

148453



S
T
A
T
E

R
E
P
O
R
T
I
N
G

A
N
D

E
V
A
L
U
A
T
I
O
N

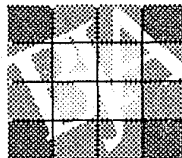
P
R
O
G
R
A
M

DEVELOPING PERFORMANCE MEASURES FOR CRIMINAL JUSTICE PROGRAMS

February 1994

148453

Assessment and Evaluation
Handbook Series No. 2



Publication Funded by
Bureau of Justice
Assistance

About the Bureau of Justice Assistance

The Bureau of Justice Assistance administers the Edward Byrne Memorial State and Local Law Enforcement Assistance Program to support drug control and system improvement efforts focused on state and local criminal justice systems. The Bureau's mission, directed by the Anti-Drug Abuse Act of 1988, is to provide funding and technical assistance to state and local units of government to combat crime and drug abuse. Through funding and technical support, the Bureau assists the states in managing the growing numbers of anti-drug programs and the rapidly increasing volume of drug cases entering the criminal justice system. It also identifies, develops, and shares programs, techniques, and information with the states to increase the efficiency of the criminal justice system, as well as provides training and technical assistance to enhance the expertise of criminal justice personnel. The Bureau accomplishes these mandates by funding innovative demonstration programs, some of which are national or multi jurisdictional in scope; by evaluating programs to determine what works in drug control and system improvement; and by encouraging the replication of successful models through linkages with the Formula Grant Program and other resources.

The Director of the Bureau is appointed by the President and, upon confirmation by the Senate, serves at the President's pleasure. The Director establishes the priorities and objectives of the Bureau and has final authority to approve grants, contracts, and cooperative agreements. In establishing its annual program, the Bureau is guided by the priorities of the Attorney General, U.S. Department of Justice, Office of Justice Programs, and the needs of the criminal justice community.

148453

**U.S. Department of Justice
National Institute of Justice**

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Justice.

Permission to reproduce this ~~document~~ material has been granted by

Public Domain/OJP/BJA

U.S. Department of Justice

to the National Criminal Justice Reference Service (NCJRS).

Further reproduction outside of the NCJRS system requires permission of the ~~document~~ owner.

This report was prepared under cooperative agreement #92-DD-CX-K026, provided by the Bureau of Justice Assistance, Office of Justice Programs, U.S. Department of Justice. The points of view or opinions stated in this document do not represent the official position or policies of the U.S. Department of Justice.

The Bureau of Justice Assistance is a component of the Office of Justice Programs which also includes the Bureau of Justice Statistics, the National Institute of Justice, the Office of Juvenile Justice and Delinquency Prevention, and the Office for Victims of Crime.

**DEVELOPING PERFORMANCE
MEASURES
FOR
CRIMINAL JUSTICE PROGRAMS**

**Assessment and Evaluation
Handbook Series No. 2**

NCJRS

OCT 4 1994

ACQUISITIONS

**Prepared by
Jerry M. Hatfield**

Development and Publication Funded by the
U.S. Department of Justice, Bureau of Justice Assistance
State Reporting and Evaluation Program

February, 1994

Table of Contents

Preface	1
I. Introduction: Evaluation in Perspective	2
A. Designing For Measurement	4
II. Comparison With "Assessing The Effectiveness of Criminal Justice Programs", Assessment and Evaluation Handbook No. 1	9
III. Establishing Performance Indicators	10
A. Principles	10
B. Maximizing Measurability	11
C. The Process	12
D. Assignment of Values	14
E. Assignment of Weights	15
IV. Performance Measurement	16
A. Frequency of Measurement	16
B. Techniques of Measurement	17
V. Performance Analysis	19
A. Programmatic Attainment	19
B. Indicator Validity and Reliability	20
VI. Summary of Benefits	21
Appendixes	22
1. QPA Rating Chart (Blank)	23
2. QPA Rating Chart (Completed)	24
3. Sample Substance Abuse Treatment Plan Using QPA Format	25
4. Formulae for QPA Calculations	26
5. Sample Substance Abuse Treatment Plan - Completed and Scored	28
Explanation	29

Preface

This Handbook was prepared as a companion to "Assessing The Effectiveness of Criminal Justice Programs", Assessment and Evaluation Handbook Series No. 1. Handbook No. 2 is designed to continue in the direction of providing a relatively simple format for program design which will allow for program evaluation. It is the assumption of this Handbook that clear and specific program design statements and descriptions will allow for accurate and efficient program evaluation efforts.

I. Introduction: Evaluation in Perspective

Evaluations of criminal justice programs vary from broadly descriptive to specific. Some practitioners perform process evaluations, other focus on outcome or impact evaluations, and others perform both. There are no right or wrong approaches, no right or wrong answers, and no choice is intrinsically better than the other. One's approach to evaluation is dependent on a number of variables, as described below:

A. Adequate Financial Resources

Those who philosophically commit to evaluation must also commit financially. The size of the evaluation budget may reflect simply how much work may be underwritten, or it might be a more subtle reflection of priorities. Unless the sponsoring agency commits adequate financial resources, evaluations will not achieve useful results.

Assuming limited evaluation budgets, it is unwise to commit to performing a large number of comprehensive process and outcome evaluations. There won't be enough resources, time, or an adequate product, and the result will be the perception that "evaluation just isn't worth it", when in fact the resources were spread so thin as to render the results useless.

It is probably more advisable to focus limited resources on either one program or project, or alternatively to prepare each of many programs for evaluation by clarifying and specifying each program's goals, objectives, and performance indicators. Using this broad approach, each program may be "evaluated", albeit even in an informal manner, and if additional resources become available, one or more may be examined more thoroughly. By following the format provided in this Handbook, the practitioner may apply basic techniques to each of many different types of programs to prepare them for future evaluation.

B. Purpose of the Evaluation

Criminal justice program evaluation may have many and varied purposes. An evaluation may be used to test the viability of a unique and innovative program, it may be used to determine the advisability of continuing a program, or it could be used as a marketing tool to gain support for continuing a program. Since all professionally performed evaluations must be viewed as unbiased analyses without preconceived outcomes, those who manage and commission evaluations must be prepared for answers which may be contrary to their intuitive notions. A "popular" program may have some weaknesses or not be meeting its primary goals. It must be clear from the beginning that one function of an evaluation is to give program managers insight into how the program may be modified to work more effectively.

C. Type of Program

The type and nature of any given program may have some influence on the decision to evaluate it. Some programs seem to naturally lend themselves to what we think of as an "easy"

evaluation - that is, some programs have easily quantifiable goals and therefore are easier to measure. It is relatively simple to measure the amount of narcotics diverted or a number of arrests. Alternatively, there are a number of other programs with "softer" goals which appear more difficult to measure. Among these programs are treatment and rehabilitation programs, domestic and family violence programs, and prevention programs. The challenge to adequately evaluate these types of programs becomes greater now because of changing federal directions and priorities. It is important to examine these programs even when it appears that the data are "soft" and the technology eludes us. One of the purposes of this Handbook is to provide a method for more clearly describing such programs and quantifying their goals so that they may be evaluated.

A. Designing For Measurement: A General Model For Criminal Justice Program Design

1. Introduction

It is generally acknowledged that the most accurate programmatic assessments may be accomplished when those assessments are based on carefully designed programs. It is the attempt of this Handbook to approach this topic by beginning with the most global perspective and ultimately narrowing it down to the most specific. This approach will provide for some uniformity and universality of not only goals, but also program structure.

This section of the Handbook suggests a generic framework into which almost any categorical program may fit. It does not suggest specific methods of goal achievement or particular objectives, and is not intended to be in any way exclusive of local creativity or initiatives.

What it does do is suggest a logical format which will ultimately allow management at all levels to assess programs more easily and uniformly than before, while at the same time allowing for comparisons across programs and program categories. This method allows for the assessment of:

1. A single program
2. A group of similar programs
3. Similar missions/purposes across different programs.

2. The Design Process

This Handbook will describe a number of steps designed to allow the program designer to move from the broadest and most universal criminal justice goals to the most narrow performance indicators. In this way, each program's goals, objectives and performance indicators will relate in some way to federal priorities, statewide missions and purposes, and ultimately to other local criminal justice programs. This process begins with a review of the Anti-Drug Abuse Act of 1988 (ADA).

One of the functions of the Department of Justice's Bureau of Justice Assistance (BJA) is to distribute to the states Byrne Program block grant funds which have been appropriated by Congress under the Anti-Drug Abuse Act of 1988 (ADA). The Act offers, in a very broad way, some guidance in the ultimate purposes of the federal assistance. (The Act also offers twenty-one Authorized Purpose Areas in which programs may be funded, but offers the states wide latitude in purpose area selection, based on the state's strategy.)

STEP 1: Determine the broad purposes of the ADA.

A review and broad interpretation of the ADA reveals that the following purposes emerge from the ADA. (Other reviewers may perceive other purposes, and those could be included here without impacting on the process itself.)

1. Systems Improvement
2. Increased Coordination

STEP 2: Determine the purposes of your State Criminal Justice Authority.

State Administrative Agencies (SAAs) are generally guided by a Policy Board, the state legislature, or some other such body. Through its strategy development process, the SAA puts forth its broad missions and usually describes categorically how Byrne block grant funds are to be employed. The missions of the SAA may be added to the above purposes described in Step 1, and together will serve as the baseline purposes against which all programs will be compared and measured.

STEP 3: Determine how subgrant goals will compare to Step 1 and 2 purposes.

In this step, each subgrant's goals are reviewed to determine how they compare to the broad missions and purposes determined in Steps 1 and 2. These goals are then inserted in the appropriate cell of the matrix described below.

STEP 4: Complete the program comparison matrix.

In the following matrix, missions and purposes are listed across the top. In this example, the two purposes from the ADA are listed, and two other blocks are provided for other missions or purposes determined by the SAA.

The column on the far left of the matrix may be used in one of two ways. It may be used to list the categories funded by the SAA or it may be used to describe categories and specific programs/projects funded. In the matrix cells will be listed each program's goals. In this way, if all the goals are combined from left to right, this list will represent all the goals of a particular category. If the goals are combined from top to bottom, those goals will represent a summary of goals related to a mission or purpose. By using this matrix as a guide to assessment or evaluation, either approach may be used - the assessment of progress of a category or the assessment of progress of a mission or purpose.

The value in using this method is that assessment and evaluation become more orderly and comprehensive, and program designers and managers will have a more clear view of how each program relates to the stated missions and purposes of both the federal legislation and the leadership of the state's criminal justice system.

Design For Measurement Matrix

	Systems Improvement	Increased Coordination	Crime Reduction	Other
<u>Police</u> 1. 2. 3. 4. 5.	Police program goals which relate to Systems Improvement	Police program goals which relate to Increased Coordination	Police program goals which relate to Crime Reduction	
<u>Courts</u> 1. 2. 3. 4. 5.	Court program goals which relate to Systems Improvement	Court program goals which relate to Increased Coordination	Court program goals which relate to Crime Reduction	
<u>Corrections</u> 1. 2. 3. 4. 5.	Corrections program goals which relate to Systems Improvement	Corrections program goals which relate to Increased Coordination	Corrections program goals which relate to Crime Reduction	

3. Clarification of Missions and Purposes

The process suggested here is one of categorization and clarification. The previous matrix describes an approach to categorization of missions, purposes, and program goals. The process of determining where in the matrix each program goal will fit requires some clarification of these missions, purposes, and goals. The following guidance offers some assistance in this process of clarification, and is important in the development of measurable goals, objectives and performance indicators. It is a process which may be applied to review a program's goals, or to programs applying for funding.

A. Systems Improvement

Define the "system" to be improved - is it a segment (Judicial) or segments (Judicial and Treatment) of the system, or is it the entire criminal justice system?

Address professional networking - similar professionals improving the system by sharing tasks, etc.

Address the larger system - Executive, Judicial, and Legislative branches

Address policies and protocols of agencies - need for revision, etc.

Education, in-service training for systems improvement

How will improvement be evidenced?

When will improvement be evidenced?

Who will effect the improvements?

B. Increased Coordination

With whom will coordination increase?

Within an organization

With all components of the CJS

With components outside the CJS

With national efforts

With agencies being impacted

Who will effect the increased coordination?

What is it that will be increasingly coordinated?

Information

Money

Other resources - personnel, training, etc.

When will the coordination occur?

Why is increased coordination important?

Is there an untoward effect anticipated?

How, specifically, will coordination increase, and how will you know when it has happened?

C. Crime Reduction

As with most missions/purposes, while the measurement of crime reduction may be quite easy, attributing that change to a source or cause is clearly more difficult. It is critically important that the following issues be considered:

- Current trends on incidence and prevalence
- Current awareness
- Awareness of public safety
- Satisfaction surveys - general population and victims
- Recidivism

4. The Establishment of Goals Within the Context of the Criminal Justice System

A program planner's perception of what is desirable or realistic in the CJS may not necessarily be that of those who administer and control that segment of the system. For example, establishing a substance abuse treatment program inside a prison may be a virtuous idea, but one which may not fit (or actually conflict with) the ultimate goals or policies of those who administer the prison. Or alternatively, even if all policy makers agree that it is a virtuous idea, the prison system may not be environmentally ready for such a move.

Virtuous ideas may not be realistic ideas. It is always important to compare a program's goals with those of the segment of the CJS in which the program will function. If this examination is done, then the probability or potential for a program's success is increased if the sponsoring system sees the program as integral to its goals. This determination is accomplished through consensus program development and thorough examination of an organization's existing missions, policies, and goals.

II. Comparison with "Assessing the Effectiveness of Criminal Justice Programs" Model (Handbook Series No. 1)

The above publication sets forth a process of preparing and describing a program's activities which allows for future evaluation. The first three steps described in that Handbook are as follows:

1. Establish Clear Goals
2. Establish Clear Objectives
3. Describe Program Activities/Strategies

This Handbook (No. 2) assumes that those activities have been adequately completed and describes in greater detail the next three steps in the process:

4. Establish Performance Indicators
5. Measure Performance Results
6. Analyze Performance Results

The next three sections of this Handbook will describe in detail those three steps.

III. Establishing Performance Indicators

A. Principles

As described in the Handbook Series No. 1, a performance indicator is described as "an explicit measure of effects or results expected. It tells to what extent an activity has been successful in achieving, or contributing to, an objective." As program planners and designers move from mission statements to goals to objectives, the levels of specificity become greater. When performance indicators are written, they represent the final and greatest level of specificity. If goals represent philosophy and objectives represent actions, then performance indicators represent anticipated results. The following principles apply to the development of performance indicators:

1. Indicators must follow from, and be directly related to, objectives. Therefore, for each objective, a number of indicators must be written that describe what we anticipate or plan to be results of the actions described in the objective.
2. Indicators must be specific and clear enough to allow for measurement by someone not intimately involved in the development or management of the actual program. They must also be reasonably attainable, given the design of the program and whatever constraints may exist.
3. Indicators may describe not only an exact result expected, but may also describe degrees or gradations of achievement, and thus may be measured incrementally.
4. Indicators will describe each activity of the program, but some may be more important than others. Relative weights may be assigned to various indicators to adjust for this.
5. Objectives may be seen as the daily activities of those involved in the program, and indicators may be seen as what was accomplished at the end of the day, week, month, or year. Objectives will answer the question "How did you spend your time?" and indicators will answer the questions "What was actually accomplished, how well, and how often?"
6. Perhaps the most important principle of performance indicator development is the necessity for group consensus. There are a number of benefits to achieving group consensus in indicator development, but first it is important to define the group.

The group which should achieve consensus should include the funding agency (usually the SAA) person assigned to monitor or oversee the program. If an evaluator is involved in the process, s/he should also be included in the group. And finally, the project manager must be involved. Optionally, others may be invited to participate in the process, including people who

have operated similar programs, even in other jurisdictions. Others who operated programs with similar populations, even if not using the same methodologies, might be included to add a different perspective. Additionally, other evaluators from other agencies or those who have examined other criminal justice programs might be involved. The minimum size is three, while the maximum is probably around six or seven. Too few people will result in little useful consensus, while too many will make the process unnecessarily cumbersome.

Another variation in the makeup of the consensus group is to include a representative of a client group. This approach would be restricted to those programs which intend to provide some type of service to a client group. Although this approach might present some coordination difficulties and philosophical divergence from standard criminal justice programs, it would probably return increased validity of measurement.

The major advantage in using a consensus group to establish performance indicators is the breadth of perspectives brought to the table. Clearly, using people from the SAA and the program itself will broaden each person's perspective and knowledge, and make the final performance indicators realistically attainable. The agreement fostered through this process also helps to make the final assessment defensible. If broad consensus has been achieved, then each representative feels greater responsibility and ownership, and assessment becomes itself a cooperative and inquisitive venture, resulting in the diminution of peoples' and agencies' defensiveness. Those who participate in the process will be less threatened by the results.

Another advantage in consensus building is a step toward true systems development. As criminal justice administrative managers and program managers work more closely, then there is the opportunity for the criminal justice system as a whole to become more coordinated and for critical relationships to build a foundation for future cooperative work.

Yet another advantage is in program planning, advocacy, and marketing. Often the people who can best assist program planners are those who manage actual programs. If true group consensus is achieved to the degree that the two parties work together, programs can be more easily improved and marketed as promising, effective, or as models. This type of marketing can prove useful in gathering support for budgetary proposals at the local, state, or federal levels.

B. Maximizing Measurability

Specificity and clarity are the keys to maximizing the measurability of performance indicators. If all the following standard questions are clearly answered in the performance indicators, then the program will have increased measurability.

1. Who will be responsible for the performance? This question will focus on which staff will perform which functions. An organizational chart will help to illustrate who will supervise the staff or activity. This approach not only provides for greater performance, but also provides staff with greater job definition and accountability.

2. What exactly will be attempted? A good test for this question is whether or not the indicator is understandable by someone not familiar with the program. It should be clear enough for someone to perform the task will little or no further information.

3. When, or over what period of time, will an action take place? A graphic illustration of a milestone chart will help staff to visualize when various actions will occur, and will serve as a useful tool in program accountability. Also, it is important to note how much time is allocated to a particular activity. This allows for some measure of relative efficiency and could lead to cost-effectiveness measures.

4. How, or by what methods, will an activity occur? There are many methods which can be used to achieve the same results. Which methods will be employed?

5. Where will an activity occur?

6. Why will an action occur? Although this question will do little to increase measurability, its inclusion helps to further define a program. If an indicator statement ends "in order to..." it offers staff a reason to do something and adds some degree of clarity.

C. The Process

Quantified Program Assessment

Quantified Program Assessment (QPA) is a highly mechanized method of describing and measuring various performance indicators. QPA draws upon a method known as Goal Attainment Scaling developed at the University of Minnesota and used in the mental health field. The adaptation known as QPA was developed by Systems Development Associates and further refines the method for application to the criminal justice field.

The components of QPA include

1. The development of Primary and Secondary Performance Indicators:
2. The assignment of numerical values on a five-point scale:
3. The assignment of weights for indicators to reflect their relative importance:
4. The calculation of goal attainment scores.

These components of QPA are integrated into the more generic subject of performance indicator development in this Handbook in order to more completely describe the process.

Primary Performance Indicators

As earlier described, a performance indicator is the final level of specificity to be described. However, in order to provide for the measurement of gradations of achievement, this Primary Performance Indicator (PPI) can be further divided into Secondary Performance Indicators (SPI).

The PPI will describe an anticipated result in clear and measurable terms. For example, if a substance abuse treatment program had as one of its objectives to interview and screen potential clients being released on parole, the PPI could be stated as follows:

"Interview and assess all parolees released from Kent County for determination of admission eligibility."

In this example, the key words are "interview", "assess", and "all". All of those key words qualify and specifically describe the event, and if left as the only performance indicator, could easily be measured. However, left as it is, either the program will interview "all" potential candidates or not, producing a purely dichotomous result. It becomes important to further define this PPI because even if some candidates were interviewed, then obviously some activity did occur. Similarly, if all candidates were interviewed but not all were assessed, some recognition should be given to at least partial attainment of that performance indicator.

One alternative would be to split that PPI into three others, one describing just the interview, another describing just the assessment, and another using the word "most" instead of "all". A more clear alternative is to write Secondary Performance Indicators which further define the Primary Performance Indicator.

Secondary Performance Indicators

The Primary Performance Indicator will describe the "Expected Level of Outcome." This is what the consensus group has decided would be normally expected. It is what you expect the program to accomplish. In order to account for variations on that outcome, two measures will be written which describe "Somewhat Less Than The Expected Level of Outcome" and "Much Less Than The Expected Level of Outcome". Similarly, two more measures will be written to describe "Somewhat More Than The Expected Level of Outcome" and "Much More Than The Expected Level of Outcome". These four new performance indicators are known as Secondary Performance Indicators (SPIs).

To continue with the previous example, it has now been rewritten to appear in the following format, including four SPIs and one PPI.

SPI (Much More ...): Interview all parolees released from Kent County to determine admission eligibility.

- SPI (Somewhat More ...): Interview 75% of parolees released from Kent County to determine admission eligibility.
- PPI: Interview and assess 50% of parolees released from Kent County to determine admission eligibility.
- SPI (Somewhat Less ...): Interview and assess 25% of parolees released from Kent County to determine admission eligibility.
- SPI (Much Less ...): Interview and assess less than 25% of parolees released from Kent County to determine admission eligibility.

If it seemed important, variations could also be written which would provide for interviews only, assessments only, or some combination of the two.

The process of further defining a performance indicator has now been accomplished by describing gradations of achievement. This is done with the assumption that partial under- or over-achievement will always occur and should be measured as part of program evaluation.

D. Assignment of Values: Use of the Scale

As previously mentioned, the gradations of achievement described above are also assigned numerical values on a five-point scale. These values become important during statistical calculations of overall goal attainment.

<u>For This Performance Indicator</u>	<u>This Value Is Assigned</u>
Much More Than The Expected Level of Outcome	+2
Somewhat More Than The Expected Level	+1
Expected Level of Outcome	0
Somewhat Less Than The Expected Level	-1
Much Less Than The Expected Level	-2

Using this scale, and assuming all performance indicators are weighted equally, if the program performed at exactly the "Expected Level of Outcome", the total score (using the

simplified formula described in Appendix 4) would be "0". However, the ultimate score attained by the program has little intrinsic value by itself. Its value lies in an examination of how the score was attained. This will be described later in the "Performance Analysis" section.

A sample indicator form appears in Appendix 1.

E. Assignment of Weights

As previously mentioned, some indicators of a program's accomplishments may be more important than others. For example, in a substance abuse treatment program, (See Appendix 3) a client's participation in a job training program may not be as immediately important as reducing or eliminating drug and alcohol usage. If this were to be decided in a collaborative manner, reduction or elimination of substance use would be weighted more heavily than participation in job training. The determination of how much more or less important one indicator is than another is a relatively subjective decision, and provides another reason for collaborative indicator development.

The assignment of exact weights is done in a manner in which one weight relates to another. The "middle ground" weight may be determined to have a value of 10. If another indicator were twice as important, it would be weighted with a value of 20. Similarly, if another indicator were weighted as half as important, it would be assigned a weight of 5. As a general rule, weights should vary in a range of 5 to 20, 10 to 40, etc.

Relative levels of importance can be established using a simplified method of weight determination which established a median weight of 10, with two other weights of 5 and 20, with 5 representing "half as important" and 20 representing "twice as important". Alternatively, one could establish the range of 10 - 20 - 40 based on the same principle.

The exact numerical value of the weights is not critically important. More important is how the weights relate to each other. Whatever value is determined for the varying weights will be automatically considered when score calculations are performed. (See next section on Performance Measurement)

The assignment of weights is optional. Although using this step in the process does provide for greater accuracy, earlier research indicates that equal weighting will lose little information. If time and resources are minimal, it is clearly more important to focus the evaluative energy on establishing clear goals, objectives and performance indicators, rather than the somewhat more complex process of assigning weights.

IV. Performance Measurement

A. Frequency of Measurement

One of evaluation's most useful features is its ability to provide program management and funding sources with current information regarding the operation of a program. Much evaluation has been criticized for producing results which are not timely, and therefore render diminished utility. The methods described in this Handbook provide for frequent and efficient measurement, and therefore provide all those interested with timely and useful results.

It remains true that a final evaluation of a program, based on observations over time will continue to produce the most valid and reliable analysis, and the methods described in this Handbook will contribute to that type of assessment. But it is also true that program managers and funding agencies have legitimate and more immediate needs for timely information which describes the ongoing progress and achievements of a program. Importantly, the two needs must not be confused. It is quite possible that short-term results may not be a predictor of long-term results. Experience tells us that many new programs will undergo a developmental process which in its early stages may not be exemplary of its ultimate long-term achievement. Therefore, whatever short-term outcomes are measured in a given program must be viewed within this context, and be used for the purposes of fine-tuning and program modification, not as a final judgement of its worth.

The first assessment performed based on performance indicator achievement should be done approximately one month after the program is fully operational, meaning when all staff are hired, trained and working toward the program's goals. This one-month assessment is performed to determine the usefulness and accuracy of the indicator statements. It will probably not take longer than one month of operations to assess this usefulness. The first month of operations is a critical phase, one during which most program staff and management begin to see their functions more realistically than could have been seen during the planning phase. It will not be unusual to make modifications to the indicator statements during this phase. The consensus group input is important at this point in order to provide balance. There will be some indicators which are simply not negotiable, and should not be altered, regardless of their potential for achievement. There will be others, however, which may be modified as a result of the one-month adjustment period. At the end of this first assessment, the indicator statements should generally be fixed with little, if any, further modification.

The next and following assessments should occur quarterly. This three-month period will provide for somewhat more valid and slightly more reliable results which will allow management to broadly predict the level at which the program will ultimately function. It is worth restating here that these assessment techniques are to be viewed with an objective inquisitiveness, a process designed to allow for modification. It may be that indicator achievement is not possible with the resources initially planned for the program. This analysis may not impact on the ultimate worth of the program, but rather points to the need to add resources in an attempt to

provide for the realization of the goals. Any evaluator must be cautious to avoid judgements which are too quick or too critical. Ongoing performance and long-term data will always provide for the most accurate assessments.

An advantage in performing frequent assessments is that they provide management with the opportunity to observe exactly what allows for and prevents indicator achievement. This form of evaluation, known as "Path Analysis" offers information which explains how and why events occur, and offers some degree of predictability of events, based on the program's actions. This method should also be planned as a part of the program's evaluation, and requires a rather sophisticated statistical model to calculate. It is essentially a "road map" of the program's progress and actions which notes and measures critical events and decision-making points, and is useful in planning future similar programs.

An annual assessment may occur after the program has been in full operation for one year. This assessment will obviously provide more reliable data than those previously done, but in the case of a program with multi-year goals may not represent a "final" evaluation. As periodic assessments are performed, a pattern of achievement scores may emerge, and will be useful in making overall observations of progress.

B. Techniques of Measurement

If a program's goals, objectives and performance indicators have been carefully constructed, then measurement will be a relatively mechanical process. However, because there is no way to completely eliminate subjective judgement from any decision-making process, measurement should rely again on the consensus group. Regardless of the clarity of a performance indicator, there may be varying interpretations of the degree of achievement. This variation may be minimized by careful wording of the performance indicators, but will probably never completely eliminate varying opinions.

Most performance results will necessarily be based on data provided by the program staff. These results are then applied to the primary and secondary performance indicators. The consensus group will decide, based on information provided by the program itself, which performance indicator most accurately describes the level of achievement. The numerical value of that performance indicator then becomes the score attained, and that score will be factored into the formula which calculates the program's overall score.

There are three alternative scoring mechanisms which may be used. The first and simplest method is to combine all numerical scores attained, and divide that number by the total number of scales. The result is simply an average score and will fall somewhere between "-2" and "+2", with "0" representing the "Expected Level of Outcome". In using this particular method it is important to remember again that any outcome score has little intrinsic worth. Its real worth lies in its comparability. These scores may be compared with previous scores or with other similar programs with similar indicators. The formula for this calculation appears in Appendix 4.

The second method for score calculation is somewhat more complex, but provides for a more accurate measurement of the relationship between indicators. Typically, the score attained measuring just one indicator will be somewhat different than if that indicator is measured with many other indicators. In other words, working toward the attainment of indicator #1 will usually have some impact on the attainment of indicator #4 and will change the results. This phenomenon is accounted for in the second form of measurement. This formula appears in Appendix 4 with definitions of where scores are to be inserted in the formula. This method is to be used for unweighted/equally weighted scales.

The third method uses the same formula as the second method, but differs in that varying numerical values representing weighted scales are inserted in the formula. This formula also appears in Appendix 4. A sample substance abuse treatment plan, weighted and scored, also appears in Appendix 5 with an explanation of scoring.

V. Performance Analysis and Interpretation

A. Programmatic Attainment

As described earlier, for each goal in a program there will be numerous objectives, and for each objective there will be numerous performance indicators. Using the methods described in this Handbook, the resulting scores allow for a variety of analyses and interpretations.

1. The first type of assessment to be made is the relative score attained for each indicator. A comparison of indicator scores will assist in determining which elements of a particular objective were higher (and therefore more "successful") than the others. If indicators were also weighted it is possible to reconsider the pursuit of some indicators. Perhaps a low weight combined with a low attainment score could point to the relative inefficiency of pursuing that indicator. Alternatively, consistently high scale scores across a number of clients (as in the case of the Treatment Plan example) might indicate that it is unnecessary to invest time in pursuit of an indicator which may not require any investment of time.

It is again important to note that the worth of attainment scores lies in their comparability. A particular score will yield little intrinsic worth, but rather provides the opportunity to compare that score with others in the same program.

2. The second type of assessment to be made is in combining all the attainment scores for all the performance indicators under a particular objective. This will allow for the comparison of one objective with another which may provide some measure of how "successful" various objectives were. Again, the score itself is not to be confused with a final judgement. It is much more important to use the scores to examine the underlying reasons for the scores. Numerous possibilities exist in this examination. Perhaps the resources were inadequate, perhaps the objective was simply too ambitious, etc. It is this type of discussion, brought forth by the scores, which will be valuable in future program modification or design.

3. The third level of assessment is similar to the second, except that it focuses on combining the objectives' scores which exist under their common program goal. As above, the same value exists in comparing one goal with another.

4. The final level of assessment considers all scores for all indicators, objectives, and goals of the program and results in a total programmatic score. This score, because it is so broad and complex, will probably render the least useful score, except in its comparison with other similar programs. Critical to this comparison, however, is that if similar programs are to be compared, they must be constructed using similar goals, objectives, and performance indicators.

B. Indicator Validity and Reliability

Two important statistical measures of accuracy are validity and reliability. The reliability of a measure refers to the degree to which a measure (in this case, a performance indicator) can be trusted to produce consistent results during repeated application. Validity refers to the degree to which a procedure actually measures what it purports to measure.

The system of measuring goal attainment described in this Handbook has not been subject to the necessary examination which would produce measures of validity and reliability in the study of criminal justice programs. While validity and reliability studies have shown acceptable results in other programmatic applications, it is uncertain how these applications would compare to criminal justice programs with different indicators.

Additional research is necessary to establish the validity and reliability of the application of this system to criminal justice programs. Until that research is performed, the major value of using this system of evaluation lies in the clarity and specificity of goals, objectives, and performance indicators, and in the comparability among indicators, objectives, goals, and programs.

Another direction of future research is the development of an inventory of performance indicators for criminal justice programs. This inventory could be based on 1) a determination of those types of programs which seem most challenging to evaluate; 2) those which receive large amounts of funding; or 3) those which are identified as pilot or experimental. This inventory would be useful in helping to standardize criminal justice program goals, objectives, and performance indicators and would provide the basis for future validity and reliability studies.

VI. Summary of benefits of this process

1. *Increased Precision of Measurement*
2. *Low Maintenance*
3. *Integrates With Established BJA Methodology*
4. *Group Consensus Promotes Non-Threatening Results*
5. *Strengthens Advocacy and Marketing Positions*
6. *Applies Across All Programmatic Lines*
7. *Strengths and Deficiencies Quickly Identified*
8. *Applies to Both Process and Outcome Evaluations*

Attribution

The author wishes to acknowledge the early research of Kiresuk and Sherman, whose work in 1968 was the basis for the principles upon which Quantified Program Assessment is built. The citation for that work is as follows:

Kiresuk, T.J. & Sherman, R.E. Goal Attainment Scaling: A General Method for Evaluating Comprehensive Community Mental Health Programs, Community Mental Health Journal, 1968, 4, 443-453.

APPENDIXES

1. Quantified Program Assessment Rating Chart
(Blank Form)
2. Quantified Program Assessment Rating Chart
(Completed for Sample Drug Treatment Program)
3. Sample Substance Abuse Treatment Plan
Using QPA Format
4. Formulae for QPA Scoring/Calculations
5. Sample Substance Abuse Treatment Plan,
Weighted and Scored, with Explanation

**Quantified Program Assessment
Rating Chart**

Program Title:

Authorized Program Area:

Program Goal:

Program Objective:

(Weight = ____)

**Much more than the
expected level of outcome.
(Score: +2)**

Performance Indicator:

**Somewhat more than the
expected level of outcome.
(Score: +1)**

Performance Indicator:

**Expected level of outcome.
(Score: 0)**

Primary Performance Indicator:

**Somewhat less than the
expected level of outcome.
(Score: -1)**

Performance Indicator:

**Much less than the expected
level of outcome.
(Score: -2)**

Performance Indicator:

**Quantified Program Assessment
Rating Chart**

Program Title: Substance Abuse Treatment for Parolees

Authorized Program Area: 13 - Identify and Meet Treatment Needs

Program Goal: 1 - To provide structured participatory counseling to parolees focused on improving each of six life dimensions.

Program Objective: 1.A - Treat parolee to achieve drug-free status

(Weight = ____)

<p>Much more than the expected level of outcome. (Score: +2)</p>	<p>Performance Indicator:</p> <p align="center">Attends 90 AA/NA meetings in 90 days.</p>
<p>Somewhat more than the expected level of outcome. (Score: +1)</p>	<p>Performance Indicator:</p> <p align="center">Attends AA/NA at least once a week, has a sponsor</p>
<p>Expected level of outcome. (Score: 0)</p>	<p><u>Primary Performance Indicator:</u></p> <p align="center">Provides consistently clean urines, no reported abuse and keeps each treatment appointment</p>
<p>Somewhat less than the expected level of outcome. (Score: -1)</p>	<p>Performance Indicator:</p> <p align="center">Submits dirty urines, reduces level of abuse, some missed appointments</p>
<p>Much less than the expected level of outcome. (Score: -2)</p>	<p>Performance Indicator:</p> <p align="center">Submits consistently dirty urines, returns to prior level of substance abuse, and/or overdoses</p>

SAMPLE SUBSTANCE ABUSE TREATMENT PLAN USING QPA FORMAT

	1. Residential	2. Health	3. Employment/Educ.	4. Relationships	5. Substance Abuse	6. Violence
(+2) Much More Than Expected Level	Owens home	No hospitalizations, illness or accidents	Enrolled in school	Married or cohabitating	Attends AA/NA, 90/90	In counseling for anger, control
(+1) Somewhat More Than Expected Level	Has apartment with lease	Has had complete physical exam	Enrolled in job training program	Monogamous relationship, frequent contacts with friends	Attends AA/NA weekly or more, has sponsor	Fewer arguments, reduced frequency of anger, control
(0) Expected Level of Outcome	Has permanent address and phone number	HIV negative, treated by physician	Steady, continuous employment	Steady, continuous relationship, contacts with friends	Clean urines, no reported abuse, keeps appointments	Arguments resolved non-violently, reduced verbal abuse
(-1) Somewhat Less Than Expected Level	Moves every month	Treated for new disease, no HIV test	Moves from job to job	Occasionally socially active	Dirty urines, reduced level of abuse, missed appointments	Suicidal thoughts, involved in violence, verbal/physical abuse
(-2) Much Less Than Expected Level	Homeless	HIV positive or no test, hospitalized	Unemployed	Frequent promiscuity, few or no friends	Dirty urines, returns to prior abuse, overdose	Perpetuates assaults, attempts suicide

I. SIMPLIFIED MEASUREMENT FORMULA FOR CALCULATION OF EQUALLY WEIGHTED QPA SCALES

$$(x_1) + (x_2) + (x_3) \dots$$

Number of Scales

x_1 = Score attained on Scale 1

x_2 = Score attained on Scale 2

x_3 = Score attained on Scale 3

... = Other scores attained on other Scales

Explanation: The scores assigned on each scale are combined, and the sum of those scores are divided by the total number of scales/performance indicators.

II. FORMULA FOR WEIGHTED AND EQUALLY WEIGHTED SCALES

$$50 + \frac{10 (w_1x_1 + w_2x_2 + w_3x_3 \dots)}{\sqrt{.7 [(w_1)^2 + (w_2)^2 + (w_3)^2 \dots] + .3(w_1 + w_2 + w_3 \dots)^2}}$$

A. Calculations for weighted scales

x_1 = Score attained on Scale 1

x_2 = Score attained on Scale 2

x_3 = Score attained on Scale 3

w_1 = Weight assigned to Scale 1

w_2 = Weight assigned to Scale 2

w_3 = Weight assigned to Scale 3

B. Calculations for equally weighted/unweighted scales

Using the same formula, insert the Scale scores where appropriate, but instead of using weights, insert the value "5" for each "w" where appropriate.

SAMPLE SUBSTANCE ABUSE TREATMENT PLAN USING QPA FORMAT - WEIGHTED AND SCORED
(Shaded areas indicate level of attainment)

	1. Residential (w = 5)	2. Health (w = 10)	3. Employment/Educ. (w = 5)	4. Relationships (w = 10)	5. Substance Abuse (w = 20)	6. Violence (w = 20)
(+2) Much More Than Expected Level	Owns home	No hospitalizations, illness or accidents	Enrolled in school	Married or cohabitating	Attends AA/NA, 90/90	In counseling for anger, control
(+1) Somewhat More Than Expected Level	Has apartment with lease	Has had complete physical exam	Enrolled in job training program	Monogamous relationship, frequent contacts with friends	Attends AA/NA weekly or more, has sponsor	Fewer arguments, reduced frequency of anger, control
(0) Expected Level of Outcome	Has permanent address and phone number	HIV negative, treated by physician	Steady, continuous employment	Steady, continuous relationship, contacts with friends	Clean urines, no reported abuse, keeps appointments	Arguments resolved non-violently, reduced verbal abuse
(-1) Somewhat Less Than Expected Level	Moves every month	Treated for new disease, no HIV test	Moves from job to job	Occasionally socially active	Dirty urines, reduced level of abuse, missed appointments	Suicidal thoughts, involved in violence, verbal/physical abuse
(-2) Much Less Than Expected Level	Homeless	HIV positive or no test, hospitalized	Unemployed	Frequent promiscuity, few or no friends	Dirty urines, returns to prior abuse, overdose	Perpetuates assaults, attempts suicide

EXPLANATION OF SAMPLE SUBSTANCE ABUSE TREATMENT PLAN USING QPA FORMAT - WEIGHTED AND SCORED

Assignment of Weights

In this example, each of the six scales was weighted as follows:

1.	Residential	5
2.	Health	10
3.	Employment/Education	5
4.	Relationships	10
5.	Substance Abuse	20
6.	Violence	20

It was determined by the consensus group that a value of "10" would represent the "middle ground". Further, it was determined that both the "Substance Abuse" and "Violence" scales were twice as important as the middle ground, and would therefore be weighted with the numerical value of "20". It was also determined that the "Residential" and "Employment" scales were somewhat less important than the middle ground, and they therefore were assigned weights of "5".

This process of weighting points out the importance of the consensus group, and indicates how experience, the client group and the moral values of the participants enter into the weighting process. Some other group might consider the "Health" scale to be the most important since the absence of good health will have a negative impact on all other scales.

This theoretical treatment program intends to serve parolees. It may be assumed that since this population has been incarcerated that their level of substance abuse involvement is more serious than a population which has not been incarcerated. If the program were to serve those with no history of incarceration, then the scales might be weighted in other ways. For example, in a more traditional outpatient substance abuse treatment program, the clients may present fewer incidents of violence. In this case, the "Violence" scale would probably be assigned a lower weight, signifying less importance.

Assignment of Attainment Scores

In this example, Client A's progress was assessed at the end of his treatment, which in this case was twelve months. The consensus group was led by the client's counselor, since the counselor had the greatest degree of contact with the client. The counselor presented the consensus group with the reasons for his/her assignment of scores. It was determined that on the "Residential" scale, the client moved an average of once per month over the past twelve months. This scale, then, was assigned a value of -1. All the other scales were assigned scores similarly.

Calculation of Scores

Using the formula which appears in the previous Appendix, the scores and weights are inserted into the formula as follows:

$$50 + \frac{10 [(5)(-1) + (10)(0) + (5)(1) + (10)(-2) + (20)(1) + (20)(0)]}{\sqrt{.7(5^2 + 10^2 + 5^2 + 10^2 + 20^2 + 20^2) + .3(5 + 10 + 5 + 10 + 20 + 20)^2}}$$

Further calculations produce:

$$50 + \frac{10 [(-5) + (10) + (5) + (-20) + (20) + (0)]}{\sqrt{.7(25 + 100 + 25 + 100 + 400 + 400) + .3(5 + 10 + 5 + 10 + 20 + 20)^2}}$$

The next calculation produces:

$$50 + \frac{100}{\sqrt{.7(1050) + .3(70)}}$$

The next calculation produces:

$$50 + \frac{100}{\sqrt{756}}$$

The final score attained is: 53.64

(If all 6 scales were equally weighted the score would be 47.42 in a range of 19.02 to 80.98)

About the State Reporting and Evaluation Program

The Bureau of Justice Assistance (BJA) established the State Reporting and Evaluation Program (SREP), a State-based program with an orientation toward establishing Federal, State and Local partnerships, to assist in implementing the reporting and evaluation requirements of the Anti-Drug Abuse Act of 1988. Through SREP, BJA provides technical assistance and training to the State and local offices and agencies responsible for implementing, monitoring, and evaluating violent crime and drug control programs funded under the Drug Control and System Improvement Formula Grant Program. SREP is coordinated for BJA by the Justice Research and Statistics Association (JRSA).

The SREP project is designed to:

- meet States' needs for technical assistance for the development of drug control strategies and the development of State monitoring plans;
- provide technical assistance and training on drug control project performance monitoring and evaluation;
- publish reports for State and local audiences on special topic areas related to drug control program performance monitoring and results of evaluations; and
- disseminate reports and information to the States and territories as a result of BJA and SREP activities.

A National Planning Group, comprised of State and local representatives from the criminal justice community provides input to the project. The National Planning Group plays a critical role in the development and implementation of the SREP projects, and also plays an integral role in the development of national indicators for performance monitoring. Since 1987, JRSA has worked with BJA and the States to establish data collection and analysis projects. JRSA and the States have produced numerous reports and technical assistance products covering many criminal justice programs and themes, including: multijurisdictional law enforcement task forces, innovative rural programs; crime laboratory enhancement programs; county-level trends in drug arrests, convictions, and sentencing; State citizen surveys on drug use and control; drug offender processing; and forecasting for criminal justice policy analysis.

The State Reporting and Evaluation Program is a unique program that focuses primarily on enhancing States' monitoring, reporting, and evaluation capacities. States participate in all aspects of the SREP project from planning and development to the implementation and delivery of technical assistance and training services. The project is designed to provide a forum for States to share information and to receive the assistance they need to develop and implement effective monitoring, reporting, and evaluation systems.

For more information about the
State Reporting and Evaluation Program contact:

Robert A. Kirchner, Ph.D.
Chief, Program Evaluation
Bureau of Justice Assistance
633 Indiana Avenue, NW
Washington, DC 20531
(202) 616-3455

Kellie J. Dressler
Project Manager
Justice Research and Statistics Association
444 North Capitol Street, NW
Suite 445
Washington, DC 20001
(202) 624-8560