("not at all satisfied"), a partially filled circle ("somewhat satisfied"), or a completely filled circle ("very satisfied"). Methods of analyzing questionnaire data are discussed farther below.

Strategies for Assessing Who Is (and Who Is Not) Being Reached by an Intervention

We suggested earlier that process evaluations can also be used to assess the extent to which an intervention is reaching the target population. The importance of this type of process evaluation cannot be overstated. If the intervention is designed for a particular subgroup within a community (e.g., people experiencing persistent symptoms of psychological distress) but is reaching only people who are showing few signs of distress, it is unlikely to be effective and may represent a poor expenditure of limited resources. Similarly, if an intervention is designed to serve a community in its entirety but people don't participate in it, or participate only sporadically, the intervention is unlikely to show positive results. Although low participation rates may suggest a problem with the design of an intervention, they may also reflect factors that having nothing to do with the program's design. Such factors might include negative inaccurate perceptions of the project, a lack of trust between community members

and the project's staff members, a lack of effective recruitment and advertisement efforts, or conflicts with other community or family commitments. We return shortly to a discussion of obstacles to participation.

To assess the extent to which an intervention is reaching the target population, it is necessary to answer a number of related questions:

- Who is the target population for the intervention?
- Who is actually participating in the intervention?
- What constitutes "participation"? (e.g., attendance at 75% of the meetings, activities, or sessions)
- What percentage of the target population can be realistically expected to participate in the intervention?
- Of those who could benefit from the intervention, who is *not* participating? What are the obstacles to their participation, and how these obstacles be overcome?

Who Is the Target Population for the Intervention?

This is a relatively straightforward question. Who do you want to reach with your intervention? The whole community? Parents with school age children? Widows? Adolescents? New or expectant mothers? Unemployed men? Schoolteachers? Families? Individuals experiencing high levels of war-related trauma?

Once you have designated the target population, it is ideal (although not absolutely necessary) to try to assess the number of people or families in the community who comprise this target group. Obviously, if your intervention targets the whole community, this is simple: The number of people or families in the community is the same as the number in your target group. However, because many interventions have components that serve specific subgroups, it is helpful to know the size of those subgroups, as this allows you to estimate a proportion of the subgroup that your intervention should reasonably be able to reach. For example, if you are developing a project to assist children disabled by political violence,

and you estimate that there are roughly 100 such children in the community where you are working, you might aim to involve 50 of them with your intervention during the first year (the actual numbers you designate will depend on the resources available and the nature of your intervention). Having set a realistic goal of reaching 50 children, you have set the stage for a relatively straightforward evaluation to determine whether you have achieved your goal at the year's end.

Monitoring Participation: Who Is Actually Taking Part in the Intervention?

There are several strategies available to assess who is and is not participating in an intervention. The simplest approach for monitoring participation is to keep an attendance record or *participant log*. For community-wide interventions, this might involve having a staff member observe and document the number of different people (or families) who use a particular setting during a given period of time. For example, for 2 hours each day over the course of a week, project staff could document the number of people who use a community playground or attend a community center. It is also possible to count or make a list of the people who participate in a particular community activity. Another strategy is to randomly survey community members to ask whether they are familiar with the project, and whether they have participated in any of the project activities.

For smaller group activities, such as social support groups, it is relatively simple to keep an attendance log that documents who attends each meeting of the group. For projects that ask participants to engage in certain tasks between group meetings (e.g., visit other group members, practice certain new skills), a record can be kept at each meeting of each participant's between-session completed tasks.

Why is it important to keep track of who participates in an intervention? First, this allows us to assess whether we are reaching those people for whom the intervention was designed. Second, keeping a participant log allows us to keep track of how regularly people are participating in the various intervention activities. This will be important later on when we conduct outcome evaluations, as there is usually a strong relationship between the degree to which people participate in an intervention and

the intervention's effectiveness. Stated differently, those people with the greatest degree of participation tend to experience the greatest benefit from a project.

Imagine a support group offered to distressed widows. Some of the women attend all 10 meetings of the group, whereas other women attend less regularly, some showing up every other week and others attending only one or two meetings. We can expect that those women who attend more of the meetings will experience a greater benefit from the intervention. *Only by keeping a participant log will we be able to keep track of each woman's level of participation in the program.* When evaluating the effectiveness of the support group, our primary aim is to show that it is effective for those women who actively participate in it. What constitutes "active" participation? There is no set rule, although generally participation below 50% (e.g., attending less than half of the meetings of a support group) is regarded as partial or nonattendance. It would, for example, be a valuable finding if we could show that women who attended at least half of the sessions showed a marked increase in psychological well-being, and those who attended all of the sessions experienced an even greater improvement.

Identifying Prospective Participants Who Are not Being Reached by the Intervention

It is relatively easy to keep track of who *is* participating in an intervention. It is somewhat more challenging to identify those individuals or families who might benefit from the program but are *not* participating in it. If we know the approximate size of the target group (e.g., the number of widows in the community), we can easily compare the number of people from the target group that *have* participated in the intervention with the total number of members of the group. This will give us a good idea of the number of potential participants who have not yet gotten involved in the project.

Often, however, we don't know the size of the target group. For example, we may not know how many individuals are struggling with depression or trauma in a community, and we may not have the time or resources to carry out a community-wide assessment. In such cases, how can we know whether we are reaching most of those people who might

benefit from our intervention? One useful approach is to ask program staff (who are normally members of the local community), as well as program participants, to identify other community members whom they believe might benefit by participating in the intervention.

Identifying Obstacles to Participation

There are numerous reasons why community members might not participate in a mental health or psychosocial intervention. They might be concerned about a negative stigma associated with participating in the program; they might need to work during the time that program activities are offered; they might not be aware of the intervention or may have an inaccurate idea of what it involves; they may be concerned about issues of confidentiality; or they may not have anyone available to watch their children while they participate in the program. Most such obstacles to participation can be readily overcome with a bit of flexibility and creativity on the part of program staff.

There are a couple of simple strategies for identifying the specific obstacles that affect participation in the community you work in. One approach is to ask people who *are* participating to talk about why they think other community members are *not* participating. Another approach is to either informally or through formal interviews or focus groups ask *non-participants* about the reasons they have chosen not to take part in the intervention. Once the obstacles to participation have been identified, program staff can then work with community members to find ways of overcoming those obstacles, and thereby expand the reach (and thus the impact) of the program.

OUTCOME EVALUATION

Whereas process evaluations provide the information we need in order to know if our programs are being implemented—and implemented in the ways in which we have planned—we need to gather additional data to know if our programs are having their intended effect. Collecting information for the purpose of assessing the effectiveness of a program is referred to as *outcome evaluation*. Outcome evaluations can be difficult to

conduct under the best of circumstances and may feel impossible to carry out under the fluid and often chaotic conditions of conflict zones and refugee camps. The authors do not portend to have solutions for all the complex challenges that arise when assessing the effectiveness of programs being carried out in these turbulent contexts; however, suggestions are offered that may help in designing and conducting more useful and informative outcome evaluations under these less than ideal conditions.

Outcome evaluations are designed to answer the questions:

- How well did the program achieve its goals and objectives?
- Who benefited most from the intervention or what components of the program had the greatest impact?
- Did the program have unintended consequences (positive or negative)?
- What was learned that would inform future interventions or other similar programs?

There are many ways of answering these questions and a variety of issues to consider when designing your outcome evaluation. The following sections highlight some of these issues and suggest strategies for handling challenges that can arise when designing the evaluation, choosing appropriate methods and measures or analyzing, interpreting, and reporting the results of an outcome evaluation.

Strategies for Evaluating the Effectiveness of Ecological Interventions

Outcome evaluation begins with a series of questions that need to be answered in order to know the degree to which a program has been effective in meeting its expectations. If the expectations for the program have not been clearly articulated, it will be very difficult to design a successful

evaluation. This may sound self-evident but it is surprising how much difficulty many program staff have explaining exactly what their program's goals and objectives are. *Goals* are the general aspirations of a program and are often stated in fairly broad terms (e.g., to reduce post-conflict distress and increase feelings of well-being in a particular displaced population). Program *objectives* are the specific methods used to achieve the program's goals and need to be stated in more precise and *measurable* terms (e.g., organize 20 12-week adult therapy groups to reduce symptoms of depression and anxiety and increase social connectedness among participants). Objectives that are too broad or vague do not readily lend themselves to evaluation. Frequently programs have multiple objectives and there should be *outcome indicators* associated with each individual objective. Outcome indicators are the specific items that are used to judge the success of a program and should be directly tied to the goals and objectives. An example of program goals and objectives is found in Box 1 (next page), which comes from the intervention with Sierra Leonean refugees described in chapter 2 of this volume.

In general, the term *effects* refers to immediate program outcomes while *impacts* refers to the more enduring long-term outcomes (the usage differs by field). We use the terms somewhat interchangeably in this chapter as many of the strategies described are useful for both.

There are many good articles and books covering the basic principles for conducting outcome evaluations (e.g., Fetterman, Kaftarian, & Wandersman, 1996; Kazdin, 1992; Linney & Wandersman, 1991; Wholey, Hatry, & Newcomer, 1994); therefore we do not go into detail on 'the basics' here, but rather focus on applying some of these basic ideas to the evaluation of interventions in high adversity contexts like post-conflict zones.

Considerations in Designing an Outcome Evaluation

One of the fundamental reasons that many program evaluations fail is simply poor planning. Although significant efforts are made to develop detailed intervention plans, relatively little time and effort is put into designing the evaluation components. Outcome evaluation, like research, will produce much more meaningful results when it is well de-

signed and conducted in as rigorous a manner as possible. While there may be limits on the rigor that can be achieved in the contexts in which

TABLE 10.1. Objectives and Goals From a Project With Sierra Leonean Refugees*

Goal	Objective	Outcome measure
1. To provide psycho-education and mental health services to Sierra Leonean adults living in Guinean refugee camps.	1. Implement eight, 12-week therapy groups (four male & four female) for 8 to 12 adults per group in each refugee camp in which the program is operating. Groups will be designed provide psychoeducation, reduce post-trauma symptoms and increase social functioning and ability to engage in the tasks of daily living.	1.1 Symptom checklists for depression, anxiety, somatic symptoms and PTSD administered at intake, & 1, 3, and 6 months (quantitative) 1.2 Measure of social connectedness and community involvement (quantitative) 1.3 Behavior checklist of indicators of involvement in daily activities (quantitative). 1.4 Participant and community focus groups organized four times per year (qualitative)
2. To create a cadre of peer counselors who are able to provide culturally appropriate mental health services to their fellow refugees	2. Professional expatriate psychotherapists will provide long-term applied training (topic-specific trainings, in-session modeling, pre- and post-therapy training and ongoing case supervision) to refugee peer counselors (ratio of approximately 10 trainees for each senior staff). Training will continue throughout employment with the program and be adapted to the changing needs and abilities of each staff.	2.1 Written examinations on basic counseling skills and the impact of war on individuals, families and communities. Conducted following initial month of training and then on a semi-annual basis (quantitative/ qualitative) 2.2 Quarterly supervisor ratings on counseling skills, ability to incorporate training into practice and job performance (qualitative /quantitative)

Note: These goals and objectives are taken from the project described in chapter 3. They are selected examples from a larger set and do not represent a complete evaluation design.

many ecological interventions operate, there are ways to improve the odds of ending up with useful results.

Begin by designing the outcome evaluation at the same time you develop the intervention plan for your program. Make it part of your proposal or implementation plan from the outset. All too often outcome evaluation is an afterthought for ecological mental health interventions and psychosocial programs that is added on after the program has been developed and is in operation or when additional funding is sought. It is much harder to develop and implement a successful evaluation post-hoc. As you go through the process of developing evaluation questions and selecting measures keep returning to the basic questions: What was the program designed to achieve? How will you know if it was effective?

Collecting Data

Determining that a program has had an impact requires several types of information: baseline data, follow-up data and, if possible, comparison or control group data. *Baseline data* is collected from the target population on all of the dimensions of interest prior to the start of the intervention. Baseline data can be collected as part of an overall needs assessment or as a specific evaluation activity. This is the data with which all future assessments will be compared, and as such, all questions of interest must be included.

At a minimum there needs to be at least one *follow-up assessment* in which the measures administered during the baseline assessment are given again. Usually, a follow-up assessment is conducted at the end of each specific program activity or intervention cycle. Sometimes, however, there are multiple assessment points during or following an intervention.

Finally, it is impossible to make any conclusive statements about the effectiveness of a program without collecting baseline and follow-up data from community members who did not participate in the intervention, but who are, in all other ways, similar to program participants. Even if your follow-up assessment indicates a significant positive effect among participants, for example a large drop in symptoms, you cannot, with any certainty, attribute the change to your intervention or interpret its meaning without a comparison or *control group*. It is possible that during

the period of your intervention, symptoms dropped in the entire community and the effects seen among program participants were not due to your intervention but to more general phenomenon like the passage of time or reduced tension in the region. It is even possible that symptoms could have dropped to a greater degree in the entire community than among your program participants, in which case the program would actually be impeding recovery. Control group data is the standard against which program results are compared in order to judge their significance.

Control groups can be created in several ways. Once the target audience has been identified, participants can be randomly assigned into groups—those who will participate in the intervention and those who will not. For obvious reasons this approach can be difficult to use in many ecological mental health interventions. However, if the program can only serve part of the target audience at a time, people can be randomly placed into sequential cycles of the intervention. If you have several types of interventions within one program, you can randomize the order in which people receive the different interventions or have some people receive several interventions at once (e.g., group therapy and life skills training) and compare them to those who receive single interventions.

All of these methods can be very difficult to organize and manage in the contexts we are discussing and finding a method for including a random "non-intervention group" or control group in the evaluation design may not be possible. It can be enough of a challenge to gather evaluation data from program participants in these settings and the added burden of identifying and following an adequate comparison sample can overwhelm program resources. Yet there are other ways to try and build a case for program effectiveness. As previously noted, you can use internal indicators like attendance records to determine if people who participated more frequently received greater benefit from an intervention than those who attended less often. This method was used by one of the authors in a recent program evaluation and it could be shown that higher attendance was related to greater symptom reduction among refugees receiving group therapy. Examining the attendance records also indicated that those with the most severe psychological symptoms at the baseline assessment were the most likely to drop out. A preliminary follow-

up with these drop-out cases indicated that they were not ready to address their problems in a group setting and some requested individual sessions where they could discuss their problems "in private."

Other methods for substantiating effectiveness include gathering follow-up data on clients for whom you have baseline information but who did not participate in the intervention at all (a zero attendance group). We have found people very willing to participate in follow-up assessments even when they have not attended any parts of the intervention. This strategy also provides information about why some people chose not to participate (process evaluation data) and provides a chance to re-engage these people in the program. It is useful to collect data that allows for examination of possible differences between those who engage and those who don't (e.g., demographic information like age and gender or experiential data like higher rates of traumatic exposure). Another method is to conduct baseline assessments with a cohort of participants prior to their engaging in an intervention (e.g., 1 month prior) and then re-assessing them when the program is about to begin. This pre-intervention or 'wait period' data can be compared with data collected at a similar period into the intervention and/or with data from another similar cohort who were assessed at the beginning and one month into the program (this is a "waiting-list" type strategy).

Planning for the Unpredictable

Another important design consideration in these contexts is to be prepared for change. If you know you are working in a continually changing or fluid environment, which is frequently the case when intervening with displaced populations, build this into your evaluation design. Although you cannot plan for every unexpected turn of events you should design your outcome evaluation around the reality of the situation as you have come to know it. Some strategies include:

- Keep measures brief so that information can be collected quickly. Assessing a few key indicators for each of the domains of interest will be less cumbersome for staff and participants than lengthy all-inclusive assessments. Brief measures also lend themselves to more frequent assessments.

- Shorten the time between data collection points. This is helpful if your program is suddenly interrupted and it will also allow you to analyze more precisely when significant changes occurred.

- Evaluate the effectiveness of specific techniques or components of the program or intervention as well as the overall impact. Knowing how individual parts of the intervention are succeeding can often be more informative than measures of overall effectiveness. It allows you to compare interventions and to analyze which parts of the program are having the most (or least) impact. This strategy also protects against ending up with no evaluation data, if the whole program ends prematurely or parts of the program are terminated.

- Have a system in place for tracking participants. In highly mobile situations such as refugee camps, it is helpful to collect good contact information, such as a friend or relative who might know where to find a participant or information about where someone intends to go if the situation changes (e.g., the camp is closed down). We have found this data crucial in locating the "zero attendance comparison group" discussed earlier.

- Have a plan in place that describes how confidential information will be protected or destroyed if staff needs to quickly relocate. For example, who will be responsible for evaluation data in an emergency situation? It is always better to destroy sensitive or confidential data if you are forced to evacuate than to leave it behind or transport it in an unprotected or unregulated manner.

Including Community Members in the Evaluation Process

An outcome evaluation will produce more meaningful results if you find ways to involve people from the local community in the evaluation process. We believe that this is an extremely important part of develop-

ing an appropriate evaluation strategy, and one that is quite often overlooked. Outcome evaluation should be a participatory process with as much local input and involvement as possible (Dalton et al., 2001; Fetterman et al., 1996). It is becoming more common for program developers to elicit community input when designing new *interventions* (usually during the needs assessment phase); however, few programs appear to seek local participation and input when designing the evaluation components of their programs.

Community involvement can have a significant impact on the success or failure of an evaluation. When community members are engaged in developing an outcome evaluation from the beginning, they feel ownership of the process. It is easier to get the community to *participate* in the evaluation if they have been involved in designing it. Through a cooperative design process, program staff and community members have an opportunity to develop an outcome evaluation that holds meaning for everyone concerned. It creates an environment in which they can begin to see evaluation as an important and valuable part of the program and not an added burden carried out only as a requirement of funding agencies and management at the "home office."

In addition, local participation can aid in developing more meaningful assessment questions and provide valuable feedback as to whether questions brought in from the outside will be understood by participants. They can tell you if the methods you have chosen to gather information will be accepted by the community or provide information that can impact the timing of your evaluation. If you know, for example, that it is inappropriate for women to meet alone with a man for an interview or that few people will be able to participate in a follow-up assessment because it is scheduled during a local festival, you can alter your design accordingly.

But it is important to do more than simply elicit local input. We suggest, whenever possible, hiring local staff to implement the evaluation or, at a minimum, include people from the community as regular members of your evaluation team. In our own programs, we have trained local staff to conduct interviews, facilitate focus groups, administer checklists, analyze data, and write reports. This often requires more time and effort spent recruiting, training, and supervising; however, the benefits are well worth the investment. In part, this is another way of empower-

ing the community you are serving. By creating local expertise in designing and carrying out program evaluation you are contributing to the sustainability of the program. There are direct benefits for the program as well. Along with being cultural resources, people are often more comfortable talking to someone from their culture who speaks their own language and understands the context for their answers.

That being said, it is also important to be aware of possible differences between the local staff you hire and the larger community. Often local staff are hired because they have more education, are multilingual or have worked previously for psychosocial programs. These differences in status or education may create barriers between local staff and program participants. ~~Do~~ We cannot assume that having someone from the community conduct an assessment will always result in more accurate information. A number of years ago, one of the authors was part of a longitudinal study examining the long-term psychological consequences for Khmer adolescents who lived through the horrors of the Pol Pot regime. He interviewed a young man who reported having had significantly high levels of post-trauma psychological symptoms since arriving in the United States several years earlier. When the interview was finished, the author briefly reviewed the assessment data reported by this young man from the previous summer and noticed that at that time he had indicated having had almost no symptoms. When questioned about this, the youth replied that during the previous assessment there had been an interpreter present, a Khmer man who was well respected in the community. He said he had not wanted to reveal his problems to someone from the local Khmer community but was willing to discuss them alone with the author, as he was "an American" and so it was alright.

We have found that it is not uncommon for people to be guarded with their comments when working through an interpreter due to concerns about confidentiality or underlying tensions created by differences in age, gender, religion, or culture. Common language does not necessarily translate into common beliefs or acceptance. The general point is that you cannot assume a specific intervention technique or methodology will work with any particular population or in any specific context until you do your homework and then try it out. You need to work with the community to understand how to adapt evaluation methods to local beliefs and conditions. Sometimes there are obstacles that you cannot do

anything about and simply must live with but it is important to know what they are.

It is also helpful to keep in mind the skills of those who will be conducting the evaluation. What kind of training will they need to ensure that you get valid results? How much supervision will be required? It is best to keep the design and techniques within the ability of the staff carrying out the work. It is also good to avoid developing a plan that requires university trained researchers if local community members with minimal formal education will be responsible for overseeing and conducting the evaluation. It is also important to provide clear instructions for each step in the evaluation process with clearly defined responsibilities and lines of supervision. Manuals can be useful resources and are helpful for keeping the evaluation on track.

External Consultants

Finally, if you do not have the internal expertise to design your outcome evaluation, bringing in an outside consultant can be very helpful. Find someone who will take the time to understand your programs objectives as well as the population you are working with and the context in which the program will be operating. Outside experts who are called in for quick consultations sometimes bring along their own agenda (and sometimes their own measures) that may not be the best fit for answering your particular program's questions.

Choosing Measures of Program Success

Selecting appropriate measures for evaluating the effectiveness of an ecological mental health intervention can be challenging and there are a variety of issues that can influence the choices you make. It is important to begin with clear evaluation questions and then look for the methods and measures that will provide the answers—not the other way around: Do not let the measures dictate the evaluation. Some of the issues you will want to consider when selecting measures include:

- **Who does the evaluation need to inform?** You may want to start by considering the audience(s) that your evaluation must satisfy. These audiences, sometimes referred to as stakeholders, include your program staff (field staff responsible for implementing the program and staff from the home office), funding agencies, other groups who may want to replicate your program and the local community. Often these groups have different needs or agendas which may require collecting different information.

- **From whom should information be collected?** Many evaluations limit themselves to collecting data from program participants. However, evaluation results will be more informative and useful if you employ multiple measures and collect information from a variety of informants. For example, if you are trying to assess the effectiveness of a program for children, you will have a more complete picture if you can gather information from the children, their peers, caregivers, and other adults, such as teachers, who know the children in particular contexts. Program effects may be more evident in one setting than another (e.g., in a structured situation like school than in less structured situations like home or play settings). Important information can be missed when data are gathered in only one setting or from only one group of respondents.

- **Who are the targets of the intervention?** The methods you choose to evaluate the program will differ depending on the population(s) being targeted. For example, are you interested in the program's effect on individuals? families or systems? or community wide effects? Many programs are interested in measuring the program's impact on several (or all) of these levels.

Types of Information to Collect

There has been a tendency for ecological mental health interventions, and many psychosocial programs, to use measures of psychological

symptoms as the sole method for establishing program success. Although measuring symptoms may be important and useful, there are other types of information that can also help establish program effectiveness. Types of data to consider collecting are:

- **Measures of psychological distress**: There are numerous standardized measures of psychopathology (many based on ICD or DSM taxonomies). The most frequently used by ecological mental health interventions and psychosocial programs assess depression, anxiety, somatic symptoms and post-traumatic stress disorder. Most often these measures will need to be translated, validated and adapted for use with new populations. The alternative is to develop new measures based on local problems and expressions of distress.

- **Process evaluation data**: Information gathered for the purpose of process evaluation can also be valuable when answering questions about program effectiveness. As described earlier, participation data and attendance logs can be useful for outcome evaluation as well.

- **Functional adaptation**: It can be useful to include measures that assess important areas of daily functioning, such as social adaptation, ability to parent children or engage in meaningful community activities in an outcome evaluation. Change that takes place in these domains can be as or more meaningful to participants than changes in psychological symptoms, and as such, they can be important indicators of program effectiveness. However, because there are no established measures of adaptive functioning, and because these behaviors are so culturally and situationally determined, you will need to develop your own measures (for an excellent approach to developing culturally appropriate measures of functioning, see Bolton & Tang, 2002).

- **Referral or service utilization data**: Gathering information on how well program participants are able to follow through on referrals made for them to other programs, or are able to utilize

other services available to them, can be helpful when evaluating program success. For example, it may be useful to know if people who have participated in your mental health intervention are more likely to enroll in skills training or job placement programs than they were prior to the intervention (or more likely to enroll than a comparison sample).

- **Ongoing outside events**: It is important to gather information on significant events that impact the local environment. Conditions such as increasing levels of violence or changes in access to important resources such as food, water, and medicine, while possibly unrelated to your specific program, may have a significant influence on your program's effectiveness. A detailed log of events that have taken place concurrently with an intervention can often prove to be crucial when interpreting the evaluation results.

There is another set of considerations to be made concerning the *nature* of the data to be collected for the evaluation. Two common distinctions that can impact your measure selection are qualitative vs. quantitative and emic vs. etic.

Qualitative Versus Quantitative Measures

Evaluation methods are frequently categorized according to whether they provide us with numbers (quantitative data) or words (qualitative data) and there has been an ongoing debate over the relative usefulness of each type in evaluating program effectiveness. *Quantitative methods* provide numerical data that can be analyzed with statistical methods. The quantitative techniques most frequently used in psychosocial outcome evaluations are checklists, surveys, and structured interviews. The numerical data derived from these techniques can be used for hypothesis testing (e.g., we expect that people who attended more than 50% of the intervention activities will have significantly more social connections at the 3-month follow-up assessment than those who attend less than 50% of the activities) and summarizing results for reports (e.g., percentage and frequency data).

Qualitative methods, on the other hand, provide narrative data that are primarily descriptive and interpretative, and highlight the unique or individual characteristics of the target population in its natural context. Methods include unstructured interviews, focus groups, and observation, as well as techniques like video taping, group mapping, role plays, and drama. Qualitative techniques have traditionally been associated with attempts to understand program effects from a local perspective, whereas quantitative techniques have been associated with a more scientific or outsider's perspective.

Recently, these traditional distinctions have been breaking down. Programs are developing quantitative checklists and surveys that reflect local beliefs and concerns and are using qualitative methods to collect baseline and follow-up data to assess changes in understanding or knowledge. For example, one of the authors conducts numerous brief interviews during the needs assessment phase of program development to gather information on local concepts of well-being and individual problems that have resulted from living through the war (a qualitative technique). The data generated through these interviews is then used to create new baseline and follow-up measures for assessing program effectiveness. These new checklist style measures (a quantitative technique) will be based on local perceptions.

Traditional survey and interview style techniques for evaluating program effectiveness often combine qualitative and quantitative methods as well. Sometimes a specific quantitative question like, "Are you separated from your family?" (yes/no) are followed by a qualitative question like, "How has this affected you?" The reverse strategy is also used. A mental health intake assessment might ask, "Tell me why you came to our program today?" or "What kind of assistance do you hope to receive?" These open-ended qualitative questions can be followed with a variety of quantitative questions that are of specific interest to the program (e.g., a list of post-trauma symptoms or common reasons people give for participating in the program).

Both types of approaches have their strengths and when used together can provide richer results. Recently, one of the authors was analyzing symptom data from a community mental health intervention and found a consistent and statistically significant relationship between age, depression, and anxiety; while being young was related to higher anxie-

ty scores, being older was associated with higher levels of depression. This quantitative finding was interesting, but it took qualitative data, gathered through interviews with program participants, to give meaning to the finding. Younger refugees described feeling as though life was passing them by while they sat in the camp; they should be going to school, starting a business and building a future for themselves. They expressed impatience with their situation, and were anxious to get on with their lives. Older refugees, on the other hand, expressed little hope of being able to start over again or rebuild their lives when they returned home from the camps and were more focused on the tremendous losses they had experienced.

Emic Versus Etic Measures

Another area of consideration when selecting evaluation measures is the degree to which program staff and/or other stakeholders are interested in understanding the intervention's effectiveness from an *emic* (insider or local) or *etic* (outsider or observer) perspective. Generally, when you use standardized measures or quantitative assessment tools that were originally developed for use with a different population than the one with which you are working, you are taking an etic approach. For example, assessing a construct such as depression among Filipinos or Vietnamese using a measure originally developed for use with middle class White Europeans would represent an etic approach. On the other hand, when you use measures or techniques that were locally derived, or methods that were designed to understand the impact of a program as community members perceive, experience, and understand it, you are taking a more emic evaluation approach. It is easy to see why qualitative techniques are often associated with emic approaches while quantitative techniques are linked with etic approaches. As demonstrated in the previous section, however, this generalization is not always valid.

As ecological mental health interventions have become more culturally sensitive, there has been a tendency to view emic methods as *good* assessment techniques while etic methods are characterized as *bad*, or at least less desirable techniques. As usual, the reality is more complicated than this and there are a variety of reasons why a particular program might choose to include one type or the other. For example, a program

might choose to include etic measures out of a desire to compare their results with those from other programs—in which case it is helpful to use common measures. In fact, the desire to generalize the results of one program to other populations (and vice versa) is a common underlying reason why some evaluators prefer standardized measures. It is difficult to compare results across programs or populations when the data collected are by definition specific to the local community and context, which is the case with emic data.

Another commonly cited reason for using existing or standardized measures is pragmatics. It is simply much easier and efficient to use a measure that has already been developed and found to be useful with other populations. Many program staff feel that they just do not have the expertise or resources to create their own measures.

In addition, funding agencies sometimes ask programs to include particular measures of program success as a requirement for funding. Almost always these are quantitative, etic measures that have demonstrated their effectiveness in other settings. They may not, however, be a good fit for evaluating a particular population or program. If faced with this situation, we suggest countering with a well thought out alternative method. Based on the experience of colleagues, funding agencies are often willing to accept substitute methods when a sound alternative plan is proposed and are becoming more receptive to methods that incorporate emic perspectives.

Historically, the outcome measures used by most ecological mental health interventions and psychosocial programs have been etic in nature, but there appears to be a growing tendency to approach evaluation from a more emic perspective. More often in recent years, programs are interested in understanding and measuring their effects as they are experienced by the participants. There is an awareness that outcome evaluations are much more meaningful if they include local perspectives. Methods that have traditionally been used by anthropologists, such as social mapping and narrative techniques are being incorporated into outcome evaluations. Some of the techniques previously described, such as eliciting community input into the evaluation process and including community members on the evaluation team can help keep a program from becoming too "outsider focused."

Using Existing Measures or Creating Your Own

After you have developed your evaluation questions, identified the domains which need to be assessed, and considered the various measurement issues raised in the previous sections, you need to decide on the actual methods and measures you will use. There are basically three approaches available. You can choose to use existing measures, adapt existing measures for use with your specific population, or create new measures. Each of these approaches has its strengths, weaknesses, and challenges.

Some of the strengths of using existing measures were presented earlier: the knowledge that they have been useful in other settings, the ability to compare findings with other programs, and the efficiency of not having to create measures. But there are problems as well. It is important to remember that a measure developed and standardized on one population is not necessarily valid for use with another. There is no such thing as a measure being "somewhat valid"—it either is or it isn't. Rarely have traditional measures been validated for use with the populations being served by ecological psychosocial and mental health programs for refugees. Nonetheless, many programs, including our own, have gone ahead and used existing measures in these contexts (for the reasons stated earlier) and with the time-limited nature of many interventions and the lack of resources for evaluation, this will likely continue to be the case into the foreseeable future.

There are a number of things that can be done to make existing measures more useful in new settings. Begin by translating the measures into the local language. The most commonly used strategy for this involves translation and back-translation (Behling & Law, 2000; Brislin, 1970). The first step is to have one person, fluent in both the language of the original measure and the local language, translate the measure into the local language. A second person, with the same skills, is given the translated version only (they do not see the original measure) and they translate it back into the original language. This "back-translation" is then compared with the original measure and any discrepancies between the translated and original version are resolved by the two translators. It is important that the two translators fully comprehend *the meaning* of the

items in the measure if they are to arrive at a meaningful final product. We recommend adding an additional step to the measure translation process by having a group of bilingual staff and community members read the back-translated measure (for language and meaning) with the task of achieving group consensus for the items. It is also helpful to then pilot the translated measures with a small group of community members to ensure that all items are readily understood as intended. A failure to take these precautionary steps can lead to unexpected (and problematic) results. For example, ~~O~~one of us was involved with the adaptation and translation of a conventional measure of children's behavioral problems, for use among indigenous Guatemalans living in refugee camps in southern Mexico. To our surprise, on an item assessing audio hallucinations most of the parents reported that their children *often* heard voices when no one was actually present. We had spent enough time in the camps to know that most children were *not* psychotic (none were, in fact). Upon further exploration, ~~however,~~ it became clear that children heard the voices of people not actually present ~~in the home~~ because the houses were made of cornstalks or loosely bound boards through which the voices of neighbors could easily heard. Piloting the measure and closely examining the results allowed us to identify and correct a significant gap between the intended and actual understanding of an instrument item.

Interview style measures should also be translated into the local language—even if the staff administering them are bilingual and can translate on the spot. Unless there is a translation to go by, different interviewers will often ask questions in somewhat different ways. Even the same interviewer will translate questions differently on different administrations.

When intervening with populations of limited literacy, written measures can also be given interview style. Pictures or diagrams can be used in place of, or along with, typical rating scales to overcome literacy problems or comprehension issues related to cultural differences. For example, a card with a picture of four glasses with varying amounts of water has been used to help people rate symptoms on a *never* (empty glass) to *always* (full glass) Likert scale. In another instance, children used a card with a sketch of a person crying on the right and a person smiling on the left (connected with a line) to indicate how they felt over the pre-

vious week. A grid was laid over the card and the children's answers were coded depending on where in the grid their mark fell.

The third alternative, creating new measures, has two primary advantages: You can create measures that specifically address the evaluation questions of interest to your program, and you can collaborate with the local community and develop measures that capture their feelings, beliefs, and expressions of distress. If your program is interested in assessing areas for which there are few, if any existing measures, for example increased trust, social connectedness, or the ability to handle daily domestic responsibilities, there will be little choice but to create new measures. But even if the program is interested in evaluating post-traumatic psychological symptoms, like depression or anxiety, you may want to create new measures that reflect local post-traumatic feelings and experiences.

There are a variety of ways to create measures that reflect local feelings and beliefs. Focus groups can provide insight into the local perspective when creating new measures. Another method that is gaining popularity is to gather information from the community in a standardized way, for example through a series of brief interviews or surveys. The interview or survey data is then examined for common themes or high frequency responses and the results of this analyses are used to create the new measures, but ones that are based on local concepts (a method of this kind was described earlier in the 'qualitative vs. quantitative' section; see also Bolton & Tang, 2002). There are a variety of resources available to assist with the process of scale development (e.g., Dawis, 1993; DeVellis, 1991).

Analyzing, Interpreting, and Reporting the Results of an Outcome Evaluation

If an outcome evaluation has been well designed and implemented the analysis and reporting phase is often fairly straightforward and rewarding. But once again, the context and conditions can create added challenges which should be considered.

Data Management

The results of an outcome evaluation can only be as good as the data on which it is based. It is important to manage the data collection and data entry processes closely. We have known psychosocial projects that got to the analysis phase of their evaluation only to find that problems in the quality of the data precluded meaningful analyses or interpretation. Most of the time, if the problems had been identified earlier in the process, they could have been easily rectified. Some of the ways to avoid problems include meeting regularly with the evaluation team to ensure that everyone is conducting the assessment in the same manner. It can be useful to have evaluation team members sit in on each other's assessments. It is usually wise to enter evaluation data in an ongoing way (i.e., do not wait until all the data are collected to begin entering and examining them). In one of our own programs, assessments were left to accumulate with the intention of entering the data over a brief period (the assumption was that it would be easier to supervise and retain consistency). Without warning, there were a series of violent rebel attacks, the refugee camps were evacuated and, in the process, almost 9 months of evaluation data were either lost or destroyed. The only data that remained were those that had been already entered into the computer and stored on discs, which could be easily transported.

At times, the relatively unregulated conditions of field work can lead to a more casual approach to data management. It is important to use the same care in safeguarding confidential data that you would use if you were operating in a university or clinical setting. Set up clear rules surrounding access to, and handling and storage of sensitive information; and review these processes, and the reasoning behind them, with the entire evaluation team. The rights of program participants to privacy should always supercede the evaluation needs of the program.

Analyzing Data

Programs choose between having local staff analyze data in the field and sending the data off-site to have it analyzed. Frequently local program staff feel that they lack the expertise to analyze their own data on-site, thus they choose to send it on to more skilled program evaluation

staff, either at their home office or to external consultants, who can use more sophisticated approaches in analyzing the data. The appeal of this approach is obvious. But there are advantages to conducting the outcome evaluation analyses locally. It keeps the program staff involved in the process and avoids the situation where the data they have worked hard to collect gets sent away with no further word; or the results are returned with no explanation of how they were generated. Analyzing data in the field helps take the mystery out of the process and contributes to building local capacity by giving community team members access to it.

When analyzing data in the field it can be helpful to keep the techniques simple, and, once again, staff will do better with step-by-step instructions and a how-to manual. Many basic summary statistics like averages, frequencies, and percentages, as well as impressive graphs and figures, can be generated with basic spreadsheet programs like Microsoft Excel. Local program staff are much more likely to have experience using these programs than statistical software packages. However, one of the authors has had great success recruiting and training refugee staff to use statistical software and now has research and evaluation staff capable of running many comparative and predictive statistics in the field.

Interpreting the Results

Sometimes program results are self-explanatory and do not require much interpretation. Other times, however, outcome evaluation results need context to give them meaning. Once you have analyzed your data and have the basic results, it is helpful to discuss them with your staff, program participants, and other members of the community. Ask them what they think about the findings. Do they reflect their experience of program impact? Were there additional program effects that were not captured by the evaluation results? This provides an opportunity to explore unintended effects of the program or to look for explanations for surprising or unexpected results. As previously mentioned, it can be enlightening to examine the results of the outcome evaluation in relation to other concurrent events that may have impacted the success of your program.

Finally, it is important to remember that there can always be at least three reasons for negative findings from a program evaluation. First, for a variety of reasons the program may not have been effective and negative results may reflect the unfortunate reality of a problematic program design. As difficult as this can be, it is one of the important reasons for conducting an outcome evaluation—to learn if our interventions are having their intended effects. Second, the program may have been well designed, implemented, and possibly effective under other circumstances but "outside factors" were undermining the program effectiveness. And finally, negative findings may be the result of a poorly designed or implemented outcome evaluation. That is, the evaluation itself may have been flawed, thus providing inaccurate results. It is always important to take time and explore the reasons behind any set of evaluation results.

Reporting Evaluation Results

Most of the time, outcome evaluation data and the reports generated from it, flow in one direction: from the program to the home office or management level staff, and then on to the funding agencies. Rarely, in our experience, do results get reported back to the local staff who provided the intervention (and often collected the data), the participants, or other community members. However, rapid feedback of results to these stakeholders is important. If the program is having positive effects, the reports can provide encouragement for local staff and build support for the program in the community. But regardless of the actual findings, these are the people who have been collecting the data or participating in the evaluation, and they need to know the results of their efforts.

Take the time to adapt the reports for each audience so that the information is meaningful for them. Pictures and graphs can be useful where literacy is a consideration. One of the authors recently sat in on a community presentation by a program where many summary statistics were presented orally. In a follow-up discussion we learned that only one person in the audience knew what a percentage was, and that while everyone had smiled and nodded politely during the talk, very little of it had been understood.

Although report formats and styles vary considerably, there is some basic information that should be included in the evaluation section of a report.

- The outcome evaluation section of a report frequently begins with a summary of the findings from the process evaluation: was the program implemented as planned; did the activities take place; was the target audience reached, etc. Too often, however, process evaluation data are *all* that is reported for ecological mental health and psychosocial interventions working in adverse contexts. Although process evaluation data can tell us a great deal that is useful about how an intervention was implemented and who it did and did not reach, process data, by themselves, cannot tell whether an intervention was or was not effective.

- It is helpful to provide a brief summary of the methods that were used to assess program effectiveness. This will save the person reviewing the report from having to refer back to the original proposal to make sense of the results.

- Provide a complete summary of your findings. Highlight and elaborating them specifically for each audience.

- Describe any problems that were encountered in the evaluation process and how you intend to address these concerns.

- Finally, state your plans for any ongoing or future evaluations of the program (e.g., post-intervention follow-up assessments or community impact evaluations).

CONCLUSION

For the sake of presentation, program evaluation has been presented in this chapter in a somewhat linear fashion. In reality, the processes that

were described are interrelated and iterative, both informing and dictating each other as the evaluation is designed and implemented. Good outcome evaluation is not static but should be an active process that is constantly re-examined and adapted to meet changing needs of a program. We have an obligation to our programs, our funders, and the participants in our interventions, to do a better job of evaluation and to pass what we learn on to others. There is increasing attention to the sustainability of interventions in post-conflict regions. Ecological mental health projects and psychosocial programs have often focused on empowering local staff with training and intervention skills; however, local staff also need the skills necessary to effectively evaluate their work. We hope that the suggestions considered in this chapter are helpful for those individuals and organizations who desire a better understanding of the effectiveness of their programs. Success begins by making evaluation a priority.

REFERENCES

Bolton, P., & Tang, A. (2002). An alternative approach to cross-cultural function assessment. *Social Psychiatry and Psychiatric Epidemiology, 37,* 537-543.

Dalton, J., Elias, M., & Wandersman, A. (2001). *Community psychology: Linking individuals and communities.* Belmont, CA: Wadsworth/Thompson Learning.

Dawes, R. (1994). *House of cards: Psychology and psychotherapy built on myth.* New York: Free Press.

Dawis, R. (1993). Scale construction. In A. Kazdin (Ed.), Methodological issues and strategies in clinical research (pp. 193-213). Washington, DC: American Psychological Association.

Dean (focus groups), D. (1994). How to use focus groups. In J. Wholey, H. Hatry, & K. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 338-349). San Francisco: Jossey-Bass.

De Jong, J. (2002). Public mental health, traumatic stress and human rights violations in low income countries. In J. de Jong (Ed.), *Trauma, war, and violence: Public mental health in socio-cultural context* (pp. 1-92). New York: Kluwer Academic/Plenum Publishers.

DeVellis, R. (1991) *Scale development.* Newbury Park, CA: Sage.

Enett, S., Rosenbaum, D., Flewelling, R., Bieler, G., Ringwalt, C., & Baily, S. (1994). Long-term evaluation of drug abuse resistance education. *Addictive behaviors, 19,* 113-125.

Fetterman, D., Kaftarian, S., & Wandersman, A. (Eds.). (1996). Empowerment evaluation. Thousand Oaks, CA: Sage.

Kazdin, A. (Ed.). (1993). *Methodological issues and strategies in clinical research.* Washington, DC: American Psychological Association.

Kruger, R. (1994). *Focus groups.* Thousand Oaks, CA: Sage.

Linney, J., & Wandersman, A. (1991). *Prevention plus III: Assessing alcohol and other drug prevention programs at the school and community level: A 4 step guide to useful program assessment.* Rockville, MD: US Department of Health and Human Services, Office for Substance Abuse Prevention.

Miller, K. (1999). *Program evaluation for HIV/AIDS prevention.* Chicago: Midwest Hispanic AIDS Coalition.

Miller, K. (Co-producer), Billings, D. (Co-producer), & Farias, P. (Co-producer). (1995). *Playing to grow: Creative education programs for children* [Instructional video]. (Available from K. Miller, Psychology Department, SFSU, 1600 Holloway Avenue, San Francisco, CA 94132)

Miller, K., & Billings, D. (1994). Playing to grow: A primary mental health intervention with Guatemalan refugee children. *American Journal of Orthopsychiatry, 64*, 346-356.

Wholey, J., Hatry, H., & Newcomer, K. (Eds.), (1994). *Handbook of practical program evaluation*. San Francisco: Jossey-Bass.