Center for the Study and Prevention of Violence

Program Evaluation Overview

Kristi Jackson
University of Colorado at Boulder

Kirk Williams
Center for the Study and Prevention of Violence

Delbert Elliott
Center for the Study and Prevention of Violence

July 1996

F-1455
Center Paper 009

Center for the Study and Prevention of Violence
Institute for Behavioral Sciences
University of Colorado, Boulder
Campus Box 442
Boulder, CO 80309-0442
Phone (303) 492-8465 Fax (303) 443-3297

Program Evaluation Overview

Table of Contents

Program Evaluation Overview

1. Introduction

What Is An Evaluation?

Evaluation is a method of inquiry that attempts to document the formation, implementation and outcome of a service or program. Stacks of text books and articles on evaluation are available in any library, but they tend to be highly technical. This overview will not provide you with the kind of analytical specifics which are found in such texts.

The goal here is to provide a general framework for understanding the basic components of evaluation research and to provide program managers with a preliminary sense of what to expect from an evaluation. This overview will not train you on how to conduct an evaluation project by yourself. Instead, it will provide you with enough information to begin shaping an evaluation plan. The overview will also suggest a variety of resources that you can access to help manage the more technical elements of program evaluation. With this information, you will be able to engage in a more well-informed collaboration with an evaluation researcher and hopefully produce a more helpful evaluation plan.

The overview will also assist you in examining the evolutionary course and context of a program. Depending on the maturity of the program, the evaluation can address 1) developmental issues of program start-up, 2) program monitoring and modification, and 3) the culmination of program activities in particular outcomes. The overview will provide program managers with tips on when to use this tool we call "evaluation," when NOT to use it, and how it can be modified to address

program needs. The overview will also focus on the context surrounding the program, so that an evaluation can be broadly and properly understood.

An evaluation always occurs within a specific context, and program managers play a crucial role in framing and describing that context to evaluators. As the contextual experts in the evaluation process, program managers are able to help modify this research tool in order to examine the most pertinent aspects of program monitoring. Not surprisingly, by providing well-informed guidance to an evaluator, a program manager will maximize the effectiveness of the evaluation while simultaneously enhancing his or her relationship with the evaluator.

Why Is Evaluation So Difficult?

Evaluations can easily be perceived as unpleasant but necessary tasks. Beginning with tests and class projects in elementary school, students learn to maneuver through evaluations in order to obtain a stamp of approval. If this stamp is denied, then all previous work has been for naught. Because the responsible program director -- like the good-intentioned student -- wants his or her projects to succeed, it is natural to identify evaluation as a hurdle that must be overcome with this stamp of approval. From the perspective of most program directors, evaluations lead to one of two conclusions: Did we pass or did we fail?

Students and professionals tend to learn the most, however, from teachers and mentors who provide concrete feedback about particular pieces of work, rather than simply a passing or failing grade or a series of red check-marks on an employee evaluation form. When given specific information about particular elements of a project, most people understand evaluation as something

more than a stamp of approval. In the long run, whether the individual is a student, nurse, stock broker or cabinet maker, the jumping of hurdles becomes much less useful than specific pieces of information that enhance performance. So is the case with program evaluation: By designing an evaluation that generates specific information, the evaluation process is transformed from a dreary but necessary hurdle into a highly useful investigation. A thoughtful evaluation can avoid the pitfalls of the "hurdle-mentality" that attempts to prove the worth of a program, and can instead focus attention on the desire to learn, adjust, and improve.

Students and professionals tend to notice that when an evaluation is based on one single exam or project at the end of the year, anxiety runs high. In contrast, when evaluation occurs gradually and relies on regular, less grandiose measures of progress, apprehension is usually reduced and the process, itself, seems less disruptive to daily activities. Similarly, by incorporating evaluation into the framework of the program so it occurs in an ongoing fashion within daily operations, the process is less likely to carry a threatening aura and more likely to offer valuable planning and decision-making information. An evaluation can become an ally to program effectiveness, regardless of the findings, especially when the evaluation is constructed to provide specific information and is woven into the daily operations of the program.

In the ideal circumstances, evaluation designs are generated long before a program or intervention is up and running. This allows program managers to gather baseline information prior to program implementation. Funding issues, political constraints, the nature of the program and lack of foresight serve as some of the many reasons why this rarely happens. Although evaluation designs should be incorporated into the planning stages of the program in order for the process to run more smoothly,

evaluations should not be feared even if a program is already up and running; a helpful evaluation design remains possible.

How Useful Can An Evaluation Be?

When evaluations are constructed properly, they can increase knowledge about a particular facet, reaffirm benefits, point to improvements, guide decision-making, bring program staff together to reshape or reaffirm goals and create a dialogue among community participants. Ultimately, this allows programs to adapt and develop according to changing needs and circumstances. Even though a program is functioning at a high success rate, an evaluation plan can help keep the program on track by identifying changing needs more quickly than intuition or professional "judgment calls."

Appreciating the usefulness of an evaluation requires an understanding of both objective data and subjective intuitions. Each has strengths and weaknesses that must be recognized when making management decisions (Gabor and Grinnell, ____). Objective data generated by formal evaluation tends to provide the most credible information from the perspective of funders, scientists and other external reviewers. Although decision-making is often based on objective data, most program managers acknowledge that certain data may be missing, ambiguous, contradictory, or may fail to take contextual factors into account. In such circumstances, decision-making relies on subjective information such as clinical impressions, past experience and intuition.

Subjective measures are an important supplement to the examination of objective data. In fact, the logical processes involved in the analysis of objective data do not necessarily contradict intuitive

understandings. Some argue that intuition IS logical thought; the result of data subconsciously assimilated and processed. Furthermore, program managers cannot rely on a careful analysis of data for ALL of the most minute actions taken during the daily tasks of program management; sometimes they must rely on past experience and professional "judgment calls."

The danger, of course, is that human beings tend to place great trust in their intuition, but unlike objective data, subjective impression are not easily checked for errors of interpretation. Therefore, when possible, major decisions about program resources and modifications should rely on objective data. This claim is not intended to convince program managers to abandon their intuitive judgments, but it serves as a warning against some of the perils involved in predetermined, intuitive patterns. Ideally, a combination (and hopefully corroboration) should be made between these two strategies of gathering information and making decisions.

Do We Proceed On Our Own, Or Hire A Consultant?

Evaluations can be conducted by existing program staff as a completely internal process, or they can be conducted by outside evaluators who have minimal contact with program staff. Not surprisingly, there are advantages and disadvantages in both strategies, and program managers must consider these dynamics prior to deciding on an appropriate strategy.

The primary advantage of an internally driven evaluation is that the needs and priorities of program staff are more likely to be addressed. By using staff members who are familiar with the program, specific contextual factors can be considered and examined. Some agencies prefer internal evaluations because they doubt that an external evaluator will be able to identify all of the pertinent

factors in assessing the program. While the internally driven evaluations are perhaps better able to address contextual specifics, this ability can also threaten the integrity of the assessment. An existing staff member will often face pressure from a variety of sources within the agency and may have a vested interested in certain outcomes of the evaluation. Even when an internal evaluation is conducted with integrity, such evaluations often encounter severe criticism; the *potential* for favoritism and other types of contamination will always provide critics with ammunition.

External evaluations are less likely to encounter these criticisms because the external evaluator presumably has few (if any) political and financial ties to the program. Obviously, there are exceptions to this because an external evaluator may rely on favorable outcomes for future work or may bring certain values to the evaluation that predetermine the focus of the study and the methods used for analysis. In general, however, funding agencies and other programs are more likely to acknowledge the credibility of an external evaluator than an internal evaluator. External evaluators generally possess more knowledge about designing evaluations and analyzing data, often using their experiences with other programs to increase the accuracy of the evaluation.

Scriven (1976) provides three alternatives to the purely internal and purely external evaluation. One strategy is to add external evaluators as consultants to keep internal evaluators on track. The next is to separate the evaluation funding from the program budget so that debates over program resources become less problematic. Finally, an evaluation advisory board can be created as a buffer between the program director and the evaluator to adjudicate disputes, assign responsibilities and serve as a resource to the program and the evaluation.

When considering these and other alternatives to the purely internal and purely external evaluation, it will be helpful to keep in mind two questions: Who does the evaluator report to (the program director or an administratively separate individual)? And where do the funds for the evaluation come from (the program budget or an independent funding agency)? Again, the answers to these questions need not always fall in the extremes, but the political and financial structure of the evaluation can have a significant impact on the findings and therefore must be considered when constructing an evaluation. These and other ethical questions are addressed in Appendix A.

To get a better handle on the political and financial pressures involved in the evaluation, it will be helpful to examine and diagram the relationships of various community participants carefully. Certain people and certain positions of authority can influence evaluation politics, economics, availability of staff and appropriate technology. Before embarking on an evaluation, the program manager must consider the impact of policy makers, program sponsors, evaluation sponsors, program management, program staff, evaluators, program competitors, clients, program participants, etc. When all of the community participants have been identified, examine the various reasons they are involved and the use they intend to make of the data. By identifying these structural dynamics, program managers are more likely to establish a firm foundation for generating a collaborative and effective evaluation design.

2. The Utility of Various Types of Evaluations

The Range of Evaluation Types

Gabor and Grinnell (____) identify five different types of evaluation: Planning, process, evaluability, outcome, and efficiency. An understanding of these types, and the range of information provided by them, will enable a program manager to narrow down the apparently monumental task of evaluating every element of an entire program. At one end of this spectrum, the "Planning Evaluation" carefully examines the specific needs of a population prior to establishing an intervention or prevention program. At the other end of the spectrum, the "Outcome Evaluation" determines whether a program is meeting (or has met) its preestablished goals and objectives. In general, these types use increasingly rigorous "hard science," but all are important and must be selected according to the particular context or research question.

Before proceeding with a description of these types of evaluations, you should keep in mind that these distinctions are for the purpose of discussion. The "different" types often blend into one another, and an evaluation plan can incorporate unique combinations of each; they are not mutually exclusive or incompatible. However, an understanding of the unique characteristics of each type will allow a program manager to design an evaluation that is appropriate to the context of the program and the stage of program development or implementation. Once again, an understanding of these types enables a program manager to serve as the contextual expert who advises a professional evaluator in order to maximize the utility of an evaluation design.

Planning Evaluation

To determine the feasibility or need for a prevention or intervention program, an evaluation might be employed to determine the problem in the population that presumably needs assistance. This assessment identifies the nature of the problem, demarcates the specific population affected, suggests a combination of risk factors or protective factors, and often identifies resources needed or available to address the problem. Although such an assessment can occur after a program is initiated to keep up with demographic trends and other changes, such evaluations are usually best suited to the early stages of program development, before the program is in operation. If the administrative structure of a program carefully monitors referral rates, sources of referral, requests from other agencies and new dilemmas presented by clients, then needs assessment can be conducted in the routine operation of the program.

This evaluation may occur, for instance, when an urban area experiences a wave of youth violence. As policy-makers and social services direct their attention toward the issue, they might employ a planning evaluation in order to examine the locus of the problem. Suppose many urban residents perceive the violence as a gang-related problem, but an analysis does not support this view. Data indicate instead that while youth violence has escalated dramatically, only a small percentage is gang-related. Initial plans for an intensive gang intervention, therefore, are modified due to the findings and a conflict resolution curriculum is developed for the public school system at large. Without the planning evaluation that should accompany program start-up, a misguided intervention could have easily been initiated. The planning evaluation allowed policy-makers to check the

common assumptions about youth violence and uncover important information about the nature of the problem.

Process Evaluation

To obtain a better sense of HOW a program achieves it's goals, or the extent to which an intervention reaches the intended population, a process evaluation can be initiated. This evaluation might focus on a particular piece of program implementation, or it may examine the sequence of events leading up to service delivery. The decisions made, who made them and the criteria on which decisions were made are all open to examination in a process analysis. Intake procedures, screening processes, client monitoring, the "dosages" of the intervention strategies, referrals, and follow-up can all be explored with an evaluation that focuses on process. To embark on such an analysis, the client's path through the entire program must be clear and the services provided at each stage must be carefully identified along with the dosages or amounts of these services that are delivered. Key activities and transition points serve as helpful markers in this type of evaluation. Process evaluations will be impeded if little or no systematic collection of data is already in place. The evaluation is likely to be more beneficial and less disruptive within an existing systematic data collection or management system.

A process evaluation might be helpful, for instance when examining the services delivered to alcoholics through a halfway house. Instead of exclusively examining outcomes, as in the example above, this assessment takes stock of various points in service-delivery. By analyzing the most common sources of referral, for instance, the program manager might become better informed about

liaison relationships with other agencies and specific points where these relationships could be improved. Such an assessment could also compare different types or amounts of counseling services to different categories of clients in order to refine the duration or amount of these treatments. In sum, a process evaluation among this population could help the manager of the halfway house better understand and refine several different elements of the treatment program.

Evaluability Evaluation

An evaluability evaluation might be conducted to determine whether a program can in fact be evaluated. Such an assessment indicates whether the objectives of the program can be measured in any meaningful way. This type of evaluation deals directly with the specific circumstances surrounding a program and examines whether or not it is ready for an evaluation. In other words, this evaluation looks at the program and asks: What existing elements might impede or enhance the utility of an evaluation?

A program's objectives are often the initial focus of this evaluation because as a prerequisite of finding out whether an evaluation can be conducted, the program manager and the evaluator must be clear about what the program is attempting to achieve. Failure to state objectives clearly will result in unreliable findings during any type of evaluation. An evaluability evaluation under these circumstances will caution the program manager to address the lack of clarity before proceeding with the desired evaluation. A frequent dilemma faced by evaluators is the lack of clarity and precision in objectives. If a program focuses on only broad goals or mission statements (e.g. reducing school

conflict) without specifying the objectives (e.g. to cut the number of referrals to the Principle's office in half), then the program cannot be evaluated properly.

In addition to the absence of specific objectives, an evaluability evaluation may reveal that a program simply is not mature enough to put it through a particular type of evaluation. For instance, a comprehensive rehabilitative program for pregnant drug users may include a wide range of activities. Family therapy, drug and alcohol rehabilitation, vocational training and prenatal care may all be important and intensive elements of the program. If a program manager attempts to engage in an outcome assessment before all of these elements have been formalized, however, then the evaluation will be of very little value. An evaluability evaluation may reveal that another type of evaluation is more appropriate (see, for instance, the process analysis above), one that will fine tune the delivery of services and prepare for an outcome evaluation at a later date.

Outcome Evaluation

An outcome evaluation can be employed to determine the degree that a program is meeting its objectives. This assessment establishes the amount of change that occurs. For instance, an outcome evaluation among incarcerated juveniles who receive a intensive treatment program as part of their sentence may assess rates of success and recidivism when the youth are released from the penal institution. Because the focus on this assessment rests in the outcomes of the program, it will disclose rates of change, recidivism or attrition, but will explain little in terms of how a program achieves its goals and objectives. Outcome evaluations are very helpful, however, when a program manager has already gathered outcome data and wants to measure the effect of a specific change in

the program. In addition to this, an outcome evaluation can be used to assess the length of time after a client leaves a program that the intervention appears to be effective.

To assess the effectiveness of a treatment program for violent youth, a battery of tests might be employed when the youth enter the prison. These tests detect the existence of drug and alcohol addiction, self-efficacy, cognitive abilities and other factors associated with recidivism. By administering these tests again after the 3-6 years of mandatory treatment, and by comparing these results to a comparable population who did not receive treatment, program managers can examine the effectiveness of their intervention. Furthermore, after their return to civilian life, the tests might be employed once again to examine the long-term impact of the program. Although this outcome assessment does not necessarily indicate what elements of the program are more or less effective, or how various treatments are able to influence the youth, it does provide a measure of the overall results of the program.

Efficiency Evaluation

An efficiency evaluation is often employed in which administrative time, facilities, equipment and other resources are examined to determine whether a program achieves its objectives at a reasonable cost. The goal of this evaluation is often to assess whether costs can be reduced without loss of effectiveness or what cuts would be least disruptive to program management and effectiveness. Efficiency evaluations seem to be particularly difficult in social services because it is often difficult to reduce client outcomes to dollars. Many benefits of the program may be intangible and some of the costs might be indirect. An evaluator must be careful about the assumptions underlying the

definitions of costs and benefits. Nonetheless, decisions often need to be made regarding the financial management of a program and with limited funding for social programs, some type of efficiency assessment is often required if the program is to be sustained.

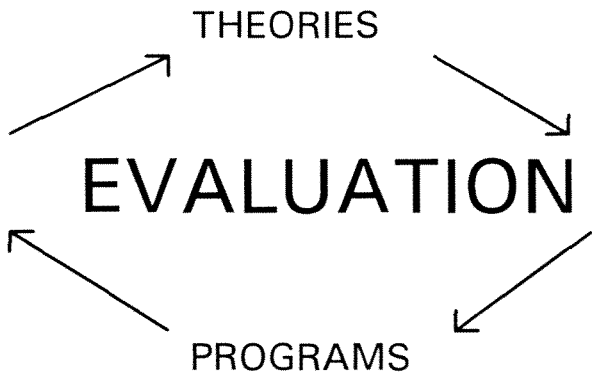
In the case of an anonymous, state-run HIV testing site, an efficiency assessment may be initiated in order to deal with anticipated funding cuts. Such an assessment could ask clients how they heard of the site in order to better understand which publicity attempts have been most successful. Comparisons between televised public service announcements, phone book advertisements, poster campaigns, newspaper inserts, and referrals from other agencies can be compared to determine which methods should be retained or eliminated. If the agency discovered, for instance, that few clients appeared for HIV testing after reading a poster/flier about the agency, then funds from the production of posters could be flagged in preparation of the impending funding cuts. In sum, an efficiency assessment can help measure costs and benefits to maximize the over-all effectiveness of the testing site. For more detailed information regarding measurement strategies in these various realms of evaluation research, see Appendix B.

3. Theories, Goals And Objectives

Programs Shaped By Theories And
Theories Shaped By Programs

Propositions suggest that specific acts or elements lead to particular results. Propositions provide the basic building-blocks for social intervention. For instance, if physiological addiction to alcohol occurs more quickly at younger ages, then interventions should be geared for potential adolescent addicts. A series of propositions can always be uncovered when exploring the foundation of a social service program, and together these propositions comprise one or more broader theories about the causes and solutions to various social maladies. As a basic definition, a theory is an interrelated set of propositions.

Theories can inform program managers about the best way to construct a program, and at the same time, the information generated by a program can be used to reshape broader theories. In each of these processes, evaluation serves as the link; evaluation can be constructed to test broader theory, or it can be constructed to elucidate the functioning of a program.



Although these two foci of evaluation are inextricably linked, most program managers rank the examination of the program as more important than the theoretical foundations of the program. While program effectiveness may be the ultimate goal, it is important to acknowledge that all programs are constructed around some sort of theoretical framework and that the management of the program follows the structure of these theories.

Some programs are more explicit than others about their theoretical grounding, but in almost all cases the theory (or theories) is implied by the goals and objectives of the program. The specific patterns may only be uncovered through a careful analysis, just as the grammatical structure of a foreign language can be uncovered through a careful analysis. Because the connection between theory and program exists regardless of whether it is acknowledged or not, the program manager who understands the theoretical foundations of a program and the importance of theory in driving an evaluation, is more likely to reap the many benefits of an evaluation. In the ideal situation, when theory is used to build a program, the theoretical elements are simply revisited during the evaluation.

Although proposals for funding and mission statements often possess the theoretical rationale for a program, a literature review can also help identify pertinent theories. Such a review can, for instance, provide propositions on how client and service-provider relationships influence outcomes, or how various durations of treatment affect long-term changes. Even when a program possesses specific theories at the outset, by the time an evaluation occurs it is wise to revisit the literature and to note recent findings that may change the focus of the evaluation and may modify the direction of the program.

Tips for finding recent theory include 1) involvement in professional organizations and conferences where recent developments are discussed, 2) an examination of programs that have recently been funded to identify their theoretical rationale, 3) a review of professional journals that discuss theoretical trends in the field, 4) a library search of recent research findings, 5) a library search in different but related fields (for instance, an intervention program that works with child witnesses of domestic violence might examine the post-traumatic stress literature related to military veterans or search for information on the influence of violent television programs). When time does not permit these types of searches, a program manager can also turn to consultants, often those involved in evaluation research, to obtain the most up-to-date theories for particular types of social service programs.

Exploring and shaping the theoretical foundations of a program can be an exciting and creative process and reminds program staff of some of the reasons they are involved in their field. These theoretical groundings need not emerge as unchangeable and all-encompassing; similar programs might have equally effective -- although different -- rationales. Each program can be based on a unique combination of theories that give it special character. The most notable program managers are those who articulate and test the underlying theories of their program at the same time that they are analyzing the effectiveness of the program. These program managers are much more likely to be involved in the reshaping of broader theory, and they are more likely to be among the resources that other programs use when examining their own effectiveness. Finally, a theoretically grounded program manager is more likely to generate a "credible" evaluation from the perspective of outsiders and funding agencies.

Linking Theories To Objectives

To demonstrate the connection between theories, goals and objectives, this section turns to the example of a week-long intensive intervention program that is designed to keep 12 at-risk youth from dropping out of high school. Based on the theory that youth who receive after-care follow-up from such programs are less likely to drop out of school, one goal may be to develop an after-care program. Two examples *of program objectives* are: 1) To match youth with local mentors and 2) To maintain ties among the 12 youths so they can provide each other with an additional support system. These objectives can be further refined into specific, measurable action steps.

For instance, specific steps in achieving objective number one (matching youth with mentors) might be A) Recruit 12 mentors from local churches and social service agencies at least one month before the program begins, B) involve these mentors in the intensive week-long program so that they can establish bonds with the youth, C) assign a staff member to work with the mentors, D) provide the mentors with weekly feedback and referral assistance, E) organize monthly group meetings for mentors so they might share some of their struggles and successes in working with the youth.

Specific steps in achieving objective number two (maintaining ties among the youth) might be to schedule a weekly meeting for the youth in which time is reserved for A) sharing their struggles B) obtaining encouragement from staff members and from each other, C) presenting their new knowledge about various study skills and D) allowing them to reflect on the effectiveness or shortcomings of a particular study skill at the following meeting.

The chain from theories to goals to objectives to action steps requires that specificity increase at each link, and measuring the objectives requires that they be specific enough and observable enough

to be examined. Theory can provide a general direction, goals can make these theories operational, objectives can make the goals observable, and action steps can make them measurable. Theories, goals, objectives and action steps are often confused and this makes the evaluation process much more difficult. If goals are not specified and narrowed into objectives, then attempts to measure them will likely result in disagreement. Different staff members and different evaluators will identify different methods of measuring goals and therefore will emerge with different results.

The theories, goals, objectives and action steps should be built into the program before an evaluation begins, although a skilled evaluator can assist in clarifying all four. Program managers will find that the articulation of these items within staff meetings and mission statements can increase staff cohesion and commitment and can limit the time and money spent on external evaluators. Before beginning an internal or external evaluation, some attempt should be made to generate these four types of statements. An evaluator's time is spent much more wisely when some attempt has been made to do so.

Specific Tips for Writing Goals, Objectives & Action Steps

Goal statements should also be agreed upon by community participants. This includes funders, evaluators, staff members and program planners. By engaging these constituencies in a discussion of theories, goals, objectives and action steps, the community participants are more likely to reach a consensus and therefore less likely to object to the evaluation process or methods. An annual discussion of these elements, when conducted constructively, tends to create greater cohesion

within the program and helps prepare the program for changes based on, for instance, changing demographics of the target population or changes in the funding structure of the program.

Another, equally important element in the articulation of goals is that they should be based on a realistic assessment of time and resources. Goals articulated in requests for funding are often loftier than their achievements and force a program manager to list shortcomings when, in fact, a degree of progress has been made. Unforeseen obstacles, such as the reduction in resources, can also undermine the intended effects of the program. In the example of the at-risk youth, what if the staff member who serves as a liaison to the mentors is suddenly taken ill or quits? Some buffer for these unexpected hurdles should be considered when articulating program goals.

Finally, the language of goals and objectives must be clear and consistent. Terms such as "mentor," "at-risk," and "maintain ties" should be defined prior to the formal articulation of goals so that all people involved in the program and in the evaluation are clear about their meaning. Terms to use in goal statements are those related to the target population, the specific problem areas of the population, the type of intervention strategies used, and the intended outcomes.

Several resources can be used to identify goals and objectives in addition to funding proposals and mission statements. Quarterly reports and record summaries may also be useful, because these documents provide information on trends and unexpected characteristics of the population or the program itself. In addition to these program documents, sources may be obtained outside of the agency such as parent or community concerns, the goals articulated by similar programs, and state or federal mandates. By examining all of these sources, a comprehensive, clear and consistent set of goal statements can be achieved.

When goals and objectives have been sufficiently articulated, specific action steps for achieving them must also be identified. The most important rule in generating action statements is that each statement should include only one purpose or aim so that it can be evaluated as a single unit. By limiting action steps to a single concept, positive outcomes in one area will not be minimized by less dramatic outcomes in another.

Strong and specific verbs should also be used to articulate action steps as well as a specific time-frame for achievement. For instance, recruiting 12 mentors for the at-risk youth at least one month prior to the week-long intensive program is an action steps that contains a measurable change within a specific time period. The change must be clearly geared in a specific direction and within a certain time frame in order for a measurement of the action steps to carry any meaning.

While articulating action steps, a distinction should also be made between a client-related objective and an agency-related objective. A client-related objective would be, for instance, to increase the average grade point of the students remaining in school at the end of the academic year by point five. Client-related action steps can be further subdivided into 1) knowledge, such as the understood usefulness of various study skills, 2) attitudes, such as self-esteem, perseverance, or frustration, and 3) behaviors, such as the actual use of various study skills or a reduction in truancy rates.

Agency action steps can be divided into 1) client issues such as recruitment and retention, 2) staff issues such as training and evaluation, and 3) resource issues such as funding and equipment replacement. With agency action steps and client action steps, efforts should also be made to reflect different levels of attainment. In other words, these action steps should follow steps of increasing

developmental complexity (in the case of clients) or organizational growth/change (in the case of agencies). Without identifying various stages of anticipated change, positive results of the program may go unnoticed, or the degree of change may not appear as large as it should.

Although a multitude of action steps may be identified in the process of connecting theories with goals and action steps, time and available resources often prevent thorough evaluation of every one. To determine which action steps should ascend to top priority and which should be put to the side, consider the action steps that are the most specific and concrete, because they are the most likely to be measured with ease. Next, consider the funding and resources available for evaluation and determine the greatest number of top-priority action steps that can be evaluated with the least amount of resources. Finally, consider which action steps must be measured before other action steps can be measured. These primary action steps should also be carefully considered because they are likely to be the most crucial for program development.

The Limitations Of Consensus Decision-Making

In the ideal setting, the key players in program management and evaluation will generally agree on the theories, goals, objectives and action steps that shape the daily functioning of the program. Rather than hope for this ideal, it may be wise to assume at least some disagreement and to begin facilitating discussions on the structure and priority of the possible theories, goals and action steps. Although many program managers prefer the process of consensus decision-making through round-table discussions, several potential problems with this format need to be acknowledged and alternative approaches considered.

Experience with round-table consensus reveals that, depending on the dynamics of the group, several factors can inhibit healthy discussion. The first and most frequent dilemma occurs when an individual or subgroup with greater supposed authority directs decisions toward a special interest. Whether intended or not, variations in authority can influence the flow of participation and the presentation of alternative views. This same dilemma emerges when differential abilities to persuade result in the exclusion of certain ideas and the magnification of others; the loudest voice often wins. In addition to these dilemmas, participants may be unwilling to abandon previously held opinions or reluctant to acknowledge a change of heart for fear of appearing soft, indecisive or rash. Under these influences, people often resist their inclination to move toward consensus. Finally, the bandwagon effect of majority opinion can compel people to agree when, in fact, they do not. Three techniques that might be considered as an alternative when consensus seems to be falling short of its ideal are the Delphi technique, the Q-Sort, the nominal group technique and the adversarial proceeding. A brief description of these techniques can be found in Appendix C.

The Context Of Discussions Among Community Participants

While consensus, the Delphi technique, Q-Sort, Nominal Group technique and adversary proceedings provide a range of alternatives for discussing theories, goals and objectives, they do not exhaust the range of possibilities. The creative alternatives for generating discussion are limitless. Before deciding on a technique, program managers should take stock of the existing political climate within the agency, the various coalitions influencing the operation of the program, and the community participant's styles of making decisions and dealing with conflict. These *contextual factors* must be

considered in order to make the appropriate decision on which technique (or combination of techniques) is best suited for the group.

An important point to remember when formulating the structure of this dialogue is that an appropriate strategy cannot be determined without carefully examining the unique dynamics of the group. A trained mediator is often the best person to assess and guide these dynamics. A second point to note is that total agreement is rarely achieved, and at some point a manager must make decisions that do not please ALL of the various constituencies. Even so, however, the careful construction of an appropriate format for dialogue reaffirms to participants that their opinions are valid and that these opinions can be expressed. If participants believe that their perspectives have been genuinely considered, and if the points they raise are somehow acknowledged and addressed, then significant progress has been made to bring people closer to agreement and understanding. Finally, the process of coming to agreement is one that occurs more naturally if discussions of priorities are a regular and frequent part of program management, not simply a response to the need for evaluation.

4. Now What?

As the major considerations involved in an evaluation project have been addressed here, the goals have been two-fold. First, this overview has attempted to provide you with a good, general foundation in evaluation research, so that the initial (and ongoing) meetings with evaluators will seem more manageable and cooperative. This particular goal does not seek to train program managers, but simply to broaden their information on evaluation research so that the process, itself, seems less foreign, and so that program managers will feel prepared to contact an evaluator. The second goal has been to improve the relationship between the program manager and the evaluator in order to maximize the integrity of the evaluation. When program managers and evaluators are able to generate a collaborative design, the study, itself, will reap more helpful information.

To conclude the efforts of this program evaluation overview, the following questions have been generated as a way of summarizing the discussion. These questions are not prerequisites, but they are simply tools you may want to use to prepare for an evaluation. Furthermore, it is not necessary that you answer every question in complete detail. The questions have been provided only so that you have a framework to sift through some of these issues, and in some cases you will not be able to answer the questions until you meet with an evaluator. Again, the intention here is to provide you with enough information so that you can ask well-informed questions and can communicate your contextual expertise to the evaluator.

Things To Consider Before Planning An Evaluation:
A Review

1. *Introduction*

   A. *Are you ready to add some objective data to you subjective perceptions of the effectiveness of your program?*

   B. *Are you willing to work with an evaluator cooperatively to begin the evaluation process?*

   C. *Are you able to see evaluation as something more positive than a "hurdle?"*

2. *The Utility Of Various Types Of Evaluations*

   A. *Based on the five types of evaluations described here (planning, process, evaluability, outcome and efficiency), do you have a sense of what type (or combination of types) is the most appropriate to your current needs?*

   B. *List each type you would like to use and the reasons for using it.*

3. *Theories, Goals and Objectives*

   A. *Can you articulate a comprehensive, theoretical rationale for you program?*

   B. *What are the basic theoretical components of your program?*

   C. *Where might you look for additional sources of theoretical work (e.g. libraries, consultants, colleagues, etc.)?*

   D. *What are your programmatic goals (remember, these should be grounded in your underlying theories)?*

   E. *What are the <u>specific</u> objectives linked to each goal?*

F. *Where can you obtain help in refining these goals and objectives?*

G. *Moving Community Participants Toward Agreement*

i. *Generate a diagram of the key players in the evaluation process.*

a. *List their perspectives on the program that might influence the evaluation.*

b. *List the stances they take on major program issues.*

c. *List the specific needs they want to fulfill with an evaluation.*

ii. *Given this diagram of the dynamics among community participants, generate at least three possible strategies for bringing these community participants together in a discussion that will hopefully bring them closer to agreement/understanding.*

iii. *Where can you find some well-seasoned facilitators to help you explore your options for managing community participants?*

Appendix A:

Standards And Ethics

Overview

Ethics and professional standards help ensure that the evaluation will contribute to the well-being of the client and the effective functioning of the program. They should be considered in the context of three separate realms. The first involves all community participants in the program. These individuals must become aware of the political pressures driving their interests and discuss these issues with the evaluator. It is the responsibility of the evaluator to become sensitive to the constraints placed on individuals based on their role within the program. Although a skilled evaluator will attempt to address the various needs of community participants and bring these needs to a general consensus, when the community participant interests are in direct conflict, it is up to the evaluator to establish priorities. In the midst of contradictory stances, the evaluator must maintain the integrity of the evaluation process.

Another realm in which ethics must be examined is the personal bias that can place limitations on the evaluator. An evaluator may be unknowingly influenced by the social ideologies of the time, personal preferences regarding evaluation methodologies, and the professional values of program evaluation. He or she must therefore carefully examine contextual constraints in the process of evaluating a program. Because it is impossible (and perhaps undesirable) for an evaluator to become completely objective and emotionless, the values and predisposition's he or she brings to the study should be described from the beginning as explicitly and honestly as possible.

With a general sensitivity to these two realms of ethics and standards, it is important to turn to a more detailed description of the third realm, the realm of standards that have been widely agreed on by experts in the field. These are the basic standards of evaluation research, and everyone involved in the evaluation (program staff, managers, other community participants, evaluators, and even clients) must be clear about some of the basic ethical guidelines established for evaluation research. These standards, described below, relate first to the process of protecting the privacy of clients. Next, a separate set of guidelines ensures that the evaluation, itself, provides usable information, is feasible to conduct, does not unfairly focus on one source of data collection, and can be conducted with a high degree of accuracy.

Informed Consent, Anonymity And Confidentiality

One of the most important ethical practices in an evaluation is to inform adequately all participants and anyone who may be directly affected by the study. Participants should not be deceived, they should be aware of the implications of their participation, and no harm should come to them as a result of their participation. The classic example used to elucidate the importance of these standards is Milgram's (1963) study of obedience.

In this study, Milgram took a subject and placed him/her behind a control panel that allegedly delivered a shock-therapy (negative reinforcement) to other human subjects. The real subjects (those sitting behind the control panel) thought they were simply administering an experiment and did not understand their role as subjects until the experiment was over. As the false subjects demonstrated more and more error in their task to memorize a series of words, instructions were given to the real

subjects to increase the shock level. Even when the false subjects were screaming for the experiment to stop, many of the real subjects continued to administer the shocks as instructed. Although this study revealed a great deal about obedience, it violated the ethical principal of informed consent.

Informed consent means more than simply a subject's voluntary agreement to participate. It also means that he or she understands what will happen in the study and is aware of the possible consequences. Because they did not participate in an experiment that provided informed consent, the individuals in Milgram's study were not only duped into participating, but many of them were also actually harmed by the experiment. When these individuals were made aware of their role in the experiment and forced to examine the implications of their own actions, many became deeply disturbed. Most of them felt used and some suffered post-traumatic stress.

Two additional standards that are often used with participants are confidentiality and anonymity. When a participant is guaranteed confidential participation, he or she is made aware of the select number of researchers that will be able to match the individual with the data provided. These researchers vow not to reveal or imply the identity of the participant to outside sources. Many studies are required by their funding source to ensure confidentiality and face heavy penalties for violating this standard. Before beginning an evaluation, it is therefore important to consider what strategies might be used to adequately protect the anonymity or confidentiality of the subjects.

For instance, if researchers intend to enter a neighborhood and interview residents about the gang activity in the area, they will, in most cases, know where the interview respondents live. In order to provide protection for these individuals, the researchers assign a number to each respondent, and list only this number on all cassette tapes of the interview and all documents that refer to the interview.

After the interview has been transcribed, the researchers further protect the interview by assigning a fictitious name to each interview. In this manner it is the interviewers, alone, who know the identity of the respondents. They do not reveal the addresses or other identifying information of the respondents, and always refer to them by their fictitious name.

Unlike a participant whose identity remains confidential, the identity of an anonymous participant is never revealed, even to the researchers, and cannot be traced. Many people involved in the evaluation process do not make a clear distinction between anonymity and confidentiality, so great care must be taken to identify and reiterate the professional standard being used clearly. As a general rule, measures are usually taken to ensure confidentiality rather than anonymity, because some contexts make it impossible to ensure anonymity, and others require some form of identification in order to check background or history (through arrest records, for instance).

Utility, Feasibility, Fairness And Accuracy

Although informed consent, anonymity and confidentiality are important standards to consider while planning an evaluation, several other standards should also guide the process. The Joint Committee for Standards of Educational Evaluation (Gabor and Grinnell, ____) have identified four general areas, each of which must be assessed before, during, and after the evaluation. These standards are utility, feasibility, fairness and accuracy.

The standard of utility states that in order to proceed with an evaluation, it should be of potential use. No matter how untainted, accurate, or creative an evaluation strategy might be, if there is little reason to believe that results will be put to use, then the process should be halted. For instance, if the program to be evaluated is extremely unique and is a one-shot program that will not be repeated, there is little reason to assume that an evaluation will be helpful. Or, if a program is run under the centralized power of an authoritarian figure who is determined to proceed along a particular course, regardless of the evaluation results, than an evaluation will also be fruitless.

Second, an evaluation must be practical to implement based on the available resources including equipment, financial support, and the skills of the evaluator(s). If a program is simply struggling to survive and to deliver the services it has promised, there may not be enough time or energy to engage in a helpful evaluation. The threshold of staff tolerance might not allow for an additional task, and there may not be enough resources to hire a consultant. Or, there may be someone within the agency who can formulate a quantitative study when the program actually needs the expertise of a qualitative researcher. All of these circumstances may indicate that an evaluation is not practical to implement.

Next, an evaluation must proceed fairly (or with priority), so that a balanced picture of the program is presented, free from an undeserved influence of any individual or group of community participants. The methods described earlier for bringing community participants together can help minimize the bias produced by an overly-powerful community participant. This process also helps ensure that the range of information sources is carefully explored.

Finally, an evaluation should be technically adequate to provide accurate results, and any methodological limitations should be addressed by the evaluator. For instance, if the only control group available occurs through a matched sample rather than a random sample, the evaluator must be clear about the implications of using this strategy. Every decision in the formation of an evaluation design contains certain limitations and benefits, so each of these decisions must be made carefully and a consulting agency or professional evaluator can help clarify these issues.

In sum, the standards of utility, feasibility, fairness and accuracy should be followed. Again, these standards should be considered throughout the process, from the clarification of objectives to the selection of a control group, to the dissemination of results. Finally, in addition to this realm of basic standards, the ethics involved in the realms of community participants and evaluator biases must also be considered. When a program manager and evaluator can produce a design by being sensitive to basic standards and their own biases, the evaluation will undoubtedly carry greater integrity.

Dissemination Of Results

Methods for distributing evaluation results are typically seen simply as administrative tasks, and many of the community participants in the evaluation will simply assume that dissemination will

occur. If the dissemination of results was simply another technical piece in the organization of a program evaluation it would not be included in this section on ethics. The distribution of evaluation information is an important element to address here, because this part of the evaluation can have a dramatic impact on the power relations of those involved in the program. The way that information is disseminated can often influence the way results are used by community participants or the way public policy strategies are generated to address the findings.

Program managers must acknowledge that to the extent information is a vehicle for changing power relationships within or among institutions, evaluation can be an intensely political activity. It is important, therefore, for the process of disseminating results to address political structures rather than to be controlled by them. An evaluation that is significantly influenced by political pressure runs the risk of justifying decisions that have already been made, serving only as a public image maintainer, or placating funders with unuseful -- though allegedly sufficient -- information. Failure to plan the process of information dissemination carefully could result in any one of these dilemmas, even though the evaluation, itself, was completed with integrity.

To plan for information dissemination appropriately, thought must be given to the various community participants who will want access to the information. A thorough list of these individuals or groups should be generated prior to commencing the evaluation, and some brainstorming should occur regarding the kind of detail each of these community participants will want. For instance, in the case of a school-based violence-prevention curriculum, the final document produced for a parent-teacher newsletter is likely to be very different than a document prepared for the agency that granted money for the evaluation.

In addition to variations in technical sophistication, different documents may be produced on different pieces of the evaluation. The Superintendent of schools may want to know the precise steps involved in every lesson plan, while the local Black Caucus may want to know about the research implications for race relations. Special care must be taken when generating these target-specific documents because different community participants will be receiving slightly different information. The most important thing to remember is to highlight the findings consistently such that the strengths and weaknesses of the program (and the evaluation) are identified in every document. While some documents may possess a high degree of supporting detail and others may provide simply a summary of results, all must provide this basic summary of findings.

In sum, to ensure an ethical system for the dissemination of results, 1) identify the various types of documents that will be produced regarding evaluation results during the early stages of the evaluation. 2) provide a basic summary of results in every document, 3) attempt to reach as many community participants as possible through this information, and 4) provide these community participants with the type of information they need and understand,

Reasons Not To Do Evaluations

It seems only fitting that an appendix on the standards and ethics of program evaluation be attentive to the reasons that a program manager or an evaluator may decide that an evaluation cannot be conducted. The five main reasons for deciding not to complete an evaluation are: 1) The program is extremely unique, will not be repeated, and therefore results will be of little utility. 2) As a result of the "treatment" the clients will be placed in a position to influence adversely the health, safety, or well-being of themselves or others. 3) There is no possibility that the results will be acted upon in

decisions about program installation, continuation or improvement. 4) There is no body interested or informed enough to carry out the evaluation effort required, or other basic resources need to complete the evaluation are unavailable, and 5) The program may not be ready for an evaluation because, for instance, an outcome evaluation cannot be appropriately conducted on a young program that is still dealing with program start-up modifications. This last point is especially important to note because even when evaluation resources are available and positive attitudes about the utility of an evaluation prevail, it may still be inappropriate to engage in an evaluation. Again, the context of the evaluation must be considered carefully.

Ethical Conflicts

To conclude this appendix on ethics it is important to highlight the fact that ethical standards are not always clear. While it may be easy to define a particular act as unethical (e.g. falsification of test results or exposing subjects to a potentially harmful treatment), many ethical dilemmas emerge simply because of competing ethical standards and it may become difficult for a researcher to choose between the two principles. For instance, in a study designed to learn more about the culture of pregnant drug-users, researchers guarantee their subjects complete confidentiality in order to ensure the accuracy of the data and the cooperation of their subjects. The investigators promise that no matter what is said during the course of the interview, no information will be "leaked." Because the interviews are likely to contain some form of information regarding illegal activities -- and may perhaps even identify drug dealers -- the guarantee of this confidentiality is vital.

However, during a routine interview one subject expresses such depression and frustration, it is clear she intends to kill herself. She explains possible methods and places and indicates the suicide will occur within the next few days. At this point the researchers are faced with a terrible choice: Should they maintain their promise of confidentiality or should they attempt to protect the woman from harm? This illustration raises two extremely important questions that must be addressed: 1) Before a study begins, what are the *detailed* guidelines that are provided for such circumstances? and 2) When faced with an ethical conflict, what are some unobvious solutions that might be employed to reduce the potential for harm?

In relation to the first question it is obviously important to gather ALL possible information from funding sources and sponsors regarding their standards PRIOR to entering the field. Some of these standards may be guided by legal precedent and issues of liability, others may be standards dictated by institutional values. In the case of institutional values, a funding source that has Catholic ties may have very different expectations for dealing with suicide threats than one with ties to the Hemlock Society (an organization supporting euthanasia). The first source for information on ethical standards must begin with the sponsoring institution(s). Several additional sources may also be contacted for guidance, including The National Science Foundation; Public Health Service; Sigma Xi (a national society of research scientists); the American Association for the Advancement of Science; and the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (a DHEW commission that published the Belmont Report in 1979).

In relation to the second question, it is important to prepare for ethical conflicts and to plan some methods for diverting them. For instance, if researchers in the current example are provided

with some intervention skills and can present the subject with phone numbers of community services that might be able to intervene in a suicide, they can perhaps redirect the subject to obtain the help she needs. In this case, neither ethical principle is violated: The researchers do not reveal her identity or circumstance to anyone and they also help to prevent harm. When faced with such an ethical dilemma it is extremely important to brainstorm alternatives to the two choices and to creatively attempt to manage the problem. Another alternative to this dilemma would be a release form that the subjects sign, indicating that all information will remain confidential UNLESS the information indicates the potential to harm oneself or others. Again, it is extremely important to fully investigate the ethical principles that will be guiding a study before the study begins.

Appendix B:
Measuring Change

The Context of Measurement Decisions

When the goals and objectives of a program have been clearly articulated and when efforts have been made to bring community participants closer together, it is time to begin measuring. One of the many benefits to the processes of clarifying objectives and working with community participants is that many of the contextual specifics of the program become evident; key players, resource limitations, anticipated benefits of the evaluation and unique pieces of program design emerge to help guide the measurement strategies. Unless there is a desire to replicate programs and evaluation studies in order to verify previous findings, it is rare to find two, identical programs. Therefore, it is quite unlikely that a particular evaluation design will be the best option for two different programs.

To illustrate this point we might turn to the current debate among social scientists about the merits and shortcomings of qualitative and quantitative designs. Qualitative designs are often used by anthropologists and sociologists to obtain a "thick" description or an insider's perspective on a topic. This technique often begins with field notes and unstructured conversations with the population under study. Suppose a researcher wants to know how various white supremacist groups obtain the resources they need for their activities or how they recruit members through informal relationships. This researcher will most likely gain confidence among one or more of these groups and perhaps work with them to understand more about their methods and activities. Through direct observation and conversations, a description of this group's activities will eventually emerge. It is the

researchers job to sort through these modes of gathering information and to identify patterns, correlations, and even causes of various activities.

With quantitative designs, however, the goal is often to stay distantly removed from the subjects under study so that they can be more objectively analyzed. Suppose a different researcher also seeks information about a white supremacist group, but instead of examining the intimate relationships within and between organizations, he or she wants to learn about the voting patterns of group members and how these patterns differ from non-members. This study might include a confidential survey completed by known members of a group (or several groups) as well as nonmembers. This survey might ask them a variety of questions about voting frequency, party affiliation and membership in other special interest groups.

Like the qualitative approach, the data from this quantitative survey could reveal patterns, correlations, and perhaps even causes of various activities. However, the specific context of the research -- including the unique type of information sought -- compelled the researchers to generate different methods of measurement. The debate between qualitative and quantitative approaches is often fired by a misconception that once method is inherently <u>better</u>. An experienced evaluator understands that each method has advantages and disadvantages and that the method chosen must be based on a thorough understanding of the research context.

Because most evaluation designs are currently conducted in quantitative frameworks, this overview focuses on quantitative techniques such as the selection of a control group and the implementation of pre and post tests. Even so, as program managers become more familiar with the multitude of options in generating quantitative research designs, the importance of understanding the

context of a specific program will become more and more apparent. For instance, a researcher may decide to compare two groups at one point in time to compare their differences, or decide to examine an experimental group at different points in time to better understand long-term effects or developmental changes in the subjects under study. Because the goal of this overview is not to drown the reader with technical specifics, a detailed list of options will not be forthcoming. However, it is extremely important that a program manager has a general sense of the range of options. To this end we will provide a description of basic components in quantitative designs: Pre and post tests, control groups and sampling strategies.

Before And After

Before participants receive a treatment or begin a program, a test is often administered for some type of assessment. This test can be a means of selecting people for a particular program, to check assumptions that have been made in planning a treatment, or to determine the effects of an intervention by comparing early scores with scores obtained upon completion of the program. To assess the long-term effects of a program, this same assessment can be made several months or several years after treatment to determine the duration and degree of the change, should any change occur.

Because some surveys have withstood the test of time, program managers often seek these -- and other -- instruments because they have been used in similar settings to assess similar topics. Standardized tests can be found on a wide range of topics from anger management to IQ rates to drug addiction. Although these instruments are often sought by program managers because they have usually experienced a series of refinements and because the scores can often be compared to other populations, the importance of context again emerges as a cautionary flag. While standardized tests are generally more convenient, they are not always appropriate. Demographic differences and other discrepancies between studies might have an effect on the subjects and therefore an apparently unusual score might be nothing more than a difference between the characteristics of two, different populations; caution must be taken when comparing scores of two, separate populations with the same standardized instrument.

Another potential limitation of a standardized instrument is that the particular quality being measured might not appropriately match the goals and objectives of a unique program. Standardized

measures are usually accompanied by information that describes the specific clinical information obtained with the instrument. The most helpful measurement packages also include warnings about what the instrument does not measure. Care must be taken to ensure that the information gathered is appropriate to the focus of the program. Information is also provided with some instruments on strategies for selecting a sample, administering the questionnaire, scoring, and interpreting the findings. While these elements may initially appear attractive, care must again be taken to ensure that the methods provided with the instrument are appropriate to the target population and the information that is sought.

The two, general sources for locating standardized instruments are commercial publishers and literature produced through professional organizations. Most of the commercially produced instruments appear first in professional journals and books, although searching for instruments in this literature can sometimes be challenging. When available, computer data banks often provide the most expedient and thorough searches for instruments, although such databases are not common in some fields and may include only some of the relevant information.

In lieu of such databases, a consulting agency can provide guidance in the selection of an instrument and may help clarify or summarize the advantages and disadvantages of existing instruments. A consulting agency can also help develop an instrument that will be designed specifically for the program under study. While basic information on existing instruments is helpful to know, before making a decision on an instrument, it would be extremely prudent to discuss the matter with an evaluation researcher. Such a person can help explain how well the instrument

measures what it says in measures, or the range of variation in scores that can be anticipated within a population.

Control groups

After careful selection of an instrument to be used before and after a program or treatment, the issue of a control group must be addressed. Although a pre and post test can provide a measurement of the effects of the program, change among the participants can result from characteristics other than the program content. For instance, a high school might add a component to their life skills program that attempts to help youth realize some of the long-term implications of their actions. Although graduating seniors might demonstrate better reasoning about long-term implications than juniors who have not taken the life skills class, this difference may simply reflect the normal, maturation of the participants; the juniors and seniors might exhibit these results regardless of whether the seniors took the class or not.

In addition to normal maturation, the simple act of testing participants can also influence the results. The first round of testing might sensitize seniors to the instrument such that scores will improve the second time; improvement can be generated through familiarity with the instrument even though the life skills program is completely ineffective. By testing another group of comparable seniors who do not receive treatment, these factors can be taken into account. If a control group is pre and post tested with only moderate change in scores, while a treatment group is pre and post tested and indicates significant improvement, then the evaluation is more likely to reliably indicate positive, influential components of the life skills program.

Ideally, experimental and control groups must be identical in composition (demographics), experiences (the same activities during the time of treatment, other than the treatment, itself), and predispositions (established attitudes or behaviors). The best way to ensure that the control and experimental groups are identical is to randomly select participants from a larger population and then to randomly assign them into either the experimental or the control group. For instance, consider a conflict mediation class that is implemented in a school for adolescents who have been repeatedly involved in fights. Among the 75 students who are candidates for the program, 25 are to be selected for the class. Among the 75 students, therefore, 25 should be randomly selected for the conflict mediation program and 25 should be selected as a control group.

Like the group of students who participate in the class, the students in the control group receive both pre and post tests. Because the students are randomly selected from the at-risk population and because they are randomly assigned to either experimental or control groups, it is quite likely that any differences among students are equally distributed. The random selection of participants and the random assignment to control and experimental groups helps ensure that the groups are identical because it washes out other differences that influence results but were not considered in the creation of the design. Success at washing out the differences between randomly selected and assigned groups can also be examined with the pre-test; in this type of design, the pre-test can help verify that the two groups possess similar characteristics prior to treatment or program implementation.

During the efforts to establish a control group, care must be taken that ethical principles are not violated. In some circumstances, withholding treatment may be unethical and therefore the group initially established as the control may receive treatment at a later time. In addition to providing

treatment to a control group at a later date, an intervention program can be staggered to admit a cohort every three months or so, such that cohorts can be compared to each other at various stages of the program (including a pre and post test). One final alternative is to use another program as a control group. In this manner, the effects of two programs will be compared to each other. Although this framework can provide some helpful information and ensures that treatment is not denied, caution must be taken in the selection of a comparison group because the variables influencing different outcomes can be difficult to trace. Once again, a consulting service can help sort through some of these options with you.

Ethical dilemmas (such as deprivation of treatment) and pragmatic dilemmas (such as a small population) may prevent the construction of a true control group. In such circumstances, the next best option is to create a slightly less reliable control group by locating individuals who share the same characteristics as the experimental group. When a control group is intentionally matched with an experimental group, the assurances of random selection and assignment are no longer provided. However, a carefully matched control group certainly provides more reliable information than an evaluation that fails to use any type of control group at all.

Sampling

The process of selecting the control group and the group that receives treatment must be carefully planned. As already mentioned, the best way to select each group is to randomly select these groups from the population of interest and then randomly assign them to one of the two groups for pre and post testing. Programs are rarely designed and implemented to afford random selection

and assignment procedures, so alternative strategies must often be considered. For example, two alternative approaches are snowball samples and matched samples. In a snowball design, one or more participants are identified and this initial group provides references for additional contacts. This strategy may be used when a topic is highly sensitive and difficult to ascertain without gaining some sort of acceptance within a population. For instance, if the study intends to explore the economic relationships within gangs, several gang members will probably have to come forward, participate, and then provide the researchers with other members who might be interested in the study. Although this strategy provides certain advantages, it risks selecting only a certain "type" of gang member (one willing to reveal pertinent information) and therefore may not reflect the entire gang population. Caution must be taken when using a snowball strategy that inferences to the entire population are not made.

Matched samples are used when it is not feasible to divide a population into experimental and control groups. In these situations it may be possible to identify another population with similar characteristics and to treat this second population as a control group. For instance, if gun control laws are passed in a city and policy makers want to examine the effects of these laws, then a city with similar demographics, economy, and crime rates might be chosen as a control. This strategy, like the snowball sample, may provide certain advantages, but presents several dilemmas. First, because the city used as a comparison is a completely different city, important characteristics may be unintentionally left out of the matching process; the city used as a control group may have the same racial composition, but a much larger portion of the whites might be immigrants from Eastern Europe. These immigrants might possess vastly different approaches to violence and conflict than their white

counterparts in the experimental city and therefore bias the results. In addition to this dilemma, the matched sample risks being unable to control for events that occur in one city, such as a natural disaster, that do not occur in the other city.

In sum, whenever possible it is advantageous to generate control and experimental groups that are both randomly selected from the larger population and then randomly assigned to either the control or experimental group. This sampling design is likely to generate greater credibility than nonrandom sampling procedures such as snowball sampling and matched samples because any potential differences between control and experimental groups should be evenly distributed between groups. When random selection and assignment are not possible, however, one of the alternative strategies should be considered; it is better to have some sort of control group, if carefully designed, than no control group at all.

Appendix C:

Moving Community Participants Toward Agreement

The Delphi Technique

Although the above dilemmas do not always emerge in round-table discussions, it is important to be aware of the potential and to identify alternative methods for moving participants toward agreement. One such method is the Delphi Technique (Stecher and Davis, 1987) which can be used to mold as well as gather opinions. This technique is best suited for groups of 50 or less and takes the form of an anonymous, written dialogue. Participants are asked to articulate and rank a series of items such as goal statements. Responses are gathered, compiled and redistributed for further comment. When participants can examine what other people have constructed and become aware of the rationale for differing priorities, they may be more likely to compromise or alter their original stance. Several rounds of this technique often serve to bring a group closer to consensus and also prepare the participants for a round-table follow-up. Of course, this technique carries dilemmas of its own, including the potential for participants to feel overwhelmed with this addition to their paperwork, or stifled by the lack of face-to-face discussion.

Q-Sort and Nominal Group Technique

Two additional strategies for arriving at group agreement are the Q-Sort and the Nominal Group Technique (Stecher & Davis, 1987). The Q-Sort is similar to the Delphi Technique, although it provides participants with additional face-to-face discussion. With this strategy, a list of items is generated, and each item is placed onto a single card. These items could be the goals of a program

or the types of clients that receive services at the agency. Several of these decks are created, and then each individual in the discussion is given a deck.

Instructions for sorting the deck could be to place them in piles according to their priority, with no more than three, separate piles, or to identify the five most important cards and to place these in rank order. After this task has been completed, individuals are paired up and asked to identify their sorting differences and to solve them. These pairs later combine into groups of four, eight, and so on. While this approach does provide for some discussion and negotiation, it is less likely to generate a situation where "might makes right." By articulating perspectives and opinions in small groups, members who often remain silent are able to reiterate their points in a larger group.

The nominal group technique also provides a format for discussion, and is best used in groups of five to ten people. Members generate individual lists of options in response to a question, and then one item is read from every one's list in a round-robin fashion until a complete list of options has been generated. This system "forces" every member to contribute an idea and subsequently engages more members in a discussion of the options. During discussion, questions may be asked in order to clarify particular options, but attacking or advocating an idea is discouraged. After this clarification session, individuals rank the options on their own, and these rankings are tallied as a group. Decisions can be made on a majority vote, or further discussion of the options may be advised prior to a second round of individual tallies.

The Adversarial Proceeding

Another alternative for coming to a general agreement on theories, goals and objectives is to establish an adversary proceeding (Stecher and Davis, 1987). This method is based on the legal structure that encourages the presentation of opposing views and mediates these views with a jury or hearing officer. The rules can be more flexible than common legal proceedings and the process, itself, is not intended to replace other forms of dialogue. By allowing for the competitive consideration of proposals -- or at least a devil's advocate -- this process openly acknowledges a difference of opinion and provides the structure through which one view is deemed more appropriate. Although rhetoric and legal technicalities can render the process inefficient and unproductive, it serves as a welcome alternative to groups of people who view debate as an important learning and decision-making tool.

Sources

In addition to the references identified within the body of this overview, several of the materials listed below served as resources although they were not directly cited or quoted.

Anderson, S.B. and Ball, S. 1978. *The Profession and Practice of Program Evaluation.* San Francisco: Jossey-Bass.

Babbie, E. 1992. *The Practice of Social Research, 6th Ed.* Belmont, CA: Wadsworth Publishing Company.

Committee on Science, Engineering, and Public Policy (National Academy of Sciences, National Academy of Engineering, Institute of Medicine). 1995. *On Being A Scientist: Responsible Conduct In Research.* Washington, D.C.: National Academy Press.

Cyphert, F.R. and Gant, W.L. 1973. "The Delphi Technique" in E. House (Editor) *School Evaluation, The Politics and Process.* Berkeley: McCutchan Publishing Corp.

Fitz-Gibbon, C.T. and Morris, L.L. 1987. *How To Design a Program Evaluation.* Newbury Park: Sage Publications.

Gabor, P.A. & Grinnell, R.M.

Herman, J.L., Morris, L.L. and Fitz-Gibbon, C.T. 1978. *Evaluator's Handbook.* Beverly Hills: Sage Publications.

House, E. (Editor). 1973. *School Evaluation: The Politics and Process.* Berkeley, McCutchan Publishing Corporation.

Milgram, S. 1963. "Behavioral Study of Obedience." *Journal of Abnormal & Social Psychology.* 67:371-78.

Morris, L.L., and Fitz-Gibbon, C.T. 1978. *How To Deal With Goals and Objectives.* Beverly Hills: Sage Publications.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. *The Belmont Report: Ethical Principles and Guidelines for the Protections of Human Subjects of Research.* Washington, DC: U.S. Government Printing Office.

Owens, R. R. "Educational Evaluation By Adversary Proceeding" in E. House (Editor) *School Evaluation, The Politics and Process.* Berkeley: McCutchan Publishing Corp.

Rossi, P.H., Freeman, H.E. and Wright, S. R. *Evaluation: A Systematic Approach.* Beverly Hills: Sage Publications.

Scriven, M. 1976. "Evaluation Bias and It's Control" in G.V. Glass (Editor) *Evaluation Studies Review Annual, vol 1.* Beverly Hills: Sage Publications.

Stecher B.M. and Davis, W.A. 1987. *How To Focus An Evaluation, 2nd Ed.* Newbury Park: Sage Publications.