

100 YEARS  
of work  
FOR EQUAL  
JUSTICE  
has not made  
IT A REALITY

LOAN DOCUMENT  
RETURN TO:  
NCJRS  
P. O. BOX 24036 S. W. POST OFFICE  
WASHINGTON, D.C. 20024

SUMMARY REPORT  
OF THE  
DEFENDER EVALUATION PROJECT

Roberta Rovner-Piechenik  
Alan Rapoport  
Martha Lane

36073 c.2

**nlada**

National Legal Aid and Defender Association

FRANK N. JONES  
Executive Director  
OFFICERS  
President  
JOHN G. BROOKS  
Boston  
Senior Vice President  
WILLIAM R. KLAUS  
Philadelphia  
Chairman, Defender Committee  
ROLLIE H. ROGERS  
1575 Sherman Street  
Denver, Colorado 80203  
Chairman, Civil Committee  
ALAN W. HOUSEMAN  
220 Bagley  
Detroit, Michigan 48226  
Secretary/Treasurer  
FRANK N. JONES  
Washington, D.C.

Board of Directors  
BARBARA A. BABCOCK  
Palo Alto  
ADAMSON BOWMAN  
Honolulu  
JOHN J. CLEARY  
San Diego  
THOMAS CRAVEN  
Winston Salem  
GREGORY R. DALLAHE  
Seattle  
JOHN C. EMERY, JR.  
Birmingham, Mich.  
JAMES D. FELLERS  
Oklahoma City  
CASSANDRA M. FLIPPER  
San Francisco  
DONALD FONTAINE  
Augusta, Me.  
LOUIS O. FROSZ, JR.  
Jacksonville

RITA GEIER  
Seattle  
CHARLES D. GILL  
New Haven  
THEODORE A. GOTTFRIED  
Springfield, Ill.  
MARYELLEN H. HAMILTON  
New Orleans  
MORGAN D. HARRIS  
Las Vegas  
ROBERT J. HELLRUNG  
Seattle  
ANTHONY R. HENRY  
Washington, D.C.  
WILLIAM P. HIGHAM  
Washington, D.C.  
ANDREW HOURIGAN, JR.  
Wilkes-Barre  
R. WILLIAM IDE, III  
Atlanta  
LEONARD S. JANOWSKY  
Los Angeles

EARL JOHNSON, JR.  
Los Angeles  
SAM D. JOHNSON  
Austin  
C. LYONEL JONES  
Cleveland  
BETTYE H. KEHRER  
Atlanta  
BROOKSLEY E. LANDAU  
Washington, D.C.  
SIMON LAZARUS, JR.  
Cincinnati  
NORMAN LEFSTEIN  
Chapel Hill, N.C.  
PAUL LIGDA  
Fairfield, Calif.  
WILBUR LITTLEFIELD  
Los Angeles  
ANTHONY P. LOCICCHIO  
Honolulu  
ROGER A. LOWENSTEIN  
Newark  
DANIEL M. LUEVANO  
Los Angeles  
JOHN L. MAXEY, II  
Jackson, Miss.  
MARTIN MENDELSON  
Springfield, Ill.

ROBERT W. MESERVE  
Boston  
ELLSWORTH MORGAN  
Newark  
MALVINE NATHANSON  
Boston  
JAMES R. NEUHARD  
Detroit  
ROBERT C. OBERBILLIG  
Des Moines  
REVIUS D. ORTIGUE, JR.  
New Orleans  
CHARLES PARKER  
New Haven  
DAVID S. PESHKIN  
Des Moines  
SHELDON PORTMAN  
San Jose  
CRUZ REYNOSO  
Albuquerque  
ELIZABETH RINDSKOPF  
New Haven  
MITCHELL ROGOVIN  
Washington, D.C.  
SIMON ROSENTHAL  
San Francisco  
STEPHEN I. SCHLOSSBERG  
Washington, D.C.

JEROME J. SHESTACK  
Philadelphia  
SHELVIN SINGER  
Chicago  
CHESTERFIELD SMITH  
Lakeland, Fla.  
JOHN E. SMITH  
Miami  
HERBERT D. SOLL  
Anchorage  
MYZELL SOWELL  
Detroit  
ROBERT L. SPANGENBERG  
Boston  
JUSTIN A. STANLEY  
Chicago  
MARTIE THOMPSON  
New York  
MAYNARD TOLL  
Los Angeles  
VALERIE VANAMAN  
Cambridge, Mass.  
STANLEY C. VAN NESS  
Trenton  
JAMES B. VENTANTONIO  
Newark  
ROBERT L. WERNER  
New York

J. ALBERT WOLL  
Washington, D.C.  
PETERSON ZAH  
Window Rock  
VINCENT J. ZICCAROI  
Philadelphia  
Vice Presidents  
GEORGE W. ALLEN  
Ft. Lauderdale  
GARY BELLOW  
Cambridge, Mass.  
MARY LOUISE BUTLER  
St. Louis  
JOHN W. DOUGLAS  
Washington, D.C.  
BETTY B. FLETCHER  
Seattle  
PHILLIP H. GINSBERG  
Seattle  
F. WILLIAM McALPIN  
St. Louis  
JOHN D. ROBB  
Albuquerque  
MARNA SUSAN TUCKER  
Cherry Chase, Md.  
THEODORE VOORHEES  
Washington, D.C.

# nlada

National Legal Aid and Defender Association  
Suite 601, 2100 M Street, N.W., Washington, D.C. 20037

*This study was prepared by the National Legal Aid and Defender Association (NLADA) in compliance with Grant Number 74-NI-99-0049 from the Law Enforcement Assistance Administration, U. S. Department of Justice. Opinions expressed are those of the authors and do not necessarily represent the views of NLADA or of the Law Enforcement Assistance Administration.*

## Defender Evaluation Project Staff

Roberta Rovner-Pieczenik, Director  
Alan Rapoport, Deputy Director  
Martha Lane, Associate Director for Management Systems  
Kathy Bradt, Administrative/Research Assistant  
Kate Lenski, Secretary

LOAN DOCUMENT  
RETURN TO:  
NCJRS  
P. O. BOX 24036  
WASHINGTON, D.C. 20024  
S. W. POST OFFICE

## S U M M A R Y   R E P O R T

### Defender Evaluation Project

Roberta Rovner-Pieczenik  
Alan Rapoport  
Martha Lane

NCJRS

SEP 1 1976

ACQUISITIONS

Assisted by:  
Kathy Bradt  
Kate Lenski

June 1976

Defender Evaluation Project  
National Legal Aid and Defender Association  
Suite 601, 2100 M Street, N.W.  
Washington, D.C. 20037

National Institute of Law Enforcement  
and Criminal Justice  
Law Enforcement Assistance Administration  
Grant Number 74-NI-99-0049  
U.S. Department of Justice

RETURN TO  
FOR DOCUMENT  
JANUARY 1971  
U.S. DEPT. OF JUSTICE  
WASHINGTON, D.C. 20535

TABLE OF CONTENTS

Introduction . . . . .	1
Development of the Evaluation Design . . . . .	3
Development of the Self-Evaluation Manual . . . . .	19
Evaluation Design - Concerns and Findings . . . . .	20
Self-Evaluation Manual - Concerns and Findings . . . . .	36

## INTRODUCTION

The Defender Evaluation Project (DEP), operating under LEAA grant number 74-NI-99-0049 to the National Legal Aid and Defender Association took as its tasks:

- (1) The development of an evaluation design and materials which could be used to evaluate offices of the public defender;
- (2) The development of a self-evaluation design and materials for offices of the public defender; and,
- (3) The testing of the evaluation design in the field.

During the preceding five years, a small number of defender office evaluations have been performed by different groups with varying degrees of competence. These evaluations provided a good base from which DEP could build. But as in other fields, early evaluations suffered from being ad hoc, without particular care having been taken in planning a sound methodology. This resulted in evaluations which: (a) were inconsistent in quality, i.e. highly dependent upon the capabilities of the specific evaluators; (b) were varied in substance, i.e. highly dependant upon the concerns of specific team members; and (c) omitted any references to the evaluation methodology used, i.e. its strengths and limitations. It was in response to these early limitations that DEP worked, attempting to develop a design which minimized the impact of individual evaluators and their idiosyncrasies, and maximized an approach which contained some measure of validity and reliability.

The Handbooks to be used by an evaluation team which is independent of the particular office to be evaluated, were designed to be credible to defenders and the evaluation community, easily used, and applicable to the large proportion of existing defender offices. The methodology used for developing these Handbooks sought close contact with and acceptance by defenders in the field. The mechanical ease which has been built into the design is the product of the continual refinement of materials over the course of three test evaluations. The potential cost of each evaluation has been minimized by paying close attention to evaluators and their assessment of the point at which enough information is gathered upon which judgments can be comfortably made. The relevance of the design is demonstrated by the fact that small (1-5 attorneys) and medium-sized (5-30 attorneys) offices represent approximately 90% of all existing defender offices. The flexibility built into the design allows for an evaluation of quality of representation, and/or a management analysis of office operations, and/or a statistical analysis of patterns of adjudication and sentencing.

The Manual to be used by a defender office to evaluate its own performance was designed to: (a) raise the level of awareness of defender offices about the professional standards which they are expected to meet; (b) be easily used by defender offices; and, (c) suggest an approach an office might use in evaluating and reorganizing the office to meet professional standards.

DEP had virtually no model upon which to build. Self-evaluations in other fields were either self-serving or lacking. In response to this vacuum a Manual was developed which attempted to accomplish the above purposes; it was further refined by comments of defenders. Its value will lie in its perceived usefulness by defender offices across the nation.

The report which follows: (a) reviews the overall development of the evaluation Handbooks and Manual, and (b) addresses specific design elements about which comments and findings are appropriate.

## DEVELOPMENT OF THE EVALUATION DESIGN

### The State-of-the-Art

The first stage in the development of the evaluation design to be used by independent consultants consisted of reviewing prior evaluations, feasibility studies, technical assistance projects and management studies which focused on offices of the public defender. This provided DEP with perspective on the nature of each previous study, the content which was highlighted, and the format of the undertaking.

The studies were quite informative to DEP, and of undoubtedly considerable help to the requesting agency. From a research/evaluation standpoint, however, certain problems were apparent:

1. Reports were inconsistent in quality;
2. The content of similar studies (e.g. topics in management) varied;
3. Little or no explanation of the methodology utilized was presented;
4. Conclusions and recommendations were made with minimal supporting information;
5. Statistical analyses of import were largely absent from reports;
6. Policies regarding the highlighting of recommendations were not apparent; important information and recommendations were "hidden" within the body of the report; and,
7. Situations encountered in defender offices were rarely interpreted within the context of the larger criminal justice system.

It was DEP's intention to create a design and develop materials which would minimize the above problems.

Information pertaining to accreditation and evaluative materials was sought from such diverse organizations as: the National Institute of Mental Health; the Juvenile Judges Institute; the American Hospital Association; the American Correctional Association; Middle States Association of Colleges and Secondary Schools; Judge Advocate General's Office of the U.S. Army; National District Attorney's Association; Office of Management and Budget.

### Advisory Board

Suggestions for Advisory Board members were solicited and received from the NLADA Defender Division staff, American University Institute for Law and Social Justice, the Defender Committee, and others within and without NLADA. The Board was selected with three purposes in mind: (a) obtaining criminal justice expertise which would aid us in the substance of an evaluation of defenders; (b) obtaining management expertise which would aid us in the development of the management analysis component of the evaluation; and (c) obtaining social science expertise which would guide us in the development of a reliable and valid design. In addition to these needs, each of the Board Members selected was "close" to evaluations; either designing or participating in them (Appendix A).

### Goals, Objectives and Criteria for Compliance

A necessary feature of any evaluation is the existence of one or more goals toward which the program to be evaluated is working. An early decision was made to focus the evaluation design on national, rather than local, Goals and Objectives. It was felt that this approach would result in a design with unlimited scope, and would lead to easier comparisons among programs if a defender office accreditation program ever became a reality.

Statements and standards on optimum defender office performance were reviewed: (a) The American Bar Association Standards Relating to Providing Defense Services (A.B.A.); (b) the National Advisory Commission on Law Enforcement and the Administration of Justice (N.A.C.); (c) the American Bar Association Standards Relating to the Prosecution Function and the Defense Function (A.B.A.); (d) the National Legal Aid and Defender Association's Second Discussion Draft of Proposed Standards for Defense Services (NLADA); and, (e) the draft report of the Task Forces of the National Legal Aid and Defender Association's Commission on Defense Services (NLADA). The American Bar Association's Code of Professional Responsibility was also consulted.

The result of this review was a decision to highlight three overall Goals for defender offices: (a) efficient and effective service delivery; (b) high quality representation of clients; and, (c) overall improvement of the adversary process and criminal justice system. Each of these goals was divided into a series of more specific Objectives toward which an office was expected to work. Each Objective was subdivided into a set of Criteria for Compliance which represented desired performances which, if accomplished by an office, would indicate compliance with an Objective.

The Goals and Objectives were formulated and revised with the assistance of the DEP Advisory Board and public defenders nationwide. The Criteria for Compliance were based on the review of standards for performance which were formulated by the A.B.A., N.A.C., and NLADA. These were altered during the three test evaluations. The entire set of Goals, Objectives, and Criteria for Compliance appear in Appendix B.

#### Performance Indicators

Once defender office Goals and Objectives were established, DEP sought items which would "indicate" full or partial compliance with an Objective or its determining criteria. Telephone interviews were conducted with defenders who had previously been members of an evaluation team, as well as with judges and prosecutors. The DEP staff reviewed Indicators which appeared in former studies of defender offices, as well as those which were being used by other NLADA Defender Division projects. An entire day of the first Advisory Board meeting was spent developing a list of Indicators for each Objective. During field visits to offices,, defenders and other members of local criminal justice systems were asked to comment on Indicators s/he would use to discover whether Objectives were being met. On the basis of these activities, a realistic list of Indicators of performance was devised for each Objective to guide team members in the collection of data.

A major effort was made to obtain client input. DEP staff contacted numerous sources to obtain advice in this regard. The Director of Rehabilitation for the D.C. Public Defender's office, and the Director of the Bureau of Rehabilitation in Washington, D.C. allowed DEP to work with their respective staff to obtain input from clients. DEP developed a questionnaire for clients which was critiqued by these staff and revised accordingly. The revised questionnaire was then used to interview approximately 30 present and former clients of the District of Columbia Public Defender (Appendix 3). Questionnaire responses were incorporated into final Indicators lists.

DEP staff decided that a variety of Indicators were needed to provide the information necessary to make an accurate assessment of compliance with any one criterion or Objective. Indicators varied from the highly qualitative (e.g. case outcome statistics) to the purely descriptive (e.g. geographical location of the office). No artificial weighting system was created to rank or prioritize Indicators. The cumulative impact of the variety of types of Indicators gathered by different methods was viewed as the "preponderance of evidence" which the consultant evaluation team was to assess (Appendix D).



### Site Reviews

Site visits to five defender offices were made by the professional staff for purposes of: (a) questioning individuals in the "field" about the appropriateness of staff ideas and preliminary materials; (b) observing operations within offices; and, (c) gathering information on defender office files and their management procedures. Each site differed in size, setting, local criminal justice system rules and procedures, structure and scope of representation. The fact that each office visited was considered by defenders around the country as a "high quality" office was their main similarity. Although DEP originally planned visits with geography in mind (i.e. nine sites, from Vermont to California), redundancies in the information retrieved at each site led the staff to cancel four visits. Diminishing returns were evident.

### Former Evaluator Feedback

A meeting was chaired by DEP staff with defenders who had formerly been a team captain or member of an evaluation or technical assistance team to discuss their perceptions of the critical concerns of any evaluation. The following points were made by participants:

1. The focal points of the evaluation, for both the evaluators and evaluatees, should be agreed upon in advance of the team visit. This would enhance credibility and efficiency;
2. An evaluation should not be compromised because of lack of time or money. When funding problems exist, priorities for the evaluation should be selected. For example, the statistical study might be omitted;
3. A management analysis is an important aspect of an evaluation;
4. A fair sample of clientele served by the defender should be interviewed to determine client satisfaction;
5. Preparation for the team visit is of great importance. There was insufficient time and materials, in the past, to adequately prepare team members for the on-site visit;
6. The team captain and team members should be trained for their respective roles in the evaluation;

7. Daily debriefings by the evaluation team should occur prior to dinner;
8. The design should be flexible enough to enable team members to "pick up" on unforeseen problems;
9. The first draft of the report should be done on-site. This would assist the team in reaching a consensus on conclusions and recommendations;
10. Since the evaluation is a measure of defender performance, the team captain should be a defender (and have some voice in the selection, or veto, of team members);
11. A permanent staff member should be present during all evaluation site visits to help provide uniformity in approach, and aid the team members.
12. Opinions about the office being evaluated should be based upon on-site findings, rather than pre-held attitudes;
13. For an evaluation to be credible, certain key people in each community must be interviewed;
14. A standard format for the final report should be considered, stressing sufficient back-up information to substantiate conclusions and recommendations;
15. Interview questions should be keyed to specific interviewees;
16. The credibility of observations was questioned;
17. Important information could be gathered in an informal setting; an end-of-the-day eating and drinking session with the personnel from the Defender Office could be an asset to the evaluation if specifically structured into the evaluation.

In addition to this meeting, comments from a variety of former evaluators were gathered during telephone conversations which helped to shape the design.

#### Evaluation Site Selection

Mail and personal solicitation for evaluation test sites resulted in applications from 14 individual offices and 3 state-wide defender systems. These applications were followed-up by

DEP staff telephone conversations with defenders around the country about "what is known" about the application sites.

The main purposes of using test sites included:

1. Exploring the feasibility of the evaluation procedure;
2. Exploring the reliability of the evaluation instrument;
3. Exploring the validity of the evaluation instrument; and,
4. Exploring the use of non-defenders as evaluators.

Twenty-five test evaluations are needed to accomplish these purposes; ten evaluations can only hope to make a start in the right direction. The three evaluations for which DEP had money could only be highly exploratory in nature.

In light of DEP's concerns for reliability and validity, (see p.15), and its desire to test the materials in offices of different size and scope, the following considerations in site selection were highlighted:

1. An office which, by independent assessment of authorities, was deemed "excellent" in the overall delivery of legal representation;
2. An office which, by independent assessment of authorities, was deemed "poor" in the overall delivery of legal representation;
3. An office which was of unknown quality prior to its evaluation; and
4. Offices which varied in size from 5 to 20 full-time staff defenders.

The sites selected for the test evaluations according to the site selection criteria were Monticello, New York, Las Vegas, Nevada, and Columbus, Ohio. The Legal Aid and Defender Society of Sullivan County (Monticello, New York) was a five-attorney rural office whose chief defender was new, and about which little was known by defenders nationally. The defenders appeared before a large number of justice of the peace courts, as well as a county felony court. The office utilized paralegals. The Las Vegas public defender office consisted of 18 attorneys and was recognized by defenders nationally as a stable, long-established office providing good representation with a wide scope of services. It had a special training program for staff members and took part in diversion and pre-trial release programs. A truncated docket study had been completed by another NLADA defender project against which DEP could compare its statistical study. The Legal Aid and Defender Association of Columbus, Ohio

had been evaluated by a community group during the previous year and was felt to be giving poor representation.

In short, these offices covered the site selection criteria in their variation in size and "quality". It should be pointed out, however, that few specifics are known about defender programs around the country by other defenders. Gross generalizations about "good" or "bad" offices usually have little solid bases upon which to rest; only isolated bits of information are known to defenders (e.g. office "A" has a good time-keeping system).

#### Team Selection and Composition

Throughout the project period an evaluator "pool" was kept. A list of names of prospective evaluators was developed following discussions with former evaluators, individuals "known" to known individuals, writings which made the author sound "attractive", and meetings and institutes attended.

A few ground rules for team selection were used: teams would be mixed in composition between novice evaluators and "old pros"; non-attorneys would be used on every team; different types of individuals within a particular profession would be "tested". For example, the latter concern was implemented by using three different management analyst types: a public defender known to have a well-managed office; a nationally known administrator of a large civil law firm; and, a university professor whose background was in the social services and who regularly consulted on office management issues. The "community/consumer" type also was varied, consisting of: an individual active in community affairs, whose organization influenced the workings of the defender office in her community; the director of a rehabilitation program attached to a defender office; and, an ex-offender "jail-house lawyer" who was working as an appellate specialist for a federal defender.

Each Team Member was chosen according to his/her expertise and interpersonal skills, as well as the anticipated requirements of the site to be evaluated.

For the small office, it was anticipated that a three-person team would suffice; for the medium-sized office, a five-person team.

The Team Captain for each evaluation was a criminal attorney. Two had previously captained teams; one was a "new" Captain who had participated in previous evaluations. All were current or former defenders.

Each team was composed of individuals having the following

skills: legal (for purposes of credibility and rapport with the local criminal justice community, and in light of the need to assess legal competence), management (since one task of every evaluation consisted of an analysis of management activities within the defender office) and community/consumer (to be able to communicate as a knowledgeable non-attorney with the client and citizen community). In addition, the logistics of the evaluation itself (e.g. xeroxing, training team members, remaining in the defender office to receive changes of interview plans) necessitated an individual with administrative/training skills who was not a member of the evaluation team.

#### Team Captain Training

Once the evaluation test sites were selected, and teams chosen, a one-day Team Captain's meeting was held. Training included:

1. A general introduction to evaluative research; that is, concepts, principles, design methods, data gathering approaches, analysis and interpretation;
2. A detailed examination of the Handbooks, and their approach to the evaluation of defender services; that is, the format, assumptions and strategies; and,
3. Consideration of those skills which are essential to good field work and which are discussed in the Handbooks; particularly, the art of interviewing and observing.

#### Evaluation Format

The format for the evaluation of defender offices developed as follows:

1. An evaluation request from a defender (or other party) is processed, verified, and the necessary letters and information exchanged;
2. The Chief Defender is mailed pre-evaluation profile (PEP) materials to be completed prior to his/her first meeting with the Team Captain and Staff member. These materials request information on the defender office, the criminal justice community, and the general community;

3. An evaluation team is selected. One member is asked to take the responsibility of being Team Captain;
4. A pre-evaluation site visit is made to the defender's office by the Team Captain and Staff. The purposes of this visit are varied: (a) to review the completed PEP forms; (b) to pinpoint key individuals to interview during the evaluation; (c) to obtain copies of relevant materials (e.g. statutes); (d) to get a "feel" for the temper of the office and criminal justice system by touring the appropriate offices and courts and meeting with local criminal justice officials; (e) to arrange for the case file/docket statistical analysis; and, (f) to inform the defender about the evaluation goals and format;
5. An interim period occurs during which statistical studies are performed, on-site interview appointments are scheduled, and materials for Team Members are assembled and mailed;
6. The on-site evaluation by Team Captain and members takes place, lasting from five to seven days. An orientation training day is held for team members prior to beginning the on-site evaluation. The evaluation period consists of daily interviews and observations in the defender office, criminal justice system, and general community, and late afternoon debriefings among team members. A management analysis of office operations is made. Agreement is reached by team members concerning conclusions and recommendations. Differences of opinion which can not be settled are added to the final report as addenda.
7. The final report is written by the Team Members, edited by the Team Captain, reviewed by the Chief Defender office, and finalized.

The entire process can take from three to six months.

The format was based upon a DEP review of previous evaluations in the defender field and other areas, an approximation of costs for an evaluation spread over differing time periods, suggestions derived from a meeting with former defender evaluation team captains and members, and DEP staff calculations of the number of interviews and observations evaluation teams of differing sizes and duration could accomplish on-site in a given number of days.

### Assessing Compliance

Each Team Member was asked to assess the extent to which a Defender Office met each Criteria for Compliance. Based, hopefully, upon team consensus, Team Members would "rate" the Office on a four-point "scale". The words "rate" and "scale" were used with great reservation.

A rating system, of any kind, is a way of taking a qualitative description (e.g. poor, fair, good, excellent) and giving it a numerical counterpart (e.g. poor = 1, excellent = 4) on a continuum. It is always arbitrary whether a description will be divided into four, five, or ten parts (i.e. numbers). The Criteria checklist is similar. It takes a series of descriptive terms to assess extent of compliance with Criteria and arrive at a common understanding among Team Members about what each term means. A four-point scale was chosen for simplicity of use:

Never/occasionally = less than 25% of the time.  
Sometimes = between 25% and 50% of the time.  
Frequently = between 50% and 75% of the time.  
Usually/always = over 75% of the time.

It is important when using any rating system that: (a) each Team Member has the same understanding of what a given rating will mean; and (b) each category to be rated (e.g. sometimes, frequently) can be distinguished from each other.

Once descriptions are reduced to the appearance of a scale, however, there is a tendency to think of the items as numbers with exact meanings; that is, to think of a "sometimes" as one unit higher than an "occasionally"; "usually" as two units higher than a "sometimes". Once this occurs, it is a simple step to then give the terms numbers (i.e. 1, 2, 3, 4) so that they can be added or divided.

Although all of this can be done, Team Members were not asked to do it for purposes of a final numerical summation of a Defender Office's degree of compliance on a particular Criterion or Objective. First, while for purposes of each of discussion Team Members might arbitrarily agree that "sometimes" will mean a 25-50% of the time, it is almost impossible to assess an accurate percentage for each category from the variety of data gathered for most Objectives. Some of the data gathered will be of a statistical nature, while other data will not. Second, it is as difficult to discriminate precisely among different categories along an entire scale (e.g. "sometimes" from "frequently") as it is to discriminate percentages within each category.

It should also be pointed out that the Criteria within a

given Objective may cover a variety of concerns. For example, the Objective of "zeal" takes into consideration the varying activities of efforts to challenge the criminal justice system, and efforts to be sensitive to client needs in non-legal matters. To conclude in a Report that "Defender Office 'X' is in compliance with Objective 2, 75% of the time," would disguise more than it would illuminate.

Given this perspective, it was decided that it would be both misleading and inaccurate to "add" ratings given on a series of Criteria within one Objective to come out with a "total" or "average" rating for each Objective.

For the Final Report, the term "degree of compliance" was used only when referring to individual Criteria. The concept loses validity and meaning when extended to a discussion of a particular Objective and was not used in that connection.

The use or non-use of a rating system is a matter of judgment and personal prejudice. There is no doubt that it is less cumbersome to list numbers than to talk about a multiplicity of variables which interact to make a Defender Office function in a particular fashion. It was DEP's best assessment, however, that for any evaluation to have meaning for Defenders, or any party reading it, any rating to be done should be applied solely at the Criteria level. When an Objective was to be discussed, it would always be in specifics; that is, in the context of complying with specific Criteria.

#### Gathering Data

No one technique of data collection is relied upon in the evaluation design. Rather, the type of information sought as indicators of defender office performance is to be gathered through a variety of techniques.

Statistical Analysis. Statistical analysis has two broad functions: (a) to summarize a large amount of information by using numbers to make the information more manageable; and, (b) to generalize about a large population on the basis of a sample drawn from this population. In the evaluation of a defender office, a statistical analysis of closed defender case files enabled DEP to summarize patterns of case handling and case outcome, and to explore relationships among a few case variables. A similar analysis of the court docket enabled comparison of defender activities with those of the private attorney and assigned counsel. These statistics were not used as an "evaluation" in their own right, but were viewed as a starting point from which interviews and observations could take their cues.



Observation. Observation becomes a scientific technique only when it serves a formulated purpose, is planned carefully, and is recorded systematically. The major asset of observation is that an individual's actual behavior is noted. One need not try to predict that individual's behavior from what s/he says it will be or has been. In addition, observation is not dependent upon an interviewee's ability or willingness to articulate responses. The major limitation of this technique, however, is that the observer is never sure that the behavior s/he is witnessing is typical of the situation, or whether some of it has been "altered" for the observer. It is for this reason that observation was not relied upon too heavily to "prove" a point, but was used in supplemental fashion or to interpret information gathered through other techniques. In the evaluation of a defender office, Team Members were asked to observe the conduct and activities of the defender during interactions with the court and his/her clients.

Interviews. The interview, in contrast to observation, has the advantage of retrieving a great deal of information in a short period of time. Not only can the evaluator ascertain facts about behavior, but s/he can also probe into beliefs about office policies, reasons for beliefs, feelings, standards of action, and past behavior. Interviews conducted during the on-site period of a defender evaluation should be flexible and adaptable to the specific situation. The interview format developed by DEP specified the focal concerns of each interview and was meant to be a guide for the interviewer. It was important that each Team Member have the leeway to use questions that s/he deemed appropriate to a particular respondent or situation. To accomplish this, questions -- their content, sequence and working -- were not fixed. This placed a minimum of restraint upon the interviewer, but a maximum amount of pressure to know ahead of time what s/he was seeking from a particular interview and interviewee.

Content Analysis. Content analysis is a method of studying and analyzing communication in a systematic, objective and quantitative manner. Instead of observing people's behavior directly, or asking them to respond to questions in an interview, content analysis looks at "communications" that people produced or which have been produced about them, and asks questions of the communication. Newspaper accounts of the criminal justice system were the subject of a content analysis. DEP made the assumption that the communications analyzed either: (a) reflected reality; or (b) offered a perception of reality. Both were important for evaluation purposes. A content analysis of newspaper articles was undertaken by Staff prior to the on-site period, and a summary of findings mailed to Team Members.

Case Study. The case study is an approach to gathering data which views any unit as a whole. This technique's major asset lies in its ability to make a process, which incorporates

the interaction of many variables at the same and different times, more understandable. It also allows time to be condensed, so that a full case, from pre-trial to post-conviction, can be reviewed in a short period of time. On the other hand, this technique suffers the problem of having its users generalize from a few cases to many, making faulty assumptions that the few cases reviewed are representative of most or all cases handled. The case study technique was used to scrutinize in detail a small number of cases handled by most defenders interviewed or observed.

In addition to the above techniques, factual data was gathered on the criminal justice system and community which served as background information against which other information was interpreted.

#### Management Analysis

Management may be defined as the process of achieving desired results through organizing the efforts and tasks of people. Starting with goals which are based upon a careful appraisal of current circumstances and future needs, sound management identifies the role that each person in an organization is to play, motivates individuals to work to achieve these goals, and establishes a means for checking progress on an on-going basis.

An analysis of the management at the test sites was performed. Topics included in the analysis were: (a) caseflow management; (b) personnel; (c) training; (d) resources; and, (e) public relations. Statistical analyses, observations, interviews, case file reviews, and a review of all printed materials were undertaken by the management analyst.

#### Reliability and Validity

It is one thing to construct a measuring instrument (i.e. set of Handbooks) for the evaluation of defender offices, and quite another to construct one that "works"; that is, is both reliable and valid. Although the Handbooks attempted to build reliability and validity into their construction, it should be stated at the outset that more evaluations are needed, using a design which is specifically constructed to assess the instrument's reliability and validity.

To be reliable, the Handbooks should produce the same results each time they are applied. For example, the Handbooks would be considered reliable if the judgment of one evaluation concerning the extensiveness of "political control" over the :

defender was again made on a second evaluation (with no change of circumstances in the defender office). Whether the Handbooks allow for this would be an estimate of their reliability; that is, results would be stable. On the other hand, different evaluators using the same Handbook to assess "political control" in the same office at the same time, should also reach the same conclusions. Whether the Handbooks allow for this would be a different type of reliability; that is, results would be equivalent. Similar conclusions should also be reached independent of an evaluator's previous familiarity with the Handbooks or with defender programs.

To be valid, the Handbooks should measure what they claim to measure. For example, in the area of "political control", a defender office functioning under extensive political control should be distinguishable, by using the Handbooks, from an office without such control. The rating given an office on that one Objective should reflect the true measure of that Objective in a particular office.

Since it is difficult to know in advance an office's true position on any one Objective, there is no direct method of determining validity. In the absence of direct knowledge, the validity of any instrument is usually judged by the extent to which results are compatible with other relevant evidence. What constitutes relevant evidence, then, is of key concern. There are a variety of approaches to validating an instrument, with different researchers categorizing them somewhat differently. For our purposes, these divide easily into assessing: pragmatic validity (which includes concurrent and predictive validity), which asks whether an instrument is useful as an indicator or predictor of some behavior or character; and construct validity, which asks whether the theory underlying an evaluation instrument is valid. Pragmatic validity necessitates a reasonably valid and reliable criterion with which to compare ratings on the instrument. Construct validity cannot be adequately tested by any single procedure, but necessitates evidence from a number of sources, including correlation with another instrument, the internal consistency of items, and stability over time. Face validity, on the other hand, is based directly upon the behavior of the type in which an investigator is interested; an instrument (in this case, our Handbooks) is judged relevant for what it is trying to measure "on its face".

The concepts of reliability and validity are inextricably linked together in many ways. If we knew, for example, that a measuring instrument had satisfactory validity, we would not worry about its reliability. If an instrument is valid, that is, it is reflecting the characteristics which it is supposed to measure with a minimum of distortion, there is little reason to investigate its reliability (i.e. the extent to which it is influenced by transitory factors). On the other hand, if an instrument is of unknown validity, it is difficult to know whether it is reliable.

The concepts of reliability and validity are both important. However, there are a large number of variables incorporated in the evaluation instrument to test in a limited number of site evaluations. Neither of the issues of reliability or validity could be resolved with three evaluations. It was most efficient, then, to use each test evaluation for a variety of purposes. Thus, DEP chose to look at the problem in another way.

An instrument which has limited reliability and/or validity is subject to a series of problems (or errors) which influence the results obtained. In the ideal situation, the Handbooks and their application will uncover true differences (between programs) in the characteristics the evaluators are attempting to measure. The following list comprises some of the possible sources of errors in conclusions reached, under real circumstances, by a group of evaluators using the same Handbooks:

1. Differences in other characteristics of the program which affect the group's measurement of the characteristics in question;
2. Differences due to transient situational factors;
3. Differences due to personal factors;
4. Differences due to variations in administration;
5. Differences due to the sampling of performance measures;
6. Differences due to lack of clarity of the measurement instrument;
7. Differences due to mechanical factors; and,
8. Differences due to factors in the analysis.

To minimize these problems, and to gain confidence in the substance, format and application of the evaluation design, the following steps were taken:

1. Feedback was obtained from consultants (i.e. the evaluation teams and statistical analysts) on: (a) the logistical, administrative and mechanical aspects of implementing the evaluation design; and (b) the substantive (i.e. performance) measures used to distinguish programs on the relevant characteristics;
2. Feedback was obtained from defenders and all interviewees at the sites visited on evaluation format and content;

3. DEP avoided evaluating an office whose performance would be influenced by a recent transient situation (e.g. influx of riot cases) or by a personal problem (e.g. ill health of the chief defender);
4. The design called for evaluators to use different types of data to assess the same characteristic;
5. The design had uniform on-site procedures (e.g. same classes of persons were interviewed, the same data collected, the same issues explored) and a uniform reporting format; and,
6. Consultants were carefully selected, trained, monitored and supervised throughout the course of an evaluation.

Through this approach, DEP refined the mechanical and administrative aspects of implementing the evaluation design, added those performance measures which reflected the characteristics in which we were interested, and deleted those which did not prove helpful. It can not be overemphasized, however, that the validation of any instrument, including this one, takes much careful planning and many more test evaluations.

## DEVELOPMENT OF THE SELF-EVALUATION MANUAL

DEP's major concern in the construction of a Self-Evaluation Manual was to create a simple-to-use booklet which raised important issues to a high level of visibility in defender offices around the county. With this in mind, the core issues of the evaluation design (previous pages) was extracted, i.e. the Objectives towards which each office should be working. Simple questions about whether a defender office was, or knew it was, performing certain activities were posited. Questions which were answered in the affirmative pointed to areas of strength. Questions which were answered "uncertain" were followed by suggestions for gathering the appropriate data. Questions which were answered in the negative pointed to areas of weakness and were followed by a general management approach for reorganizing the office.

As in the evaluation design to be used by an independent consulting team, the questions raised by the Manual are general enough to be appropriate for a large proportion of defender offices. They do not rely on any one set of national standards, but are a synthesis of the elements most common to all.

It is important to recognize that a Self-Evaluation Manual is distinctly different from a technical assistance handbook. The former is concerned with raising issues; the latter is concerned with techniques of how to resolve the issues.

The Self-Evaluation Manual was internally critiqued by NLADA's National Center for Defense Management and mailed to approximately 60 defenders around the country for comments. The revised manual incorporates the suggestions received.

## EVALUATION DESIGN - CONCERNS AND FINDINGS

### Goals, Objectives and Criteria for Compliance

The major distinctions between DEP-developed Goals, Objectives, and Criteria for Compliance (GOC), and the A.B.A., NLADA, and N.A.C. standards are as follows:

1. DEP synthesized national standards posited by attorney groups into a small number of GOC's with which the larger proportion of defenders agree;
2. To avoid confusion with nationally recognized standards for defender services, the DEP design refers to "Criteria for Compliance" to describe specific defender performances expected;
3. DEP made analytical distinctions between GOC's which do not appear in other sources, differentiating between general statements of ideals and specific attorney activities;
4. Management activities were not viewed as "standards" in the traditional sense, but as tools depended upon by an office to achieve its GOC's.

There is no doubt that the number of variations of GOC's possible is limited only by the boundaries of imagination and the computer. The process DEP went through to develop its own approach, however, offers confidence in its acceptance. The literature review, the commentary of the Advisory Board, and the changes made during test evaluations increased and reflected subtleties of defender performance and continually redefined appropriate and inappropriate activities. The design is particularly flexible; any set or combination of current or future national standards can be substituted for some or all of the GOC's.

Each test of the evaluation design resulted in changes in language in the GOC's. Commentaries are provided prior to the GOC's so that underlying intent will be obvious.

DEP has learned that the Chief Defender should be fully aware of what an evaluation is and whether that is what is desired. It is the job of the evaluator, early in the process, to discern whether an office is requesting or in need of evaluation, technical assistance, a management analysis, or feasibility study. It is also the role of the evaluator to supply the



defender office with information about the content, form, and projected length of the evaluation so that a decision can be made by the defender about its relevance and desirability. An evaluation is often requested for less than the apparent reasons. The Chief Defender may want to prove his/her funding agency that additional operating funds are needed; s/he may be interested in public relations for him/herself or office in the community; s/he may want a few techniques to make office operations more efficient. An early understanding of what the evaluation can and cannot do will avoid misunderstandings in the future.

Evaluators found the GOC's easily understood and implemented. In fact, experienced evaluators commented that one of the major failings of former evaluation attempts was a lack of direction (i.e. no clearly understood goals, objectives, criteria), which resulted in the inefficient use of time and a random assortment of topics covered.

It should be pointed out that the GOC's selected for DEP's evaluation design will generate some disagreement around the country. For example, the design does not take a stand for or against horizontal representation; it does not specify an ideal caseload; it does place a burden on the defender to take an active role in the criminal justice system in his/her community. Further, DEP took the perspective that the client of a public defender should be given the same high quality representation as that given the affluent client of a privately retained attorney. This places an extra burden upon some defenders who are already overwhelmed by a high caseload. It was DEP's feeling, however, that GOC's should consist of what defender offices should be working toward, rather than those activities which offices are minimally able to carry out at present.

#### Indicators

The relationship between Indicators and Criteria for compliance, Objectives and Goals must be clarified and explained to evaluators. The term "Indicator", itself, is relatively foreign to attorneys; it should be related to the term "evidence" to have a meaning. Since some previous evaluations did not emphasize the need to marshall evidence to fully support all findings and conclusions, the approach of accumulating bits and pieces of evidence before conclusions could be drawn was new to some evaluators. DEP noted, and attempted to alter, a tendency on the part of some evaluators to take one Indicator (e.g. one interviewee opinion on a topic) as the answer to a question, rather than seeking a variety of Indicators (e.g. opinions, observations) and sources (e.g. 3 interviewees) to substantiate their conclusions.



It is impossible, and undesirable, to enumerate all potential Indicators for all GOC's. Defender office settings and circumstances vary enough so that what is applicable in one office may not be so in another. The Indicators selected for inclusion in the Handbooks are those which are frequently found in the larger proportion of defender offices. They have been pared to a manageable number. It cannot be overemphasized, however, that new Indicators which are particularly relevant to the evaluatee office will emerge in the course of every evaluation. They should be sought and utilized.

Indicators should be viewed by the evaluation team as guides to the information to be collected. The gathering of such information is only one step in a process of judgment which must rely on their professional expertise for selecting, assimilating and synthesizing this information gathered.

There is no one-to-one correspondence between Indicators and GOC's. This is a very difficult concept for evaluators to understand. For example, a statistic on types of sentences given clients over the previous three months may contribute to understanding disparities between defenders and privately retained attorneys (the Objective of "Competence"), to the defenders' search for sentencing alternatives (the Objective of "Zeal"), or to the defender's relationship with community-based organizations (the Objective of "Community Education"). Although the Handbooks attempt to organize potential Indicators according to their most likely GOC relationship, the categorization is for organizational purposes. The overlap among GOC's must be highlighted.

On final consideration, it was appropriate that DEP did not impose "numbers" on defender offices. National workload standards were not established; a "correct" percentage of times the Defender Office goes to trial with cases was not given; an "appropriate" number of dismissals and acquittals was not suggested. This is not to say that these statistics were omitted from consideration at each site. While they were calculated and used by evaluators, it was with the understanding that "right" numbers were not known. The circumstances at each site must be viewed as unique. Unlike numerical terms which appear in other defender standards (e.g. N.A.C. report), all statistics gathered in this design are reviewed as points for further inquiry, and not as Criteria for Compliance in their own right.

An unanticipated consequence of the Indicator lists was the "scientific" appearance they gave the Handbooks. Evaluators, presented with these previously thought-out bits of evidence and ideas, took the entire evaluation process seriously. Although the structured direction of the Handbooks offered was resented by a few "seasoned" evaluators at the start of the evaluation, by the end of each site visit the structure imposed on the team was considered of great help in reaching final conclusions and recommendations and in writing the Final Report.

## Methods of Gathering Data

The test evaluations were especially revealing in terms of the strengths and weaknesses of the methods used to gather data. DEP is hoping to minimize the latter (i.e. weaknesses) in its Handbook revisions.

Team Members were consistently impressed with the type and comprehensiveness of the data they were expected to gather and utilize for their evaluation. A common feeling expressed at the conclusion of each evaluation was that, in contrast to previous evaluations, they now had the supporting documentation upon which to base their conclusions and recommendations.

Background Information. The three test evaluations revealed that the pre-evaluation preparation (PEP) materials mailed to defender offices for completion were too extensive, not within the office's ability to provide, and not necessarily needed by the evaluation team. The approach DEP took in requesting material once again proved the adage that "too much is as bad as too little." The problem would have been minimized had there been a greater time interval between the request for an evaluation and the pre-site visit. The consequence of closely budgeted DEP time was to exaggerate the problem and, to DEP's advantage, make it highly visible.

To illustrate: Court and prosecution case statistics requested were overly detailed. Attitudinal questions regarding "political climate" or minority group activities were more appropriate during the pre-site meeting discussions. Some defender office statistics requested were unavailable; kept only by offices which were required to do so by their local funding agency.

Given three test evaluations against which to compare reactions and responses to the PEP materials, changes have been considerable: materials have been shortened; some statistical detail has been omitted; information about agencies outside the defender office has been limited; attitudinal questions have been revised, and many shifted to on-site discussions.

Statistics. The time pressures under which DEP worked did not allow for a full testing of the relevance and usefulness of a statistical analysis of defender files and the court docket. The analysis was not available to the first evaluation team until after the on-site visit was completed. The second evaluation team received statistics which were hurriedly analyzed by Staff during the course of the evaluation. At the third site, the analysis was presented to the team on the first day of the evaluation, during its training session. Its format, however, did not allow for easy comparability between the defender and private attorney.

These experiences have taught DEP a great deal about the conditions under and methods by which analyses should be conducted in the future. The codes developed for both the defender case file and court docket studies have been considerably shortened in the revised Handbook. Omissions include items which: (a) were not readily accessible in files; (b) were irregularly noted by the defender or court; (c) entailed independent judgment on the part of the coders; and (d) were not as important to the evaluation as initially perceived. These changes will undoubtedly affect the time needed for data collection, the accompanying cost for coding and keypunching, and the reliability of the information gathered.

The initial program for the data, i.e. the Statistical Package for the Social Sciences (SPSS), has been revised to reflect coding changes. An early revision took place between the first and second evaluations to allow for an easier analysis of the data.

A minimum of a month is needed to undertake the statistical studies. It is doubtful that this time period can be meaningfully reduced because of the unorganized state in which many defender files are found. Whichever group is used for the job, only a small number of employees should be involved in the data gathering and coding process; this increases coder reliability.

A major issue for any future evaluation which utilizes a statistical analysis is one of "quality control". DEP is suggesting that a Staff person, or someone attached to the evaluation team, be at the site during the first few days of data gathering to train coders and be available for questions. This same person should insist upon reliability checks by the contractee throughout the coding operation, in addition to conducting such checks him/herself toward the end of the coding process. The contractor must assume responsibility for checking coder reliability and be penalized if the programmed data appears unreliable. Although suggestions for reliability checks were in the Handbooks used by the three site contractors, they were not made explicit in the DEP contract and appear to have been omitted in the contractor's work. Time pressures were a contributing problem, as was the lack of demands placed by the DEP staff.

It was also advisable that Staff, and/or contracting supervisor, review the types of files the coders will be reviewing, in order to more intelligently train them for their job. The form and extensiveness of case files and notations varies considerably from jurisdiction to jurisdiction. To adequately judge the extent and nature of some of the problems to be met by the coders, the supervising agent should make a preliminary assessment of the state-of-the-files.

It is suggested, further, that more than one type of

charged offense be reviewed in the statistical analysis aspect of the evaluation. In two of the evaluations, all charge types handled by the defender office were sampled. A sample, therefore, ranged from assault through robbery through burglary. In one case, 15 major categories were delineated. In the third evaluation the Team Captain suggested sampling one typical charge category, in both the lower and superior courts. While this was logistically easy, the Team Members did not feel that comparisons on one charge, between defenders and privately retained attorneys, were valid enough to warrant generalizations about each group's performance. On the other hand, the evaluations which covered all charge types were too large and unwieldy for good comparative analyses. It is suggested that three frequently handled charge types in the lower and superior courts be considered for inclusion in any sample, for both easy comparison and sampling. The three charge types selected should vary according to defender practice in the specific office being evaluated.

It is also suggested that Chief Defenders be given a copy of the statistical data. In each case they expressed interest and felt that they could benefit from the information.

The mere act of undertaking a statistical analysis seemed to have a positive, educational value for the defender office. In one case, the Chief Defender used his own staff to gather data which was not on the DEP code, but which he suddenly realized would be of help to him in a budget hearing. In another office the Chief Defender requested a copy of the code so that he could use it selectively at a later time.

From a practical standpoint, the statistical analysis should be an optional part of the evaluation. Attorneys, given the best of training in the uses and abuses of statistics, persist in distrusting them unless they themselves have had complete charge of their collection and analysis. This is a prejudice which is and will be difficult to counteract in future Team Members. From a research/evaluation standpoint, a statistical analysis provides an excellent starting point for in-depth interviews and, when reliable and valid, an important type of data which can be used in complementary fashion to the qualitative data gathered.

Observation. Attorney team members placed greater significance on observations as a data gathering technique than did DEP. Although guides to observation under various situations were well-structured in the Handbook, DEP was initially skeptical of relying on observation (see p.14). The attorney Team Members felt, however, that under the hectic conditions under which defenders typically work, the problem of "altering" behavior is minimal. Their own expertise as attorneys, they contended, could be counted on to uncover such ploys if they occurred.

The viewpoints of and arguments put forth by the Team Members were well taken. DEP has adjusted its sample on-site time schedule to ensure that observations are given a greater share of time during the evaluation.

Interviews. Interviews, as expected, comprised the greatest proportion of time during the on-site evaluation. Feedback from anonymous interviewees on questionnaires mailed directly to DEP substantiated that questions were relevant and Team Members objective.

One problem encountered was questioning certain individuals on too many areas. This resulted in long interviews and interviewer and interviewee fatigue. By the second evaluation, however, this situation was rectified by limiting interviews to no more than four Objectives. This strategy was successfully continued in the third evaluation; the Handbooks reflect the revisions.

Interview formats have been altered to reflect the four areas covered per interviewee. In addition, their initial bulk has been limited to no more than two pages per interviewee. These revised formats place before the Team Member the relevant Criteria for Compliance for each Objective covered, other questions deemed relevant, specific Indicators per Objective, and observations and record searches (where appropriate).

Clients should be viewed as one source of input and information, but the evaluation of a defender office cannot and should not rely heavily on client/consumer attitude. Interviews with former and present clients, in prison, on the street, or in half-way house facilities, conducted during the development of the evaluation design were difficult to assess in terms of the validity of the comments made. "Was your attorney in contact with you weekly?"; "Did your public defender do a good job for you?" There is little doubt that given the best of defenders, if a client is "doing time", he/she feels that a "paid for" attorney would have gotten them a "better deal". While this does not hold true in every case, it was uncovered frequently enough to have DEP conclude that client interviews should be used as one of a number of data sources. Client viewpoint must be carefully analyzed to separate out highly subjective opinions (e.g. "good job") from fact ("visits weekly").

Records. The record reviews suggested in the evaluation design (e.g. jail visitor logbooks, defender calendars) were data sources which appeared to enrich the team's informational base, yet which would have been omitted without specific inclusion in the Handbooks. Although the data gathering instructions accompanying these suggested sources of data were not rigorously followed, the sources themselves were used and added material upon which to base evaluation conclusions.

It was originally suggested by DEP's Advisory Board that the evaluation team review 5-10 case files of each attorney in the defender office. This was a practical impossibility, in terms of time and money, and unnecessary from the viewpoint of attorney Team Members. Although evaluators were instructed in the Handbooks to review all of these cases, teams used their discretion in terms of the number of cases reviewed. Typically, each defender interviewed, which may not mean every defender in the office, opened his/her files to a cursory review and was questioned in detail on two or three selected files. Team Members used their discretion and reviewed that number of cases they felt "comfortable" reviewing; that is, the number needed to satisfy them about the quality of representation offered. This, of necessity, will vary from office to office, and from team to team. It worked well from a logistics standpoint and appears to have merit from an evaluation perspective.

#### Evaluation Logistics

Time Frame. The time frame established by the evaluation design had to be cost-conscious and consider the constraints upon potential Team Members. A Defender Office can not be expected to devote \$50,000 to an evaluation which might take one month on-site. While disgruntled evaluatees might feel that a one-week on-site period in a medium-sized office is too short a period for a "good" evaluation, the design had to weigh validity against cost and arrive at a format and time intervals which satisfied both needs. A five to seven day on-site period appears to suffice for the small and medium-sized office, respectively. The cost of a typical evaluation appears in Appendix E.

The difficulties involved in finding a team which can give from five to seven consecutive days to an evaluation are great, but they can be resolved. DEP discovered that it is difficult, if not impossible, to assemble a team for longer than a seven day on-site evaluation; and that having all team members together on-site throughout the evaluation appeared critical for maximizing the evaluation's reliability and validity. The daily interchange of ideas and the consensus reaching process among Team Members is essential to an acceptable evaluation.

Based upon DEP's three test experiences, a defender office evaluation, from request through final report, should take approximately four months. This Time period could be shortened by a month if a statistical analysis was not undertaken. While the design's suggestions for report writing allow each Team Member to leave the evaluation site with a group consensus on conclusions, recommendations and supporting documentation, time off-site is needed to flesh out the report, have it reviewed by other Team Members, edited by the Team Captain, and mailed to the Defender Office.

Team Size and Evaluation Length. The choice of team size and length of on-site evaluation for the three sites visited worked well: for the small office, 3 team members, on-site for 5 days (1 day for team training, 3 days for evaluation, 1 day to reach consensus and outline for the report); for the medium office, 5 team members, on site for 6-7 days (1 day for team training, 4-5 days for evaluation work, 1 day to reach consensus and outline the report).

While the present design format and team structure provides a good yardstick against which to structure future evaluations, it cannot be considered appropriate for all cases. A rural office which covers a large geographical area might entail more on-site time merely because of the travel involved. A medium-sized office which is concerned primarily with a management analysis might need a team with more management than defender expertise to remain on-site a shorter period of time. In short, for what the DEP design can accomplish under "normal" office conditions, the time and manpower allotment is sufficient. Variation in the use of the design under differing conditions would entail design alterations.

Team Composition. Thirteen individuals were chosen as evaluators according to DEP's original guidelines (e.g. experiential background, innate ability, and interpersonal skills), and varied in composition for purposes of DEP observation. The following outlines team composition:

	<u>Small Office</u>	<u>Medium Office</u>	<u>Medium Office</u>
Team Captain	Chief Defender, former evaluator	Pvt atty, former Chief Def, and former Team Capt	Administrator, former Chief Defender and Team Captain
Management	Chief Defender, former evaluator	University professor, management and criminal justice specialist	Administrator of civil law firm, mgt consultant
Community	Citizens group representative	Paralegal in Def Office, Former offender	Dir, Rehab unit in Def.Off, former Soc Wkr
		Administrator, Former defender	Appellate def, Former team member, eval experience
		Appellate def, prior team member	Defender, no eval. exper.

The following observations on team selection reflect DEP's experience:



1. Teams functioned well without consisting entirely of attorneys. Diverging from past evaluations, each team had one or two non-attorneys. Feedback from attorney team members was that the difference in perspective the non-attorneys brought with them, although often the basis for long discussion, was healthy and should be encouraged. Staff observations substantiated this.
2. Defenders/former defenders performed well as Team Captains. While there is no group against which to compare the performance of these attorneys, staff who observed each evaluation felt it was desirable that a Team Captain be intimately familiar with problems unique to defenders. This led to easy rapport with the evaluatee and gave credibility and legitimacy to the entire team.
3. Differing types of management analysts are capable of performing well on the evaluation team. The common denominator among the three used by DEP was a thorough understanding of management principles and a familiarity with the criminal justice system. Workstyles, however, differed, and these differences were considered in the Handbook revisions to accommodate two distinct styles of management analysis.
4. The "community person" should be chosen with the demographics of the defender population at the site to be evaluated in mind. DEP selections for this role varied considerably, but in each evaluation the individual chosen performed well. Sensitivity to the "consumer's" perspective, and the ability to establish rapport with individuals and groups who have information in this area, are essential to this role.
5. The mix of experienced and novice evaluators worked well.
6. Attention should be paid to the writing and editing skills of potential Team Captain. It is unlikely that a potential Team Member would refuse a position on an evaluation team on the basis of poor writing skills; it is likely, however, that some Team Members will have poor facility with the written word. The Team Captain, then, must be capable of rewrites.

DEP was particularly satisfied, at the conclusion of the three evaluations, with the positive feelings of evaluators toward their experiences. Each Team Member stated that s/he would want to participate in future evaluations.



### Evaluation Format

Team Training. Team training is more important than most Team Members like to think. Left to their own devices, most evaluators would prefer to choose their own direction, reach their own conclusions, and have them accepted as reliable and valid en face. Team training is viewed as an inconvenience. From the staff's viewpoint as observers, however, and from the opinion of most of the evaluators towards the end of the evaluation period, team training did have important consequences:

1. Each team member, from the outset, shared a common understanding about the scope and purpose of the evaluation, and about the respective roles to be played;
2. Each Team Member shared the same understanding and focus of the evaluation itself (i.e. the Goals and Objectives used to assess a defender office);
3. The evaluation design, to be an efficient method of organizing people and time, was clearly understood by all; and
4. It allowed a Team Captain to become acquainted with his members and assess their abilities.

There are other functions which team training performs. For example, it reviews substantive information of which evaluators should be aware before undertaking the evaluation (such as information on the PEP materials), but which they cannot be counted upon to have read or assimilated. It also allows staff to highlight evaluation (vs. legal) concerns, such as the need for gathering and utilizing data to support evaluation conclusions. The integration of management and other concerns is an area which should be discussed with Team Members who will be viewing different aspects of the same occurrence and will be writing up a common evaluation report. Team training, by its very nature, should stress the interdependency of each team member in the search for consensus.

DEP did not come close to its "ideal" training session until the third evaluation. On the basis of the three evaluation experiences, it is suggested that one full day (approximately five hours) be set aside for training. The Handbooks contain suggestions for team training procedures and substance.

Training a large group of evaluators in one session could be seen as more efficient than the approach presently undertaken, in terms of staff time and numbers of trained potential evaluators. However, when other practicalities are considered (e.g. the availability of trained evaluators at the desired evaluation time, the time needed on-site to review the Handbooks

and local system materials), it is difficult to choose one form of training over another. Circumstances will have to guide the ultimate decision.

Scheduling Interviews. During the pre-site visit of the Team Captain and Staff to the Defender's Office, a list of potential interviewees was gathered according to requirements previously determined by Staff. The few minor problems the evaluation teams encountered were resolved on-site and appear in the revised Handbooks. The main points to be highlighted are as follows:

1. At the time an appointment is made, Staff must verify that the person being considered as the interviewee is the most relevant individual(s) in the agency. Evaluators on-site periodically discovered that "the person down the hall", while in the same agency (e.g. probation), was often the more knowledgeable individual for interviewing purposes.
2. Constant supervision should be kept over the interview appointment schedule as it is being completed. Lack of such surveillance resulted in an overabundance of interviews with bailiffs in one site and an inordinate number of police interviews in another.
3. The daily interview schedule should allow for ample debriefing time at the end of each day. Interview schedules constructed for the first evaluation ended late each day and resulted in debriefing sessions which lasted late into the evening. Schedules for the third site evaluation minimized the problem.
4. Each interview format should be manageable in terms of the scope of items covered and the mechanics of its use. The initial interview formats prepared proved too lengthy for a one-hour interview and too cumbersome in paper quantity. Revisions were made on both successive evaluations and a highly manageable final format appears in the revised Handbook.

The third evaluation, from the standpoint of logistics, went very smoothly. DEP suspects, however, that an additional three to five evaluations would be helpful to further refine the administrative and logistical aspects of an evaluation.

Daily Debriefings. The daily debriefing sessions are a particularly significant part of each evaluation. These sessions, observed by DEP Staff at the evaluation sites, accomplished what

was anticipated for them. They helped the team develop a comprehensive picture of the defender office, criminal justice system and community. They provided for the exchange of information and insights among Team Members. They enabled the Team Captain to monitor the progress of the evaluation, coordinate the activities of the Team Members, and alter the direction of the evaluation as appropriate.

Logistically, refinements were needed, and made, during each evaluation to enable sessions to make the most efficient use of time and information. The exchange of information among Team Members was done orally, rather than both orally and written, as originally planned. Each day's interview assignments were made the evening before, rather than on an originally planned weekly basis. By the second evaluation it was clear that daily discussions could not repeat the cumulative learnings of the previous days; they should only add to or comment upon prior debriefings if new information arose. In short, the "normal" process and content flow of each debriefing session throughout the three evaluations was understood by Staff only as it occurred, and changes were made accordingly during the evaluations and in the Handbooks.

Exit Interview. The exit interview with the Chief Defender changed somewhat from the original. DEP staff had initially anticipated that this interview would be no more than a courtesy "thank you"; the evaluation team had not yet met to reach consensus and conclusions and could not be expected to discuss these with the defender. Two evaluation teams, however, chose to use this time to ask the Chief Defender to discuss and clarify substantive issues about which the evaluators needed further explanation and resolution. The last evaluation team conducted this session only after a short team meeting which was used by each member to decide what information s/he lacked or was unclear in his/her area of evaluation responsibility. These techniques worked well and were incorporated in the final evaluation design.

#### Integration of Management

The management analyst on each team was particularly important to each evaluation. This individual brought a perspective and set of questions which were generally lacking in other Team Members. An unanticipated consequence of this individual's presence was the amount of information shared with Team Members which they felt would be helpful in their respective offices.

Each management analyst (defender/manager, professor and criminal justice consultant, civil law office manager) felt comfortable with the basic design. Working styles, however, differed considerably. The revised Handbooks, therefore, have

abandoned the original checklist approach -- while retaining the original content -- and present two alternative approaches to data collection which evolved from the site experiences.

1. Objectives Approach. This approach is designed for someone who is very familiar with the Objectives and work of the defenders. The approach divides each Objective into the management components of planning, organization, administration and control to guide the collection of data.
2. Functional Approach. This approach is designed for the individual who is more familiar with management than with defenders. The collection of data is not done by Objective, but by functional areas of management: paper flow, personnel, training, resources, public relations.

Each management analyst is encouraged to review both approaches and choose the one best suited to his/her own style; sections of both, however, may be combined since the same information is covered in the two approaches.

The manner in which management information was originally integrated into the three test evaluations created different problems at each site. Part of the problem lay with the design, which called for certain areas to be evaluated by both the management analyst and the Team Members assigned to specific Objectives. When this situation arose, the same interviewee was being seen at least two times, with similar information being covered. This made daily debriefings needlessly redundant and led to a questioning of roles among Team Members. Part of the problem also lay with the Objectives themselves, and the needed interdependence and intertwining of management and legal issues to fully understand Defender Office performance. These problems were largely resolved in the revised Handbook by better defining the scope of each Team Member's work, advising the Team Captain to avoid unnecessary overlap in debriefings, and outlining a final report table of contents which attempts, for organizational purposes, to separate management and legal issues.

#### The Final Report

Rating the Office. DEP's initial hesitancy over the strict use of ratings was verified in each team's approach to the problem. Although given the scale with four gradations, each team felt its best use lay as a guide to discussion, something to help maintain the focus of the evaluation on specific Criteria, and as a guide to writing the final report and reaching general conclusions about Objectives. The rating scale was not found useful, however, as a quantitative summary of performance.

Researchers/evaluators who are committed to the use and importance of numerical summaries will view DEP's approach with unease. This should be taken as healthy disagreement in a field where subjectivity and value judgments play a larger role than often assumed.

Report Logistics. The original evaluation design called for the Team Captain to write the entire final report based on outlines submitted by Team Members prior to leaving the evaluation site. This was structured to save time (e.g. a Team Member being late in submitting his/her report) and maintain consistency in the writing and approach. The three evaluations revealed, however, that each Team Member knew his/her area very well by the end of the evaluation period and that it would be difficult for any other member to translate another's knowledge. Not only were Team Members unexpectedly interested in writing their own section (to be edited by the Team Captain), but the Team Captains felt that it would be difficult to take full responsibility for the final report, given the realities of their workloads and responsibilities in their offices. Responsibility for the evaluation report, then, was delegated to Team Members according to each member's assigned area of concern. The Team Captain was responsible for providing background information, a methodology section, and a final editing for style and substance.

The process by which consensus was reached on Criteria and Objectives worked well. The final day on-site was reserved for Team Member discussion on issues, reaching a consensus on office "ratings", and developing a skeletal written outline of conclusions, supporting documentation, and priority recommendations.

Reliability and Validity of the Final Report. Evaluation sites were selected for many reasons, one of which was to provide an independent criterion against which to judge the validity of the final reports. Las Vegas, Nevada was viewed by defenders around the country as a "good" office; Columbus, Ohio, a "poor" office; Monticello, New York, an unknown quantity. These gross assessments, known only to DEP staff, were substantiated by a rough assessment of whether Team Members felt an office was meeting the 11 Objectives evaluated. DEP staff categorized Team Member conclusions on each Objective, for each site, on a three point scale. The following table resulted:

<u>Site Visited</u>	<u>Objectives Met</u>		
	<u>Totally</u>	<u>Partially</u>	<u>Not At All</u>
Las Vegas, Nevada	7	3	1
Columbus, Ohio	3	5	3
Monticello, N.Y.	7	1	3

While the measurement system used by DEP Staff is certainly gross in nature, Team Member conclusions in the final reports appear to validate the responses of defenders around the country that Las Vegas, Nevada has a "good" office, and the Columbus, Ohio has problems.

It is interesting to note that the "halo" effect did not occur. That is, in an office viewed by Team Members as generally "good", problems were still obvious; the "good" rating didn't distort the team's conclusions about all Objectives. In a similar manner, the Columbus, Ohio rating was not altogether poor; good points were noted.

DEP attempted to validate the findings of its statistical studies in two of the three sites against statistical information gathered earlier in the year by another NLADA project, the Indigent Defense Systems Analysis Project. Unfortunately, the validity of the IDSA statistics could not be validated and many questions of similar nature in both studies were coded differently.

There is some evidence of the overall reliability of the design:

1. Mail feedback received anonymously by DEP staff from interviewees agreed that questions asked were relevant and that the interviewers were professional, objective and open-minded;
2. Although Chief Defenders took their opportunity to correct factual inaccuracies and team misconceptions, only one Chief Defender felt the need to attach an addendum to the final report. The addendum primarily highlighted the fact that problems found in the office were there because of the previous defender, and that as a relatively new defender, a number of significant changes had already been started; and,
3. The design was closely followed by Team Members at each evaluation site.

A few Criteria for Compliance, however, have been made more precise in the revised Handbooks, as were Team Captain instructions about the use of these Criteria as guides to their discussions; both changes stemmed from DEP's comparative review of the three final evaluation reports.

To conclude, there is little evidence to doubt the general reliability and validity of the evaluation report produced. Continued testing, however, might uncover specific areas needing refinement.

## SELF-EVALUATION MANUAL - CONCERNS AND FINDINGS

The Self-Evaluation Manual was met largely with positive responses from defenders. The following were typical comments:

- questions are good
- indicates excellent insight into the workings of a public defender office
- comprehensive and covers the majority of problem areas that confront defenders
- will be a very important management resource that can withstand program changes
- discussions of management and administration are logical and easy to follow
- practical solutions to obtaining information needed

Some of the respondents stated that their office was implementing some suggestions which appeared in the Manual, and that the Manual raised issues which they realized they hadn't given enough consideration in the past.

Criticisms of the manual included:

- too general in approach
- not very practical
- many of the suggestions would require data gathering resources which few defenders have or will have
- suggestions disregard limitations of time, money, and manpower placed upon the defender

To a great extent, the criticisms of the manual reflect: (a) defender confusion about the difference between a self-evaluation manual and a technical assistance handbook; and (b) a real, yet unmet, need in the defender world for specific forms of help in responding to specific problems. Perhaps the most discouraging aspect of the criticisms received was the concern in the field with finding tools to maintain the status quo in offices which are underfunded. The manual's attempt to raise issues, and to remind defenders of their full responsibilities seemed impractical to some in the face of already existing obstacles.



It is DEP's feeling that a self-evaluation begins a process of establishing expectations and asking defenders to routinely ask themselves whether their office is meeting them. The next step in the process of upgrading the performance of defender offices would be a series of technical assistance handbooks, each focusing on select issues which are problematic to the large proportion of defenders.



**END**

7 tables/more