

**EVALUABILITY ASSESSMENT:  
A TOOL FOR PROGRAM DEVELOPMENT  
IN CORRECTIONS**

**Patricia Van Voorhis  
Professor  
Division of Criminal Justice  
University of Cincinnati**

**&**

**Kelly Brown  
Research Assistant  
Division of Criminal Justice  
University of Cincinnati**

Monograph Prepared for:

National Institute of Corrections,  
Attn: Dave Dillingham  
PO: #6012



## INTRODUCTION

Thirty years ago, an immense growth in government sponsored social programs dramatically changed the character of social agencies and the programs they housed. With offers of government grants and new policy initiatives, the Great Society optimistically set out to solve its social problems, including crime. But while agencies experienced numerous growth opportunities, the influx of new resources also delivered clear changes in society's expectations of public sector programs. The revolution was to be a scientific one. Policy makers, practitioners, and social scientists, alike, expected to learn from the experimental programs they encouraged, selecting those "that worked" for further development and funding. The new opportunities were conditional; agencies were expected to submit their new program models to evaluation research. Thus, where once public programs were afforded autonomy from government oversight, a "hands off approach, of sorts, they were now held to increasing standards of accountability.

Within 10 years of the emergence of the Great Society initiatives, evaluation research emerged as a growth industry (Rutman, 1980). A large technical and scholarly discipline appeared, and by the late 1970's every federal agency was mandated to conduct or assist in evaluations of new or existing initiatives (GAO, 1977).

The optimism of the 1960's was short-lived, assuaged, in part, by results of the very research it encouraged. Published during the early 1970's, the results of the early evaluation studies were devastating. In corrections and community corrections, Robert Martinson's conclusion that correctional rehabilitation efforts had "no appreciable effect on recidivism"

(Martinson, 1974) were echoed by other studies. Taken together, the first generation of evaluation studies nearly dismantled any further efforts to provide meaningful interventions for offenders (Cullen & Gilbert, 1982; Palmer, 1992). Similar findings were describing other social endeavors, including education and mental health (Mann, 1972). Ultimately, the impact of the “nothing works” message helped to fuel political disenchantment with public programming. Doubts about whether these programs worked amplified growing concerns for the costs of government spending (Cullen & Gilbert, 1982).

Although one could fault the high costs of wasteful government or mean-spirited conservative funding agendas for the new found skepticism over government’s ability to solve social problems, a number of social scientists, chose to take another look at the early evaluations. And then upon closer examination, several scholars noted that the research had made premature conclusions in asserting that the programs had not worked (Gottfredson, 1979; Palmer, 1975). In depth critiques of the correctional evaluation research, for example, revealed clear flaws in the research methodology of most studies; strong conclusions of any sort simply were not warranted (Palmer, 1975; 1978).

In addition to methodological and technical problems with the research, it should have been clear to researchers and programmers alike, that some of the evaluated programs had been too difficult, if not impossible, to evaluate---but they were evaluated, anyway. In fact, many of the evaluations described poorly designed programs which evidenced unclear goals and no clear understanding of what activities would produce the desired results (Nay & Kay, 1982; Wholey , 1979; Wholey , 1983). Often the evaluations attempted to measure the effects of activities that were not actually taking place, because the program had incurred problems

with their implementation (Nay & Ray, 1982; Rutman, 1980; Scanlon, Horst, Nay, Schmidt, and Waller, 1979). In addition, little attention was devoted to the process or service delivery models of the programs. “Black box” evaluations reported outcome/effectiveness of programs that were not described (Nay & Kay, 1982; Rutman, 1980). When such evaluations produced successful findings, the programs could not be replicated. When such programs failed, it was unclear whether they had not worked or had not been implemented according to design. Simply put, much of what reportedly “did not work” in actuality, “did not happen” (Van Voorhis, Cullen, & Applegate, 1995). Other outcome evaluations were regarded as useless and wastes of valuable resources (Nay & Ray, 1982; Rutman, 1980; Wholey , 1979; Wholey , Nay, Scanlon, & Schmidt, 1977). Few actually generated program improvements or were found relevant to program administrators’ needs (Wholey , 1979).

Although not in time to prevent these research findings from impacting on public funding agendas, the massive costs of evaluating interventions and programs prematurely were becoming clear. In the broadest sense, these costs included: a) serious threats to the future of the programs or initiatives being evaluated; ineffective programs were not likely to be refunded; b) inaccurate knowledge of what “worked”; the flawed research seriously compromised the development of a larger knowledge base; c) inability to identify program changes that would make programs more effective; and d) a waste of research monies; outcome evaluations are time consuming and expensive. And on the larger policy level the cumulative effect of several such evaluation produced the most onerous cost of all--the withholding of resources for dealing will serious social problems---the very ammunition needed by some to say that we need not deal with some social problems because nothing

works.

To overcome these problems, a research procedure known as the “evaluability assessment” was devised in the early 1970’s, most notably by scholars at the Urban Institute in Washington, D.C. (Rutman, 1980; Smith, 1989; Wholey, 1979). As the term suggests, evaluability assessments (EAs) were designed to precede an outcome or comprehensive evaluation, to determine whether a program is “evaluable”. Essentially, EA helps to determine whether a program has been planned and implemented well enough to be evaluated. Moreover, if the EA determines that the program is “evaluable”, it is likely also to suggest some of the parameters for the evaluation, e.g., criteria and effectiveness measures. If the study determines that the program is not “evaluable”, an EA should direct program personnel to areas of the program that need further development. In such instances, the EA will forestall a more costly and potentially damaging evaluation of the program until program adjustments have been made (Schmidt et al., 1979; Smith, 1989; Wholey, 1983; Wholey et al., 1977). A list of the benefits of EA’s appears in Table 1.

Depending on their findings, EAs can result in any of the following organization decisions (Smith, 1989):

1. To conduct a comprehensive outcome evaluation of the program.
2. To change the program.
3. To do nothing/take no further action.
4. To stop the program.
5. To ignore the EA.

In the pages that follow, we focus on the second option, using EAs to facilitate

Table 1: The Purposes of Evaluability Assessments and their Value to Program Administrators and Practitioners.

---

- o Clarifying a programs goals and objectives.
  - o Diagnosing goal confusion among stakeholders, and preventing staff, clients, administrators, funders, and others from working at cross purposes.
  - o Assessing the feasibility of the programs goals and objectives and determining which program objectives are most attainable.
  - o Identifying program priorities. Which objectives are most important?
  - o Maximizing programs use of research monies, by discouraging a costly comprehensive evaluation until ready.
  - o Diagnosing program's strengths and weaknesses.
  - o Improving program performance.
  - o Examining the logical consistency between program components. Is there an explicit theory which links needs to program procedures in a logical manner, reflective of cause-effect solutions. If not, the EA may suggest one where problems are linked to objectives, and specific activities/services for achieving the objectives.
  - o Assessing the extent to which the program design fits the knowledge base of the discipline (are interventions being used that have been known to fail in the past).
  - o Making recommendation for the larger evaluation or determining if an outcome evaluation is possible.
  - o Determining the information needs of stakeholders, thereby assuring that a larger evaluation will obtain the appropriate information needs of the stakeholders.
  - o Saving agency resources, by making improvements in a timely manner and forestalling expensive but unwarranted evaluations of a program that is not ready to be evaluated.
  - o Facilitating the development of a "learning organization," where agencies and programs plan futures according to accurate diagnostic information regarding present day operations. Evaluability assessments, then serve as aids to program development and planning.
-

program planning, development, and change. We intend this monograph to serve as a guide to practitioners in community and institutional correctional programs and to encourage the use of evaluability assessments as tools rather than “win/lose” propositions (option 4). Indeed, most of the questions addressed by EAs center on how well the program has been planned. The essence of this monograph, then, is on how to plan and develop the type of program that is likely to produce a successful comprehensive evaluation. Such an evaluation would examine both the integrity of the programs services and their effectiveness (Rossi & Freeman, 1989).

The monograph will also move beyond the early technology of evaluability assessments to models that better fit present-day correctional programming. We note that correctional evaluability assessments of today reflect the results of a second, more successful, wave of evaluation studies, published from the mid-1980's to the present. Reviews of the more recent research (meta-analyses) show us that correctional intervention is more successful than portrayed in the earlier research (Andrews, Zinger, Hoge, Bonta, Gendreau, & Cullen, 1990; Andrews & Bonta, 1994; Gendreau & Ross, 1987; Lipsey, 1991; Palmer, 1992; Van Voorhis, 1987). These meta-analyses also identified the types of programs and program strategies that were most likely to produce successful outcomes. Understandably, some recent evaluability assessments have examined programs in terms of whether or not they demonstrate these principles; are they utilizing, in other words, some of the program dimensions known to be effective and avoiding those established as ineffective (Gendreau & Andrews, 1994)? This monograph highlights some of those most recent efforts. We begin with a demonstration of how evaluability can be a function of program planning.



## **PROGRAM PLANNING AND EVALUABILITY: SOME OF THE PROBLEMS**

The link between program planning and evaluability can be stated quite simply. Poorly planned programs cannot and should not be evaluated; when they are evaluated, the results of such evaluations can be misleading and policy makers may be prone to conclude that a program does not work, when, in fact, the program may not have operated according to a sound design, or it may not have operated at all. In all likelihood, many of the perceived failures in correctional treatment probably occurred during the program planning. In order to show how this occurs, we portray a series of planning and implementation stages which parallel program evaluation activities.

We start with the assumption that the mind set of a good evaluation (evaluability) and the mind set of a good program plan are similar. For example, both the evaluators (as well as those who assess evaluability) and the staff/planners must address several key questions:

1. Who are our clients and what problem characteristics are we treating? Who, in other words, is our target population?
2. What intervention fits the needs or problem areas of this target population?
3. Why did we choose that particular intervention?
4. What are our goals and objectives?
5. Is there a logical and theoretical relationship between these goals/objectives, the needs of our clients, and the interventions we have chosen?
6. How will we know when we have implemented this intervention according to design?
7. How will we know whether the intervention was effective?

If these questions have not been asked and determined in advance of program

implementation, there is no way that an evaluation will be able to answer them at a later day, after clients have completed the program. Additional problems will occur if stakeholders have different views regarding these issues (Smith, 1989). For example, if staff, policy makers, and program funders, have different ideas concerning the program's mission and its criteria for success and want different information from an outcome evaluation, the success of such an evaluation will be compromised, as will the evaluator's ability to carry out the evaluation. We review some of the above questions in turn to illustrate how insufficient attention to planning affects evaluability .

### The Importance of Knowing Who We Are Treating?

An overview of recent criminological research shows that, the causes (or risk factors) of criminal behavior and recidivism are well known. Moreover, the meta-analyses show us that most program successes in reducing recidivism are achieved by targeting: a) high risk rather than low risk offenders, and b) specific risk factors or causes of crime (Andrews et al., 1990). Alternatively, programs may fail to reduce recidivism when they are administered to a vaguely described group of offenders or to some problem that is not a cause of crime. Some of these more common, albeit popular, "mistargets" include self-esteem, poverty, and intellectual functioning.

Apart from consideration of individual correlates of crime are individual characteristics such as personality (e.g., psychopathy), intelligence, and motivation, characteristics that affect one's ability to participate and succeed in a program. Andrews, Bonta, and Hoge (1990) incorporate these under their "responsivity principle". Warren (1983) and Palmer (1978) referred to them earlier according to principles of differential treatment or matching.

When we don't fit interventions to risk and responsivity factors, programs are likely to incur the following evaluation results or problems:

1. A program fails, because it is targeted to individuals who can't benefit from the program.
2. A program works with some and not with others. Our successes are canceled out by our failures. The program looks bad, when it really did work with some. Moreover, we will obtain the information needed to identify the characteristics of individuals best suited to the intervention.
3. The program is a true failure because it did not provide a service which targeted a problem that is related to crime. Offenders, then, might improve on the targeted problem, but the improvement will not produce any changes in their future offending behaviors.
4. A program actually harms some or all participants who are ill suited to the intervention.

The Value of Planning an Intervention Which is Empirically or Theoretically Sound, and Implementing It According to Design

This problem has two facets: (1) employing an intervention strategy that is known to be effective or to have some logical, theoretical reason for being effective; and (2) monitoring the implementation or the program services to determine whether the services are being delivered as designed. This is the problem of the "black box" which is mentioned in many treatment and evaluation texts. In observing these problems, we could also say that the program had limited treatment integrity.

There are many examples of these types of concerns:

1. The program is not delivered according to design. This can occur when earlier printed, rhetorical, accounts of the program fail to describe the program in enough detail to facilitate its replication. It also occurs when staff fail to implement or execute all of the clinical dimensions of an intervention (e.g., a

social learning program which does not incorporate principles of good role modeling).

2. There is no clear intervention design; we are instead delivering some global treatment, e.g., counseling, case management, job skills, rather than an explicit treatment process that is known to be effective.
3. The program has conformed to the latest panacea---a program that has intuitive appeal but no theoretical or empirical promise.
4. Staff don't understand the intervention. They were not a part of the planning process; they were improperly trained; and/or they do not have or follow a treatment manual. They simply don't have enough information to administer the services correctly or consistently.
5. The organizational and political climate is too confusing to support a successful intervention.
6. Budget cuts create a situation where programs are asked to do the impossible - - keep the program without the funds.
7. The "dosage" is inadequate. It may be a good intervention, but the amount of client participation time is insufficient.
8. Clients did not or could not attend.
9. No attempt was made to determine whether the program is properly designed or operating according to design.

The most tragic outcome of such events occurs when an evaluation creates the impression that "nothing worked" when, in fact, "nothing happened". But there are other problems with neglectful interventions. For one thing, they cannot be improved or replicated because there is no way of doing so with a program which we do not understand.

### How Will We Know When We Have Succeeded?

In the correctional literature, a common method of answering this question is with a general

measure of recidivism, such as revocation, rearrest, or reconviction. While recidivism is not an unimportant measure of program success, it must be recognized that programs are not the only forces influencing offender recidivism (DiIulio, 1993). Most recently, correctional agencies and programs have been encouraged to think in terms of what it is they hope to accomplish in a shorter time frame, by the time offenders are released from program supervision (Petersilia, 1993; Van Voorhis et al., 1995).

As we view the treatment evaluation literature, however, we notice limited attention to **intermediate objectives** (i.e., what the program hopes to accomplish with whom by the end of program participation). Ideally, the intermediate objective must relate to change on some individual characteristic that is a “cause” of crime; this characteristic should also be one of the causes or risk factors our program targeted; it is also a “criminogenic need” (Andrews & Bonta, 1994).

Failure to develop and attend to intermediate objectives or for that matter, failure to develop any clear, attainable, goals can result in the following:

1. Stakeholder confusion over the program’s mission.
2. A costly evaluation of a program whose mission and goals are unattainable. This occurs in a number of ways, through vague goals, insufficient resources, and insufficient interest. Finally, some programs exist on paper, as in the case where media attention is more valued than actually committing resources to the lofty program goals.
3. Confusion in research and policy-making circles concerning what is a realistic measure of success. Some say a 10 percent reduction in recidivism is a failure (e.g. see Whitehead & Lab, 1989) others maintain that even a 10 percent reduction can translate into meaningful cost benefits (see Gendreau & Ross, 1987). Still others assert that true measures of success are also found in the more intermediate attainments, such as more education, new skills, and

improved employability (Palmer, 1978).

4. Inability to evaluate the program on the criteria most relevant to what it actually did! The program ignores the very standard that is most relevant to its service and most likely to show success---the intermediate objective. In this way programs may be setting themselves up to fail
5. Failure to recognize an important risk factor for recidivism, since change on criminogenic needs are predictive of future recidivism (Andrews & Bonta, 1994).

It is clear that the best place to resolve these problems is during program development. At the same time, these are the very problems that evaluability assessments are designed to detect and correct prior to the execution of a full scale evaluation that could have devastating costs.

## **EVALUABILITY ASSESSMENT MODELS**

In this section we review the process of conducting an evaluability assessment, including questions, information needs, and main actors involved. We begin with the earlier, more generic models of the 1970's and move to a model more specifically geared to correctional programming, one which is also reflective of recent correctional research.

### A General Model of Evaluability Assessment

A synthesis of the early evaluability models (e.g., Rutman, 1987; Wholey, 1979, 1983; Wholey et al., 1977) reveals a sequential series of activities, ranging from preparatory work to final feedback and reporting to program and agency staff. Typically, the assessment is conducted by an outside evaluator.

Questions specific to each step of the evaluability assessment are listed in Table 2.

A good deal of preparation must occur in advance of actually conducting the evaluability assessment. Prior to conducting any interviews or observations, evaluators routinely devote ample time familiarizing themselves with the nature of the program being assessed and determine the general parameters for the EA. They attempt to accomplish a clear understanding of the purpose of the evaluability assessment; understandably, they hope to find that the identified purpose of the EA is one that is shared by other stakeholders. The evaluators also attempt to secure assurances that staff, clients, managers, and other key stakeholders are committed to the EA and agree upon the activities needed to achieve its successful completion (Smith, 1989). At this point, evaluators also will want to know if stakeholders are in a position to make the programmatic changes recommended by the EA. Finally it is essential that the evaluators know what agency activities and purposes constitute “the program” to be evaluated (Wholey, 1979).

Smith (1989) recommends the early formation of an EA work team that is composed primarily of implementation staff, but may also include agency management, if needed. Smith’s vision of the EA work team was one whose responsibilities included: a) the identification of stakeholders for interviews, b) the review of interview questions to be asked during the EA, and c) input on the final evaluation report and its recommendations. One benefit of this high level of staff involvement is to improve their program development skills as well as their interest and appreciation for a more comprehensive evaluation. There are some problems with Smith’s conception of the work team, however. First, a number of programs cannot afford the staff release time needed to staff such a team. In addition, EA work teams

Table 2: General Issues of Evaluability Assessments

---

INITIAL QUESTIONS:

1. What is the purpose of the evaluability assessment?
2. What agency or organizational activities and resources comprise “the program”?
3. What is the nature of stakeholder commitment to the evaluability assessment?
4. Who should be interviewed in the process of conducting the EA.

QUESTIONS FOR DATA GATHERING PHASES

1. What is the program by design (the intended program)?
    - a. What are the program’s goals and objectives?
    - b. Is there goal consensus among stakeholders?
    - c. What is the theoretical model for the program?
    - d. What program activities are supposed to follow from the theoretical model?
    - e. What resources are planned for implementing and operating the program?
    - f. What are the intended outcomes of program services.
  2. How does the program actually operate?
    - a. How are client needs assessed?
    - b. What services are provided?
    - c. How are staff trained for program implementation and operation?
    - d. Is there an implementation plan for the program (if it is a new program)?
    - e. What concerns, if any, do stakeholders have for the program?
  3. Questions pertaining to any future evaluations.
    - a. Have procedures for measuring attainment of program objectives been identified?
    - b. What information needs do program stakeholders have? What uses would they have for a larger evaluation?
    - c. What level of interest is there among stakeholders in a larger program evaluation?
-



Table 2, Continued

---

### CONCLUDING QUESTIONS FOR THE EVALUATION REPORT

- a. Do stake holders agree about the expectations, goals, and objectives of the program? If not, what areas of disagreement were detected and what is the likely outcome of these disagreements?
  - b. Are staff qualified and properly trained to operate the program?
  - c. Is there a coherent, logical program theory?
  - d. Do the actual program services fit the designed program model/theory?
  - e. Are there any unintended side-effects from the program?
  - f. Can the program design be validated? Is it actually operating according to design?
  - g.** Is the program plausible? Are services being delivered in sufficient amount and quality? Are goals attainable through identified resources?
  - h. How well do program services and client needs reflect the designed program theory?
  - i. Are program objectives and services clear enough to facilitate a program evaluation?
  - j. Would a larger evaluation have any utility to program stakeholders?
  - k. What is the agenda of a larger program evaluation?
  - l. What evaluation measurements and questions are feasible?
  - m. What measures and evaluation criteria are acceptable to stakeholders?
  - n. What evaluation data are obtainable for the agreed-on evaluation measures?
  - o. Is the actual program likely to achieve agreed upon criteria for success?
  - p. What programmatic changes would enhance program success?
-

could quickly compromise the benefits of an independent external evaluation. External evaluators, in comparison to “in-house” evaluators are widely assumed to be less biased by organizational patterns, perceptions, and preformed opinions (Weiss, 1972). Still, Smith’s concerns are well founded. EAs are impossible to conduct without some cooperation from administrative and front line staff, and the work teams could successfully counter the staff suspicion that sometimes surrounds EAs. Fortunately, there are alternatives to accomplish the same benefits, e.g., through ample interview time, and opportunities for staff to review and comment on the final report.

After identifying the scope of the evaluation, the parameters of the program to be evaluated, and the individuals to be surveyed, the next step is to secure a picture of the ideal program, or the program according to its design. The evaluability assessment, then, entails an understanding of the “program on paper”. At a later point in time, this program design will be validated through interviews with staff and observations of the program in operation.

Key to this stage of the evaluation is the construction of the program’s theory (Wholey, 1979). Here evaluators attempt to identify logical reasons for why the stakeholders believe that program activities will achieve their desired ends. There must be plausible causal links between client needs, program activities, and intended goals; evaluators will seek to map these links. Smith (1989) indicates that the theory need not be an established psychological or sociological theory of causation. Logical, rational links between problems, objectives, and services will suffice. <sup>1</sup>

Accounts of the ideal program are secured through reading program and agency documents (Smith, 1989) as well as through an initial series of interviews with stakeholders,

particularly the managers (Wholey, 1979). At this point, evaluators seek a sense of management's purpose and expectations in implementing the program. Much of the information about the program's goals and objectives will represent management's ideas concerning performance indicators that may ultimately prove useful in a larger outcome evaluation of the program. At this point, resources, program assumptions, and progression of services/activities also will be identified.

The information furnished by the program managers and designers subsequently will be cross-referenced through questions of staff and field observations of the program. Conducting the EA in this two-tiered process affords an examination of the agreement among stakeholders about the program's purpose (goals and objectives). At the same time, moving from the ideal model to the actual model is intended to provide an opportunity to see whether the program is operating according to its design. In some instances, EAs observe a program where managers and staff alike have not articulated clear goals, objectives, and purposes. In such cases, there is no clear program model and thus no way to assess the fit between the ideal and the actual model. Another too common scenario, involves situations where staff resistance impeded efforts to put the program in place. Finally, a program whose plan is too rhetorical or grandiose on paper for available resources is not likely to operate according to its plan. In such instances, the final EA report is likely to recommend further planning.

Another focus of evaluability assessment concerns plausibility. Can this program be implemented and operated on the resources earmarked for the program? Will the program be able to identify and attract the target population it was designed for? Are the staff qualified to deliver the services as planned? Will they receive training prior to implementing new

assessments or services? Are the goals, objectives, and expectations of program planners realistic?

We must also examine the plausibility of the outcome evaluation. By this point, it should be obvious that an unclear or poorly implemented program should not be evaluated and if it is, anyway, the evaluation is likely to be misleading. It may also be deceptive, if readers of the evaluation are not informed that the program was not successfully implemented at the same time they are told that it was ineffective!

Finally, even if the program is well planned and implemented, an outcome evaluation is not possible without the information or data needed to conduct the evaluation. Is it possible to formulate clear indicators/measures of success? Does the agency have the capability of efficiently generating the data needed for the evaluation? These issues must be resolved prior to embarking on a large-scale evaluation. In most instances, agencies with a viable management information system or clear record-keeping process, will be viewed more favorably than those without one.<sup>2</sup>

### Evaluability Assessments in Corrections

EAs in corrections could, of course, proceed in a manner identical to the process outlined above. In this section, however, we will focus on an evaluability method that is even more relevant to the demands of correctional programming. In doing so, it is important to recognize that the current technology of correctional EA goes beyond simply fitting the questions outlined in Table 2 to a correctional model. The more recent EAs hold programs to a standard of operating from the knowledge base of a massive body of correctional evaluation research generated from the 1970s to the early 1990s. As noted earlier, these

studies were assessed in a series of meta-analyses and reviews (e.g. Andrews, Zinger, Hoge, Bonta, Gendreau, & Cullen, 1990; Andrews & Bonta, 1994; Gendreau & Ross, 1987; Lipsey, 1991; Palmer, 1992; Van Voorhis, 1987) whose primary benefit was the identification of the characteristics of effective treatment programs, also known as the “principles of effective intervention” (Andrews & Bonta, 1994; Gendreau, 1995).

Because the evaluability assessment model presented above predates the release of the major portion of the evaluation research, it is easy to see why the earlier evaluability assessments expected little more than program consistency, integrity, and a record-keeping function. In contrast to earlier EAs, more recent evaluability assessments also ask whether the program utilizes a service delivery model that is known to be successful with offenders, or is it repeating a failed intervention from the past? In doing so, EAs now assess whether programs utilize what has been learned about ideal target problems, the types of programs that work, and organizational characteristics of effective programming (e.g., staffing, resources, amounts of service delivery, and governance structures). In sum, evaluability assessments not only examine whether a program has a well articulated target population and program objectives, but whether the program is targeting the individual problems most likely to reduce recidivism with the services found by previous research to be most effective.

One evaluability assessment which incorporates the principles of effective intervention is the Correctional Program Assessment Inventory (CPAI)(Gendreau & Andrews, 1994). In reflecting the lessons of several meta analyses, the main considerations of the CPAI are that programs:

1. Target high risk offenders. Previous evaluation research notes that the most successful outcomes are achieved with high risk rather than low risk offenders.

Low risk offenders, in fact, sometime do worse in offender treatment programs, than similar offenders placed in a comparison group.

2. Target “criminogenic needs”. In other words, programs are advised to attend to those individual characteristics known to be associated with crime and recidivism. Criminogenic needs are also called risk factors. When they are reduced in correctional interventions, we a note reduced likelihood of future recidivism. In evaluating a program on this criteria, it would also be important to observe the program using reliable and valid assessments of the criminogenic needs and risk.
3. Address responsivity. In addition to targeting risk factors, programs should screen out those clients who would not have the ability to participate. Depending on the nature of the program, relevant responsivity considerations include intelligence, anxiety, verbal ability, motivation, and more recently, cultural appropriateness.
4. Use interventions known to be effective. The meta analyses note that some programs are particularly well suited to offenders, while others are not.
5. Attend to treatment integrity. Build in certain quality assurance measures to assure that the program is implemented and operated consistently according to design.

Criminogenic needs are identified as crime correlates throughout the literature and research on crime causation. In consistency with this literature, and with the authors’ own meta-analyses of the predictors of criminal behavior (Gendreau, Andrews, Coggin, & Canteloupe, 1992) and recidivism (Gendreau, Little, & Goggin, 1995)) the CPAI lists the following as appropriate criminogenic needs (program targets) for offender populations:

1. Change attitudes, orientations, and values favorable to crime.
2. Change antisocial feelings.
3. Reduce antisocial peer associations.
4. Reduce problems associated with alcohol/drug use.
5. Reduce anger/hostility.
6. Teach prosocial alternatives to lying, stealing and aggression.
7. Increase self-control, self-management and problem-solving skills.
8. Encourage constructive use of leisure time.

9. Improve skills in conflict resolution.
10. Promote more positive attitudes and performance toward work and school.
11. Resolve emotional problems associated with child abuse (physical, emotional, sexual).
12. Promote family affection/communication.
13. Improve family problem solving.
14. Promote family monitoring and supervision.
15. Resolve deviant sexual arousal/attitudes/behavior.
16. Provide low-pressure, sheltered environments for mentally disordered offenders.
17. Focus on harm done to victims.
18. Relapse prevention: teach clients to recognize high risk situations and to deal with them.

The evaluation research of the past three decades, and the meta-analyses of those studies, clearly finds patterns suggesting that some types of programs work better with offenders than others. Assuming that the program is targeting criminogenic needs with these interventions, the following have proven to be most successful:

1. Social learning interventions (e.g., anti-criminal modeling and skills development) .
2. Cognitive-behavioral approaches (e. g . , cognitive therapy, cognitive skills training, problem solving therapy, rational emotive therapy, self-control skills, self-instructional training, stress inoculation training, criminal personality groups).
3. Radical behavioral approaches (e.g., classical conditioning such as desensitization, operant conditioning such as token economies and contingency management models).
4. Targeting specific criminogenic needs (e.g., treatment of such specific types of crime as sex offending, substance abuse, and violence).

In addition, the meta analyses note that some types of programs have not been helpful and have sometimes been found to be harmful to offender populations. These include:

1. Non-directive, person-centered therapies.
2. Psychoanalysis.
3. Group milieu therapies.
4. Chemotherapies.

5. “Punishing smarter” models (e.g. boot camps, drug testing, electronic monitoring, shock incarceration.)<sup>3</sup>

Researchers have found that effective programming includes more than the choice of target problems and the interventions they are matched to (Andrews & Keisling, 1980; Palmer, 1992). As a result, several items on the CPAI examine such organizational and program components as:

1. Whether the program offers enough service to program clients. A number of factors can impinge upon a program’s ability to provide adequate “dosage”, including poor attendance and limited resources.
2. Appropriate job qualifications for program staff, managers, and administrators.
3. Leadership involvement in program design and delivery. Some level of leadership involvement is preferable to a model where leadership and management are isolated from program design and operations.
4. Staff involvement in program decision-making. Top-down leadership generally has not been found to be effective in the past.
5. Provisions for quality assurance. The program should have procedures for assuring the integrity of its service-delivery model, that the program has been implemented according to its design. One important provision would be a program manual, providing a clear implementation and service delivery description for the staff assigned to operate the program.
6. Record keeping and management information systems. These of course are relevant to the program’s ultimate evaluability.
7. Follow-up services. Programs with relapse-prevention models are more successful.
8. Evaluations and monitoring. The program collects data on client recidivism. Learning organizations, that seek feedback about how they are doing are simply more successful than those that never ask.

In viewing these criteria, one might fault the expectation that programs be wedded to



the familiar. Is the CPAI, in other words, encouraging programs to replicate known interventions, thereby discouraging innovation? Would experimental programs be viewed in such a negative light as to discourage future innovation and experimentation? The answer to this question should perhaps consider the degree to which the experiment departs from the knowledge base of the criminology and corrections discipline. Indeed, the history of panaceas in corrections furnishes us with a sordid list of program targets that had absolutely nothing to do with criminal behavior along with interventions that were popular for reasons other than the fact that they were appropriate for offenders. Most important, while panaceas or “magic bullets” have been common to correctional program endeavors, their evaluations results have been dismal (Palmer, 1992).

Just the same, in all likelihood, there will be new theories of crime, and it would be reasonable to expect to find them reflected in correctional interventions especially once the theory is found to have empirical support. In response, just as the earlier evaluability assessments sought to identify a program theory of operation, concerns for theory shown in the CPAI seek to determine if the program has a good reason for selecting the intervention and the targets it selects. Is the program demonstrating, in other words sound understanding of the logical links between causes of social problems (program targets) and their solutions (program interventions)?

As for the criminogenic needs, future criminologists may identify new ones or determine that some like self esteem are criminogenic for some types of offenders but not for others (Van Voorhis, 1995). But one needs to reflect on the fact that criminogenic needs are based on crime correlates, and crime correlates have show tremendous stability over time,

place, and population variabilities. Even so, one can expect that the CPAI will be an evolving tool; it has undergone several revisions to this point, and one can expect future versions to be current with emerging criminological research.

The CPAI uses a scoring system for each of the program dimensions listed above, but in actual practice programs are more likely to request a CPAI for diagnostic purposes, to identify program strengths and weaknesses (Van Voorhis et al., 1995). In this sense, it is an extremely valuable tool to incorporate into organizational consulting or technical assistance activities aimed at identifying areas for program refinement and improvement.

Whether the evaluation is for program improvements or a “win/lose” ordeal, may in itself be a consideration for an evaluability assessment. Indeed, criminal justice agencies have been faulted for not furnishing the type of organizational climate that facilitates good planning and appropriate use of evaluability assessments and evaluations. Indeed, the organization that learns from its mistakes is qualitatively different than the one that leaps from one failed panacea to another. This concern moves us to the next section, where we move to suggestions for planning the type of program that will do well in an evaluability assessment.

## **A MODEL FOR SUCCESSFUL PROGRAM PLANNING**

Program success on an evaluability assessment is dependent upon both organizational characteristics and the planning and program development skills of the staff. The most important directions for programs are to assure that the program: a) has a healthy organizational climate for program development and planning, and 2) utilizes a rational planning

model.

### The Organization Climate

The first questions we might ask in this regard concerns, how safe is it for staff and administration to evaluate their program? One does not have to look far to observe that evaluation results have been misused to undermine the continued existence of correctional interventions. They can be extremely vulnerable to political interests and are often misinterpreted. Evaluations are too frequently conducted with the understanding that negative evaluation results will sound the final bell for a program. Small wonder that program staff may not want to cooperate with evaluation research. A more constructive approach is sorely needed.

There are, in fact better reasons to evaluate programs---reasons that are far superior to the win/lose stance that has characterized so many evaluation enterprises of the past. If properly conducted, evaluations can help a program improve its planning and effectiveness. In doing so, an evaluation can engage staff in a constructive share of the planning. Staff involvement in planning and program articulation can be a source of staff motivation.

We can expect substantial improvements in the climate of a program evaluation by simply rejecting the “win-lose” perspective on evaluation research which sees programs as either “working” or “not working.” In fact, if the evaluation study is comprehensive enough, it will produce more information than whether the program worked. It may, for example, identify specific program components or services that failed (rather than the entire program); it may show us that the program worked for some types of offenders but not for

others; it may tell us whether the program achieved a proper service “dosage”. Thus, the evaluation may become an important source of feedback, leading to program adjustments rather than to program obliteration (Rossi & Freeman, 1989).

We might also create a fairer climate for the program by promoting more wide scale appreciation for realistic standards of success. Indeed, researchers tell us that the most we can hope for, with the most effective program designs to date, is a 20% improvement in recidivism rates for an experimental group over a comparison group (see Lipsey , 1991; Palmer, 1992), perhaps 40 to 50 percent among the most optimistic reviewers (see Andrews & Bonta, 1994; Gendreau, 1995). While some have argued that these figures indicate failure rather than success (see Whitehead & Lab, 1989), even the conservative 20% success rate translates into impressive cost benefits (Gendreau & Ross, 1987; Lipsey, 1984).

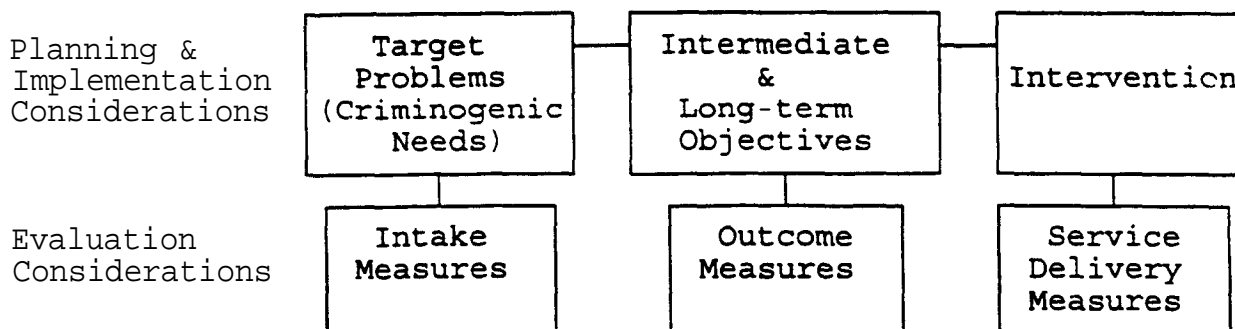
Organizations also discourage planning and evaluation efforts when they are unable to provide adequate financial and organizational support. Staff may not be able to break free from ongoing responsibilities to meet in planning or evaluation sessions. The program may not have the desire or funding to conduct an evaluability assessment prior to launching a larger evaluation. Finally, there may be insufficient financial resources to train staff. Sometimes, programs want rapidly assembled results that can be shown to their stakeholders (e.g., referral sources, funders, and political constituencies). And while they may be able to produce an evaluation quickly, the result may be an evaluation that: a) is based upon remote measures of program success, b) cannot be used to learn much of anything about effective programming, and c) is even more likely, by virtue of the program and evaluation design flaws, to portray the program as a failure.

Perhaps the most important issue concerns how well the program has been planned by the administrators and staff, and how well program staff are then able to articulate various program components to evaluators. As a recent panel on Violence Prevention for Young Adolescents observed, evaluations must be planned at the earliest stages of program design (Wilson-Brewer, Cohen, O'Donnell & Goodman, 1991). Here, the mind set of a good evaluation and the mind set of good program planning are strikingly similar. The program that cannot describe the population it serves, or is not delivering a clear process of treatment, or is not targeting the treatment to a cause of the problem being treated, usually will not succeed. We address these issues in the next section.

### The Planning Process

Much of the evaluators' focus in conducting an EA builds from the program's original plan. To illustrate, Figure 1 represents a planning structure that shows the interrelationship between knowledge of the target population (or target problems), preparation of objectives that fit those problems, and selection of intervention strategies that address the target problems and are also known to be effective. Figure 1 also depicts the important interface between program planning and program evaluation activities. This interface becomes an important issue for the evaluability assessment.

Figure 1



Briefly, program planning begins with a thorough understanding of the target problems and the characteristics of individuals who need to be served in order to address those-problems. Program selection/screening criteria follow from this understanding.

Once program staff have a thorough understanding of the problems they are treating and the responsivity considerations of their clients, they prepare objectives that can reasonably be met within the current program resources. A program for violent offenders, for example, may endeavor to improve anger management skills. There will be intermediate objectives, pertinent to what the program hopes to accomplish prior to release, and long-term objectives, such as post-program reductions in recidivism. Selection of a service delivery strategy follows. An Anger Management Program (Novoco, 1975) would fit the objective mentioned above.

Simply put, there must be a systematic planning process which evidences a logical interrelationship between the components identified above. Services must target risks factors for crime. Services must also fit the objectives of the program. Most importantly, the planning process should be grounded in the literature pertaining to “what works” as well as

in the psychological and criminological theories of what caused the problem behaviors to begin with. Choice of target problem and choice of programmatic service can and should be informed by existing literature on criminology and correctional treatment.

We offer some suggestions for keeping to such a planning model:

1. Provide for interaction between case management and counseling staff, planners, and evaluators as a part of an overall planning/evaluation model.
2. Create a climate where staff work from a base of: a) What is the problem?; b) Who are our clients?; c) What are our objectives?; d). What intervention will achieve these objectives?; and e) How will we know when we have achieved our objectives?
3. Consider evaluation and planning skills in the hiring of staff, even those who are not hired for research purposes. Most undergraduate programs in social sciences require relevant course work. Consider in-service training for current staff who have not been so trained.

### Improving Our Knowledge of Who and What We are Treating

Figure 1 indicates a box for “target problems”. This is where any programmatic intervention begins. Objectives, screening procedures, selection criteria, the choice of interventions and ultimately measures of success all relate back to the problems we are addressing. If we do not have sufficient knowledge of the target population, then, all other program components are adversely affected.

**Assuring that the program is targeting a criminogenic need and translating that criminogenic need into an intermediate objective is one of the most crucial tasks of program planning.** If we have chosen a target problem (criminogenic need) related to crime, substance abuse, sex offending behavior, or aggression, we also have a dynamic risk

factor. If the client shows improvement by the end of the program on the criminogenic need, his or her risk of recidivism declines---there is in other words, an interaction effect between risk (as measured by criminogenic needs) and treatment success on recidivism (Andrews & Bonta, 1994).

In addition to target problems, programs should plan at this point for responsivity considerations. How shall we deal with clients with emotional problems or high anxiety or low intelligence?

In order to fully implement plans for risk, needs, and responsivity, programs will need to make optimal use of systematic assessment and classification procedures. The most common assessment for use in criminal justice settings are risk assessment instruments. In order to maximize a program's potential for ultimately reducing recidivism, programs are advised to select dynamic measures of risk pertinent to the criminogenic needs they target. Alternatively, assessments of specific criminogenic needs would be useful to those programs targeting a specific risk factor, such as criminogenic thinking errors (see for example the Pride in Delinquency Scale, Shields & Whitehall, 1991)(Van Voorhis et al., 1995; Van Voorhis & Brown, forthcoming). Assessments of responsivity include personality-based classification systems (e.g., see Van Voorhis, 1994 for a review), or IQ assessments (e.g., the Shipley Institute of Living Scale, Shipley, 1983). Responsivity, however, is highly dependent upon the program.

Evaluators will incorporate these assessments into the evaluation analyses.

Improvement on criminogenic risk factors or attainment of intermediate objectives are important program outcome considerations. Another important question for the analysis is



the issue of the relationship between the attainment of intermediate objectives and recidivism. Unfortunately, evaluations seldom ask how their program successes do in the long term. Sometimes clients in experimental groups show only a small reduction in recidivism rates over their counterparts in a comparison group, or no differences between the groups. But what about the clients who participated in the full program, especially those whose dynamic risk score were reduced? If the program has been well planned and implemented, these clients should have even lower recidivism rates, and the research shows this to be the case (Andrews & Bonta, 1994).

#### Improving Our Ability to Deliver High Quality Interventions According to their Designs--Treatment Integrity

In providing for treatment integrity, two directions must be pursued at the planning stage of the program: a) there must be a sound reason for selecting the treatment model, and b) the selected model must be administered according to its design. With respect to the first issue, asking the question, Where did the idea for this program model come from?, would be extremely important. With growing knowledge about “what works”, it makes little sense to follow a hunch. Thus, before asking whether a program does work, we must ask whether there is any reason why it should work. Programs grounded in a theoretical model (theory of crime) or found in previous controlled evaluations to be effective really should work, so we at least should start at that point.

To this end, the recent meta-analyses and reviews of the correctional treatment evaluation literature are invaluable. This is especially true since there is now substantial consensus concerning the most successful treatment models for offender populations. As

noted earlier, they are: a) behavioral, b) cognitive behavioral, c) social learning models, and d) multi-modal models (see Andrews, Zinger, Hoge, Bonta, Gendreau, & Cullen, 1990; Gendreau & Ross, 1987; Lipsey, 1991; Palmer, 1992; Whitehead & Lab, 1989). And least successful programs for offenders include: a) in depth therapies (psychoanalysis and person-centered), b) deterrence-based models (e.g., scared straight; boot camps, shock incarceration), c) chemotherapies, and d) group milieu therapies.

**It is also important to note, at this point, that supervision and strict custody are not considered treatment; indeed, supervision and custody, in themselves, do not direct an intervention to a criminogenic need for purposes of changing behavior--no matter how much media or political attention they receive.** This was learned, rather poignantly, two years ago when researchers at the University of Cincinnati conducted reviews of intermediate sanctions, and found that only those with treatment components as well as supervision components showed any appreciable impact on recidivism (Cullen, Wright, & Applegate, 1995).

Even when adopting a sound program design, planners often fail to communicate a clear picture of the clinical or procedural components to the staff who must administer it. Such oversights would include programs that attempt to “provide positive role models” without indicating why and how good role modeling is done; or programs identified simply as “educational programs” or “counseling programs”. Such programs say little about what specific services are being administered and how they are being administered. To address this concern, written treatment manuals, are essential to maintaining treatment integrity and consistency. Good treatment manual are written for the staff and provide “how to”

descriptions of intervention services. In addition, they seek to provide staff with a clear understanding of program policies, services, and how they are to perform those services. Importantly, the treatment manual should be too detailed to serve as a marketing document, or the program brochure.

The fact that programs seldom develop and use treatment manuals may partially implicate the research and scholarly community who write evaluation studies. Seldom are program accounts written in such a manner as to facilitate replication and the development of a treatment manual. The few exceptions include, but certainly are not limited to: Palmer's and Warren's accounts of the Community Treatment Project (see Palmer, forthcoming; Warren, 1983; Vicky Agee's Closed Adolescent Treatment Program (Agee, 1979); Fabiano, Robinson, and Porporino's (1991) work on Cognitive Skills Training Programs; Don Andrews' accounts of effective social learning approaches (Andrews & Bonta, 1994). Arnold Goldstein's treatments for aggression (Goldstein & Keller, 1987); work at the University of Utah on family therapy (Barton, Alexander, Waldron, Turner & Warburton, 1985); Novoco's Anger Control Training (1975); Yochelson and Samenow's Criminal Personality Programs (a cognitive program that targets criminal "thinking errors")(Yochelson & Samenow, 1976). We highly recommend these readings.

A closely related issue concerns whether a well-selected program is operating according to design. Specifically, are all of the clinical dimensions of the intervention present? For example, a routine discussion among family members lacks the clinical dimensions of some of the family therapy models that are known to be effective with some types of offenders (Van Voorhis, 1992). Ongoing clinical observations from treatment

directors, clinical directors, or consultants may be crucial to preventing this problem. If the intervention is not following correct clinical practice, it would be a disservice to say it didn't work when, in fact, it didn't happen. Ultimately, a program's treatment integrity is a matter for the evaluability assessment, but internal monitoring, prior to submitting the program to outside review, would be a sound course of action. In this regard, program management or a program team should examine their: a) level of organization (e.g., is every one administering the same program?), b) consistency in staffs identification of target problems, c) the coherence of the chosen intervention, and d) use of intermediate objectives that link criminogenic needs and services.

To this point we have talked as though the best thing a program could do was to implement a model that is known to be effective with a known cause of the problem. However, other programmatic factors such as level of staff training, number of hours of program participation, attendance, and staff interpersonal skills are equally important (Palmer, 1992). These are addressed in the principles of effective intervention (Andrews & Bonta, 1994; Gendreau, 1995) as well as in the CPAI.

### Measuring Program Effectiveness

We have noted the importance of conceptualizing program effectiveness in terms of intermediate objectives. It follows then that one clear measure of program success is attainment of intermediate objectives. If we develop programs in terms of intermediate objectives that link effective interventions to known causes of crime (criminogenic needs), we are likely to improve programs in the following ways:

1. We give the program its best chance of producing effective evaluation results.
2. We have a program that targets appropriate causes of crime with effective interventions.
3. We measure program success in ways that are optimally relevant to the services we delivered.
4. We measure program effectiveness on criteria that we have the most control over.
5. We target those individual characteristics that, if changed have the greatest chance of reducing recidivism.
6. We use valid measures of success. Measurement problems are more likely with the long-term recidivism measures than with measures relevant to intermediate objectives.
7. The attainment of intermediate objectives impact on risk.

ultimately, then programs should plan to reduce criminogenic needs. Programs for aggressive individuals should be held to the goal of reducing risk factors for violence. Substance abuse programs should be held to the task of reducing the risk factors for drug abuse or alcoholism.

Determining whether or not intermediate objectives have been met requires a program to state such objectives in quantifiable (measurable) terms. The evaluator's or program staff's success in developing an appropriate measure will depend upon the program's statement of a quantifiable objective. Rhetorical, promotional objectives cause problems for measuring, but this is usually not apparent by the time the program is evaluated, because by that time, a vague objective will have already caused a good deal of programmatic confusion as well.

Use of intermediate objectives would clearly exclude the following as criteria for program success: (a) number of clients who completed the program, (b) client surveys of how well they liked the program, and (c) attainments on a characteristic that is not a criminogenic need.

This discussion does not exclude the importance of setting a long-term goal of reducing recidivism. However, it would be highly advantageous to first examine program effectiveness on those changes that the program had the most control over. In addition, attention to intermediate objectives links program success to targets and services in a more rational manner. Finally, intermediate objectives are relevant to recidivism, because criminogenic needs are related to crime.

For purposes of review, the suggestions put forward in this section are reassembled in a checklist format (see Appendix A).

## CONCLUSION

We have attempted, in this monograph, to describe the benefits and strategies of evaluability assessment. It is our hope that we also have encouraged practitioners and policy makers to view evaluability as a tool for facilitating optimal program development. Used wisely, evaluability assessments identify program strengths and weakness and direct further program development energies to areas where they are most needed. The evaluability assessment, however, cannot serve as a substitute for good planning. For that reason, we have devoted considerable attention to presenting the planning process that will result in a program that is ready to be evaluated. Our focus in this section has been to link

program planning to a social science planning and evaluation model that views the most important tasks of planning as: a) targeting criminogenic needs and translating those needs into intermediate objectives, b) selecting interventions with proven effectiveness, c) monitoring the integrity of the intervention to assure that it is being administered according to design, and d) measuring success.

Unfortunately, corrections is often pressured to do otherwise. On a daily basis, atheoretical “magic bullets” are put forward by the media and by political pressures; such pressures drive policy and intervention designs. Almost always the “magic bullet” prescribes an intervention that: a) does not target a cause of crime; b) has already been found to be ineffective (e.g., boot camps without a treatment component); c) ‘has no theoretical reason for working (e.g., yelling at boot camp participants); or d) defies everything that we know about crime and its effective treatment. Most of what we know about effective interventions is not currently in practice. An emerging knowledge base that shows great consistency across authors and studies is virtually being ignored (see Andrews, et al., 1990; Andrews & Bonta, 1994; Gendreau & Ross, 1987; Lipsey, 1992; Palmer, 1992; Van Voorhis, 1987) and too much of what is currently in practice does not emerge from the knowledge base of our field.

The costs, of course, are to the impressions that we create of our own efforts. Many programs appear to be hurting themselves when they plan and conduct evaluations too quickly. This occurs in many ways. First, in agreement with Palmer (1978), we suspect that many evaluations showing no treatment effect, have, instead, masked the treatment effect. That is, many programs have worked with some offenders and not others, but with no way of subdividing the research sample into subgroups that account for different levels of

responsivity, we produce successes which are canceled by our failures. Second, many program evaluations have not fully utilized the results that have maximum chance of showing success (e.g., attainment of intermediate objectives); or they produce evaluation measures which do not fit program services and clients. Third, when program disorganization results in staff working in different directions or when we are uncertain about the characteristics of program clients, we will probably have insignificant results. Statistical tests are designed to reject all but those events which are most likely to be related, and random events are not related! Fourth, many failures have resulted from treatment of factors such as self esteem, which are not necessarily a cause of crime. Finally, well-supported theories of crime and the few extant evaluation studies of effective programs, that can point us to programs that work, are not used as guides for program development. Hopefully, we have shown that all of these unfortunate shortcomings can be addressed during the planning phases of the program.



## NOTES

1. A cursory overview of the correctional treatment research, however, shows clear benefits to grounding correctional prevention or treatment programs in an empirically well-supported theories of crime and delinquency causation (Glaser, 1974; Van Voorhis et al. 1995).
2. On its surface, the evaluability assessment may appear similar to a process evaluation (Rossi & Freeman, 1982), an evaluation of the nature, integrity, and extent of program service delivery models. The evaluability assessment of whether the program has the necessary components of a larger evaluation, however, differentiates it from a process evaluation. At the same time, process evaluations often entail some data collection pertaining to descriptions of the target population served, and level and content of service delivery. EAs need not collect processing data. Similarities between EAs and process evaluations, of course, are recognizable; both typically seek to determine whether the program has a clear design, and is operating according to the design.
3. Unless the programs have contain one of the above treatment components (see Cullen, Wright, & Applegate, 1993).



## REFERENCES

- Agee, V. (1979). Treatment of the Violent, Incurrigible Adolescent. Lexington, MA: Lexington Books.
- Andrews, D. & Bonta, J. (1994). The Psychology of Criminal Conduct. Cincinnati: Anderson Press.
- Andrews, D. & Bonta, J. & Hoge, R. (1990a). Classification for effective rehabilitation: Rediscovering psychology. Criminal Justice and Behavior 17, 19-52.
- Andrews, D., & Keisling, J. (1980). Program structure and effective correctional practices: A summary of the CAVIC research. In Robert Ross & Paul Gendreau (Eds.), Effective Correctional Treatment (pp. 441-463). Toronto: Buttersworth.
- Andrews, D., Zinger, I., Hoge, R., Bonta, J., Gendreau, P., & Cullen, F. (1990). Does correctional treatment work? A psychologically informed meta-analysis. Criminology 28, 369-44.
- Barton, C., Alexander, J., Waldron, H., Turner, D., & Warburton, J. (1985). Generalizing treatment effects of functional family therapy: Three replications. American Journal of Family Therapy 13, 16-26.
- Cullen, F. & Gilbert, IS. (1982). Reaffirming Rehabilitation. Cincinnati, OH: Anderson Press.
- Cullen, F., Wright, J., & Applegate, B. (1993). Control in the community. Presentation at a conference of the International Association of Residential and Community Alternatives, November, Philadelphia, PA.
- DiIulio, J. (1993). Measuring performance when there is no bottom line. In DiIulio (Project Director, BJS-Princeton Project), Performance Measures for the Criminal Justice System. Washington, D.C.: U.S. Department of Justice.
- Fabiano, E., Robinson, D., & Porporino, F. (1991). A Preliminary Assessment of the Cognitive Skills Training Program: A Component of Living Skills Programming: Program Description, Research Findings and Implementation Strategy. Ottawa, Ontario: Correctional Service of Canada.
- Gendreau, P. (1995). The principles of effective intervention with offenders. In Alan Harland (Ed.), Choosing Correctional Options that Work: Defining the Demand and Evaluating the Supply. Thousand Oaks, CA: Sage.
- Gendreau, P. & Andrews, D. (1994). Correctional Program Assessment Inventory. (4th ed.) University of New Brunswick, St. John, New Brunswick.

Gendreau, P., Andrews, D., Coggin, C., & Chanteloupe, F. (1992). The development of clinical and policy guidelines for the prediction of criminal behavior in criminal justice settings. Unpublished manuscript. University of New Brunswick, St. John, New Brunswick.

Gendreau, P., Little, T. & Goggin, C. (1995). A Meta-Analysis of the Predictors of Adult Offender Recidivism: Assessment Guidelines for Classification and Treatment. Ottawa, Ontario: Soliciter General of Canada.

Gendreau, P. & Ross, R. (1987). Revivification of rehabilitation: Evidence from the 1980's. Justice Quarterly 4, 349-408.

General Accounting Office (1977). Finding Out How Programs are Working:: Suggestions for Congressional Oversight. Report to Congress for the Comptroller General of the United States, PAD-78-3, November.

Goldstein, A. & Keller, H. (1987). Aggressive Behavior: Assessment and Intervention. New York: Pergamon Press.

Gottfredson, M. (1979). Treatment destruction techniques. Journal of Research in Crime and Delinquency 16, 39-54.

Lipsey, M. (1984). Is delinquency prevention a cost-effective strategy? A California perspective. Journal of Research in Crime and Delinquency 21, 279-302.

Lipsey, M (1991). Juvenile delinquency treatment: A meta-analytic inquiry into the viability of effects. In T. Cook, H. Cooper, D. Corday, H. Hartment, L. Hedges, R. Light, T. Louis, & F. Mosteller (Eds.), Meta-analysis for exploration: A casebook. New York: Russell Sage Foundation.

Mann, J. (1972). The outcome of evaluation research. In C. Weiss (Ed.), Readings in Social Action and Evaluation. Boston: Allen & Bacon.

Martinson, R. (1974). What works?--Questions and answers about prison reform. The Public Interest 34, 22-54.

Nay and Kay (1982). Government Oversight and Evaluability Assessment. Toronto: Lexington Books.

Novoco, R. (1975). Anger Control: The Development and Evaluation of an Experimental Treatment. Lexington, MA: Lexington Books.

Palmer, T. (1975). Martinson revisited. Journal of Research in Crime and Delinquency 12, 133-152.

Palmer, T. (1978). Correctional Intervention and Research: Current Issues and Future Prospects. Lexington, MA: Lexington Books.

Palmer, T. (1992). The Re-Emergence of Correctional Intervention. Newbury Park, CA: Sage Publications.

Palmer, T. (forthcoming). Individualized Intervention with Young Multiple Offenders: The California Community Treatment Project, Hampton, CT: Garland Press.

Petersilia, J. (1993). Measuring the performance of community corrections. In DiIulio (Project Director, BJS-Princeton Project), Performance Measures for the Criminal Justice System. Washington, D.C. : U.S. Department of Justice.

Rossi, P. & Freeman, H. (1989). Evaluation: A Systematic Approach. Newbury Park, CA: Sage Publications.

Rutman, L. (1980). Planning Useful Evaluations: Evaluability Assessment. Beverly Hills: Sage Publications.

Scanlon, J. W., P. Horst, J. N. Nay, R. E. Schmidt, and J. D. Waller (1979). "Evaluability assessment: Avoiding types III and IV errors." In G. R. Gilbert and P. J. Conklin (eds.), Evaluation Management: A Selection of Readings. Washington; D. C.: Office of Personnel Management, Federal Executive Institute.

Schmidt, R. E., J. W. Scanlon, and J. B. Bell (1979). Evaluability Assessment: Making Public Programs Work Better.

Shields, I. & Whitehall, G. (1991). The Pride in Delinquency Scale: Paper presented at the annual correctional psychologists' winter conference. Burritts Rapids, Ontario.

Shipley, W. (1983). Shipley Institute of Living Scale. Los Angeles: Western Psychological Press.

Smith, M. F. (1989). Evaluability Assessment: A Practical Approach. Boston: Kluwer Academic Publishers.

Smith, M. F. (1990). "Evaluability Assessment: Reflections on the Process." Evaluation and Program Planning. 13:359-364.

Van Voorhis, P. (1987). Correctional effectiveness: The high cost of ignoring success. Federal Probation 5, 56-62.

Van Voorhis, P. (1994). Psychological Classification of the Adult Male Prison Inmate. Albany, NY: SUNY Press.

Van Voorhis, P. (1995). Personality and the crime paradigm: Directions from the classification research. Paper presented at the annual meeting of the American Society of Criminology, November, Boston, MA.

- Van Voorhis, P., Braswell, M., & Morrow, B. (1992). Family therapy. In Correctional Counseling, eds. D. Lester, M. Braswell & P. Van Voorhis, 155-174, Cincinnati, OH: Anderson Press.
- Van Voorhis, P. & Brown, K. (forthcoming). Risk Classification in the 1990s. Washington, DC: National Institute of Corrections.
- Van Voorhis, P., Cullen, F. & Applegate, B. (1995). Evaluating interventions with violent offenders. Federal Probation 50, 17-27.
- Warren, M. (1983). Applications of Interpersonal Maturity Theory to offender populations. In Personality Theory, Moral Development, and Criminal Behavior, W. Laufer & J. Day (Eds.), 23-50. Lexington, MA: Lexington Books.
- Weiss, C. (1972). Evaluation Research. Englewood Cliffs, CA: Prentice Hall.
- Whitehead, J. & Lab, S. (1989). A meta-analysis of juvenile correctional treatment. Journal of Research in Crime and Delinquency 26, 276-295.
- Wholey, J. S. (1979). Evaluation: Promise and Performance. Washington, D. C. : The Urban Institute.
- Wholey, J. S. (1983). Evaluation and Effective Public Management.
- Wholey, J. S., J. N. Nay, J. W. Scanlon, and R. E. Schmidt (1977). Evaluation: When Is It Really Needed? Washington, D. C. : The Urban Institute.
- Wilson-Brewer, R., Cohen, S., O'Donnell, L., & Goodman, F. (1991). Violence prevention for young adolescents: A survey of state of the art. Paper presented by the Educational Development Center, Inc. for the conference, "Violence Prevention for Young Adolescents," Washington, DC, July, by the Carnegie Corporation of New York.

## Appendix A

### Programmatic Checklist for Planning and Evaluation Activities

#### I. Organizational Support for the Evaluation

- A. Does your agency approach evaluation research for purposes of feedback that can lead to program adjustments rather than solely to program dismantlement?
- B. Does your agency evidence a realistic view of the successful rehabilitation program (20-30 percent reduction in recidivism), so that the program will not be held to results that are typically not attained in criminal justice?
- C. Do staff and administrators share responsibility for sound program planning?
- D. Does your agency maintain a planning and evaluation agenda -- a commitment to program testing and improvement.
- E. Do staff and administrators approach their work systematically, with an understanding of the interaction between target problems, objectives, interventions, and measures?
- F. Does your agency consider evaluation and planning skills in the hiring or training of program staff?
- G. Will your agency's administration support the research design of the evaluation?
- H. Are financial resources available and sufficient to conduct an evaluation?
- I. Will it be possible to hold occasional staff meetings pertinent to evaluation issues?

#### II. Understanding of Target Problems and Characteristics of Program Clients

- A. Have you identified the correlates of violence and aggression (dynamic risk factors) that can be addressed by your program?
- B. Are the dynamic risk factors selected by your program reflected in your program's selection criteria?
- C. Are clients assessed at intake for dynamic risk factors that will be addressed by the program?
- D. Are clients assessed at intake for those responsivity factors that are most likely to affect their ability to succeed in the program?

### III. Choosing an Appropriate Intervention

- \_\_\_\_\_ A. Does the planned intervention address the criminogenic needs/dynamic risk factors of your clients?
- \_\_\_\_\_ B. Did the planned intervention originate in a theoretical or empirical understanding of what causes aggression and violence?
- \_\_\_\_\_ C. Has the planned intervention been tested before?

### IV. Ensuring Program Integrity

- \_\_\_\_\_ A. Has an evaluability assessment determined that your program is ready for an evaluation?
- \_\_\_\_\_ B. Have you devised procedures for determining that your program has served the correct target population (see I. C.)?
- \_\_\_\_\_ C. Are written treatment manuals available to assist staff in the consistent delivery of the planned intervention?
- \_\_\_\_\_ D. Have staff been trained in the principles of successful treatment as well as in the clinical dimensions of the planned intervention?
- \_\_\_\_\_ E. Will data on treatment implementation (e.g., attendance lists, case management logs, surveys of clients and staff, observations of treatments) be collected?

### V. Choosing Appropriate Outcome Measures and Research Design

- \_\_\_\_\_ A. Has the program identified intermediate objectives that relate to dynamic risk factors and to the planned intervention (see II. A & III. A., above)?
- \_\_\_\_\_ B. Are the intermediate objectives stated in quantifiable, measurable terms?
- \_\_\_\_\_ C. Does your program have procedures in place for collecting data/measures pertinent to the intermediate objectives?
- \_\_\_\_\_ D. Are the measures of intermediate objectives also relevant to the planned intervention?
- \_\_\_\_\_ E. Are follow-up data on aggression and violence being collected (preferably from a variety of sources -- police records, self-report, and staff evaluations)?



\_\_\_\_\_ F. Has your program agreed to random assignment of subjects to experimental and comparison groups? Alternatively have you identified another type of comparison group that can serve as a basis for comparison with your program clients?

### Choosing an Evaluator

\_\_\_\_\_ A. Is the evaluator familiar with the research on treatment effectiveness?

\_\_\_\_\_ B. Does the evaluator understand that you need to know not just whether the program worked overall, but also what program components might be modified to improve effectiveness?

\_\_\_\_\_ C. Can data be efficiently collected? (e.g., Do data collection procedures fit reasonably into the case management process of your agency, or are they overly cumbersome?)

Source:

Van Voorhis, P., Cullen, F. & Applegate, B. (1995). Evaluating interventions with violent offenders. Federal Probation 50, 17-27.

