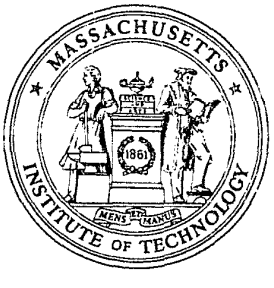
OPERATIONS RESEARCH CENTER



SACHUSETTS INSTITUTE OF TECHNOLOGY

FOURTH QUARTERLY PROGRESS REPORT

OCTOBER 1, 1978 - DECEMBER 31, 1978

AN EMPIRICAL STUDY OF METHODS USED IN CRIMINAL JUSTICE EVALUATIONS

BY

Richard C. Larson

Principal Investigator

and

Co-Director
Operations Research Center

OPERATIONS RESEARCH CENTER
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139

"An Empirical Study of Methods Used In Criminal Justice Evaluations," Grant No. 78NI-AX-0007, funded by the U.S. Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice.

Introduction

During the fourth quarter we completed the analysis of reader responses to the questionnaire used to examine our sample of 200 reports of evaluations of criminal justice programs. The bulk of this quarterly report consists of our executive summary of our findings. A fully detailed report on the first phase of research will be completed by February 15th, and will be forwarded to LEAA shortly thereafter.

Personnel

Of the project staff described in the third quarterly report, Vicki Bier, Tim Eckels and Ed Kaplan continued to work on the analysis of the evaluation reports. They each wrote up a presentation of their findings with interpretations, as did Cheryl Mattingly and Nancy Reichman. Cheryl and Nancy are no longer officially with the project, although they are both available on an informal basis to follow up the research and to aid in preparation of the final report. Meanwhile, Nancy is still involved in criminal justice research, and Cheryl is continuing to develop her interest in process evaluation.

Work In Progress

We have just completed a first draft of the questionnaires to be sent to the authors of the 200 reports ("evaluators") and the alleged consumers of those reports (program managers and funders of the evaluations effort). Upon review by Jack Fowler, a survey research specialist at the University of Massachusetts at Boston, and by the Committee on

Human Subjects at M.I.T., these questionnaires should be mailed out to respondents by March 1.

Herewith follows a summary of our findings to date, based on a reading of 200 criminal justice evaluations (this sample and the questionnaire used were discussed in earlier quarterly reports; a brief description of the computer program used is included in Appendix B).

Executive Summary

The aim of a "good" evaluation is to assess the value of a particular program. An evaluation, if done properly, must be useful either to program managers, program clients, policy-makers or other social scientists. Although any single evaluation might be hard pressed to be "useful" to all of these groups at once, evaluators should strive to meet the needs of decision-makers in at least one of them. Whose needs the evaluators may attempt to address depends on the stated purpose of the evaluation (as commissioned), the evaluators' own motivations, the nature of the program, and the evaluation's financial sponsor.

Be that as it may, we have established six criteria by which the usability of a given evaluation may be assessed vis a vis its audience. These criteria are:

1. degree of comprehensiveness;
2. appropriateness and application of methodology;
3. flexibility of design and implementation;
4. amount and quality of communication between program staff and evaluators;
5. accessibility of the final report;

6. demonstrated awareness of institutional and environmental issues affecting the program and the evaluation.

An evaluation conducted solely for the purpose of meeting legal or administrative requirements should be revealed as such when submitted to measurement against these criteria.

We are most interested in how the needs of the first three groups mentioned above -- program managers, program clients, and policy-makers— were met by the 200 evaluations in our sample. What follows is a report of the results of the first stage of research being done to a determine evaluation usability. This analysis is based on a directed reading of the evaluation reports themselves. Further questionnaires are being designed and will be sent to the authors and consumers of the evaluation to complement these results.

Inputs

Just as knowledge of inputs is necessary to evaluate a program, so knowledge of inputs to an evaluation is necessary to fairly evaluate that evaluation. Inputs that have been found to be important include the level and type of funding, staffing patterns, timing, etc.

Only eight evaluations in our sample indicated the percentage of the program budget alloted to evaluation, and only four reported the total funding for the evaluation. This response prompts us to ask if budgetary information should be included in the evaluation report itself. Although there is little comment on this issue in the evaluation literature, we note that evaluation is conducted as a service; evaluators should be accountable for their product. One should be able to

assess the value of an evaluation given known budgetary constraints. Budgetary information would also provide a useful indicator or model for someone who is planning or undertaking a similar project.

The time spent on evaluation is also an important input to consider. Relatively few evaluations in our sample documented the time spent on the evaluative effort. There were no indications that the length of time varied significantly among agencies responsible for the evaluation. This is somewhat surprising as one would expect to find longer evaluations from established research firms such as the Police Foundation or the Urban Institute. We suspect, although we did not specifically test for it, that there may be some variation by type of evaluation study. It is impossible to determine whether our sample over-represents the short-term evaluation. One could conjecture that those evaluators constrained by a short evaluation period would be more likely to highlight the time alloted as a means of rationalizing inconclusive results. Knowing the time frame puts the evaluation in context, which enhances its overall utility.

The vast majority of the 200 evaluations in our sample were conducted by people who were not associated with the program itself; only 22 percent of the evaluations were conducted by in-house personnel. Although for the most part the evaluators appeared to have some kind of social science training, generally it was difficult to pin down their background characteristics and professional affiliation. Knowing who conducted the evaluation is useful in that it may help indicate the possible biases of and constraints on the evaluators and their relations

with the program staff.

The readers* tried to determine from the evaluation reports themselves when the evaluations were planned. Fifty-two percent of the evaluations were planned during program implementation, 37 percent before program implementation, and 11 percent were planned after the fact. Examination of the planning of evaluation by agency responsible for it revealed that evaluations undertaken by State Planning Agencies were somewhat more likely to be planned prior to program implementation than evaluations conducted by LEAA at large. This may be a direct result of the function of State Planning Agencies. Almost all evaluations in our sample conducted by the Police Foundation were planned prior to program implementation. This fact most likely reflects the demonstration nature of the projects undertaken by the Police Foundation, in which evaluation is often considered as significant as the program itself. The evaluation literature suggests that evaluations should be, at the very least, in the minds of the program developers. There is, however, some danger in completely designing an evaluation before the program has had a chance to settle down. This factor proved to be a problem in several of the evaluations where program instability was cited as a reason for inconclusive results. In such instances, flexible evaluation design can be of great importance.

Examination of the 200 criminal justice evaluations in our sample revealed that slightly over half of them were designed to yield

*Throughout this report, the term "reader(s)" refers to the people on our research term who read the 200 evaluation reports. The term "evaluator(s)" refers to the authors of those reports.

information that would be useful in a broader context than just evaluating the specific program under study. In general the evaluations that were broad in scope seemed to be set up as demonstration projects or so called "state-of-the-art" studies for a particular field. Given the different goals, objectives, resources and perceptions of the evaluative purpose, we expected to see some differences in scope among the agencies conducting the evaluations. Evaluations undertaken and designed by the larger research organizations (Police Foundation, Urban Institute, etc.) were more likely to have a broad focus than evaluations conducted by LEAA and the State Planning Agencies. With the exception of evaluations conducted on team policing programs and on technical innovations, there was remarkable consistency among the subject areas in terms of the scope of the research designs. This suggests that goals, resources and the perceptions of purpose vary more by supporting agency than by subject area. This seems to confirm the importance of the institutional setting of evaluation.

Process

Reviewing the process of an evaluation is valuable in that knowledge of the evaluation process allows us to assess the legitimacy of the results and conclusion of the evaluation. Naturally this is important for making decisions about the future of the program. What we refer to as the evaluation "process" includes the design of the evaluation, its coverage of the program, the methodology used, the flexibility of implementation and interaction with program staff during the various stages of the evaluation. Thus we have included under process those questions which treat the comprehensiveness of the evaluation, the methodology used, the

flexibility of the evaluation design and procedures, and the degree and quality of communication between program staff and evaluation staff while the evaluation was being carried out.

Comprehensiveness

To be considered comprehensive, a given evaluation should focus on as many aspects or stages of the subject program as possible, from design through implementation, to effect on the client population. An evaluation which focuses on any one of these aspects in isolation is often less useful for either internal (program) or external (policy) purposes, than one which is more comprehensive (although special purpose evaluations can clearly be highly useful if carried out during the relevant stage of the program). Obviously, financial and political constraints may restrict the potential coverage to be provided by the evaluations.

Goals

Researchers in the evaluation field have often commented on the importance of clear goal specification. It is argued, first of all, that a program will operate more efficiently and effectively if the managers and staff are working toward explicitly shared aims. Secondly, for most forms of outcome evaluation, the evaluator will need a clear description of the objectives that the program is pursuing. The delineation of objectives and the measurement of their attainment are rarely simple procedures, however. Established goals may be absent at

the commencement of a program, or there may be substantial disagreement about what constitutes the "official" or "proper" goals. What is crucial in any case is that adequate communication and negotiation take place among the various actors in the evaluation process. Two questions were used in an attempt to investigate this issue:

14a. Did the evaluators consider whether program goals were clearly specified?

14b. Do you feel they were clearly specified?

Despite the fact that we were attempting (in the second question) to look through the evaluation to the character of the program itself, it was often apparent when there was disagreement over goals, or where there had never been a clearly constructed set of goals.

The majority of programs did appear to have clearly stated objectives, and the majority of evaluations did discuss and/or reiterate them. Breaking down the results by topic area, however, we discovered that these favorable tendencies occurred most often in the "police logistics" studies. Given that these programs are usually more operational and focused than programs in areas such as training or corrections, this result is not too surprising. We can see that our concerns here will be mostly with the non-logistical program areas where goals are often not displayed or considered by evaluators. This seems to be particularly the case in correction programs, in Assistance and Training Units, and in the Police-Community Interaction programs.

A rather striking result appears in the pre-trial programs, where all eight programs specified goals succinctly. It should be added that

these programs were located in a variety of areas and that the evaluators were sponsored by a number of different agencies. Considering that these programs are often similar in structure to other correctional endeavors, it may be that this was more or less a chance relationship. On the other hand, it may be that because pre-trial programs are more vulnerable to criticism due to the "presumed guilt" controversy, they tend to be more definitive about their purpose.

Related to the issue of goals is that of "target" population. Slightly more than one-third of the evaluations in our sample discussed the issue of whether the target population and the program were well matched. Another third did not seem to consider the issue; in the remaining cases the response of "not applicable" generally meant that there was not enough information given in the report for an analysis to be made, or that the program was of a different nature.

Interestingly, in our sample, the evaluators focused on discovering the population which the program was best suited to serve, as much as evaluating the program in terms of how well the intended population was being served. This may be because the target population was often not initially very clearly defined.

With the exception of pre-trial release and AVM/CAD* systems, there was little variation among substantive areas in the percentage of evaluations which considered whether the program was directed at the

---

* AVM refers to Automatic Vehicle Monitoring Systems; CAD refers to Computer Assisted Dispatch Systems.

the appropriate target population. The range was between 27 percent and 50 percent, with most scores clustering around the average of 35 percent. Not surprisingly, 83 percent of the pre-trial release evaluations considered this issue. The issue of who should be diverted or receive alternative treatment to incarceration was generally of central importance and some of the evaluations in this area were explicitly designed to focus on this issue.

The question of whether the program was directed at the appropriate target population was "not applicable" in nearly one third of the evaluations in our sample. The primary reason for this was that not enough information was provided in the evaluation to be able to answer that question. Even in cases where the evaluators made assessments, the studies generally did not provide enough information for a reader to make an independent assessment. It is important to ask why so few studies explicitly considered the client population issue. Certainly in our sample this appeared to be problematic, at least in the non-logistical programs.

Implementation

The issue of how a program was implemented, and whether this implementation corresponded to the original plan of the program, is obviously an important one for evaluation. There is always a danger in assuming that the intended program is the same as the program in operation and then attributing the outcomes to the intended program, when in fact the two may differ quite markedly. A little less than half of the

evaluations in our sample included some discussion of program implementation, or at least indirectly alluded to the fact that the evaluators had taken a critical look before assuming that the intended program was in operation. However, slightly more evaluations gave no discussion of this at all, seeming implicitly to assume that the program had been implemented as designed.

In some evaluations, where evaluators played an active role in the program, they tried to ensure that the program did conform to the model by carefully monitoring the implementation process. Others used participant-observation and interviewing techniques to see how staff spent their time or how they perceived their jobs, and how this matched the original design. A few process evaluations or comprehensive evaluations which included a process component did historical studies, following the program from its inception through the changes which came about during its operation. Those evaluations which were careful to monitor and document changes in the program were very much in the minority, however. Often, discussion of implementation was indirect or quite superficial.

There was some difference in response according to substantive area. The evaluations which seemed particularly negligent about including discussion of program implementation were the training programs and juvenile diversion programs. The substantive areas which fared better were pretrial release and team policing.

It is important for the audience of whom the evaluation is intended to have information on program activities, particularly if the program might be replicated, or if smiliar programs are being conducted and the

audience includes decision-makers not familiar with this project. Also, it is important as an indicator of how much knowledge the evaluators had of the actual workings of the program, and to what extent process and input issues were taken into consideration by the evaluators.

In our sample, a bit more than half of the evaluations did describe program activities. However, this question was answered very loosely, so that even when activities were "described", the description was not necessarily detailed enough or coherent enough to give the reader an adequate picture of what went on in the program. As noted earlier, less than half of the evaluations included consideration of whether the existing program bore any resemblance to its official design.

Some evaluations did an excellent job in presenting program activities, giving historical accounts of the projects' unfolding, including weaknesses and pitfalls as well as successes. The best descriptions seemed to either include a detailed outline of procedures for the implementation of the program, or to go beyond a generalized or idealized version and supplement the total picture with specific examples of program activities.

By contrast, many program descriptions were very brief. This was the most frequently cited problem. Descriptions often tended to be static, and did not consider program progress.

Another problem with even some of the better descriptions was that often they had an idealized, public relations flavor. There was a sense that the program could not possibly fit the description given of it. This was particularly true for those evaluations which were intended to

serve as models and to appeal to a wide audience of decision-makers who might be interested in implementing such a program.

Methodology and Methods

There are two problems to be dealt with when assessing the methods used in an evaluation. The first of these concerns the appropriateness of the methods used in view of the purposes of the study and the type of program (how well the methods were chosen). A complementary concern is how well these methods were applied.

It seems best to begin a discussion of methodology and methods by briefly classifying the studies in our sample. It turned out that the most popular evaluation type was that of experimental/quasi-experimental design. There were also a large number of narrative case studies. Somewhat fewer studies opted for general outcome evaluations with no common structure, or for a more comprehensive approach analyzing input, process and outcome components.

Evaluations of social service programs tended to be narrative case studies or input evaluations. Logistical studies were more dependent on experimental/quasi-experimental design. For certain areas of inquiry, the methodology was better chosen than for others. Those studies which scored highest on "suitability of focus" ratings were innovative methods (followed by "miscellaneous"), court programs, AVM/CAD, resource allocation and pre-trial release.

Those methodologies (or "foci") which were most often suitably selected (scored highest on "suitability of focus" ratings) were

performance evaluations with performance measures, formal models and "comprehensive" evaluations. The lowest scores were obtained by narrative case studies, input evaluations/audits and general outcome evaluations. The remaining foci all clustered around the mean score.

In general, comprehensive approaches were judged to have been quite suitable for those studies that attempted them; narrative case studies were judged as unsuitable when such an approach was attempted. In between, the experimental/quasi-experimental design and general evaluation types were judged by the readers as being only moderately suited to the purpose (which means that some were much less suited than others).

The most common method appeared to be the use of descriptive statistics. Qualitative analysis was also frequently employed, as was statistical inference and related methods (e.g., regression and analysis of variance). Unfortunately, information on combinations of methods used is not available due to the nature or the computer program used to retain questionnaire responses.

Responses to the question on adequancy of measures split about 50/50 overall. Breaking it down into categories, over 50 percent of the logistical studies received positive ratings on this question, while fewer than 50% of the social service studies received positive ratings. Two small categories were outstanding: all pre-trial release evaluations received a positive rating, as did all AVM/CAD entries.

While the majority of studies clearly documented their methods of analysis, the most common problem with methods was that the evaluators misapplied or misused common statistical techniques. The standards of

most studies seemed invariant over type of method, i.e., all techniques were applied with equally mixed results.

One final problem, tied directly to both theory and method, was that in a great many cases (72 percent of the sample) outcomes could not be directly attributed to program activities. This means that the reliability of many criminal justice program evaluations is in doubt.

When examining methodology and methods, it is important to go beyond the approach and techniques chosen and examine the sources of data used to evaluate the program. An evaluation may appear to be methodologically sound, while the choice of data sources results in information which is worthless for decision-making or program management. Where did the evaluators go to collect their data? Were the sources chosen suitable and adequate to the task at hand? To explore these questions, we recorded the data sources used by the evaluators and rated the suitability of those sources.

Few surprising or important results emerged from the ratings of data sources. Administrative records are by far the most commonly used data source, which is understandable given their availability, low cost, and probable pro-program bias. There were only minor differences in the "suitability of data sources" ratings among the various program areas.

The use of multiple data sources is one way to enhance the validity of evaluation results and provide a rich description of both program activities and outcomes from several standpoints. Many evaluations in our sample tried to use multiple sources: this was particularly the case in the social service categories, where scarcity of sources led

the evaluators to try many different approaches.

One subgrouping which seemed to have a particular problem with data sources was Training Units. In examining training program evaluations, we find that suitability scores were generally lower than average, and that frequent remarks were made about the inadequate number and type of sources in these studies. Judging from other questions as well, this particular subject area seems to be posing problems for evaluators. Perhaps evaluators are using an inappropriate methodology for these programs, and something different might be attempted. Given that these are educational programs, signs of the innovative approaches used in recent years in education, such as ethnography, "illuminative evaluation" etc., might be considered.

Once data sources have been selected, the key issue becomes how effectively they are utilized. Some evaluators simply didn't tap sources with sufficient care and precision, a problem complementary to that of misusing statistical techniques. Occasionally they did a poor job of interpreting what they collected. The ineffective use of data also stemmed, sometimes, from information overload, i.e., too much was collected for no apparent reason.

A further problem seemed to be a lack of documentation. Evaluators often failed to explain where data came from and how they were used to draw various conclusions. In particular, where "softer" techniques were used there was often no explanation of how an observation or interview was conducted, nor were representative samples of results presented. Fairly common was the assertion that"[conclusion A] is based on

interviews", yet no dialogue samples or summaries were attached.

While it appears that the majority of evaluations have stated conclusions which are supported by the data analysis, several notes of caution must be made concerning these results.

From our readers' comments it becomes clear that in many cases the conclusions were supported by the data analysis because both the conclusions and the analysis were inconclusive. Perhaps of greater interest are some of the reasons why the readers decided (when they did so) that the conclusions drawn were not supported by the data analysis. These include the following:

1. no conclusions
2. no data and or data analysis
3. data were of poor quality
4. inconsistency in the data analysis
5. improper measurement
6. no comparison measure
7. recommendations were not based on findings
8. no process component which would explain outcome
9. method problems--obvious threats to internal and external validity
10. evaluators missed obvious alternate explanations

Flexibility

Since one of the concerns of this project is with the appropriateness of existing evaluation methodologies, we were concerned about whether

inflexibility of available methods might be an obstacle to successful implementation of the evaluation. For instance, in some situations the demands of a rigorous experimental design utilizing randomized assignment might be highly impractical and difficult to satisfy: in such cases efforts to use randomized assignment are likely to result in "botched" designs, and the suppposed strength of the methodology will become a weakness. For this reason, it is desirable that a given evaluation plan be designed in such a way that it can be changed as the evaluators learn about the program from preliminary research and from interaction with program staff.

Readers' responses to the questions dealing with flexibility of evaluation design were distributed as follows:

> 25 percent of the sample were scored "yes" (meaning designs were flexible), 30 percent scored "no", and 45 percent were "not applicable."

The large number of "n.a." responses to this question reflects the fact that it was often not possible to assess how the evaluators would have or had reacted to changed circumstances; also, certain types of evaluations which were limited to descriptions or ex post facto data analysis would not really be vulnerable to such changes.

The readers' comments on this question revealed a number of ways in which evaluations may flexibly adapt themselves to changing circumstances or unexpected findings. These include greater use of open-ended rather than simple yes/no or multiple choice questions in interviews or surveys, and the willingness to shift from written to oral interviews when necessary. Also, the period of data collection may be lenghtened

or shortened. Shifting to new sources of data, exploratory data analysis and theory-building approaches may help to make sense of unexpected or apparently contradictory results. The use of multiple control groups is advocated in experimental design. Finally, simulations or formal modeling may be used to estimate information not available from the data.

Communication

Obviously, the more aware evaluators are of program changes, the easier it will be to adapt the evaluation plan accordingly. The only way this awareness can be developed is through regular interaction and communication with program staff. In addition, if, as is desirable, the evaluation is being conducted during program implementation, interim results may be put to best use by being fed-back to program staff.

The item aimed at the communication issue was "Describe any feedback between evaluators and program staff." As the responses to this question do not lend themselves to definitive conclusions, it will be necessary to investigate the issue further with the author and consumer questionnaires. Some emerging patterns may be discussed, however.

Positive responses to this question frequently contained comments about the formative functions served by this feedback. The key issue here is user-orientation. It is important to consider ways of getting helpful information into the hands of the program staff during a long-term evaluation (with the exception, perhaps, of strict experimental designs). The non-use of reports can often be traced to a time-lag, meaning that, by the time the evaluation is completed, the program may have been substantially transformed, or even completed or in some

instances discontinued.

There was a high degree of communication (although summative rather than formative) and cooperation between staffs which smoothed the way for the evaluation to proceed. In other cases there was an obvious degree of tension or conflict between staffs; in one case they "argued all the way through the evaluation."

The relatively high proportion of "N.A." responses to this question was due to a number of factors. Based on the comments, it seems that about half of the n.a. responses were in cases where the evaluation report simply didn't provide enough information for the reader to draw any conclusions. In about a quarter of the cases the feedback question was not applicable because the evaluation was actually performed by the program staff. The remaining quarter of the responses were in cases where the question was inapplicable for other reasons, such as if the report was actually an overview of work in a given field, rather than an evaluation of a particular program.

Outputs

For our purpose, the major output considerations are accessbility and use of the final report of the evaluation. There are two aspects to accessibility-presentation and availability.

One of the most basic requirements for a genuinely useful evaluation is that data be presented in a clear fashion. The information that inspired conclusions should be readily understandable and complete enough for reanalysis by the reader: this is a way of assuring the integrity

of the research.

In order to locate deficiencies in presentation we asked "was the presentation of data adequate?" The majority (63 percent) of evaluations received a "yes" on this item, indicating that presentation was not a major problem. The social service categories tended to have somewhat more problems than did police logistics. This may indicate difficulties with presentation of qualitative data: Comments made under other questions seem to confirm this observation. Some reports, however, apparently did a good job of presenting qualitative data.

Four problems seemed to be generic among those programs which "inadequately presented data". These were noted in most or all program areas, although the non-logistical are the most common offenders. These problems were:

1. consistency. It was noted several times that the data presented were not consistent with the measures and/or sources chosen by the evaluator;

2. completeness. In both qualitative and quantitative studies the presentations were often far too sparse. Summaries were often used where more detail was needed;

3. legibility. If evaluations are going to be user-oriented they should be easily accessible and readable. Although "readability" was not frequently mentioned as a problem, the presentation of too much data was. An overabundance of unneeded detail can render an evaluation almost as useless as one with no data presentation, especially from the practitioner's point of view.

4. correctness. There were some comments about improper grouping of data, incorrect calculations, etc.

If evaluations are going to be more useful, it is imperative that the presentation of data be legible, complete and correct. The

possibility should be open to the reader to reinterpret the results and confirm evaluation integrity. A good many in our sample did not meet some or all of these criteria.

Although evaluations of criminal justice programs are for the most part available from the Document Loan Service of the National Criminal Justice Research Service (NCJRS), the procedure for obtaining them is long, complex and frustrating. Managers of criminal justice programs and local government officials may have the training, patience and staff needed to follow this procedure through to the end: in many instances local government officials would not have the necessary staff or expertise.

Community leaders and concerned citizens groups very likely would not be able to use this method as successfully as someone as persistent and articulate as Dorothy Green, our project secretary (although naturally they would not need to get as many documents).

Usefulness of an evaluation in decision making was difficult to assess from the evaluation reports. Our mailed questionnaires will include several questions on this important issue.

Conclusion

Based on this first phase of research to determine the usability of criminal justice evaluation via an empirical study of their methods, several observations may be made, with the reservation that the responses to the author and consumer questionnaires may either confirm or refute these observations.

If an evaluation does not meet one or even two of the criteria set

forth in the introduction to this summary, it may still be useful to its intended audience, depending on that audience's understanding of methods and knowledge of the program. Unfortunately, many of the criminal justice evaluations in our sample fail on several counts, thus, there is doubt as to how useful they could be in decision-making. The evaluations seem to be weakest on methodology, followed by comprehensiveness and flexibility of design and implementation. Clearly, many are weak in other areas as well.

Some of the problems manifested in these evaluations are things that could be rectified at the proposal stage if sponsoring agencies know what they want from an evaluation and review proposals carefully. This is particularly true of problems in choice of methodology, application of methods, and flexibility of design. We hope to be able to make specific recommendations as to what a sponsoring agency should look for in an evaluation proposal after integrating our theoretical work on evaluation methodology development with feedback from the evaluators, program managers and funders.

Other difficulties of communication and presentation are less measurable to administrative review, but they are no less important to the usability of the evaluation. We hope to gain insights as to what might be done about them when we review responses to the author and consumer questionnaires.

PROJECT LIBRARY - PUBLICATIONS RECEIVED

October, November and December, 1978

1. "Norfolk Fellowship Foundation Inc., Massachusetts Correctional Institution - Exemplary Project," Abt Associates, Cambridge, Massachusetts, 1975. [NCJRS Document Loan Program]

2. "Topeka-Legal Services for Prisioners, Inc. Exemplary Project Validation Report," Abt Associates, Cambridge, Massachusetts, 1975. [NCJRS Document Loan Program]

3. "High Impact Anti-Crime Program, National Level Evaluation - Final Report V.2," E. Chelimsky, 1976. [NCJRS Document Loan Program]

4. "High Impact Anti-Crime Program - Performance Measures Used in the Impact Program," Siegel and Gorse, 1975. [NCJRS Document Loan Program]

5. "Behavior and Attitude Modification in a Jail Setting," Florida Inter-Agency Law Enforcement Planning Council, Metropolitan Dade County, 1971. [NCJRS Document Loan Program]

6. "St. Louis High Impact Anti-Crime Program Impact Evaluation: Planning and Evaluation Progress Report," Missouri Law Enforcement Assistance Council, 1973. [NCJRS Document Loan Program]

7. "Southestern Correctional Management Training Council - Exemplary Project Validation Report," Abt Associates, Cambridge, Massachusetts, 1976. [NCJRS Document Loan Program]

8. "Massachusettts - Police Institute - Exemplary Project Validation Report," Abt Associates, Cambridge, Massachusetts, 1976. [NCJRS Document Loan Program]

9. "Chicago - Community Education on Law and Justice - Exemplary Project Validation Report," Abt Associates, Cambridge, Massachusetts, 1976. [NCJRS Document Loan Program]

10. "Suffolk County (NY) Police Department Project Teletraining - Exemplary Project Validation Report," Abt Associates, Cambridge, Massachusetts, 1974. [NCJRS Document Loan Program]

11. "An Evaluation of Policy Related Research on Standards of Effectiveness of Pre-Trial Release Programs," National Center for State Courts, Denver, Colorado. [NCJRS Document Loan Program]

12. "Phase I Evaluation of Pretrial Release Programs - Work Product One - Issue Paper," National Center for State Courts, 1976. [NCJRS Document Loan Program]

13. "Phase I Evaluation of Pretrial Release Programs - Work Product Two - Narratives and Flow Diagrams," National Center for State Courts, 1976. [NCJRS Document Loan Program]

Project Library - Publications Received (continued)

14. "Phase 1 Evaluation of Pretrial Release Programs, Work Product Three - Evaluation Framework," National Center for State Courts, 1976. [NCJRS Document Loan Program]

15. "Phase 1 Evaluation of Pretrial Release Programs - Work Product Four - Present State of Knowledge," National Center for State Courts, 1976. [NCJRS Document Loan Program]

16. Phase 1 - Evaluation of Pretrial Release Programs - Work Product Five- Designs for Phase 2," National Center for State Courts, 1976, [NCJRS Document Loan Program]

17. Phase 1 Evaluation of Pretrial Release Programs, Work Product Six - Single Program Evaluations," National Center for State Courts, 1976. [NCJRS Document Loan Program]

18. "Police Practices: The General Administrative Survey," Kansas City, Missouri Police Department, edited by John F. Heaphy, Police Foundation, Washington, D.C., 1978.

19. "Focusing Attention on Career Criminals - An Idea Whose Time Has Come," by Joan Petersilia, The Rand Corporation, Santa Monica, California, May 1978.

20. "The Rand Habitual Offender Project: A Summary of Research Findings to Date," by Peter W. Greenwood, Jan Chaiken, Joan Petersilia, and Mark Peterson, The Rand Corporation, Santa Monica, California, March 1978.

21. "Crisis Intervention Training - An Experimental Evaluation Program in Immediate Action Decision Making," Trenhelm and Whiteneck, 1974. [NCJRS Document Loan Program]

22. "National Youth Project Using Mini-Biles - Annual Report, November 16, 1973-December 31, 1974," F.Y. Hoshiyama, 1975. [NCJRS Document Loan Program]

23. "National Youth Project Using Mini-Biles - Annual Report, 2nd Funding Year, September 2, 1972-November 15, 1973," F.Y. Hoshiyamai, 1974. [NCJRS Document Loan Program]

24. "Impact of Training on Job Related Decisions - An Evaluation of the Colorado Youth Workers Training Center," Larson and Whiteneck, 1973. [NCJRS Document Loan Program]

25. "Theory and Practice of Delinquency Prevention in the United States - Review, Synthesis and Assessment," J.P. Walker, 1976. [NCJRS Document Loan Program]

Project Library - Publications Received (continued)

26. "Evaluation of the Development and Upgrading of Basic Judicial Skills - Final Report," Liacouras, et al., 1974. [NCJRS Document Loan Program]

27. "An Evaluation of the Specialized Units Project of the Pennsylvania Board of Probation and Parole," Lewis and Clark; Kaufman, project director, Pennsylvania State University, Institute for Research on Human Resources. [NCJRS Document Loan Program]

28. "180 Degrees, Inc.," Project Evaluation Unit, Governor's Commission on Crime Prevention and Control, Minnesota, May 1974. [NCJRS Document Loan Program]

29. "Interactive Display of Polygonal Data," by P.M. Wood, California Univ., Berkeley, Lawrence Berkeley Laboratory, October 1977; National Technical Information Service, Springfield, Virginia.

30. "Adoption and Utilization of Urban Technology: A Decsiion-Making Study - Analysis and Conclusion, Part 2," by W. Henry Lambright, Syracuse Research Corporation, New York, September 1977; National Technical Information Service, Springfield, Virginia.

31. "Evaluation by Simulation of Control Strategies for a High Frequency Bus Service," by R.L. Jackson, Transport and Road Research Laboratory, Crowthome, England, 1977; National Technical Information Service, Springfield, Virginia.

32. "An Interactive Spatial Analysis and Display System," James W. Clark, Washington University, Seattle, Washington, Urban Transportation Program, July 1977; National Technical Information Service, Springfield, Virginia.

33. "Safe Streets Reconsidered: The Black Grant Experience 1968-1975. Part B: Case Studies," Advisory Commission on Intergovernmental Relations, Washington, D.C., January 1977; National Technical Information Service, Springfield, Virginia.

34. "Evaluation of the Refuse Management System at the Jersey City Operation Breakthrough Site," Jack Preston Overman, and Terry G. Statt, Hittman Associates, Inc., Columbia, Maryland, February 1978; National Technical Information Service, Springfield, Virginia.

35. "Computer-Assisted Traffic Engineering Using Assignment, Optimal Signal Setting, and Modal Split," Stanley B. Gershwin, John D.C. Little, and Nathan Gartner, Massachusetts Institute of Technology, Cambridge, Mass., May 1978.

36. Models of Spatial Process, by Arthur Getis and Barry Boots, Cambridge University Press, New York, New York, 1977.

Project Library - Publications Received

37. "The Cincinnati Team Policing Experiment: A Techcical Report Volume 1," Alfred I. Schwartz and Sumner N. Clarren, The Urban Institute, The Police Foundation, Washington D.C., 1978. (Note: Volume II is part of Volume I)

38. "The New Haven Pretrial Diversion Program - A Preliminary Evaluation," Daniel J. Freed et al., New Haven: Yale Law School, September 1973. [Private Loan]

39. Using Social Research in Public Policy Making, by Carol Weiss, Lexington Books, D.C. Heath and Company, Lexington, Massachusetts, 1977.

40. "Recent Developments in Bus Transportation," by John W. Billheimer, Transportation Research Board, Washington, D.C., 1978; National Technical Information Service, Springfield, Virginia.

41. "Evaluation of Alternative Options to Provide Public Service Delivery," by Tamara L. Sparks and Arsen J. Darney, Midwest Research Institute, Minnetonka, Minnesota, North Star Division, October 1977; National Technical Information Service, Springfield, Virginia.

42. "Emergency Medical Services: Research Methodology Proceedings of a Conference Held in Atlanta, Georgia on September 8-10, 1976," prepared for National Center for Health Services Research, Hyattsville, Maryland, December 1977; National Technical Information Service, Springfield, Virginia.

43. "Finding Out How Programs Are Working: Suggestions For Congressional Oversight," Report to the Congress by the Comptroller General of the United States, November 1977.

APPENDIX B

"Using EVAL for Data Analysis: A Primer"

During this quarter, Tom Wong wrote a user-manual for the MIT/NILECJ EVAL system, a special-purpose computer program described for the research project "An Empirical Study of Methods Used in Criminal Justice Evaluations".

The purpose of this computer program is to maintain on-line all the project's records and each of the 200 evaluation reports. In addition, it allows a user to search all these reports and questionnaires according to such variables area of work, focus of evaluation, data sources, and methodologies, and generates analysis report based on the information. (EVAL stores each evaluation report in a separate file. Each file consists of a fixed set of information about the evaluation report, for example, the author of the report, date of publication, organization that publishes the report, etc., followed by a questionnaire for this report. A 6-digit code number is assigned to each report and its questionnaire.) This system is currently implemented on the Multics Operating System at MIT.

END