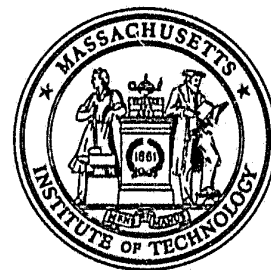
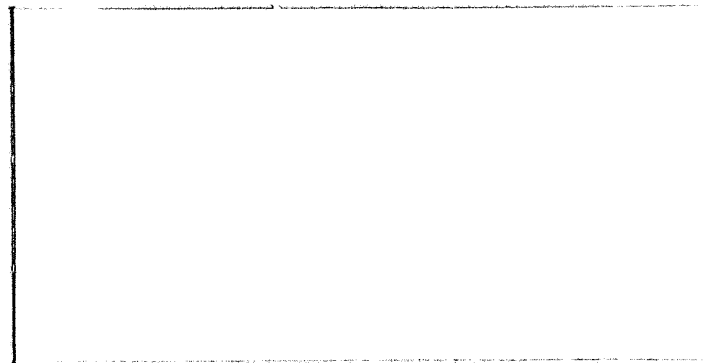


89669

OPERATIONS RESEARCH CENTER

working paper



**MASSACHUSETTS INSTITUTE
OF TECHNOLOGY**

89669

U.S. Department of Justice
National Institute of Justice

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

Permission to reproduce this ~~document~~ material has been granted by
Public Domain/NIJ/Office of Research
& Evaluation Methods

to the National Criminal Justice Reference Service (NCJRS).

Further reproduction outside of the NCJRS system requires permission of the copyright owner.

TOWARD A FRAMEWORK FOR PROCESS EVALUATION

By

Timothy J. Eckels

OR 094-79

December 1979

Prepared under Grant Number 78NI-AX-0007 from the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice. Points of view or opinions stated in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice.

ABSTRACT

A potentially powerful design for the evaluation of public programs is one which combines an assessment of program processes with an experimentally oriented analysis of program outcomes. Relatively little has been written, however, on how these two evaluation approaches can best be combined. This paper proposes a framework for sequencing, in a complementary and integrated manner, elements of process evaluation and experimental or quasi-experimental design. A number of assessment sequences are proposed, and the strengths and drawbacks of each are outlined. Five evaluation studies in the criminal justice field are then discussed in order to illustrate the potential of these sequences. The paper concludes with a summary of how an experimental design and experimental outcomes can be affected by the supportive use of process evaluation.

CONTENTS

I. INTRODUCTION	1
Organization of the Paper	1
Process Evaluation: A Definition	2
Complimentary Process Evaluation: Six General Advantages	4
II. TOWARD A FRAMEWORK FOR EVALUATION DESIGNS: THE SEQUENCES	5
A Design Notation	6
Pre-Intervention Design	7
Simultaneous-Intervention Design	9
Post-Experiment Design	12
III. THE SAMPLE EVALUATIONS	13
Neighborhood Team Policing Experiment #1	13
Team Policing Experiment #2	26
Pretrial Intervention Project	30
Inmate Self-Government Experiment	38
Specialized Parole Units Experiment	42
IV. CONCLUSION	47
Two Operational Realities	47
Processes and Outcomes	48
Summary	50
Footnotes	51-54
Bibliography	55-57

I. INTRODUCTION

A potentially powerful design for the evaluation of public programs is that which combines an assessment of program processes with an experimental assessment of outcomes. The potency of such a design can be inferred from two recent arguments in the field of evaluation research. The first is that experimental and quasi-experimental designs as evaluation techniques are plagued by a number of shortcomings.¹ These include restraints on randomization, limits to the number and kinds of variables that can be tested, inflexibility in the face of program change, and various threats to external and internal validity.² The other common argument is that evaluators have generally concentrated on program outcomes to the exclusion of an understanding of program implementation and operation. Voiced in a number of ways, the call has been for more careful assessment of those program processes which are expected to result in predetermined outcomes.³ This paper will look at the relationship between these two arguments. What we are calling "process evaluation" can be combined in various ways with an experiment, and each of these ways carries the potential to strengthen both the experimental design itself and the overall quality of the outcome evaluation.

Organization of the Paper

There will be four sections to this paper. This introduction will continue with a definition of process evaluation and with an outline of six general ways outcome evaluation can be enhanced by a partnership with a process component. In Part II, a number of evaluation sequences will be proposed. Each of these will be a variation of what Campbell and Stanley

have called experimental or quasi-experimental designs.⁴ We will see that a process component can be introduced at various times in the life of an experimental evaluation, and that certain advantages and disadvantages will result. For each sequence, the potential effects of the process study on outcome measurement will be presented. Part III will make these advantages and disadvantages more vivid by applying the evaluation sequences hypothetically to five of the studies from our project sample.⁵ Finally, a conclusion will address briefly some of the practical issues involved in this merger between process evaluation and experiments. The conclusion will also capsule the main themes of the paper.

Process Evaluation: A Definition

Process evaluation has been defined a number of different ways by researchers who see the need for more explication of those program dynamics that will supposedly lead to predetermined outcomes.⁶ Perhaps most common has been the idea that a process study -- or whatever else the author may choose to call it -- should focus on the "black box" between program inputs and outputs. Robert Yin summarizes that

...without more precise knowledge about the activity (of a program), information about the activity's effects or outcomes cannot be usefully interpreted. This problem may be most simplistically characterized with reference to the standard research paradigm, in which some causal action or activity (typically referred to as x or a set of independent variables) is believed to have some effect or impact (typically referred to as y, or a set of dependent variables) ...Although researchers may be rightfully concerned with a) the development of appropriate research designs for linking x with y, or with b) the definition of various outcome measures -- i.e., definitions of y -- inadequate attention has been given to the problem of defining x.⁷

Thus a possible definition for "process evaluation" would be a study which describes, defines, and assesses those program dynamics and operations which are to produce some given outcome.

This definition will be expanded somewhat here. While the activity leading to an outcome may be the primary focus of a process evaluation, in practice such studies can tell us much more about a program. We may discover unanticipated consequences of an intervention; we may find that the target population is receiving the program differently than we had expected; or we may find that a separate program is affecting our own in an unforeseen way. Patton discusses the possible range of a process study:

Under field conditions in the real world, people and unforeseen circumstances shape programs and modify initial plans in ways that are rarely trivial. The process evaluator sets out to understand and document the day-to-day reality of the setting or settings under study. He tries to unravel what is actually happening in a program searching for the major patterns and important nuances that give the program its character... Process evaluations look not only at formal patterns and anticipated outcomes, but also investigate informal patterns and unanticipated consequences in the full context of program implementation and development.⁸

It should be added that a process evaluation need not begin only after an intervention has begun. Decision-makers may have an interest in certain processes before an experiment begins, perhaps to modify the planned program or to adjust the experimental design. The study is not only concerned with the dynamics leading to "y" but also with the program setting -- the interests and attitudes of various groups, the history of the target population or area, competing political trends, the existence of other programs, etc. Such factors may have important implications for a program's outcomes.

For the purpose of this paper, a process evaluation is that which

describes, defines, and assesses the activities and conditions associated with an intervention and/or with the setting of an intervention. Such a study may be used to modify a program, adjust the evaluation design, change outcome measures, assess in a preliminary way the assumptions and the potential of a program, or simply to understand program dynamics. The tools for process evaluation might be ethnographic studies, surveys, interviews, various forms of observation, document reviews, or any of a number of other field methods. Analysis is usually inductive, and the evaluation is usually flexible enough to be developed as it is being conducted and as initial conclusions are drawn.⁹

Complementary Process Evaluation: Six General Advantages

With the above definition in mind, we will see that each evaluation sequence of process and experimental components promises several benefits. Briefly, these benefits can be summarized as 1) recognizance of program setting for factors which may affect the program either positively or negatively, 2) revision or confirmation of the evaluation design, 3) clarification of program implementation and operation, 4) clarification or confirmation of causal factors operating on program outcomes, 5) discovery of unintended or unanticipated effects, and 6) assessment of the program approach and its potential to produce positive outcomes.

The interim report of this project offered the preliminary conclusion that evaluations too often lack a process component.¹⁰ The attempt here will be to show just how and with what results process evaluation can complement outcome evaluation.

II. TOWARD A FRAMEWORK FOR EVALUATION DESIGNS: THE SEQUENCES

A number of authors have argued that process-oriented techniques are used too infrequently and with inadequate care.¹¹ This points to a need for guidelines. Evaluators may be unaware of the advantages to be reaped by paying more attention to program processes, and they may be uncertain about how best to apply what are usually more qualitative techniques. This section will address the lack of comprehensiveness in criminal justice evaluations by presenting a number of assessment formats or sequences. While this will not be an exhaustive set of guidelines, it will at least be a first step in promoting the productive use of complementary process evaluation. The aim will be to provide a systematic framework within which process evaluation can be designed and adapted to fit particular programs.

Selective design, careful timing, and sensitive implementation are perhaps the three most important ingredients of supportive process evaluation. While design and implementation will tend to vary with particular projects and circumstances, there are a number of generalizations that can be made about the timing of process techniques. Thus while this section will not propose "prepackaged", comprehensive designs, it will outline a number of evaluation sequences -- with their respective advantages and disadvantages, -- that are considered to be applicable on a general basis. Briefly, the process component of an evaluation can be employed 1) before an experimental intervention, 2) during an experimental intervention, but before a posttest, and 3) after the posttest but while the program or intervention is still intact. Various combinations of these sequences will also be advocated.

A Design Notation

Campbell and Stanley have introduced a number of experimental and quasi-experimental designs,¹² only a few of which are found with any frequency in criminal justice evaluations. We will be using their notation here to display various sequences of outcome and process components. A Pretest-posttest Control Group Design (True Experimental Design), for instance, is symbolized by:

$$\begin{array}{c} R O X O \\ R O O \end{array}$$

where R stands for the randomization process, O is the pretest, X is the intervention, and O₂ is the posttest. The lower line signifies the "control group" for which no intervention is introduced. An example of the more common quasi-experimental design is the Nonequivalent Control Group arrangement:

$$\begin{array}{c} O X O \\ \hline O O \end{array}$$

Campbell and Stanley also outline what they call "pre-experimental" designs, some versions of which will be included in this analysis as they are common forms in criminal justice outcome evaluations:

O X O One-Group Pretest
Posttest Design

$\frac{X O}{O}$ Static-Group
Comparison

Because there are so many possible control group arrangements, the second "line" will be omitted from the proposed formats introduced here. Thus for a fairly comprehensive use of process evaluation (P), where randomization may or may not be used, we might have:

$$\begin{array}{c} P (R) O X O P \\ P \end{array}$$

Process techniques would be "turned on" both before and after the experiment. The notation X_P indicates that the process component would also be taking place during the intervention.

Pre-Intervention Design

A process component might be applied to a program before the experimental conditions are introduced to the target group or arena. The component may be initiated either before or after the pretest (if one is included in the design) but there will be some advantage to positioning it even before that part of the sequence.

$$\begin{array}{c} P O X O \\ \text{or } P X O \\ \text{or } O P X O \end{array}$$

Pre-Intervention Design

A number of advantages will arise when techniques such as interviews and observation can commence well before the experiment begins and can conclude before the experimental conditions have begun. Many of these improvements will apply to the other sequences as well, but it is proposed that they will be the primary strengths of a pre-intervention design. A process component introduced before an experiment carries the potential to:¹³

- 1) Identify obstacles to smooth implementation. Appropriate changes can be made in the program design, or actors and institutions can be "prepared" for the program.
- 2) Provide feedback on implementation logistics. How the program is introduced to the target population may make a difference in its success.
- 3) Explore and adjust problems in the experimental design (e.g., randomization techniques).
- 4) Assess and modify the evaluation approach. Perhaps the experimental design will be deemed an unsuitable method

given the circumstances.

- 5) Identify conflicting programs or other environmental constraints. If potentially counterproductive programs are already operating in the area, they can be altered or terminated, or the new program can be adjusted accordingly. Likewise, programs which provide similar services to the control group or additional ones to the experimental group can be identified.
- 6) Judge the suitability of data sources. Are enough sources to be tapped by the outcome evaluation?
- 7) Provide feedback on the appropriateness of the target population.
- 8) Judge the appropriateness of the program approach. Evaluators may conclude that the program should not be initiated or that some other approach should be utilized.

A pre-intervention process component commencing before a pretest will command at least three additional advantages over one which is introduced afterward:

- 1) More suitable emphasis on certain program variables may be recommended. Evaluators may discover that some variables deserve more attention than others, or they may propose that new variables be introduced to the design. Likewise, decision-makers or interest groups may want information on certain program effects which have not been included in the original design. The evaluator should be aware of these and make an explicit decision about whether or not they are to be included. The pretest can then be adjusted accordingly.
- 2) Evaluators can judge the suitability of proposed measures. Are "arrests" for instance the only measure that should be used in a given situation? Whether or not the evaluator chooses to alter the design, he or she should be aware of the limitations inherent in any single measure.
- 3) The timing of the pretest may be important. The gap between the pretest and the intervention caused by the introduction of a process component opens the way for sample "contamination." Indeed, the process evaluators may themselves alter certain conditions by their activities.

These same advantages will apply to a pre-intervention strategy where no

pretest is to be used.

The overall advantage of the pre-intervention strategy --- particularly if the process evaluation does not continue beyond the intervention --- is that contamination of the intervention impacts will be largely precluded. Interviewers and participant observers will be relatively free to conduct open-ended research without fear of altering the eventual outcomes. Of course, it is still conceivable that suggestions made or techniques used during the process stage may affect the later behavior of actors under the experimental conditions. The only decisive way to avoid this is to reserve process-oriented techniques for the period following the experiment, a choice to be made after the various costs and benefits have been weighed.

Simultaneous-Intervention Design

There are a number of particular problems with quasi-experimental designs which can be averted by introducing a process element during and/or after the intervention:

O X P O Simultaneous-Intervention Design
P

As "treatment" of the experimental group begins, activities could be observed, records could be monitored, actors could be interviewed, or researchers posing as clients could be processed into the program. The component might be a continuation of previous process-oriented research, or it might begin only as the experimental conditions are introduced.

The advantages of this design include most of those outlined for the pre-intervention design with the added benefit that alterations will be more focused and the added disadvantage that they will come late in the operation. For example, through observation it may become apparent that

an outside program is interfering with the outcomes of an experiment. By observing the actual conflict during program operation, researchers will have a clear sense of the issues involved and the modifications required. But the new information may come too late to "save" the experiment.

In addition to the previously outlined points, then, a simultaneous-intervention design carries the potential to:

- 1) Detect problems associated with the experimental design (e.g., attrition, randomization, etc.) Methods could be suggested to rectify the problems or outcomes could be qualified.
- 2) Detect instability in the intervention. If the program has not "settled down" enough to bring about outcomes, the timing of the posttest could be revised.
- 3) Provide suggestions for speeding up the full implementation of the program. The process component would be a feedback mechanism, helping the program to reach a stable state.
- 4) Produce a more focused sense of implementation problems. Whereas a pre-intervention design can only conjecture about the intervention, this sequence can address the question: "Is the program being implemented as designed?"
- 5) Pinpoint problems with program operation that will be likely to affect outcomes. Evaluators with a thorough knowledge of program processes may be able to predict the effect of such problems on posttest results; appropriate adjustments or qualifications might be made with regard to the test results.
- 6) Determine whether the intervention is being received by clients as it was intended to be received. Interviews, for instance, may uncover a source of misinformation during program implementation and operation.
- 7) Confirm or nullify causality. The process component may identify cause and effect relationships and pinpoint extraneous factors that are beclouding the issue of causality. Such results could be used to alter program operations, confirm outcomes, or qualify outcomes.

- 8) Produce evidence of the "Hawthorne Effect." Interviews with program clients might raise the possibility that responses have grown more out of the existence of the experiment than the anticipated benefits of the intervention.
- 9) Uncover unanticipated and/or unintended program effects.
- 10) Raise questions about the accuracy or appropriateness of program goals selected for testing. Again, a posttest could be modified or the program could be adjusted.
- 11) Help confirm outcomes or explain negative outcomes.
- 12) Serve as a "back-up" evaluation in the event that the experiment collapses.
- 13) Otherwise guide both the program and the evaluation as the need for flexibility and modification arises.

Once again these profits of process evaluation are not the sole product of an intervention design. Some of them may emerge from different sequences, but they are the primary advantages of this particular format.

The major risk of this timing is experimental contamination, a result that might best be avoided by using only "before and after" process components. However, because there is the most to be gained from this sequence, it is highly recommended that unobtrusive research techniques and other "muting" methods be used to avert the problem. If it can be implemented successfully, simultaneous-intervention process evaluation carries the special advantages of A) being concurrent with the experimental conditions thereby avoiding the problems associated with a possible variation in program operation either before the pretest or after the posttest; and B) being able to feed formative information back into both the program and the evaluation.

Post-Experiment Design

A process component can be implemented after the posttest(s) has been administered, thereby avoiding the problem of contamination altogether.

(O) X O P Post-Experiment Design

Most of the issues addressed above would still apply to this sequence, provided the program remains in operation even after the "experiment" is terminated. Of course, the design does not lend itself to program or evaluation revision in time to make a difference in experimental outcomes. Ideally, a new cycle of experimentation would begin after this process component has fed information back into the program, but funding rarely allows such extensions. Probably the main purposes of the post-experimental design would be to qualify, confirm or nullify the experimental impacts, and to make suggestions for program modifications. The value of retrospective evaluation should not be underestimated, however. An evaluation which eventually makes a discovery about spurious causes or unanticipated effects is still several notches above an exclusively experimental approach.

There are other reasons for choosing one or some combination of these proposed sequences. One, for instance, would be cost. Perhaps funding will only allow for intensive observations during one time period.

The contention here is that process evaluation should be used as comprehensively as possible. In most situations, the ideal evaluation would use process approaches before, during, and after the life of the experiment. Then, following the recommendation of Campbell,¹⁴ the cycle of experimentation would continue once feedback had been provided by the post-experiment component. This would be the "ideal type" evaluation sequence.

III. THE SAMPLE EVALUATIONS

Next we will present a number of conjectured scenarios drawn out of actual evaluations from our sample of 200. We will explore in each case the feasibility and desirability of applying one or more of the process-experiment designs. This should serve to confirm the general applicability of the proposed designs and to clarify their various advantages and disadvantages as they are used in differing situations.

Before the examples are presented, it should be noted that the reports are not identified by title, and have not been selected because they are particularly inferior or superior in quality. At this point, we know little more about the programs than what is presented in the reports, so the narratives will be largely conjectural in nature. They should not be construed as definitive judgements of the evaluators' work. (Some assumptions made here may in fact be inaccurate.) The scenarios are designed instead to illustrate program characteristics that may be overlooked when there exists an inordinate emphasis on experimental outcomes.

NEIGHBORHOOD TEAM POLICING EXPERIMENT

We will begin with a description of a particular patrol experiment because it offers perhaps the richest supply of potential problems and shortcomings remediable through complimentary process evaluation. The basic premise of the program was a restructuring of patrol activities. In the new design, 1) police were assigned to neighborhood cars which were to answer calls locally; 2) team commanders were assigned around-the-clock-responsibility; 3) professional supervision with consultation replaced authoritarian supervision; 4) training and education on neighborhood issues

and the new style of policing was given to the officers; 5) community relations was given high priority; and 6) decentralized planning took the forefront.

The objectives of the program were:

- To control crime more effectively
- To increase community cooperation in crime control
- To improve police-community relations
- To tailor police operations to the needs of local communities
- To increase police officer job satisfaction
- To improve the working relationship between patrol officers and their immediate supervisors.

The evaluators had hoped to use an experimental design with randomized personnel, randomized beats, and sufficient baseline data. However, due to circumstances outside the experimenters' control, the program proliferated too quickly to allow for these arrangements. The authors were forced to settle for only roughly related comparison groups and no baseline data. They used a patrol survey, a citizen survey, and departmental measures at various points after the experiment had begun. The variables were analyzed with a simultaneous linear regression equation technique. Because the various precincts began using neighborhood patrol at different times, a time coefficient was included in the regression equations. A number of positive results were found, although many remain dubious due to threats to validity.

A strength of this evaluation report is the attention paid to various problems encountered during the evaluation. The authors are careful to describe the shortcomings of the document, and for that reason it lends

itself to speculation about the supportive use of process-oriented techniques. Some of the problems and potential remedies are as follows.

Ineffective Implementation

A number of problems prevented the effective implementation and stabilization of the program. These failings weakened the reliability of the conclusions and may have detracted from the outcomes. Dispatchers, for instance, failed throughout the experiment to follow guidelines designed to keep (to the extent possible) patrols within their own neighborhoods. Precinct commanders often misunderstood the team policing concept and misled their patrolmen. Contradictory orders were commonplace. In one instance, for example, patrolmen were ordered not to engage in conversation with shopkeepers or to shop in their patrol neighborhoods -- a policy which was in direct opposition to program tenets. In addition to these problems, several actors were resistant to the training and to the patrol concept itself. Unfortunately, many of these operational problems were apparently discovered ex post facto in the evaluation effort.

A pre-intervention process component (P X 0) could have addressed these implementation difficulties in a number of ways. Observations and discussions with the dispatchers, for instance, might have uncovered a basic resistance to the planned program concept. Perhaps they did not want to change old habits or perhaps there was a tacit agreement between themselves and the patrolmen that the conventional methods were preferable and should stay intact to the extent possible. If interviews were conducted before the dispatchers caught wind of the forthcoming experiment, they might have been quite candid about their feelings toward change. With such information in hand, evaluators could go back to program designers and propose

incentives, training, "bargains", or policy changes that might ease the dispatchers into cooperative service.

Likewise, the sources of resistance in the department could have been predicted and the way paved for a smoother program implementation. If officers feared a loss of autonomy and anonymity under the new program as it was designed, perhaps tradeoffs could be made or incentives offered to make the change more attractive. If commanders feared a lack of control due to the new style of authority, perhaps professionalization could be introduced by increments or perhaps alternative forms of personal security could be offered to the commanders. The possibilities are unlimited, but the pre-intervention design has the unique advantage of making such discoveries soon enough to prompt corrections in implementation policies.

A simultaneous-intervention design ($\frac{X}{P}$ 0) could conceivably make similar discoveries in time to adjust the program. Interviews with commanders might expose their ignorance about the concept or their resistance to certain policies. In this case, the interview could be more directed, probing at the problems in communication and resistance to change that have already become apparent. A special training session, higher salaries, or a slackening of the 24-hour-responsibility rule might be recommended as avenues to greater cooperation. Continuing observation or interviewing might then signify when the commanders have settled into their role and the program has stabilized.

It should be noted that the evaluators did recognize the "dispatch problem" during the course of the experiment and recommended a number of remedial policies. The definition of an "emergency call" was changed, to no avail, and by the end of the experimental period, the authors could

only suggest that further guidelines be issued and that a number of problems in data collection be resolved. An effective simultaneous-process component, however, would go beyond speculation and the issue of data collection, to a thorough understanding of organizational obstacles. Observers or interviewers would seek out dispatcher attitudes and habits. They would determine what, if any, internal agreements existed between dispatchers and patrolmen. and test the proposition that misunderstanding and misinformation were actually responsible for the problem. They might find that the problem was "organizational," or they might find evidence that a real assignment problem existed. It is conceivable for instance that "emergency demand" was high enough in surrounding precincts to warrant frequent outside assignments. (These assignments might easily escalate the problem due to the shortage of units remaining in the original precinct.)

Whatever the findings, the major advantage of the simultaneous-intervention design would be the concentration of evaluator effort on the particular problem as it arose. Observers could monitor dispatcher work directly and ask specific questions about radio assignments as they were made. If implementation problems seemed to persist, the various outcome tests could be postponed until a more intensive process component was put into play and the sources of resistance were addressed.

Information about some of the implementation shortcomings of this program would best have been gathered later in the sequence. This might have been done either through a modified simultaneous-intervention design or after the tests and measures had been administered. For instance, one program element that was designed to make leadership more democratic and professional was the use of team conferences. Guidelines instructed

commanders to hold regular meetings with their patrolmen and to encourage them "to make suggestions or criticisms for improving operation of the team." The evaluators suspected that what few regular conferences actually took place lacked significant interchange between members and lapsed into one-way briefing by commanders. Empirical data were needed on the nature of officer-patrolmen interaction. Yet there was undoubtedly the danger that observers would alter the behavior of the commanders and perhaps affect the outcomes of the Patrol Management Survey. Similarly, the activities of patrolmen on the beat might have been monitored by "ride-along-observers," but there would be no way to sort out behavior that was induced primarily by observer presence.

One possible sequence in such cases would involve introducing the process component with the intervention, but "turning it off" long enough in advance of the final tests to help preclude the possibility of a "Hawthorne Effect." Ride-along-observers might provide useful feedback and ascertain the integrity of patrol activities under the team plan. Outcome measures could be administered once as the observations ceased, and once again two months afterward. While evaluators could not be certain that patrol activities continued as planned during the latter two months, they could at least be assured that it was not observer presence that was producing outcomes.

The post-intervention design would circumvent the contamination problem with more certainty. Observers might collect data on the Team Conferences after the last set of measures was taken. This would allow evaluators to modify the report with the conclusion that the conferences were in all likelihood not conducted as designed. While implementation is not improved

with this sequence, awareness of it can at least be reported.

Whatever order process components take in an evaluation such as this, implementation problems require a thorough documentation of organizational characteristics. The Team Policing experiment and evaluation were administered with the tacit assumption that the police department would react rationally to intervention policies, producing outcomes systemically and predictably in response to various administrative guidelines. The assumption stems from what Manning calls an "administrative model of policing."¹⁵ For the evaluators, the formal organizational hierarchy was considered to be the key to invoking and understanding change. Yet as Manning points out,

...the symbolic imagery of policing as a bureaucratic-professional paramilitary organization is not entirely consistent with the actual process and patterns of social interaction that can be observed in police departments. Rational/legal models of police operation do not sufficiently reflect the range of behaviors and procedures that can be uncovered through careful field observation.¹⁶

While the evaluation attempted to explain implementation problems retrospectively, little attempt was made to identify informal networks, tacit understandings, reciprocal arrangements between actors, "unofficial" policies and procedures, or other elements common to the informal structure of most organizations. Yet the "underlife" of the police department deserved attention. It probably had more to do with implementation problems than any lack of information or basic unsuitability of policies. Don Schon describes the type of information that can be missed with an exclusive focus on outcomes:

In the Seattle police force, according to John Van Maanen, patrol car policemen have certain formal relationships with central dispatchers which are specified in the task system of the police department. Patrolmen report their whereabouts to the dispatchers and respond to the dispatcher's calls; dispatchers monitor the patrolmen, receive calls for help, and assign patrolmen to situations. But, "draped over" these formal relationships (in Kadushin's phrase) is a complex of informal understandings and agreements. Dispatchers understand that when cops go into Charlie's (an informal meeting place), they are likely to be there for two hours or more. Dispatchers will then protect those patrolmen by calling on others for assignments during that two-hour period (though they know where to reach them in case of real emergency). In return, patrolmen are continually taking cups of coffee and cigarettes to the dispatchers who are trapped for long periods of time in their little rooms and have no other access to these amenities. As a consequence of these informal relationships and understandings, the police chief's attempt to introduce a new cadre of female, civilian dispatchers produced a violent negative reaction among patrolmen. 17

For the Team Policing experiment, extensive fieldwork, either before or during the intervention might have uncovered similar informal characteristics and opened the way for more successful implementation.

The Issue of Causality

The Team Patrol Project was deemed a success on a number of dimensions, but there are several alternative explanations for the outcomes. Some of these "alternative hypotheses" were mentioned by the evaluators - who were notably candid about the shortcomings of the experiment - but few were explored systematically in an effort to either confirm or nullify them.

Because this was not a true experimental design, it is impossible to know whether outcomes were actually due to patrol reorganization or to special attributes of the police officers involved. No randomization

took place and it is conceivable that team commanders and patrolmen were selected on the basis of past performance. One acknowledged difference between the experimental and comparison groups was that of age: officers in the teams were generally younger and less experienced than those in the regular precincts. It was found that crime rates were reduced in the team precincts and that per-officer arrests were higher than in the "controls." Some statistical tests were performed by the evaluators to rule out the possibility that the team officers were policing more effectively because of their youth, but the change in crime rates and a number of other measures could still have grown from this difference.

While it would be difficult to exclude with certainty the effects of age or other personnel factors, a process component might remove some of the doubt surrounding the issue of causality. A simultaneous-intervention or post-experiment design might include interviews with various actors about their perception of outcome determinants. A large number of officers and administrators might in such a case indicate that the team success was due primarily to a few unusually outstanding commanders. Observational and administrative data could then be collected on the style and effectiveness of these men as compared to the control-group commanders. Weighty evidence from both interviews and observations might prompt evaluators to make serious qualifications about the results and to recommend more stringent experimentation. If, on the other hand, a large majority of respondents credit the patrol reorganization with the success, evaluators could be more confident about the program's outcomes.

The reduction in crime rates could also have been due to an increase in the use of volunteers by the teams, unique circumstances in the

neighborhood, or just random variation. Once again, probing for such causes would be possible with process-oriented techniques, and while absolute proof would not result, outcomes might be either confirmed or questioned. Crime patterns, for instance might be shifting from one section of the city to another due to changes in the criminal underworld, to the expanding "turf" of juvenile gangs, or to an informal "target hardening" campaign in certain neighborhoods. Such changes are not always obvious, but neither are they difficult to perceive if evaluators observe and talk with people in the criminal justice system.

The possibility of altered crime-and-arrest statistics becomes another potential "causal" factor in the patrol experiment. Team commanders could have reported crime rates differently or quietly instructed their men to do so, in an attempt to answer pressures from headquarters reminding them that they were personally responsible for reducing crime in their neighborhoods. In this case, a careful monitoring of records and comparisons of commanders' reporting activities over time might confirm or help neutralize the competing hypothesis.

In all cases, the most informative process component would occur either during the intervention or after the experiment, for only then can fieldworkers observe actual causal mechanisms in action.

The Need for Recognizance of Program Setting: The Case of a Conflicting Program

A pre-intervention design could have prevented problems that arose out of a conflicting program. A career motivation scheme designed to rotate officers regularly among precincts was operating before the teams were implemented. Personnel mobility caused by the program was

antithetical to the principle of "close community ties" inherent in the Team Policing concept. The evaluators found that far too many officers were being moved from precinct to precinct during the course of the experiment. If evaluators had been able to conduct a "recognizance" of the area (i.e. discussions with key actors, observations, briefing on administrative records, etc.) well in advance of the experiment, the problems might have been avoided by a cancellation or rearrangement of the career program.

Unintended Effects and Their Causes

One of the primary advantages of the simultaneous-intervention strategy is the potential for discovering and investigating unintended program effects. In this case, for instance, interviews with patrolmen might reveal a "rumor" that criminals are responding to the patrol reorganization not by curtailing their activities but by moving them into non-team neighborhoods. Such a trend would be difficult to prove statistically, but established informants might offer confirming impressions.

One surprising outcome of the evaluation was a significantly greater self-reported use of aggressive patrol tactics (i.e. stop-and-frisk, questioning suspicious individuals, etc.) in the experimental neighborhoods. Such a trend was naturally considered undesirable for a police-community relations effort. Interviews with patrolmen, commanders, and administrators (especially those not associated with the team beats who might be more candid) might not only lead to an understanding of this phenomenon, but also point to other undesirable characteristics of the project.

Aggressive patrol may stem from leadership demands which in turn may stem

from administrative pressures to produce more arrests and to intimidate "the criminal element." If the quality of these arrests themselves were investigated, it may be found that more "dirty" or unconstitutional apprehensions were taking place in order to inflate outputs.

These possibilities could be missed if the evaluators relied only on the "administrative model" and insisted exclusively on counting numbers. Process evaluation may be no more penetrating, given the cohesive, reticent nature of police groups. On the other hand, cooperating police officers could be used with the assurance that their information would remain anonymously ascribed. In the case of "dirty" arrests, simple court monitoring might be sufficient to show that large numbers of cases are not reaching adjudication.

Another surprise was that various measures of "citizen cooperation" during the experiment indicated that one precinct showed a marked decline between the months of April and June. The cause of such a shift could have been anything from changes in the weather to new policies about how and when informants should be used. Perhaps the police department was receiving bad publicity during that period, reducing cooperation from witnesses and other informed citizens. The only chance of obtaining such explanatory information is if evaluators are open and observant when it comes to discussions between themselves and various actors. Patrolmen might have the idea that cooperation has gone down due to militant activities in the neighborhood, another possibility that could be verified by process-oriented methods.

The "unanticipated effects" discussed here were only those which could be inferred from the presentation of the evaluation. A fundamental

problem with insular outcome assessment is that the existence of some effects may not even be suspected. Open-ended fieldwork techniques must be added to an evaluation if there is to be any hope of "ferreting out" the unpredictable. This is a primary advantage of the simultaneous-intervention and post-experiment designs, for with a greater awareness of "what is going on" while the program is in motion, evaluators stand a better chance of locating unanticipated effects.

Clarifying Program Operations

The simultaneous-intervention design also lends itself to a "formative" role. If the actual operation of the program is elucidated early enough in the experiment, program managers may make improvements either before or after the posttest (depending on the degree of experimental integrity desired). For this evaluation, the concepts of "unusually effective" and "unusually ineffective" patrol practices might have been useful.¹⁸ The actors did note that some patrolmen and some teams seemed to perform very well. From informal reports, it was learned that they had established effective community relations and had become more aware of community needs. The authors also reported the existence of extremely poor performers. Intensive interviewing of commanders, officers, and residents only in relation to these extreme groups might have proved immensely valuable. The evaluation needed a clear and detailed picture of what contributes to successful team policing and what detracts from it. This may have been best determined in a post-experiment design, given that the extent of fieldwork needed would likely have disrupted the experiment.

TEAM POLICING EXPERIMENT #2

In a similar Team Policing experiment, a greater emphasis was placed on the anticipated benefits of patrol structure. Management was decentralized and investigative operations were combined with patrol operations. Two patrol areas were reorganized and compared to traditional areas. Many of the problems just discussed were present in this experiment as well, despite the fact that baseline data were obtained in this case. Crime rates were reduced (although the authors are properly cautious about the reliability of crime statistics) and arrests increased in the experimental beats. At least three potential advantages of a process component become more clear in this experiment than in the former.

Problems with the Experimental Design

One of the requisites to an effective police patrol experiment is that the beats to be compared with each other are similar. Unfortunately, the programmers in this case were not able to match beats very satisfactorily. The two experimental beats were significantly different from the control areas on a number of demographic variables and they were both undergoing large-scale urban renewal. As the authors point out, the urban renewal itself may have been in some way responsible for the experimental outcomes.

A pre-intervention design should be implemented even before the baseline data (pretest) is collected in cases such as this. More should be known about the differences between beats before an experiment is put into motion so that decision makers can 1) change the locale of the experiment if the differences are ungovernable; 2) adjust statistical outcomes to reflect the differences; and/or 3) report the degree to which outcomes may

have been affected by the differences.

There are a number of ways such information could be garnered. In the case of the urban renewal problem, crime and arrest-rate changes could be examined in other parts of the city where urban renewal is under way. Experimenters could investigate crime rates in other cities where urban renewal has taken place. Various actors, including patrolmen, citizens, social workers, and others familiar with the area could be interviewed about their perceptions of crime change during the renovations. None of these methods would be foolproof, but they would help the experimenters to formulate a general sense of the impact brought on by the renewal. The patrolmen might report, for instance, that new kinds of crime were being created by the construction: e.g., vandalism and theft centering on the renewal sites.

If the evidence indicates that the differences between beats is creating a difference in crime rates, the experimenters might seek out more comparable beats. Or they might go ahead with the experiment but use a "Bayesian" approach to data analysis, increasing the significance levels of hypothesis tests so that they reflect the area disparities. In this way a much greater difference in crime and arrest rates would be required to warrant a "favorable" report. Whatever the case, a pre-intervention process component would give the experimenters more information to present. They may go ahead with the outcome evaluation as designed, but temper the conclusions with information on the likely impact of the urban renewal projects.

A simultaneous-intervention design could have addressed the problem of non-randomized team leaders. In this experiment, the commanders were

definitely selected for their outstanding qualifications, yet there was no effort to determine if their expertise was "making the difference." As we suggested before, interviews and discussions with patrolmen and administrators could be used to sort out the issue of causality. Interviewed subjects might make frequent references to a commander's ability to organize or motivate the team in such a way that arrests are more likely. While this can be described as a "causality" issue, it also has consequences for the design of the experiment. Program designers may act upon the feedback immediately by creating an additional, more randomized team. Results would then be staggered, but certainly more meaningful.

Unsuitable Goals or Measures

For one of the major crime categories in this experiment, it was found that while the experimental group had made more arrests, those arrests were not leading to a concomitant rate of prosecution. The authors speculate about the causes of this but never investigate fully the possibility that patrolmen were making unacceptable arrests in order to fulfill the expectations of the experiment. Other causes might have operated as well.

A simultaneous-intervention component might begin by investigating the low prosecution rate through court monitoring and interviews with prosecutors. But process evaluators might also question the worth of arrest statistics as outcome measures. It may well be in this case that "ritualization" has taken place; i.e., arrests have become an end when they really should be considered only a means to the end of crime control.¹⁹ Researchers might propose that some measure further along in the criminal

justice process be used. Perhaps only those cases which make it to court should be counted, or possibly arrests could be discounted if they are "thrown out" of court on an arrest-related technicality. Of course, the failure of cases to reach prosecution may not be all the fault of the arresting officer, so the first step would be to explore the prosecutors' apparent reluctance to pursue the cases from the experimental beats.

The issue of arrests could be unresolvable, but process evaluators may propose that a wider range of outcome measures be used to assess the program. Victim surveys, for instance, could be used to investigate police contact with the community. Are the recipients of police service more satisfied in the Team Beats? While such measures should never stand alone, they could be used to support other outcomes.

If new alternative measures are proposed, a trial period might follow the posttest. In this way, plans that emerge out of a simultaneous-intervention design could be tested and evaluated by a post-experiment component. Ideally, stages of experimentation and testing would then continue.

Another Causal Hypothesis

It was found in this experiment that one Team was making extensive, unconventional use of "mug shots" in their work. The point was raised that this "technique" might have made a difference in outcomes. Interviews with various actors might have offered clues. Did officers, for instance, think that the mug shots made a significant difference in their ability to apprehend criminals? If this seemed to be the case, the process component could then contribute to flexibility in the experimental design. A new control group might be introduced in which the men were encouraged to make

similar use of mug shots.

PRETRIAL INTERVENTION PROJECT

In the next sample, several pretrial programs within one state were evaluated with a quasi-experimental design. Clients were in most cases selected by prosecutors, defense attorneys, or project staff and diverted to a project which provided counseling, services, and referrals to existing community agencies. In general, the goals of the program were:

- 1) Savings of time and/or money to the criminal justice system through early intervention in cases which do not require prosecution for individual or societal reasons.
- 2) The reduction of recidivism through the provision of services which meet client needs and the avoidance of the potentially negative consequences of prolonged contact with the criminal justice system.

The projects were evaluated on three dimensions: Cost, Effort and Effectiveness. The "Effort" component in many ways illuminated the processes of the projects, but also excluded a fair amount of important information. Effectiveness, in terms of client status and recidivism measures, was determined by comparing clients who successfully completed the program to those which were terminated early. Data were collected on clients as they entered the criminal justice system, upon termination, and for follow-up periods covering six, twelve and thirty-six months.

The evaluation was presented and performed well, given restrictions operating on the authors, and various problems and shortcomings were described candidly. More extensive process analysis would have improved the evaluation in a number of ways.

Implementation Ambiguities

There were implementation problems that would best have been addressed by a simultaneous-intervention design. For the most part, these were difficult to predict and were most amenable to correction only after they had become clearly defined.

The evaluators mentioned in their final discussion of the programs that the referral processes may not have been entirely equitable. In one program, for instance, the public defender prompted most of the referrals, raising the possibility that non-indigents were being disproportionately excluded from the program. A simple monitoring scheme might have remedied the problem by providing client profiles to the court which in turn could have taken steps to correct for inequities. The evaluation should be credited for the amount of information that was given on the selection criteria of the various programs. This provided the reader with a basis for comparison to similar projects, at least to the degree that the criteria were actually used. A monitoring device would have confirmed the character of the diverted population.

At least two other shortcomings were bared by the evaluation, but little information was collected on their causes. First, not all of the programs reached their desired caseload sizes. Second, it was discovered that only 50% of the total project clients were being referred to community services. This figure was deemed too low, since one program mandate was to hook clients up with outside agencies. The evaluators were quite explicit about several implementation problems, but information collected on these two is insufficient to guide change. Observers and interviewers could

have been used to describe in detail the referral mechanisms of the courts and the pretrial projects.

Modifying Goals and Measures

One of the outcome measures for the project was the employment status of the clients upon termination. Many of the programs showed a favorable outcome on this measure, but others did not. Interviews with clients and program counselors might reveal a difficulty for newly trained clients in finding employment. Clients stigmatized by participation in the program undoubtedly face discouraging prospects in the employment field, especially with unemployment rates so high among poor and minorities. A simultaneous-intervention component might conclude that employment status should be eliminated or diminished in importance as an outcome measure.

Unintended Effects

Pretrial programs as a group are vulnerable to a number of negative, unintended effects. Perhaps the best way to present these undetected consequences is to formulate a "worst possible case" followed by the remedial potential of process evaluation. We will see that a post-experiment design would generally have been most successful in pinpointing the problems. While some of those effects to be illustrated here were not even likely in this particular project, they will all be presented for their relevance to other programs. Imagine, then, a pretrial project with the following characteristics.²⁰

First of all, our undesirable prototype is based on referrals made primarily by the prosecutor, who selects clients not from these cases which he/she plans to prosecute, but from his/her total caseload. As a result,

many defendants who would otherwise not have been prosecuted are now being placed into a coercive program. Furthermore, even those cases that would have been prosecuted are receiving excessive referrals. Defendants that quite obviously would have received non-custodial sentences, verdicts of not-guilty, or dismissals on technical grounds are now being conscripted into the diversion program. In effect, the "net" of state control over defendants has been widened, an outcome just opposite from that which was intended.

The problem is exacerbated by the tendency of the prosecutors and judges to see the program as a suitable "treatment" for minority and poor children. Bullington describes this pervasive tendency:

The criminal justice and social work professions share a long-standing bias against lower-class and minority lifestyles and social institutions. The capacity of the black family, in particular, to function as a vehicle for socialization and social control is repeatedly questioned by these professions. Diversionary programs are advertised as promoting interactions such as might be found in middle-class families. Thus such programs may come to be seen by judges and probation officers as particularly appropriate for youngsters from lower-class and minority cultures.²¹

In our hypothetical project, this phenomenon is inflating even more the number of clients embraced by the criminal justice system.

The expansive tendency of the intervention program is related to a dangerous lack of due process in the referral mechanism. In a number of cases, charges have not been filed and/or a probable cause hearing has not taken place. This may not constitute a problem except that the defendants are often unaware of their rights and of the possible choices facing them when a prosecutor advises: "The diversionary program would be best."

By law, the choice to enter a program must be informed and voluntary; in

practice the choice is often made exclusively by the prosecutor. Furthermore, the defendants are rarely made aware that they can quit the program at any time and regain their right to a speedy trial. Consequently, many clients are leaving the program with an unnecessary "unfavorably terminated" label. The unintended tendency of the program is to divert large numbers of clients coercively, who might otherwise have chosen to go to court.

Due process is also suffering during the background investigation of potential clients. Excessive information is collected to determine their eligibility, violating the clients' right to privacy. Investigators often solicit unnecessary details about the crime and the arrest which can then be used against the defendant in subsequent trials. Potential clients who are aware of this practice are refusing to participate in the program.

Stigma is still another unfortunate consequence of the pretrial program. While the defendants are not guilty by law, they are often perceived to be so once it is known that they have participated in a pretrial program. Various actors in the criminal justice system are treating the former clients with bias. Police tend to arrest them more often, prosecutors take them to court more often, and judges are giving them harsher sentences with the participation in mind.

As this "worst possible case" illustrates, the potential for unintended effects in a pretrial program are substantial. Gibbons and Blake suggest still other possibilities yet to be explored and confirmed by researchers:

...(O)there outcomes of a diversion program, in addition to its impact (or lack of impact) upon diverted youngsters, may be alterations in police department referral practices and police attitudes, changes in community tolerance of youthful deviance and other consequences of this kind.²²

While either a simultaneous-intervention or post-experiment design could expose many of these effects, the latter will be proposed in order to illustrate the following advantages. (In this case, the component will not follow the last "posttest," for the followup period of the project is 36 months; rather the component will be implemented after a significant number of clients have been successfully terminated - perhaps one to two years after the program begins).

First, this type of program needs time to "settle down" while habits and practices take shape. Problems such as the "widened net effect" may not emerge until prosecutors learn the boundaries (or lack thereof) of the program. Most of the problems discussed are the type that emerge only as the program matures. Second, if observers and interviewers intercede excessively during the early stages of the intervention, prosecutors (and the other actors) may respond unfavorably. Feeling self-conscious about the scrutiny, for instance, a prosecutor might underutilize the referral mechanism, dampening the positive potential of the program. Third, process evaluators will have a better sense of what they are looking for. After a year of program operation, suspicions about various abuses of the program will be generally known.

An effective process component would include interviews with all of the actors involved in the referral process. It would also entail observation of courtroom activities and conferences between the defendant and the referral agent, and participant observation where possible. The "expanded net" problem might be confirmed by combining various interview results with an analysis of some sample cases and their outcomes. It may be found, for instance, that several defense attorneys and judges are

fearful that too many clients are being pulled into the program; prosecutors may not concur, but their methods may show otherwise. They may admit that their conferences with defendants are short, that the defense attorney is rarely present, and that a decision to prosecute has often not been made beforehand. The case analysis may then uncover several program clients who stood minimal chance of conviction in a regular trial. All of this evidence might be strengthened by complaints from defendants that they would have made different choices with more information at hand.

The "due process" problems might be exposed through similar means. Observers' confirmation that the defense attorney is rarely consulted or that "probable cause hearings" are never held before diversion would be powerful evidence. Interviews should certainly include clients who are unfavorably terminated from the program. Did they know that they could have returned to trial? If the evaluators can surmount the ethical issues and inevitable personnel resistance, participant observation may be the best way to obtain indisputable information. Researchers posing as defendants could enter the system and report on the limited information divulged by prosecutors, the coercive tendencies of the system, and other violations of due process.

The stigma problem is perhaps the most difficult to document. Interviews with judges and police could again be used to support case analyses. Did a judge consider the former participation of the convicted person in a diversion program? Would the sentence have been different otherwise? The stigma problem is also more difficult to remedy; evaluators may simply discuss the likelihood of its existence and allow program designers to weigh it against the benefits of the program.

The key to this post-experiment process evaluation is that it be sensitive and unstructured enough to detect any unanticipated changes that may have occurred. Investigators should feel free to expand their work and follow "leads" as they arise. The process information can then be used to reform the program for future experimentation and to modify plans for similar endeavors.

Appraising the Program Approach

One conclusion of a process investigation might be that the concept of a pretrial intervention is misguided. Especially if the unintended effects seem insurmountable, evaluators may conclude that a formal system of diversion ends up doing more harm than good. In the process of performing interviews and observations, the researchers may find that informal mechanisms of diversion are more effective and less detrimental. As Gibbons and Blake point out, some studies indicate that substantial numbers of alleged offenders never reach the criminal justice system because they are handled effectively by parents, shopkeepers, teachers, social workers, and policemen.²³

Process evaluators might recommend that programs be implemented instead that will support these informal diversion mechanisms. They might back up their recommendation with the idea that "at the very least" steps should be taken to check the expanding domain of the formal diversionary system. Such a recommendation would ideally, of course, come before the implementation of the experiment. This would be the advantage of a pre-intervention component, designed to explore the informal means of diversion already in place. Acting on knowledge already available from former experiments, the researchers might intentionally seek out alternatives to the formal structure. They would talk with parents, police, storekeepers,

social workers, etc., in an effort to understand and describe "what is already there." The planned program might then be cancelled, altered substantially, or at least equipped with safeguards designed to reduce the foreseeable side-effects.

INMATE SELF-GOVERNMENT EXPERIMENT

The next case will be used to illustrate a number of problems that are generally associated with the evaluation of prison programs. The evaluation itself has been fictionalized slightly in order to make the recommendations applicable to similar endeavors such as inmate group counseling and behavior modification.

An inmate self-government program, the processes of which are described in only scanty detail, was tested in a maximum security prison. A pretest-posttest design was used involving a stratified random sample of 173 male inmates. Outcome measures were attitudinal, covering dimensions of social responsibility, self-esteem, self-competence, acceptance of others, and acceptance of law and authority. Evaluators found that participation consistently fostered a more positive sense of social responsibility while such attitudes, especially acceptance of law and authority, deteriorated within the nonparticipant group.

Clarifying Program Operation

The most conspicuous shortcoming of the evaluation is the lack of descriptive data on program operation. It is difficult to judge the outcomes without knowing how much autonomy was really granted to inmates and how they responded to it. Furthermore, the evaluation offers little information for program modification or replication. Observations of

governmental meetings between inmates and prison staff might reveal certain practices that should or should not be replicated in other programs. At the very least, a post-experiment process component should have been used to provide program managers with such information.

Clarifying Causality

A number of alternative hypotheses should have been forwarded by the evaluators and tested through a simultaneous or post-experiment design (depending on the risk of contamination). First of all, there is no indication that the experimental and control groups were physically separated during the experiment. The potential for "interactive effects" was high. Perhaps non-participants thought less of law and authority precisely because they were barred from participation in the program. They might otherwise have experienced the same improvement in attitude manifested by the experimentals. Second, there is the possibility that subjects answered the questionnaires not as they actually felt, but as they perceived prison officials would want them to feel. After participating dutifully in a prison program, any inmate seeking to increase his chances of release would be anxious to show an "improved attitude." (This hypothesis of unreliable testing would depend initially on the transparency of the administered questions). The evaluators attempted to show that because the inmates volunteered for the experiment, they did not feel obligated to participate, nor did they see the program as meaningless. But what may appear to be voluntary may well be perceived by the inmate to be obligatory, a fact which is well known by students of prison phenomenology.

A post-experiment process element should include interviews with the

inmates, guards and administrators, and observation of prison meetings. Information should also be obtained from program participants who wouldn't have a stake in maintaining a facade - these might be men who have been released from the prison, who have fixed sentences, who are already "informers" for the prison officials, or who can depend on other informal mechanisms to get them out on parole. Control group members should also be interviewed, for they have no stake in the program and yet are undoubtedly well informed about the "underground" attitudes toward the program. These sources of information should provide a composite picture of the degree to which the program is received sincerely and the degree to which control group members are feeling alienated.

Official vs. Operational Goals

Deutscher has commented on the tendency for actual program goals to become quite different from those that are stated officially at the outset of an intervention.²⁴ Prison programs are particularly vulnerable to this dissonance between goals. While the designers of inmate self-government viewed increasing self-esteem and social responsibility as the goals of the program, the guards and prison administrators may have viewed the program as further means to control inmates. Numerous studies of prison life have shown the complex and informal set of punishments, rewards, agreements and tradeoffs that exist between actors in the prison environment.²⁵ There is no reason to believe that an inmate self-government experiment would not be incorporated into this network. Guards could use threats of non-compliance with negotiated agreements, rewards of promised consent to inmate proposals, and various forms of bargaining, all designed to make

easier their job of maintaining control.

Techniques similar to those discussed earlier could be used in an effort to distinguish stated goals from operational goals. Even if no evidence is found that the program is being used as a coercive device, qualitative data may at least highlight friction between conflicting goals. It is inarguably a goal of the prison system to maintain control over its population. Common sense dictates that any program which is designed to provide inmates with autonomy and independence is likely to clash with the general goals of imprisonment. In describing an inmate group counseling experiment, for instance, Kassabaum notes,

(T)o the extent that obedience is the goal of correctional efforts, attempts to implement a treatment program that seeks insight into emotional determinism of conduct and increase in the sense of individual responsibility may be perceived by both staff and inmates as somewhat beside the point.²⁶

The Experimental Environment and Program Approach

A pre-intervention process component might have predicted some of the problems discussed so far. Ethnographic techniques might have presented a profile of a prison with strong informal ties between inmates, a propensity among inmates to exhibit "model," conforming behavior, and a system of punishments and rewards between staff members and clients. Such a study might have concluded that it would be extremely difficult to detect the true effect of an inmate self-government program. In fact, the researchers might also conclude that no therapeutic program within a prison setting is likely to succeed due to the tendency of inmates to be continuously labeled as criminal, and due to the contradictory, oppressive nature of incarceration itself. The criticism is even more relevant to

counseling and behavior modification programs within prisons. As the hopelessness of rehabilitation becomes apparent through process evaluation, researchers may find themselves questioning the very theories and rationales underlying such programs. Indeed, Kassabaum and his associates did just that after a thorough evaluation of a group counseling program.

The most fundamental requirement for further research on the effectiveness of prison...programs would seem to us to be a frank recognition that psychological treatment programs involve assumptions about the causes of crime, the informal and formal organization of the prison...and the nature of the postrelease experience, all of which may be quite unrealistic when applied to actual existing conditions...To the extent that prison holds a heterogeneous collection of persons, including men who have been labeled criminal without possessing abnormal emotional or personality attributes, the manipulation of such attributes, even if successful, will not affect the probability that men from prison will be again labeled criminal subsequent to their release from custody.²⁷

We have seen similar problems with the case at hand. A process component added at any stage of the inmate self-government experiment would undoubtedly have raised serious doubts about the validity of the results, and about the reasonableness of seeking inmate independence in a coercive setting.

SPECIALIZED PAROLE UNITS EXPERIMENT

In our final example, several parole units were given greatly reduced caseloads in an experiment aimed at diminishing recidivism among clients. Measures of success were also to include the improved attitudes of clients and parole agents. An interrupted time-series design was used with non-equivalent control groups of traditional parole caseloads. Questionnaires and interviews were administered and data were collected on client failure rates, but no difference was found between the parole units with differing

caseload sizes. The authors then summarized a number of past parole experiments to augment their own conclusion that caseload reductions do not make a difference, and they concluded that more information is needed on the quality and intensity of parole counseling before real progress can be made.

Experimental Design Problems

The experiment was crippled by a number of data collection problems. Some of these might have been remedied by a limited form of simultaneous-intervention evaluation. The parolee questionnaires, for instance, were filled out by clients who were handpicked by their parole agents. Then, it was later learned, the agent usually stood nearby while the client responded to what was purported to be an anonymous questionnaire. Both practices severely limited the validity of the survey results, and both practices might have been terminated if observers and interviewers had detected them earlier. This would have entailed a limited amount of questioning and observing at the time that the questionnaires were distributed, hardly enough to threaten the integrity of the experiment.

The experiment was also plagued by an excessively low return rate on all questionnaires. The consequences of this problem would have been less serious if more process-oriented evaluation had been planned. Informal discussions with parole agents and parolees might have revealed much of the information sought on the questionnaires. Admittedly, the results would be less structured and objective, but conclusions might have been more certain in light of other data.

Alternative Data Sources and Measures

Given the plethora of information available to these experimenters on the failure of past caseload-reduction experiments, a thorough pre-intervention process component was certainly warranted. Indeed, the emphasis on caseloads and recidivism rates seems somewhat narrowly focused. The evaluators should have been exploring alternative forms of assessment and improvement by implementing fieldwork studies even before the experiment was designed.

Observation and interviewing could be used to obtain a better sense of how parole agents deal with their clients. The work of unusually effective agents could have been investigated most carefully for signs of generalizable techniques. Various treatment schemes could have been developed in which parolees with different characteristics would receive distinct forms of supervision and counseling. Kassabaum summarizes the need for more work on the quality and character of parole work:

Parole outcome has been regarded implicitly as simply a function of the behavior of the parolee; this view has resulted in the neglect of the study of the parole officer as a decision maker. Moreover, the parole division has not yet been studied as a complex social organization. Thus we are led to the awareness that our data on parole success and failure do not provide a clear indicator of postrelease behavior, since we do not fully understand the nature of the parole experience.²⁸

With more descriptive information available on parole practices, a more relevant program might then be implemented. Perhaps it would include training of agents, guidelines calling for less supervision for certain types of clients, and new policies on what counseling activities should receive primary emphasis. The evaluators might recommend less supervision and control, and more help in job hunting and other forms of

aid for the parolee. Services might be oriented not so much toward changing the parolee as toward changing the mechanisms which tend to label him or her as criminal, such as frequently required check-ins with the parole office. Once an experiment with these components is initiated, process evaluators could continue to assess parole unit operations. Emphasis could be removed from recidivism rates and redirected toward counseling quality.

Causality: An Alternative Hypothesis

One interesting cause of recidivism reduction was apparently not operating in this experiment, but has been suspected in others. Martinson describes a study in which failure rates were reduced with diminished caseloads, while a look at client offenses indicated that experimentals were involved no less in crime than the controls. As the researchers discovered:

The reason that the experimentals' relatively large number of offenses was not being reflected in their failure rates was simply that the experiments' probation officers were using a more lenient revocation policy.²⁹

The agents were feeling pressured to show diminished recidivism rates, so they simply raised their threshold for parole violation.

Such a trend might only be discovered through a simultaneous-intervention design. Records would be monitored, and administrators, parolees, and agents would be interviewed. Subtle changes in revocation policies that might be ignored with an emphasis on outcomes could be "teased out" by process-oriented techniques.

Process Evaluation as a "Backup"

If process evaluation had been used fairly comprehensively in this case, the dismal performance of the experiment might have been less detrimental. As it was, the experimenters could say little more than that the design was unsuccessfully implemented and the results were inconclusive. Even barring a pre-intervention design, much could still have been learned. Rich, descriptive data on processes is much better than no information at all. Especially in this case, process data could have been used to paint a general picture of program performance and to guide future endeavors in the field.

IV. CONCLUSION

Amitai Etzioni has encapsulated many of the problems discussed here when he wrote:

Since most organizations, most of the time, do not attain their goals, in any final sense, many monographs are frequently detoured into lengthy discussions about lack of success, to the exclusion of more penetrating types of analyses.³⁰

Our attempt here has been to provide a framework for "more penetrating types of analyses." The framework is not meant to be restrictive or delimiting, for process evaluation by its very nature tends to resist inflexible systemization. Rather, the attempt has been to give evaluators a structure within which the various advantages and disadvantages of process evaluation were presented. It is hoped that the applications of these sequences to five examples has made their potential more clear.

Two Operational Realities

Two remarks about the application of process evaluation should be made. First, process-oriented techniques and experimental techniques are not always compatible. There are very real differences in how field work is conducted as opposed to experiments, and these differences can lead to substantial tension between the two evaluation components. There may be problems, for instance, created by the time-frames of the two components, or by the nature of the methodologies used. These tensions cannot be explored fully here, but the reader should be aware that the merging of process evaluation with experimentation is not always an easy task.

Second, the limitations of process evaluation should be recognized. As was seen a number of times in the narratives, a process study can be no

more certain about causality and other issues than can an experiment. What makes process evaluation such a powerful tool is its potential to help identify, or question program effects -- not to prove or disprove them. As Patton points out:

Evaluation research is only of use if one believes that some systematic information is better than none. Evaluation research has meaning only if one believes that a rough idea of the relationship between program activities and outcomes is preferable to relying entirely upon hope and good intentions. Evaluation research does not provide final answers, but it can provide direction. Thus, evaluation research does not lead to final statements about causal linkages, but can reduce uncertainty about such linkages.³¹

Processes and Outcomes

While the benefits of process evaluation outlined in this paper have been diverse, they are all linked, to one degree or another, to the outcomes of experimental design. We have seen, first of all, that a recognition of program setting can lead to program modification and more successful outcomes, or to a modified interpretation of outcomes. In the first Team Policing experiment, for instance, evaluators found that conflicting programs and policies were interfering with the performance of the new teams. Had the evaluators discovered these trouble spots before the experiment, appropriate adjustments might have led to more certain and more policy-significant outcomes.

Second, we explored a number of situations where alterations in the experimental design might have clarified or confirmed outcomes. For the second Team Policing experiment, for instance, it was suggested that a process evaluation might have contributed to a choice of control beats that were more comparable with the experimental beats. A more informed

selection of beats could have resulted in more reliable outcomes, or perhaps in different outcomes together. Likewise, the possibility was raised that the measures used in the experimental design were inadequate. Again, outcomes may have been altered substantially if different or more measures had been used.

A third advantage of process evaluation was the clarification of program operation and implementation. In the Pretrial Intervention Project there was a need for a process study to pinpoint just why some programs did not reach their projected caseload size and why only 50% of the total project clients were being referred to community services. With such information in hand during program operation, policy changes might have contributed to more promising outcomes by the time posttest data were collected.

The fourth advantage of process evaluation was the explanation of causality. For the second Team Policing Experiment it was conceivable that one team performed more successfully because of their unusual reliance on mug shots. Again, the eventual status of outcomes might have been affected had a process component discovered this practice earlier.

Fifth, a process study can discover and substantiate unintended or unanticipated program effects. In the case of the Pretrial Intervention Project, the legal "net" may have been widened to induct more, rather than fewer, clients into the legal system. If such an occurrence were to be discovered through process evaluation, the outcomes of the experiment would have to be viewed in an entirely different light. Even "successful" program participants might have avoided court processing altogether before the experiment was initiated.

Finally, we discussed the possibility that a pre-program process component could bring into question the likelihood that desired results are achievable. The study can become actively critical in this sense, perhaps challenging the assumptions, the goals, or the intended means of a planned program. In the Inmate Self-Government case, for instance, the prison environment might have been too contradictory in nature to allow for meaningful, positive outcomes. Again, eventual outcomes might have been viewed in a different light or perhaps not even pursued had a process study been conducted.

Summary

In the past two decades, social scientists have learned that experimental design as an evaluation mode can be plagued by the unexpected and the uncertain. One response to this lesson has been a call for the complementary use of process-oriented research. This paper has shown just how process evaluation can be formative, feeding information to decision-makers soon enough for meaningful changes in the program and in the components of the experimental design. We have also shown how process evaluation can help to modify, clarify, or confirm experimental results. It is hoped that these advantages can be best reaped by evaluators as they consciously apply process-oriented techniques at certain points in the experimental design.

It would be impossible to offer a step-by-step recipe for the complementary use of process evaluation. What we have done instead is to provide a set of signposts for evaluators. Perhaps choices about the systematic use of process evaluation can be made more constructively with some of these signposts in mind.

FOOTNOTES

¹Robert Boruch, "On Common Contentions About Randomized Field Experiments," in Gene V. Glass (ed.) Evaluation Studies Review Annual (Beverly Hills: Sage Publications, Inc., 1976), pp. 159-194.

Robert S. Weiss and Martin Rein. "The Evaluation of Broad-Aim Programs: Experimental Design, Its Difficulties, and an Alternative," in Administrative Science Quarterly, March, 1970. pp. 107-109.

²Boruch, pp. 159-186.

Donald T. Campbell, "Reforms as Experiments" in Guttenberg and Struening (Eds.), Handbook of Evaluation Research Vol. I (Beverly Hills: Sage Publications, 1975) pp. 71-137.

Robert S. Weiss and Martin Rein. "The Evaluation of Broad-Aim Programs: Experiment Design, Its Difficulties, and an Alternative," in Administrative Science Quarterly, March, 1970. pp. 107-109.

³Robert S. Weiss, "Input Studies and Process Studies as Alternatives and Complements to Outcome Evaluation," Unpublished Manuscript, pp. 21-24.

Michael Quinn Patton, Utilization-Focused Evaluation (Beverly Hills: Sage Publications, 1978) pp. 165-167.

Irwin Deutscher, "Toward Avoiding the Goal-Trap in Evaluation Research," in Frances G. Caro (ed.), Readings in Evaluation Research (Second Edition), (New York: Russell Sage Foundation, 1977), pp. 221-235.

Robert S. Weiss and Martin Rein. "The Evaluation of Broad-Aim Programs: Experimental Design, Its Difficulties, and an Alternative," in Administrative Science Quarterly, March, 1970. pp. 107-109.

⁴Donald T. Campbell and Julian C. Stanley, Experimental and Quasi-experimental Designs for Research (Chicago: Rand McNally College Publishing Co., 1963).

⁵Two hundred evaluation reports have been reviewed. See: Richard C. Larson et. al., "Interim Analysis of 200 Evaluations on Criminal Justice," from Research Project entitled "An Empirical Study of Methods Used in Criminal Justice Evaluations," (MIT, 1979).

⁶See References listed under Footnote #3.

⁷Robert K. Yin. "What is Citizen Crime Prevention," (MIT: Unpublished Paper), 1978, p. 28.

⁸Patton, p. 165.

⁹Patton, pp. 165-167.

¹⁰Larson, et. al. pp. 36-39, p. 70.

¹¹Deutscher, pp. 221-239.

Larson, et. al., pp. 36-39, p. 70.

Patton, pp. 175-176.

¹²Campbell and Stanley, pp. 6-70. (Further references to Campbell and Stanley will come from these pages.)

¹³While I am not aware that any author has compared the advantages of certain evaluation sequences, several authors have discussed the general advantages of process evaluation. Few of the advantages listed here (and under the other designs) could be attributed to any single writer, but credit should be attributed to several people who have discussed many of these ideas in one form or another. See especially Weiss and Rein (Footnote #1); Deutscher, Patton, Weiss (Footnote #3); and Yin (Footnote #7).

See also,

On the clarification of program implementation and operation:

Howard E. Freeman, "The Present Status of Evaluation Research," (Manuscript prepared for the 1976 UNESCO Evaluation Research Conference, Washington, D.C.)

Robert B. Everhart, "Problems of Doing Fieldwork in Educational Evaluation." Human Organization, Vol. 34, No. 2 (1975), p. 208.

On the clarification or confirmation of causal factors:

Michael Scriven. "Maximizing the Power of Causal Investigations: The Modus Operandi Method," in G.V. Glass (ed.) Evaluation Studies Review Annual, Vol I. (Beverly Hills: Sage Publications, 1976).

H. C. Quay. "The Three Faces of Evaluation: What Can Be Expected To Work." in Criminal Justice Behavior, Vol. 4 (1977), pp. 341-354.

On the assessment of the program approach:

N. Christie. "Is it Time to Stop Counting?", in Evaluation Research in Criminal Justice. Materials and Proceedings of a Research Conference Convened in the Context of the Fifth United Nations Congress on the Prevention of Crime and Treatment of Offenders. (Rome, Italy: U. N. Social Defense Research Institute, 1976), p. 270.

¹⁴Campbell, pp. 71-137.

¹⁵Peter K. Manning, Police Work (Cambridge: MIT Press, 1977).

¹⁶Manning, p. 140.

¹⁷Donald Schon, "Network Related Intervention," Unpublished Manuscript, 1977, p. 2.

¹⁸Robert Klittgaard, "Going Beyond the Mean in Educational Evaluation," Public Policy Vol. 23 (Winter, 1975), pp. 59-79.

¹⁹Amitai Etzioni, "Some Dangers of 'Valid' Social Measurement," from Social Intelligence for America's Future, ed. by Bertram M. Gross, p. 51.

²⁰These characteristics were derived from a number of publications:

Bruce Bullington, et. al. "A Critique of Diversionary Juvenile Justice," Crime and Delinquency (January, 1978), pp. 59-71.

Ronald Roesch, "Does Adult Diversion Work?" Crime and Delinquency (January, 1978), pp. 72-80.

Don C. Gibbons and Gerald F. Blake, "Evaluating the Impact of Juvenile Diversion Programs," Crime and Delinquency (October, 1976). pp. 411-420.

²¹Bullington, p. 67.

²²Gibbons and Blake, p. 413.

²³Bullington, et. al. p. 63. (Study referred to by Bullington et al.: Daniel Katkin, et. al. Delinquency and the Juvenile Justice System. (Bound Brook, Ct.: Duxbury Press, 1976).

²⁴Deutscher, pp. 221-239.

²⁵See, for example, Erving Goffman, Asylums (New York: Anchor Books, 1961).

²⁶Gene Kassabaum, et al., Prison Treatment and Parole Survival: An Empirical Assessment. (New York: John Wiley and Sons, 1972), p. 372.

- ²⁷Kassabaum, p. 322.
- ²⁸Kassabaum, p. 340.
- ²⁹Robert Martison, "What Works? Questions and Answers about Prison Reform," The Public Interest. (Spring, 1974), p. 25.
- ³⁰Etzioni, p. 52.
- ³¹Patton, p. 180.

BIBLIOGRAPHY

- Berk, Richard A. and Peter H. Rossi. "Doing Good or Worse, Evaluation Research Politically Re-examined," Social Problems, Vol. 3, no. 3 (Feb., 1967). pp. 337-349.
- Bier, Vicki. Discussions and Comments, May, 1979.
- Boruch, Robert. "On Common Contentions About Randomized Field Experiments," Evaluation Studies Review Annual, Volume I, ed. G. V. Glass, Beverly Hills, 1976, pp. 158-194.
- Buckle, Leonard, Discussions, April-May, 1979.
- Bullington, Bruce, et. al. "A Critique of Diversionary Juvenile Justice," Crime and Delinquency, Vol. 24 (January, 1978), pp. 59-71.
- Buch, Malcolm and C. Gordon. "The Advantages of Client Involvement in Evaluation Research," Evaluation Studies Review Annual, Vol. 3, ed. Cook, et. al., Beverly Hills, 1978.
- Campbell, Donald T. "Reforms as Experiments," Handbook of Evaluation Research, Vol. I. Beverly Hills, 1975. Pp. 71-137.
- Caro, Frances G. Readings in Evaluation Research (Second Edition). New York, 1977.
- Coates, Robert B. and Alden D. Miller. "Evaluating Large Scale Social Service Systems in Changing Environments: The Case of Correctional Agencies," Journal of Research in Crime and Delinquency, (July, 1973), pp. 92-106.
- Chomie, Peter and Joe Hudson. "Evaluation of Outcomes and Processes." Social Work, Vol. 19, (November, 1974), pp. 682-687.
- Christie, N. "Is it Time to Stop Counting?" Evaluation Research in Criminal Justice. Materials and Proceedings of a Research Conference Convened in the Context of the Fifth United Nations Congress on the Prevention of Crime and Treatment of Offenders. (Rome, Italy: U.N. Social Defense Research Institute, 1976).
- Deutscher, Irwin, "Toward Avoiding the Goal-Trap in Evaluation Research," Readings in Evaluation Research (Second Edition), ed. Francis G. Caro, New York, 1977. Pp. 221-239.
- Etzioni, Amita. "Some Dangers of 'Valid' Social Measurement," Social Intelligence for America's Future, ed. Bertram M. Gross. Boston, 1969, pp. 48-60.
- Everhart, Robert B. "Problems of Doing Fieldwork in Educational Evaluation." Human Organization, Vol. 34, No. 2 (Summer, 1975), pp. 205-215.

- Gibbons, C. and Gerald F. Blake. "Evaluating the Impact of Juvenile Diversion Programs," Crime and Delinquency, Vol. 22 (October, 1976), pp. 441-420.
- Gifford, Bernard. "Issues in Public Policy Analysis," Class lectures at MIT, Spring, 1979.
- Goffman, Erving. Asylums. New York, 1961.
- Hawkrige, David. "Designs for Evaluation Studies," Evaluation Research Strategies and Methods. Washington, D. C., 1970.
- Kaplan, Edward, Discussions, May, 1979.
- Kassabaum, Gene, et. al. Prison Treatment and Parole Survival: An Empirical Assessment. New York, 1972.
- Klittgaard, Robert. "Going Beyond the Mean in Educational Evaluation," Public Policy, Vol. 23 (Winter, 1975), pp. 59-79.
- Larson, Richard C. "Evaluation of Public Programs," Class lectures at MIT, Fall, 1978.
- Larson, Richard C., et. al. "Interim Analysis of 200 Evaluations on Criminal Justice," from the Research Project: "An Empirical Study of Methods Used in Criminal Justice Evaluations." MIT, 1979.
- Manning, Peter K. Police Work. Cambridge, 1977.
- Martinson, Robert. "What Works? Questions and Answers about Prison Reform," The Public Interest, No. 35 (Spring, 1974), pp. 22-54.
- Marx, Gary. "Issues in Criminal Justice." Class Lectures and Discussions at MIT, Spring, 1979.
- Mattingly, Cheryl. "Making the Case for Process Evaluation," Unpublished Manuscript, MIT, 1978.
- Patton, Michael Quinn. Utilization-Focused Evaluation. Beverly Hills, 1978.
- Quay, H.C. "The Three Faces of Evaluation: What Can be Expected to Work," Criminal Justice Behavior, Vol. 4 (1977), pp. 341-354.
- Roesch, Ronald. "Does Adult Diversion Work?" Crime and Delinquency, Vol. 24 (January, 1978), pp. 72-80.
- Rossi, Peter. "Issues in the Evaluation of Human Services Delivery," Evaluation Quarterly. Vol. 2, No. 4 (November, 1978), p. 580.
- Schon, Donald. "Network Related Intervention," Unpublished Manuscript, 1977.

- Scriven, Michael. "Maximizing the Power of Causal Investigations: The Modus Operandi Method," Evaluation Studies Review Annual, Vol. I, ed. G.V. Glass, Beverly Hills, 1976, pp. 101-118.
- U.N. Social Defense Research Institute. Evaluation Research in Criminal Justice. Materials and Proceedings of a Research Conference Convened in the Context of the Fifth United Nations Congress on the Prevention of Crime and Treatment of Offenders. Rome, 1976.
- Weiss, Robert S. "Input Studies and Process Studies as Alternatives and Complements to Outcome Evaluation," Unpublished Manuscript.
- Weiss, Robert S. and Martin Rein. "The Evaluation of Broad-Aim Programs: Experimental Design, Its Difficulties, and An Alternative," Administrative Science Quarterly (March, 1970), pp. 107-120.
- Yin, Robert K. "What is Citizen Crime Prevention?" Unpublished Paper, MIT, 1978.

Special Thanks to Professor Richard Larson and Vicki Bier for their helpful comments on this paper.

END