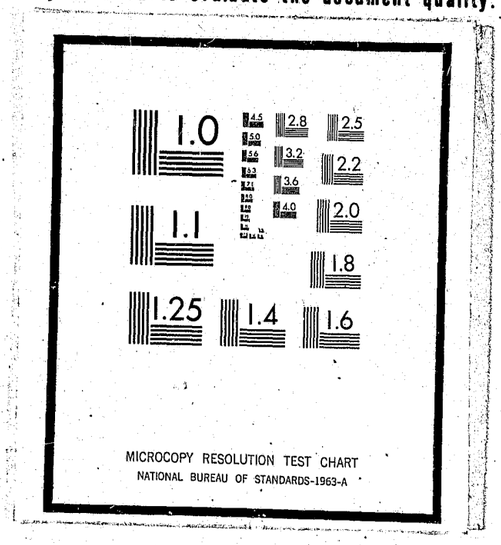
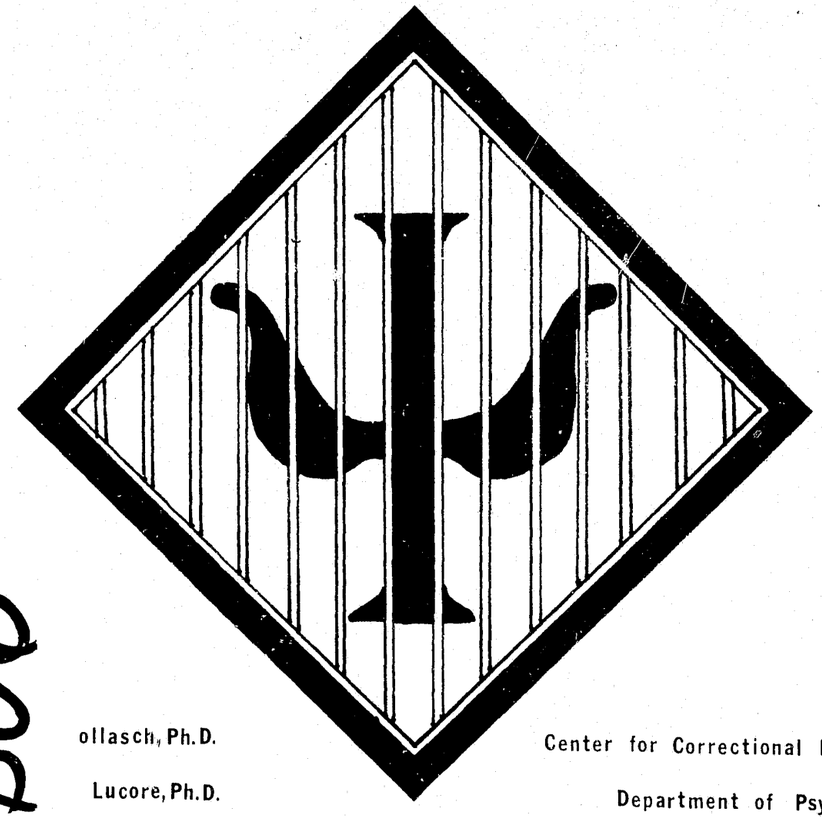


4

A GUIDE TO IMPLEMENTATION and EVALUATION of JUVENILE JUSTICE PROGRAMS



ollasch, Ph.D.
Lucore, Ph.D.
yston, Jr.
riffin

Center for Correctional Psychology
Department of Psychology
University of Alabama

March, 1974

A GUIDE TO IMPLEMENTATION AND EVALUATION OF JUVENILE JUSTICE PROGRAMS

Center for Correctional Psychology
Report Number 11

This handbook was prepared under a grant from the Alabama Law Enforcement Planning Agency (ALEPA 72-DS-1-4-1-35-1158-11) to the Department of Psychology, University of Alabama. Points of view or opinions stated are those of the authors and do not necessarily represent the opinions or policies of the Alabama Law Enforcement Planning Agency.

We are indebted to the Los Angeles County Sheriff's Department for the sections entitled: Operational Definitions of Variables, Scientific Method and Research Methology, and Beyond Program Evaluation to Agency Accountability. These sections were adopted with only minor editing from "Program Measurement and Evaluation", In: Criminal Justice System, National Advisory Commission on Criminal Justice Standards and Goals, September, 1972, pp. 215-247.

Project Staff

Sylvia Kollasch, Ph.D.
Patricia Lucore , Ph.D.
Leon Royston, Jr.
Fannie Bulls
Michael Fazzio
Lola Lawson

TABLE OF CONTENTS

PART I

THE PROGRAM

THE PROGRAM

CHAPTER 1 INTRODUCTION

Programs which are established to deal with problems of juvenile delinquency and other forms of corrections generally arise out of perceived inadequacies of the present systems for handling such problems. For example, a judge may attempt to institute some form of temporary shelter care because he/she doesn't want to keep children in the local jail and is presented with a sufficient number of runaway children or children who need shelter to merit establishing an alternate facility. A probation officer might be concerned with the fact that some of his probationers aren't responding favorably to supervision, but has only the alternative of sending them to a state training school which might aggravate the existing problems. As a result, the probation officer may attempt to establish a group home in the community, thus providing another alternative which might be less harmful than incarceration.

As many of our readers know from frustrating experiences, establishing a program has many pitfalls and calls for a sophistication in activities for which many juvenile justice personnel have little background. The purpose of the present package is to provide a handbook for such personnel in formulating and evaluating needed programs. The package is also designed to provide assistance to on-going programs for purposes of self-analysis and upgrading evaluation processes.

In general, the birth of a program takes the following steps, though not necessarily in the exact order listed below.

A. Assessment of Needs

At this stage, one or more needs of the existing system become apparent. It is best to take a careful look at the whole system to see if some other less apparent weakness might need attention. Eventually, a decision will have to be made as to whether the entire program needs to be remodeled or whether it would suffice to add a new program to satisfy the most pressing needs.

B. Scope of Program Needs

If a large area or region has similar needs, but the cost assessment runs too high for any local area to operate its own comprehensive program, a regional concept should be considered. While there are benefits in regional programs in their ability to offer more diversified services and uniformity of handling, etc; there are also benefits in dealing with juvenile offenders in their own community and having services available locally.

C. Assessment of Community Support and Resources

Since most granting agencies require some community funds to be provided, program planners must find a source for continued local funding of their program. Equally important, however, is to consult various members of the area as to the response of the community to the proposed program. Questions of availability of facilities and personnel need to be considered.

D. Development of a Philosophy of Treatment

By this point in time, program planners probably have at least a notion of their philosophy regarding the causes of delinquency and optimal means of handling juvenile offenders. This philosophy is the key to all planning, and should be given serious thought.

E. Development of Program Design, Goals and Objectives

The design of a new program is affected by financial limitations, community response, philosophy of the program planners, perceived needs of the community or area, and other factors. Future

difficulties may be avoided if fairly specific decisions are made regarding who the program will deal with, what techniques will be used, the chain of accountability, and the job descriptions of each of the personnel.

F. Assessment of Monetary Needs to Implement Design

Careful estimates should be obtained on competitive salary rates, costs of facility up keep and rental, insurance, travel, phone, food and other custodial costs (in custodial programs), backup personnel, and whatever other costs are arrived at. Cost computations should be as accurate as possible, as underestimates or overlooked expenses will haunt administrators for the entire grant period. On the other hand, you must be able to justify each expense, so overestimates are unnecessary and may jeopardize grant approval.

G. Organization of Advisory Committee and Agency Contacts

The advisory board should consist of local juvenile justice personnel (judge, etc.) and other interested citizens, preferably representing a variety of areas of competence. Contacts should be made with other agencies which may affect or be affected by the new program, and tentative working agreements should be established. Such agencies might include the school system, Department of Pensions and Securities, mental health and medical programs, and any other agencies which are considered relevant. If resistance is anticipated from the community or agencies, it would be worthwhile to consult with resisting groups for their suggestions and allow them to express their opinions and reservations. Some of the suggestions may be worthwhile and the groups may have a more favorable outlook simply because they were given an opportunity to contribute.

H. Collection of Baseline Data

This step may occur at any point, but should not be delayed until the program has begun operation. Data should be compiled for the area and populations the proposed program will serve, preferably including several years of court statistics and dispositional information. Other data collected at this time will depend on the specific goals and objectives of the new program and the evaluation research design and criteria.

I. Identification of Grant Sources and Grant Proposal Preparation

Several state and federal granting agencies have funds available for programs, depending on what type of services the program provides, eg. LEAA, HEW, NIMH, etc. Contacts with these sources will generally help to define which source is most appropriate to apply to for funding, following which, application forms should be obtained.

The grant proposal should be in a readable format (eg. outline) and should clearly state the need for the program, its objectives, philosophy, structure, treatment methods, and personnel qualifications and roles. All expenditures should be broken down into component parts, providing monthly rates for utilities, rates per mile for travel, cost per item of furniture, etc. A written justification of each expense should be included in the budget narrative, explaining briefly how rates were arrived at, what each staff member is expected to do, etc. A schedule of operations should be provided, giving deadlines for the accomplishment of specific tasks. It is usually a good idea to provide a little lead time at the beginning of the grant period to ensure hiring and orientation of personnel before deadlines begin to occur.

J. Search for Facilities and Personnel

Once the grant proposal has been submitted, a search should be initiated for appropriate facilities and program personnel. A director or administrative personnel should ideally be selected first so that they may participate in selection of other personnel.

K. Training and Orientation of Personnel

Personnel cannot officially be hired until the grant has been approved. The first week or more of program operation will usually be spent heavily in procedural planning and familiarization

of personnel with the problems involved in their work. If special training is required (eg. behavior management), qualified consultants should be utilized to provide such training.

L. Program Operation

Once the program is functioning and the staff become familiar with their roles, it is generally a good idea to have regular staff meetings to discuss special problems and to make minor modifications.

M. On-going Evaluation

For purposes of evaluation, records need to be kept up to date on whatever research criteria were determined in advance. These criteria should directly reflect the goals of the program. If no available personnel or board members have knowledge of research methodology or statistics, some funds might be included in the proposal for consulting services to assist in designing and evaluating research efforts. The present handbook may be adequate to guide an evaluation, if evaluation needs are not too complex.

The information in this handbook is intended to reply essentially to two program needs:

1. To serve as a guide for designing a coherent and internally consistent program, so that it in fact functions as it is intended to. In order to facilitate that goal, some information is provided regarding types of programs available and sources to consult (See Appendix A). These sources are not intended to be exhaustive as space does not permit intensive treatment of philosophies and methods.

2. To provide a framework for program evaluation, including information on criteria, measurement, research design and analysis of results.

CHAPTER 2 SYSTEMS APPROACH

Section 1. Juvenile Justice Programs as Open Systems

Programs which are designed to respond to a need in the community and which are funded from public sources do not have the luxury of being closed systems. They are subjected to varied pressures from their constituency to perform activities and strive toward goals which are often mutually exclusive. This fact, coupled with a diversity of backgrounds of personnel within such programs, produces a tendency to undermine the ability of a program to be internally consistent and maximally organized to perform its stated goals. If the system fails to meet minimum societal demands, it is likely to die an early death. Likewise, if goals are ill-defined, or operations do not fit the goals, the program will be more susceptible to pressures.

Figure 1 presents a systems model for juvenile justice agencies, indicating input, throughput, and output characteristics with arrows showing the relationships between elements (See Figure 1). Since the chart appears complex, explanations of the elements are included in this section.

In order to avoid any misunderstanding of the present use of the word treatment, the term is intended to reflect a research model, rather than a medical model. Treatment refers here to any type of intervention program as opposed to no intervention in the client's life.

INPUT

A. Philosophy:

The philosophical and/or theoretical position one takes when organizaing a program for juvenile corrections is of utmost importance, and will affect every aspect of program implementation.

If, for example, one takes the hypothetical position that delinquent behavior is caused by a hormonal imbalance, a program would be designed to facilitate physiological rather than psychological or environmental changes. This, in turn, will affect: 1) the type and qualifications of personnel required to staff the program - in this hypothetical system, medically and/or biochemically-oriented persons. 2) The physical plant required to house and carry out the procedures necessary. Structures to carry out medical procedures will differ from ones designed for psychological procedures. 3) The amount of money needed to finance the operations will differ from one type of approach to another (programs requiring custodial care higher). 4) The type of records kept and data gathered will depend greatly on the program's theoretical orientation towards delinquency. For evaluation purposes, records should contain data that reflect the programs objectives. 5) The methods of intervention used are obviously affected by the program's theoretical position. One would not, for example, initiate group therapy as a treatment method if the problem were viewed as hormonal imbalance.

B. Goals and Objectives:

The proposal of a program's goals and objectives is an essential and extremely important aspect of project development. Numerous considerations must be taken into account, and all factors relevant to goal development should be carefully and thoughtfully analyzed.

Many proposed social action programs incorporate expansive, all-encompassing goals, such as, "to rehabilitate delinquents" or "prevent delinquency", failing to break these concepts down into their objective specifics. In order to evaluate program effectiveness, it is necessary to have some measurable criteria for goals and objectives. Good intentions, though a worthy starting point, do not necessarily lead to positive results.

More detailed information on conceptualizing goals and objectives will be discussed in the section on evaluation criteria.

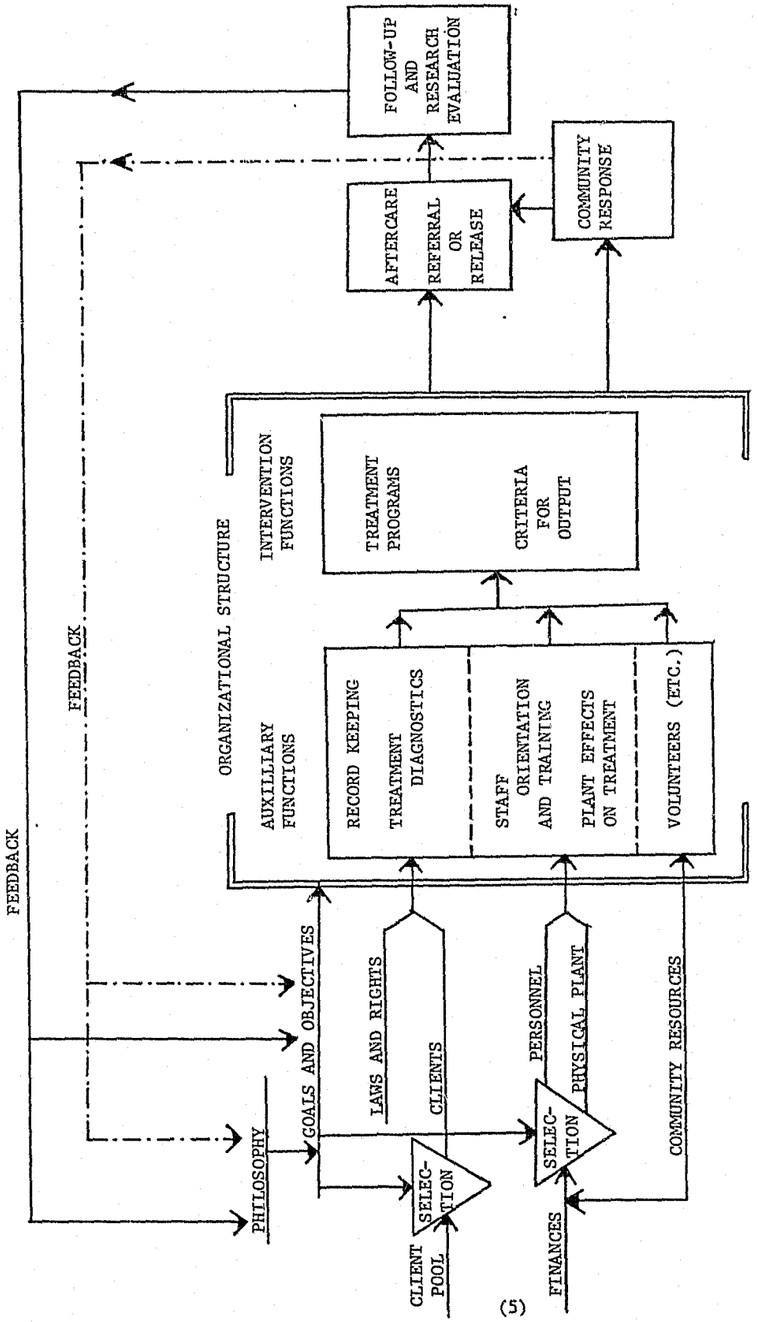


Figure 1. A Systems Model For Juvenile Justice Programs.

C. Laws and Rights:

A working knowledge of juvenile laws and the juvenile code is essential to any person working in a juvenile program no matter what the philosophy or the goals and objectives of the program. Client rights must be protected; he/she must have a good understanding of his rights as well as what alternatives and options he is entitled to under the law. It should never be assumed that the agency or program decides what a child should do without first consulting the individual in question. He should be allowed to have input into any decision involving a method of intervention in his life.

D. Client Pool and Clients:

The client pool for any project of this nature is vast, making careful screening of the numerous potential clients necessary in order to involve as few persons as possible. Labeling of people comes through association, and where mere association or involvement with a particular program (such as a detention facility) may be detrimental to a child, his association with this program should be avoided when possible.

The goals and objectives of the program determine which types of behaviors are to be treated, thus the persons who will be selected as clients. For example, if the program is aimed at runaways only, it would not be expected that a large number of armed robbery clients would be handled.

Personnel and facilities of the program would also serve to limit the kinds of problems and the numbers of juveniles who will be involved in the program. One would not introduce a severely retarded person into the program without having facilities or personnel available who could work with such a person. Thus, such an individual would likely be referred to another agency. It is a significant function of the program and its personnel to be well acquainted with sources and agencies to which individuals with specific needs may be referred.

E. Finances:

Every project is limited by its budget. When the amount of money available to a total program is fixed, the program's goals and objectives, personnel, facilities, and number of clients are a function of that fixed sum. For example, expansive goals and objectives are not possible when there is a restricted amount of money. Thus, the program goals and objectives must be correlated with the program's financial situation. In addition, salaries must come out of the budget, making it necessary to assess the number and qualities of personnel required to make the program work.

F. Personnel:

As mentioned above, several factors influence the selection and employment of program personnel. Besides those already cited, the personnel pool is affected by the program location. One will be less likely to find certain kinds of personnel in rural settings than in urban settings. This fact makes it essential to assess both program needs and personnel resources available in the particular community to obtain an idea of whether or not it will be necessary or possible to recruit personnel from outside the community. For example, it is somewhat impractical to design a program in a rural area which will require the full-time services of 5 Ph.D. psychologists unless there has been some positive indication that qualified people are willing to locate in that area. If a program is built around high level personnel, monetary reality must also be faced. The program must be willing to budget salaries which will draw individuals at the desired level.

The numbers of personnel to be employed in a program should be logically and realistically determined prior to personnel selection. In addition, job descriptions should be drawn up listing both the qualifications necessary to fill the position and the duties to be performed by the personnel hired. Personnel should have a philosophy essentially consonant with that of the program so that their activities do not undermine the goals of the program.

G. Physical Plant:

The facilities of a proposed program will be influenced by several factors. Among these are finances and available facilities, program goals and objectives, and numbers of staff and clients

to be involved in the program. Goals and objectives, client needs, and numbers of staff should be kept in mind when selecting a project site. The plant should provide sufficient space for both staff and clients and their activities. For instance, if the client population is expected to be ten to fourteen persons a day, a large building is not needed to accomodate them. When buildings are adapted to purposes outside of their original design, it is important to consider the possible effects of the structures on the treatment. For example, if an empathetic counselor is expected to play the role of jailor in a sterile lockup, his relationship with the client will be confused both for himself and the client.

H. Community Resources:

Several questions need to be asked and answered about the community where a program is to be located: Are there agencies in the community which can handle clients referred from the program? Is in-service staff training available, such as from a college or university located in the community or through consultants from one located elsewhere? What is the feasibility of employing community volunteers from civic action groups, religious organizations, or interested private citizens? Is the community amenable to locating the program here? In every case, the community is entitled to know and understand the kinds of problems, the number of clients, etc. to be helped through this program. Their interest and participation in the program should be encouraged. Program success, especially in the area of juvenile corrections, hinges greatly on community resources and acceptance of the program.

THROUGHPUT

A. Organizational Structure:

An organizational structure, though a necessary ingredient in any successful program, should be flexible enough to allow some room for individual style differences. The program goals and objectives serve as part of the basis for deciding on the type of organizational structure that the program should have. The structure should be such that individual responsibilities are defined, and that assessment of personnel performance is possible.

B. Record Keeping:

The kinds of records kept by the program is in large part determined by the direction the program hopes to take. For example, if the program is basically a behavior modification project, records of individual behavior are necessary. If the idea behind the project is one of warehousing, numbers figure prominently in the records of the program. In any case, adequate record keeping systems should be designed before the project is underway, keeping in mind that they may need to be modified and/or updated as the program progresses or as treatment techniques are changed.

C. Treatment Diagnostics:

No treatment program is equally effective with all types of clients, therefore an early decision must be made as to appropriate methods to be used with each. There are two aspects of treatment diagnostics: 1) determining if the program is suitable for the client; and 2) deciding how the client will be handled within the program. Initial diagnosis should determine whether the client has been properly screened when selected or if he can be better treated in another program. If diagnosis determines he can be served, the next step is to decide which of the options for treatment within the program is most suitable, to make individualized treatment plans, and to assign appropriate personnel to implement the plan. Diagnostic criteria may include the type and severity of the offense; the background, age, and personality of the client; and the general etiology of the problem.

D. Training:

The kind of program to be implemented determines the level and kind of pre-service and in-service training desirable for the personnel employed in the project. If the program is to involve

a certain kind of therapy, then people qualified to utilize that therapy should be employed or train those persons who are employed. In-service training is necessary to keep program personnel up-to-date on innovative techniques, literature and new theoretical positions in the particular field.

E. Effects on Treatment:

The kinds of physical facilities available influence the kinds of treatment that a program can offer. Certain physical requirements must be met in order that certain types of treatment operate effectively. If a program is to be recreational in nature, recreational facilities must be available for the client's use. In like manner, if individual therapy is to figure greatly in the programs,rooms for the individual and his therapist must be accessible.

F. Volunteers, etc.:

The resources of the community are often critical to the operation of particular kinds of programs. If a large base of community support is behind a program, the project will likely be more effective and stable. If one-to-one juvenile and adult contacts are crucial to the functioning of the program, but it is not possible to maintain such a staff-client ratio, then volunteers from the community might be trained and utilized effectively. Wherever feasible, a specific staff position of volunteer coordinator should be established.

G. Methods of Intervention:

In deciding on the methods to be used in the program, one must keep in mind the goals of the program, the kinds of personnel (staff and volunteer) and facilities available to implement these goals, characteristics of clients and their reasons for becoming involved in the program. Intervention in a person's life, when called for, must be humanely and ethically accomplished.

H. Criteria for Output:

It must be clearly understood by both staff and clients what should be accomplished in order for the client to move on to another kind of treatment or to another agency or out of the system entirely. Criteria for release or transfer should be directly related to the objectives and philosophy of the program.

OUTPUT

A. Follow-up:

Once clients complete the program, some follow-up procedures and methods should be utilized to discover what happens to them when they return to their own community. Depending on the kind of information the program needs, personal interviews or other follow-up procedures should be built into the program. Follow-up considerations are primarily for research purposes, but may also have therapeutic support value.

B. Aftercare:

Any program which involves long term custody of a youth in unfamiliar surroundings should provide or insure provision of some form of assistance upon release in reorienting the youth to his/her community. Failure to provide effective aftercare may eradicate whatever beneficial effects a program has achieved.

C. Research Evaluation:

In order for a program to maintain effective responsiveness to the needs of the area served, there must be continuous research evaluation of the program's effectiveness. Such research provides a framework for updating program goals and techniques for treatment. In order to draw definite conclusions about the program's role in any changes which occur, the evaluation procedures need to

include air-tight research designs which are carefully implemented. Research criteria and procedures are outlined in a later part of this paper.

Section 2. Systems Approach: Basic Steps in Program Evaluations

The systems approach is a general approach to problem solving and a group of scientific techniques drawn from many fields and disciplines that are used in solving particular problems. In a word, the systems approach applies the scientific method to the solution of practical problems by focusing on two familiar forces - reason and adaption. (See Figure 2)

Systems Approach as an Evaluation Tool

Program evaluation should employ procedures of the systems approach to analyze effectiveness of ongoing and proposed programs to accomplish organizational objectives.

In following this approach, the researcher should:

a. State the real need the program is supposed to satisfy;
b. Define the objectives which will contribute to satisfying the real need;
c. Define those practical limiting constraints which any proposed program must deal with;
d. Generate alternatives;
e. Select the best alternative by careful cost/benefit analysis;
f. Implement the selected alternatives for testing and evaluation;
g. Evaluate the experimental system; and
h. Based on experimental and real world results, feed back the required modifications and continue this cycle until the objectives have been attained.

The Need Statement

The researcher must develop a statement (the need statement) defining the actual problem that is facing the organization - the problem that the organization is attempting to solve effectively with the proposed or ongoing program. The steps are as follows:

1. First define the problem in generalized terms. Example: Our community needs less juvenile crime.
2. What appear to be the causes of the problem? To what extent are they currently known?
3. Who are the specific groups affected? Identify their special traits.
4. Isolate the factors that would resolve the problem: eg. to lessen juvenile crime, establish a group home.
5. Determine the type of personnel, skills and abilities that are requisite. Define in depth and mention specifically the people and the knowledge - skills necessary to fulfill the need state; i.e. better environment for the juvenile from deprived homes.
6. Re-evaluate #1. Verify the need. Is it an assumption or is it real? Is it a problem fabricated to amplify a presupposed answer?

Ironically, the need statement is difficult to evaluate. Perhaps there are too many needs or perhaps the needs arise from problems that are too large. The world "needs peace" and the nation "needs a crimeless society" but these two needs are too vast, too amorphous, to permit the problems to be defined. Better to say "the world needs a nuclear arms treaty with the Soviet Union" or that "American cities need a 20 percent reduction in theft." These needs are less extravagant, far more achievable, and well within the framework of the "divide and conquer" strategy of the systems analyst.

Define Objectives

The researcher must designate the objectives of the organization that will satisfy the real need. What does the organization have to do? What are its goals? Steps include:

1. Define how and why the objective will contribute to the fulfillment of the real need. Does it

FIGURE 2. DYNAMICS OF THE SYSTEMS APPROACH[11]

1.	Setting objectives, and subobjectives; clarifying where department should be going.	10.	Consulting with potential users, testing alternatives on trial basis; debugging.
2.	Forecasting, long-range planning; and contingency prediction.	11.	Deciding: selection of alternatives and level of service.
3.	Assessing present system; how well is it performing.	12.	Programming for chosen alternatives; "trouble shooting".
4.	Analyzing problems, needs, and gaps between present state in light of resources, constraints and forecasts.	13.	Budgeting: investing resources according to decisions.
5.	Scanning the real world experientially and via literature for new alternatives.		
6.	Inventing solutions, posible new programs; creative alternatives.		
7.	Planning alternatives and program; gaining consideration for them; developing performance criteria.		
8.	Developing alternative programs; analyze cost/benefit.		
9.	Evaluating alternatives. Do they meet objectives? Are they workable? Compare total costs.		

[11] Matthew B. Miles, The Development of Innovative Climate in Educational Organizations (Menlo Park, California; Stanford Research Center, Educational Policy Research Center, 1969), pp. 4-7.

contribute significantly? What portion or portions of the real need does the objective satisfy?
2. Describe the object in specific measurable terms that identify who, what, when, where, and how.
3. Define the minimum acceptable criteria for verfying that objectives have been achieved.
4. Describe the environment conditions under which the desired ends are to be attained.
5. Re-evaluate the objectives. Are they clear? Verificable? Quantifiable or qualifiable? Are they relevant to the real need?

As in the need statement, the definition of objectives calls for a truce with rhetoric. Objectives, if they are to be anything, must be clear, specific and measurable. Program specifications should answer these questions. What shall be done? Who will do it? When and where should it be done? And finally, how will the program be processed?

These specifications do not rule out duality, complexity or sophistication but they do void the common practice of setting goals that are too vague to evaluate, too elastic to leave traces of responsibility. The need is for thought. What does the organization desire and how does it know when it is there? Objectives are results that must be achieved. The methods for realizing these results come later in the process.

Identify Constraints

The researcher must identify the constraints--the real-world problems and issues that any proposed program must adjust to, as follows:

1. List the constraints in general categories, for example, finances, available personnel resources available plant, training, voluntary assistances.
2. List the specific constraints within each category and establish the source of those constraints
3. Label the constraints according to the nature of key characteristics. Example: economic; short-term/long-term; traditional; subject to change.
4. Rank constraints in the order of effect upon system, program, realization of need, etc.
5. Re-evaluate the constraints. Have facts been separated from assumptions, constraints from variables, intuition from bias, need from pressures? Have "pet solutions" or "built-in answers" introduced unwarranted or imagined constraints? If constraints are indeed inviolate, have they been validated as such?

It is common to link constraints with "real-world" problems. The truth and simplicity of this linkage, however, is deceptive. For there are many kinds of real worlds; and we must attend to this as we categorize the many kinds of constraints.

There are those constraints that cannot be handled at all. The elimination of delinquency would certainly simplify the problems that face juvenile corrections. But what of those constraints that are considered to be out of the organization's control but which actually are not? The most real of all the worlds is not always the one we grew up with or the one we heard repeated the most often. We must also consider the multiplicity of "real worlds" when we ponder the impact of constraints upon the analysis of our objectives. Although we are often moved to disavow it, a serious difference separates "impossible" from "uncomfortable."

Study Alternatives

The researcher must evaluate alternative systems or programs designed to achieve the objectives as follows:

1. Research and develop information on present and future conditions (current and expected state of the art). Collect information with the idea of uncovering potential means to specified ends.
2. Solicit ideas from many and varied sources.
3. List all the ideas that have been researched or suggested. At this point do not rule out the ideas that seem outlandish, impractical or unrealistic in view of the constraints.
4. Review the list of suggested ideas to determine that it is as complete and as wide-based as

possible. Solicitation only from "in-groups" or inhibiting individuals from proposing radical solutions are both common failings. Again, ideas that appear to be impractical or inappropriate should be listed.

There are many alternatives in problem solving. Careful and detailed analysis will bring some of them to the surface. But there are factors to be considered along with the mechanical steps of science; logic, for instance, and imagination. Creativity and courage are what we aim for, and we cannot come too close to the mark.

Select Best Alternative

The researcher must study all alternatives in terms of objectives as well as constraints and, after careful analysis, select the alternative or alternatives considered to be the best:

1. After consideration of all vital selection criteria, define the criteria which will be used to select the most promising program(s) or system(s).
2. Establish a method (quantitative when possible) for rating each alternative against the selection criteria.
3. Make the final selection of alterntives(s) for testing.
4. Re-evaluate the solutions. Were radical solutions penalized unfairly? Were conclusions rationalized that were already predetermined? Was there bias in the scoring system? Is there objective evidence to indicate that the alternatives selected would be effective?

Criteria which will be used in evaluating the alternatives must be determined. They are deprived mainly from the objectives and constraints; they frequently deal with performance, cost, time, risk of failure, and conformity to established policy. It is essential to establish an analytical method for selecting one or more alternatives based on the application of criteria to each of the alternatives.

This analytical method may involve relatively simple, logical thinking and written analysis or the use of more complex statistics, flow charts and decision trees, mathematical models, and computer simulations. The results of all of this is the selection of one or more alternatives for testing.

Implement Alternative

The researcher must implement the alternative to meet the specified objectives as follows:

1. Specify the activities, events, time frame, and the resources that will be needed.
2. Plan programs to evaluate the alternative (s). Pilot projects, controlled experiments, and other programs are all part of the methodology considered elsewhere in this appendix.
3. Re-evaluate #2. Was adequate data collected? Was the experiment extended for a sufficient length of time?
4. Implement the program.
5. Evaluate #4. Is the system being implemented in such a way that it has a chance to survive? Is the plan being altered without sufficient justification?

Evaluation and re-evaluation is the constant theme of the overall systems approach to program design and evaluation. Criteria, standards, and scientific methodology are the banner words. The approach is more complicated because it involved a complex intellectual process, far above the world of slogans and superficial assumptions.

No matter how good the selected alternative course of action may be, it cannot achieve the objective unless it is properly implemented. Circumstances may change rapidly during implementation, necessitating minor and/or major modifications. Wherever possible, there should be a pilot implementation of the selected alternative before it is accepted as a continuing program or method. Even where this is not practical, the first implementation of the selected alternative should be a

trial effort--planned, monitored, and evaluated. The purpose of pilot implementation is not to allow organized inertia to prove that the new approach is impossible, but rather to insure the best possible results from the new method or approach prior to full-scale implementation.

Evaluation

The researcher must undertake an evaluation of the overall system in order to determine whether the results of the program match the objectives that were initially specified, as follows:

1. Re-examine the original statement of objectives. Collect from these statements those items that are specifically measurable as well as those environmental conditions within which the objectives must be pursued.
2. Incorporating diagnostic features that provide for corrective action, develop as many reliable and valid tests as may be required to establish whether all of the objectives are being met.
3. Administer the tests to the system and interpret them both qualitatively and quantitatively.
4. At specified intervals, re-examine and evaluate the need and all elements of the system. Are the tests reliable? Is the testing being applied to what was originally specified?

After the pilot implementation of the selected alternative has proceeded to the point where results can be evaluated as planned, formal analysis of the results achieved is necessary to determine whether the selected alternative performs as well as expected. This state of evaluation of the pilot implementation of an alternative should not be considered as equivalent to normal program evaluation.

Feedback and Modification

Finally, the researcher must, based on experimental and real-world results, feed back the required modifications and continue the program until the objectives have been attained. This process includes the following steps:

1. Determine probable cause for the deficiencies in the program by checking for discrepancies between specified objectives and the results obtained.
2. Examine the entire program in order to pinpoint where the correction(s) can best be applied.
3. After developing a specific plan for correction, implement the correction during the next system (or program) cycle.
4. Begin a new evaluation and continue the cycle until specified performance is attained.

There is no reason to discontinue the evaluation as long as the program is in operation. No matter how dramatic and well-founded the results might be, the program needs to be checked for the flaws that time and the real world will inevitably produce.

Nothing is sacred. Evaluation may prove the undoing of many promising alternatives or even the very objectives themselves. This is as it should be, but fickleness or faintheartedness is another thing. Having arrived at the objectives conscientiously and laboriously, we should not quickly toss them aside or modify them without clear and cogent encouragement from the evaluation process.

A Cautionary Note

Use of the systems approach should include reliance on extra-rational qualitative factors that are not quantifiable, such as judgment, experience, and intuition.

Despite the desire to arrive at decisions on the basis of quantified values, the use of extra-rational or qualitative criteria suffuse the systems approach by both accident and design. Techniques which purport to quantify the respective merits of alternative approaches often are eagerly seized in the chimerical hope that decision making in complex environments can be simplified into value-free, low-risk "sure things." But the danger of trying to boil down decisions to patterns of numbers is that attention may fixate on peripheral derivative factors that can be quantified; while the central causal issues, unreceptive to measurement, may be pushed too far into the background.

CHAPTER 3 TREATMENT ALTERNATIVES

To a large extent the philosophy of delinquency causation determines the kind of treatment which will be used. Methods of intervention considered appropriate will vary with the perceived causes of delinquent behavior. There are a large number of theories about the causes of delinquency, some of which overlap with each other. However, they can be broadly classified by whether they indicate the cause of delinquency originates in society, within the individual, or in the child's family or school. Of course, there are probably different causes in different cases. Thus, a group treatment home relying on peer group pressure to bring about change would be most effective with clients whose delinquency appeared to fit a reference group theory and would be effectively administered and staffed by adults whose philosophy was consistent with that kind of treatment.

Table 1 shows some of the more widely used kinds of delinquency treatment methods along with the theories of delinquency on which they are based. A rather wide range of theoretical positions and treatment methods have been included. This is not meant to suggest that these are recommended either as good theories or good treatments. Certainly no one kind of treatment is appropriate for every juvenile delinquent.

Schooling, skill training, or enhancement of employment opportunities are treatment programs which assume that the delinquent behavior is the result of class status, in which case its purpose is to change that status; the result of frustrations from poor opportunities, in which case the treatment provides new opportunities; or that delinquency is directly the result of patterns of school failure. The treatment appears to be the same but the reasoning behind its application and the anticipated results are different. Similarly the use of parental substitute figures such as house parents in a group home may be predicated on the need for surrounding a delinquent child with different standards of behavior, the need for effective role models, or the need for someone to carefully monitor and control the child's behavior. In the case of each kind of treatment a variety of settings are possible. Care should be taken not to confuse the setting with the treatment and to make explicit in planning exactly what treatment method is to be used.

Table 1a. THEORIES OF DELINQUENCY AND RELATED METHODS OF TREATMENT - SOCIAL PRODUCT

Delinquency Cause	Treatment Philosophy	Treatment Methods
Social Product		
1. Lower-class culture	Introduce middle-class norms. Motivate toward higher aspirations, train for law-abiding lower-class.	1. Educational enhancement job training schooling alternatives
2. Delinquent subculture	Alternative means of satisfying needs. Opportunities to learn alternatives and places to apply them.	2. Success opportunities such as jobs, alternative schooling, vocational training, athletic accomplishment. Specific opportunities to learn other ways of reaching goals.
3. Thrill seeking gang culture	Provide alternative excitement.	3. Keep very busy, extremely active program. Channel individual into group that gets its thrills in socially acceptable ways. (Outward Bound for example).
(15)		
4. Subterranean values in surrounding culture	Surround with people who encourage other value systems. Help parents to learn to encourage suitable behavior.	4. Place in environment which discourages delinquent behavior (Foster Home or group home). Train parents to reinforce law-abiding behavior.
5. Youth Culture rebellion	Example of peers not in rebellion, association with adults who can relate positively to young, family-oriented treatment.	5. Group association with non-delinquents. Young adult associations (Big Brother, etc.) Family group sessions to reduce need for rebellion.
6. Reference Group or group process	Change direction of forces of group dynamics.	6. Group-treatment, positive peer culture, group management of such dynamics as cohesiveness and conformity.

Table 1b. THEORIES OF DELINQUENCY AND RELATED METHODS OF TREATMENT - INDIVIDUAL

Delinquency Cause Individual Product	Treatment Philosophy	Treatment Methods
1. Biological difference	1. Treatment to correct physical defect or to compensate for physical tendency.	1. Depending on diagnosis: drugs, hormonal therapy, compensatory teaching or training. Exercise opportunities for overly active delinquents.
2. Character disorder	2. Psychiatry or similar treatment to correct psychogenic problem in development.	2. Psychotherapy
3. Acquired self-concept	3. Overcome child's feelings of "badness" or worthlessness.	3. Teach parent to value child-parental effectiveness training. Success experience in learning situation--rewarded success. Planned or programmed training to assume responsibility and grow.
4. Acquired habits	4. Teach new responses, do not reward series of responses leading to delinquent acts.	4. Attack specific patterns of learned responses. Behavior modification methods such as token economy. Require restitution to victim.
5. Lack of control	5. Develop social ties, system of limits, outer control system. Encourage development of inner strength.	5. Prevention: spot children with poor controls early and give them greater system of control, more careful watching. Treatment: tightly structured control and surveillance with gradual development of inner control.

Table 1c. THEORIES OF DELINQUENCY AND RELATED METHODS OF TREATMENT - FAMILY OR SCHOOL

Delinquency Cause Family and/or School Product	Treatment Philosophy	Treatment Methods
1. Parental Role Model	1. Needs appropriate role model; strong father figure; pattern to copy.	1. Help parents to present better model (family counseling or training). Provide substitute parent figure (probation officer, counselor, teacher). Provides substitute parents. (Foster Care).
2. School preparation	2. Provide schooling which fulfills child's needs.	2. Educational opportunities other than standard academic, remedial schooling, vocational training, employment opportunities, arrange schooling so that no child becomes chronic failure.

PART II

PROGRAM EVALUATION

CHAPTER 4 CRITERIA AND VARIABLES

Section 1. Evaluation Criteria

Goals of a program may begin by sounding grand and important but so vague as to defy measurement. Eventually program goals must be made specific, exactly defined, and from them the objective, measurable criteria of success must be developed. It is all very well to talk about "rehabilitating delinquents" or the "prevention of delinquency" but before effectiveness can be determined it is necessary to make explicit exactly what is expected to change and how much it is expected to change. An appropriate criterion is some aspect of behavior which can be observed, quantified, objectively measured, and compared to other groups or to the same group at an earlier time.

The goals of the program in their largest sense relate to community expectations of crime reduction. Criteria chosen at this point may be general population changes or changes in the individual delinquent himself. In the eyes of the public the most desirable change a program evaluation could show would be a general change external to the program itself; that is, a reduction in the rate of delinquency. Strictly speaking, the rate of delinquency cannot be measured. What can be measured is some level of recorded delinquency. Appropriate criteria might be a significant reduction in the number of commitments, the number of adjudications, the number of petitions filed, the number of offenses known to the police, or the number of citizen complaints made. There are certain pitfalls in the use of each of these criteria, however. Evaluation designs based on external criteria involving a change in the rate of delinquent behavior in the population as a whole may be susceptible to the effects of a great many outside influences in the community. Changes in public attitudes, law enforcement intensity, or court severity may change the recorded rates of complaints, arrests, and adjudications while having nothing to do with the program being evaluated. Reduction of recorded delinquency rates is the most attractive criterion from the point of view of the general public but such records are rather far removed from the actual program being evaluated and are unlikely to reflect changes brought about by the program unless those changes are very large.

It may be necessary to avoid the problems involved in public records which reflect aspects of the community beyond the experimental control of the evaluation staff. A common alternative to using public records is to develop records to reflect change in the children involved in the program. The criterion in such a case is a specified amount of change in some measure of individual behavior over a given time or following a given course of treatment. The most obvious choice of a criterion behavior is the number of delinquent acts committed. Public records of an individual's delinquent behavior are one source of data; however, such records may be inaccurate or reflect the intensity of law enforcement more than delinquent behavior. One way of determining the extent to which a child is involved in delinquent behavior is to ask him directly. Carefully designed, such self-report measures can be useful, but the necessity for following up people after they leave a program can be difficult and time consuming. A further difficulty in the use of the number of delinquent acts as a criterion measure is the necessity for dealing not only with numbers of acts committed but with their relative severity. Complex methods for weighing criminal behavior are available but they require detailed information beyond what is usually recorded by police or courts.

Recidivism is a commonly used kind of criterion. Recidivism is usually interpreted to mean the recommitment of an individual, either to the same program or to some other program. It is measured by institution or court records. As a criterion, recidivism is somewhat inadequate unless efforts are made to follow-up individuals including those who leave the state, become involved in other programs, or otherwise get into further trouble but are not counted in readily available statistics. Recidivism reports also refer to some particular length of time and it is important that the span selected be realistic. Commonly used time spans for recording recidivism are twelve, eighteen, twenty-four, or thirty-six months. Longer time spans are desirable for assessing the continuing impact of a program; however, with increasing time, follow-up becomes more difficult.

TABLE 2. POSSIBLE CRITERIA FOR EVALUATION OF PROGRAM EFFECTIVENESS

CRITERIA	MEASURED BY	PROBLEMS
Primary goals; community wide change:		
1. Number of commitments reduced by X amount	1. Comparison of records over time: period must be long enough to overcome effect of seasonal fluctuation	1. Significance of apparent change other factors which may create change
2. Number of adjudications reduced by X amount	2. Court records	2. Same as #1
3. Reduction of adjudications for a specific kind of offense by X amount	3. Court records	3. Same as # 1 Reduction in one offense may be offset by increase in another
4. Number of petitions filed or number of complaints police handle reduced by X amount	4. Court records or police dept. records compared pre-program and post-program	4. Adequate records susceptible to other factors such as change in intensity of law enforcement
5. Number of known offenses of a particular kind reduced by X amount	5. Police records	5. Records reflect actual numbers inadequately Reduction in one area produces increase in another More vigorous enforcement leads to apparent increase
Primary goals; change in target persons behavior:		
1. Reduction of X amount in number of delinquent acts committed.	1. Court records Police records Self-report	1. Accuracy of records Follow-up of individual; loss of information through residence change, etc. Handling of severity of offense dimension
2. Recidivism	2. Recommitment to same or other programs Institution records or court records	2. Appropriate time period Inclusion of other programs (such as adult facilities) Different kind of offense
3. Reduction of X amount of maladaptive behaviors in specified time period	3. School attendance record Job record Parental report Self-report scales Comparison of treatment and control groups or of pretest and post test	3. Behaviors to be measured must be very carefully defined Reporting techniques using schools, parents, etc. introduce biases and inaccuracy Behaviors chosen may not be appropriate; i.e., does school attendance really relate to delinquency? Experimental design difficult
Primary goals; change in target persons attitudes:		
1. Individual has improved self-esteem by X amount over randomly selected control	1. Questionnaires Standardized attitude assessment Opinion of qualified observers	1. Measurement inaccurate Standarized tests may not be applicable Attitudes may not be related to behavior Experimental design difficult
2. Improved attitude toward authority by X amount over randomly selected control	2. Questionnaries Standardized assessment measures	2. May not be related to changed behavior Compliance with force rather than real change
3. Improved knowledge of how to get along in community compared to randomly selected control or to pretest	3. Evidence on questionnaires or interviews of objective knowledge about appropriate ways of handling needs without getting in trouble	3. Knowledge may not imply behavior Measurement devices largely unavailable or difficult to produce Experimental design difficult
4. Basic needs being fulfilled in ways approved by society compared to randomly selected control	4. Self-report questionnaires covering subjects satisfaction with the ways his needs are being met	4. Accuracy of measurment Follow-up Experimental design difficult May not be related to real behavior
Secondary goals; change within program:		
1. Success of school program Group achievement compared to control group	1. Standard achievement tests Grade records Grade placement	1. May be unrelated to main goals of program
2. Individual success or advancement compared to control group or to pretest	2. Same as #1	2. Same as #1

CRITERIA	MEASURED BY	PROBLEMS
3. Skill mastery--job training	3. Performance of skills	3. Same as #1 Job skills may not be transferable to a real job
4. Skill mastery--social skills	4. Demonstration of skills in relating to others--requires judgement of trained observer	4. May not carry over with other relationships Same as #1
5. Specific maladaptive behavior reduced--compared to baseline data	5. Record of number of times specific behavior occurs in baseline and in post test data	5. Change may not persist Change may be unrelated to actual delinquent behavior Requires careful training of staff
6. "under the skin" personality changes compared to control group	6. Professionally administered or evaluated psychological tests	6. Personality change may not be related to delinquent behavior Implies flaw in personality to begin

In using recidivism as a criterion, it is again useful to weigh different severity levels. Probation or parole violation, for example, may result in recommitment to a program without necessarily representing a return to criminal behavior and should probably not be handled in the data as recidivism.

If delinquency is viewed as an end result of other maladaptive behavior, one way of measuring success is to use a change in the rate of such behaviors as a criterion for success. Thus, a given change in school attendance or employment or changed leisure time activities can be used as the specific criterion. A basic shortcoming of the use of such criteria is that the relationship between the specific behavior and delinquency may not be as significant as has been assumed. Ideally, such a measure first requires thorough research to demonstrate its validity. Several such measures are available which have been so validated. Measures of behavior outside of institutional settings are also difficult because of inaccurate reporting or inadequate records.

The attitudes of those who have taken part in a program are even more difficult to relate directly to delinquent behavior. There are, however, a number of theories of delinquency which can be used for a basis to establish the relationship between such attitudes as self-esteem or acceptance of authority and delinquent behavior. Growing out of these theories, there are a number of rather widely used tests and questionnaires. These measurement devices are often of questionable validity as predictors of behavior.

Occasionally programs for delinquent children focus directly on improving a child's knowledge of how to get along in society. Such a goal assumes that a child who has at his command acceptable means of staisfying his needs will not resort to delinquent behavior. The effectiveness of such programs can be measured either by obtaining objective evidence that the child has learned other ways of satisfying his needs or by evidence that his basic needs are being met in other ways. Here again, knowledge may be poorly related to actual behavior. The range of knowledge and needs being measured may also be too narrow to be meaningful.

The goals of a program in delinquency treatment or prevention relate to changes in the behavior of those who have been through the program. However, on a more limited scale, it is often helpful to evaluate the efficacy of parts of a program, not as treatment of delinquency but as distinct elements of a program with their own goals. Thus, school, job training, or social skills may be measured by appropriate performance or achievement tests. In these instances, the criteria are related to a specific change in the level of knowledge or ability, and it is not necessary to make theoretical assumptions about the relationship of such criteria to delinquency. In the same way, specific behaviors which are maladaptive or disruptive within the program may become the targets for change.

When part of a program consists of psychiatrically oriented treatment, appropriate criteria are specific personality changes measured by professionally administered or evaluated psychological tests. Evidence of real personality change is very hard to produce although there are a number of indirect measures of such change used.

Selection of appropriate criteria and measuring devices are only the first steps in planning an evaluation. Selection of experimental design, selection of subject samples, and control of other variables are equally important to the overall research effort.

Section 2. Operational Definitions of Variables

To test the experimental hypotheses, the independent and dependent variables should be operationally defined at the practical level of observation. The activities or "operations" that enable empirical observation - that actually spell out the procedures used to measure the variables - should be identified during the program planning and design state.

Operational definitions are essential in research evaluation. These are definitions which assign measurable meaning to variables by specifying the activities or "operations" which, when performed, bridge between the abstract and concrete, and supply the necessary empirical conditions that are suitable for numerical categorization.

To say that variables are measured is a misnomer. A more accurate description would be to say that the properties of variables are measured. And even this qualification misses the mark.

What is measured are the indicators of the properties of the variables. "Indicator" is a convenient word for the activities identified by the operational definition as signaling the presence of the variable. For example, frequent combative behavior may indicate hostility; excessive perspiring may indicate anxiety.

This descending level of specificity - variable to property to indicator - may better understood by citing the experience of the Los Angeles County Sheriff's Department with its recently completed 3-year field experiment to scientifically measure and evaluate the relationship between stress training of law enforcement officers and field performance. Figure 11 presents the primary data collection instrument used in this experiment and the way the broad concept, "field performance," was operationally defined.

The concept was analyzed into eleven core properties such as leadership, communication, responsibility, initiative. These properties were then operationally defined into indicators that could be empirically observed and measured. For example, responsibility was defined as, "answerable for his actions; application to duties; dependability in completing assignments."

All definitions and rules are limited and cannot capture the richness and variety of reality. They are, nevertheless, indispensable to the research design, to communication, and to replication of the research.

CHAPTER 5 SCIENTIFIC METHOD AND RESEARCH

Section 1: Introduction

The fundamental system of thinking and action that permeates processes of effective program evaluation is the philosophy of inquiry, or the application of the scientific method to guide observations, measurements, and evaluations of data. This philosophy posits the procedures and techniques of scientific research for gaining knowledge. It also proposes a certain attitude - the empirical attitude that searches for and relies on objective factual observations and evidence.

In addition to its empirical base, the scientific method is systematic. Conducted according to a comprehensive plan (the research design), it not only specifies what to observe, but looks for relationships, patterns, and order between observations. It also supplies the power of self-correction via built-in controls that help verify the reliability and validity of the data attained. "Control" means the ability to isolate and assess the fluctuation of variables which are relevant to what is being observed.

Research investigations are open, explicit, and reproducible while the assumptions, values, calculations, limitations, and conclusions are documented and susceptible to testing, criticism, and refutation.

The Scientific Method as a Philosophy for Program Evaluation

Every administrator within the criminal justice system should possess a working knowledge of the philosophy of the scientific method and how this method can be harnessed to enhance the process of program evaluation. Essentially, the scientific method requires the following:

1. Reliance on facts. "Facts" refers to events which may be directly observed and replicated. ("Evidence" could be substituted for "facts".)
2. Use of systems analysis in comprehending complex phenomena. Analysis involves division of a system or program into the specific procedures and operations for purposes of assessment, design, and redesign.
3. Development of hypotheses to guide research. Hypotheses are careful explicit predictions of outcomes that can be tested against observations.
4. Depersonalization or freedom from bias and the subjectivity that characterizes common sense convictions.
5. Objective measurement. Knowledge expands in large part through the development and refinement of instruments of measurement.
6. Quantitative methods to treat data. The main concepts are operationally defined; that is, the activities performed to manipulate and measure a concept are specified in quantifiable terms. Additionally, the objective language of statistics is channeled to analyze, classify, and summarize data.

Research Procedures and Principles

Use of the scientific method as a philosophy and tool for program evaluation should be guided by the following sequence of research procedures. These procedures are constructed to insure collection and self-correction of data and to unify the data into objective conclusions. Although defined separately for clarity, some procedures overlap and often proceed simultaneously. These procedures are:

1. Statement of the problem including some history about why resolution is important. Defining the problem in a way conducive to experimentation is as essential as searching for the solution.
2. Review of the literature and previous experience to gather information about alternative strategies and solutions--what has been accomplished?
3. Development of hypotheses to guide the research and test results of the program.
4. Selection of the setting, including time schedule, staff, and budget.

5. Determination of the research design or overall plan to collect and analyze the data. Careful planning is mandatory at this point so that observations and measurements actually address objectives and produce suitable feedback.[6]
6. Selection of the population to be evaluated and the sample to participate in the experiment.
7. Identification of data to be collected plus design or selection of data collection instruments: questionnaires, examinations, interviews, observations.
8. Implementation of the experiment and collection of data according to the present design and time schedule.
9. Statistical analysis and interpretation of results.
10. Written report.

Constructed to provide objective data and conclusions about results of ongoing and proposed programs, evaluations may be said to equal research. And when guided by the requirements attendant to the scientific method, evaluation produces feedback for more knowledgeable allocation of resources.

Use of Experimentation in Evaluation of Proposed Programs

Before installation of a proposed program, a research investigation should be conducted to determine the relationship of the proposed program's output to overall agency objectives and needs. The system to evaluate the program should be designed and tested concomitantly with the design and development of the program.

Results of the investigation should identify significant program components such as inputs, processes, constraints and objectives - what the program expects to produce in terms of immediate and long-range goals.

The research should evaluate the measures of effectiveness - quantitative and qualitative - that can be monitored to appraise the direction and progress of the program toward the stated objectives.

Decisions about the value of a program depend on the methodology and measurement of accomplishments to distinguish program effects from effects of other forces interlacing the environment - to isolate what happened as a result of the program from what would have happened anyway.

Principles of experimental research provide the basis for estimating the amount and direction of program effects. Experimentation empowers empirical tests of hypotheses in a manner that strives to exclude or correct extraneous influences, thus clearing the way for reasonable inferences concerning factors of significance. In basic outline, experimentation does the following:

1. Exposes an experimental group to the experimental treatment program which is the independent variable symbolized as X. An example of X might be stress training for law enforcement officers.
2. Does not expose the control group to the independent variable.
3. Compares both groups on the dependent variable, symbolized as Y. In keeping with the above example, Y might be field performance. In this example, the purpose of the experiment would be to evaluate the effect of stress training on field performance.

Through the use of quantitative and qualitative criteria, measurements of effectiveness, and methods of statistical analysis, research experimentation compares outcomes for the control and experimental groups that were randomly selected from a common population. Random selection means that each person in the population experienced an equal opportunity to be selected for either group. The experimental program (independent variable) is applied to the experimental group and withheld from the control group. Because subjects were assigned at random (or matched on relevant traits, then randomly assigned), the groups may be considered comparable and observed differences credited to the experimental program.

A primary concern of experimentation is to seek out possible causal relationships between significant variables. These relations are made to surface by subjecting the data to statistical tests of significance. From the findings, the administrator may infer with known degrees of confidence how X affects Y. For example, he will be able to estimate wheter X and Y vary together in some correlated fashion; or wheter X causes or leads to Y; and, if so, with what frequency.

[6] Feedback refers to the total information process through which primary and secondary effects of organization actions are fed back to the organization and compared with desired performance.

This description of experimentation, popularly considered as the scientific method, contains these core components: dependent variable, independent variable, careful ignoring of irrelevant variables, and careful control of other relevant variables. When properly conducted, it leads to the "if-then" statement (See figure 3); that is, if this is the case, then that will happen. If frustration, then aggression. The then part of the statement houses the dependent variable whose effects are dependent upon how the investigator manipulates the if (independent) variable.

Experimentation and Planned Variations in the Evaluation of Ongoing Programs

To evaluate the effectiveness of ongoing programs, a representative sample of the program should be periodically subjected to a systematic investigation.

The purpose of these investigations should be to measure the degree of convergence between expected and actual outcomes, plus the relationship between program outcomes and agency objectives. Additionally, the investigation should evaluate effects of program criteria, characteristics, strategies, and structure for their contribution to objectives. Using representative samples, planned variations should be introduced into the program structure and routine procedures. Results should be evaluated for their power and utility to promote program and agency objectives.

Interweaving all stages of periodic evaluations and planned variations should run the search for more productive alternatives and more meaningful and refined measures of effectiveness.

Section 2. The Research Design

Experimentation presents various types of research methods and structures to introduce planned variations into ongoing programs, or to provide data to evaluate proposed programs against agency goals. By experimenting with planned variations, the administrator may illuminate and improve causal interactions between input, processes, and output of proposed or ongoing programs before full installation or change over. The extent and thoroughness of the effort depend on the selected research design.

Research design is the blueprint or plan for conducting the experiment and for collecting, measuring, and analyzing the data. It documents the procedures that will be followed to furnish results as economically, validly, objectively, and accurately as possible. Toward this end, it shows what problems will be addressed, what questions will be answered (with what degree of confidence), and what controls will be used to minimize extraneous influences.

Forcing decisions of crucial choices, the design poses such questions as: Through what media will data be collected - questionnaires, interviews, survey of records, observations? Would an intensive study of a small sample be more useful than that of a larger group? How can variables be quantified? How controlled should conditions be? By working through the design in the beginning, by moving from problem definition through the research design, implications of goals are highlighted, achieving a more focused approach to specific procedures.

Selecting the Appropriate Research Design

Whenever possible, evaluation of proposed and ongoing programs should employ research designs supported by scientific criteria. Design possibilities should be carefully analyzed and fitted to the needs and resources of the organization. Major considerations dictating type of design selected should include:

1. Type of evaluation desired;
2. Existence of applicable methodology;
3. Time and money available;
4. Qualified personnel;
5. Availability and accessibility of suitable comparison groups;
6. Ability to collect required data;
7. Perceptions of the administrator of how best to attain and measure the data.

Delineations of criteria for research designs abound in treatises on social science investi-

gations. Among the more feasible summaries is that of Harry Shulman (1962)[9] who suggests five general criteria:

1. Explicit statements of the underlying assumptions of processes to be studied including anticipated changes.
2. Formation of hypotheses.
3. Control over conditions of observation.
4. Analysis of the significance of findings in terms of statistical measures of reliability and validity.
5. Clear conclusions about relations between results, hypotheses, and limitations affecting these relations.

The strength or power of a design depends on the degree to which it eliminates threats to valid inferences from the data.

The "strongest" are the true experiments that include the essential basis for valid inference: random selection of subjects. This automatically rules out extraneous "explanations" of the effects of a given treatment.

Weak designs include nonexperiments that are vulnerable to problems attending absence of control groups and randomization. Quasi- or compromise designs are stronger; but without random selection, equivalence between experimental and control groups cannot be assured.

Conducting an experiment of scientific standing demands more than staff and economic resources, important as they are. An equal consideration addresses the possibility of manipulating and measuring the central variables that network the program being evaluated. Some variables are susceptible to measurement, others are not. Variables like reinforcement, training and educational methods, disciplinary practices, and working environment usually pose few problems. Measuring others like leadership, motivation, and values spotlights a continuing research dilemma that is often dodged by packaging the whole affective domain in one catchall category entitled "fog-factors" or "imponderables."

In sum, the multiplicity and complexity of evaluation suggests the need for a "situational" approach - one that matches the research design to the problem, resources, purpose, and constraints impinging on the instant evaluation process.

True Experimental Design

Under ideal, optimum, or favorable conditions the true experimental design should be applied. Conditions should include these controls: a) pretest/post test measurement of dependent variable; b) experimental group-control group; c) randomized assignment of subjects; and d) random assignment of treatment to the experimental and control groups. The true experiment as pictured in Figure 4 achieves the most strenuous approach for evaluating program accomplishments. Conducted under carefully controlled conditions, this experiment manifests the ideal of science because it conveys the greatest confidence about observed relations.

The test of whether the difference (D) between groups is attributable to the program treatment rests on whether D_1 is significantly larger than D_2 when examined by tests of statistical analysis with the level of significance decided during the planning stage. A variation on the true experiment matches subjects on important personal traits and then randomly assigns them to the two groups, with the treatment also randomly decided between groups.

Compromise or Quasi-Research Designs

In situations where the true experiment is deemed possible with the exception of the randomization of subjects, the compromise or quasi-design should be used.

[9] Harry M. Shulman, Juvenile Delinquency in American Society (New York: Harper, 1961), p. 760. Quoted in E. K. Nelson "Perennial Problems in Criminological Research." Crime and Delinquency XVII, No. 1 (1971) p. 30.

FIGURE 3. SINGLE VARIABLE EXPERIMENTAL DESIGN

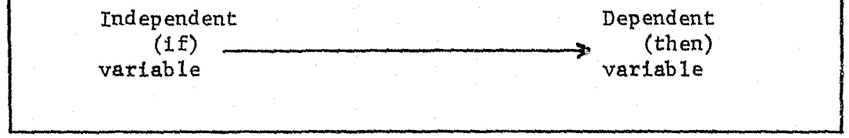


FIGURE 4. TRUE EXPERIMENT-EXPERIMENTAL GROUP/CONTROL GROUP WITH RANDOMIZED ASSIGNMENTS

	Before Pretest	Program Treatment	After Post test	
Randomly assigned subjects (Ss) — Exper group	YES S_1	YES	YES S_{1_A}	$D_1 = S_{1_A} - S_1$
Control group	YES S_2	NO	YES S_{2_A}	$D_2 = S_{2_A} - S_2$

FIGURE 5. COMPROMISE OR QUASI-EXPERIMENTAL DESIGN

	Pretest	Program Treatment	Post test	
Exper. group	YES S_1	YES	YES S_1	$D_1 = S_{1_A} - S_1$
Control group	YES S_2	NO	YES S_{2_A}	$D_2 = S_{2_A} - S_2$

The weakness of nonequivalence between groups resulting from the absence of random selection and assignment should be recognized. (The whole structure of probabilistic-statistical reasoning depends on this fundamental control).

Investigators should also recognize that difficulties attendant to randomization require frequent resort to the compromise design. To partially offset the limitations of no randomization, investigators should randomize assignment of treatments and plug in other checks to upgrade equivalence. Examples: a) select samples from the same population or use samples as similar as possible; b) confirm similar distribution in personal traits by matching of such things as age, race, education, marital status; and c) compare means and standard deviations on pretest measurements.

A commonly used design is the experimental group - control group pattern with no randomization; thus, there is no assurance of equivalence between groups (See figure 5).

Despite precautions to increase equivalence by various means, the compromise design manifests problems, all of which fall under the primary difficulty called selection. Indeed, the reason for random selection and assignment is to avoid the problem of "self-selection", that is, where subjects are selected into groups for reasons extraneous to the research purpose. To illustrate: assume that volunteers are selected for a pilot program to reduce recidivism, and other inmates are used as the control group. If the volunteers differ in traits allied to the dependent variable, the difference in results between the two groups may originate with this trait rather than the program treatment. Conversely, if all subjects are randomly selected and assigned, the selection dilemma disappears.

Nonexperimental Designs

Under most circumstances, nonexperimental designs, like the one-shot case study or the one-group, pretest/post test, should not be used to evaluate programs.

With the absence of scientific standards like randomization, experimental and control groups, and pretesting of responses, nonexperiments reflect no credibility and their results should be viewed with restraint and corroborated by additional sources of information.

Organizational conditions and shortcomings that require use of nonexperiments should be corrected and more scientific experiments employed for program evaluation (e.g. the true experiment or the quasi-experiment).

Nonexperiments, also called preexperimental designs, display two typical types: one-shot case study; and the one-group, pretest/post-test design.

In the one-shot case study (See Figure 6) there is no baseline - no premeasurement of the dependent variable. Data in the pretest cell is supplied ex post facto by conjecture and imagination. Without pretesting, scientific assessment of effects of the program/treatment is impossible.

Absence of a control group for comparison purposes negates scientific standing in the one-group, pretest/post-test (See Figure 7). To illustrate this design's weakness, consider the case of "evaluating" effects of a new academic program (independent variable) on standardized test scores (dependent variable). With Figure 6 there is no control or comparison group by which to affirm that test improvement appeared because of the new program and would not have happened anyway.

Time-Series Designs

Time-series designs should be used when major changes are expected to occur over time; for example, policy interventions, development of individuals or treatment results.

Longitudinal studies afforded by these designs should be considered as particularly suitable for sampling and evaluating effects of planned variations in ongoing programs.

Time-series experiments tend toward instability and changes may be erroneously credited to the independent variable. Statistical analysis of data also presents special problems and the usual tests of significance may yield spurious results. Thus, when dealing with data collected

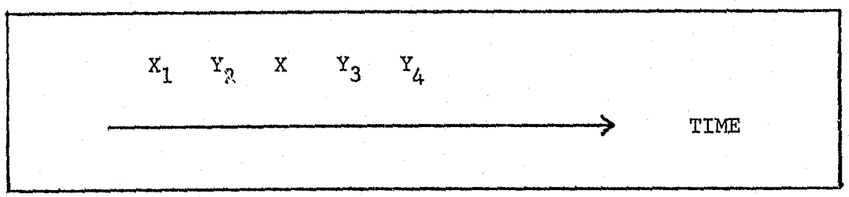
FIGURE 6. ONE-SHOT CASE STUDY ("AFTER ONLY" MEASUREMENT)

Pretest Measure	Treatment Program	Post test Measure	
S	YES	S_1	$D = S_1 - S$

FIGURE 7. ONE-GROUP, PRETEST/POST TEST DESIGN

Pretest Measure	Treatment Program	Post test Measure	
S	YES	S_1	$S = S_1 - S$

FIGURE 8. LONGITUDINAL TIME-SERIES DESIGN



over time, investigators should identify and monitor the uncontrollable variables that influence the data; for example, maturation, social forces, economic shifts. Results from analysis of time-series data should be carefully reviewed before a "significant" interpretation is attached; especially when the results agree with the hypotheses.

The time-series design is exemplified by Figure 8 where the symbol Y signifies the dependent variable being measured and X the program treatment or independent variable.

Time-series designs alleviate the problem of separating reactive effects induced by pretest measurements from those induced by the program treatment.

Reactive effect speaks to increased subject sensitization to the problem because of the pretest. Just measuring subjects can change their responses, posing the need to sift out of post test data those effects produced by the treatment from those produced by the pretest. This problem can become acute when measuring responses to controversial issues. Suitable designs are available using different but similarly constructed samples for each test with the treatment applied to just the post test group.

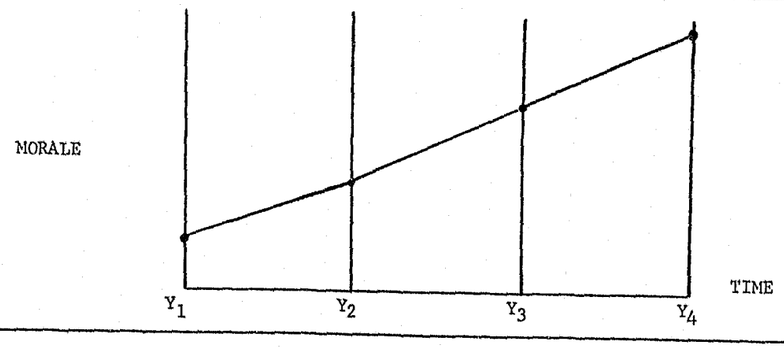
If a reactive effect prevailed in the Figure 7 design, it should surface at Y_2. This measure can be compared with measurements taken at Y_3. If an increase appears at Y_3 over the increase noted at Y_2, it can reasonably be credited to X, the treatment.

Take a hypothetical example using a probation department. Suppose an administrator starts a new policy to improve morale. Before activating the policy, he measures morale with anonymous questionnaires on two occasions three months apart, Y_1 and Y_2. He then starts the new policy. Morale is measured on two subsequent occasions, Y_3 and Y_4. Assume that results show the trend graphed in Figure 9.

According to the graph, X affected morale over and above the effect of time. Morale was increasing before X, but it rose sharply afterwards. The only problem: Did something else happen between Y_2 and Y_3 to cause the rise in morale? Was morale improving to some degree by the mere act of measuring it? This is the perennial question of time-series designs when control groups are not used. Data can be confounded by many other influences.

Variations on the time-series design include adding more observations, using one or more control groups, or adding more X's, more experimental interventions.

FIGURE 9. PRETEST/POST TEST TIME-SERIES ANALYSIS OF DATA COLLECTED AT 3-MONTH INTERVALS



Section 3. Measurement

Levels of Measurement

Data can be used in many forms, but in experimentation it is reduced to numbers so that the mass of observations can be analyzed and evaluated. The rules selected to assign numbers to observations are the criteria that define the level of measurement, the numerical scale employed in the experiment, and the statistical operations used to analyze the data.

The four levels of measurement in their ascending order of power are the nominal, ordinal, interval, and ratio scales. The higher the level, the more information there is about the phenomenon. For this reason, investigators should strive to use the highest possible level of measurement in a given situation.

Administrators should possess a working knowledge of the properties and uses of the four levels of measurement. Briefly, they are:

a. The nominal or classificatory scale refers to the simplest level where numbers are used to distinguish persons or traits. Example: male-female; married-single. The essence of this scale is classification. There is no numerical value attached to the number and they do not represent a value or amount of anything. Nominal measurements rest on two rules: All members of a set are assigned the same number; and each set has a different number.

b. The ordinal scale should be used when the observations can be related and rank-ordered into possessing "greater than" or "less than" amounts of the attibute under study. This scale distinguishes one object from another and also tells whether the object contains more or less of the trait than other objects in the set. The scale provides no information on the amount of difference between objects. It also possesses no absolute quantities, nor equal intervals between the numbers. For example, because the numbers are equally spaced on the scale does not mean that the underlying properties they represent are equally spaced also. Radical differences in correspondence may go undetected. The ordinal scale possesses no true zero point, so there is no way to detect when the object contains none of the property.

c. Interval scale refers to a level of measurement that should be used when the distance between any two numbers is equal and of known size. This scale possesses all characteristics of the nominal and ordinal scales plus numerically equal distances. This means that equal distances on the scale depict equal distances in what is being measured.

d. The ratio scale is the most sophisticated level and one rarely attained by social research. It possesses all properties of the other scales plus absolute zero at its point of origin. This addition enables application of all arithmetic operations to data.

This scale is used for physical measurements such as length, time, and weight. With its absolute zero, statements of ratios are meaningful; for example, a 6-inch time line is twice as long as a 3-inch line.

Some form of a ratio scale should be used when the investigator needs to state relationships between variables as products; for example, an individual's preference for a given event equals the product of its utility to him and his expectation of it happening.

In the process of program evaluation, much time is devoted to observing. Observations supply the raw material by which to test the hypotheses. The first step in reducing the mass of data into a format understandable by the human mind is to redefine the observations into a numerical form that can be handled statistically. This is done by the measurement method; and the degree of information achieved is dependent on the level of measurement utilized.

Properties of the four levels are summarized by Figure 10. Each ascending level contains properties of the lower levels, and the higher the level, the more information is available.

FIGURE 10. PROPERTIES OF THE FOUR LEVELS OF MEASUREMENT.

Scale	Properties
Nominal	Classification
Ordinal	Classification + order
Interval	Classification + order + equal units
Ratio	Classification + order + equal units + absolute zero

Steps in the Measurement Process

Any measurement process should include at least these steps:

1. Determine the objective, the purpose of the program. Without clear objectives, it is impossible to set standards to evaluate performance.
2. Decide relevant factors. These are easy to define when dealing with physical systems; not so with social systems where evaluation routinely proceeds with limited information.
3. Select key indicators of factors; indicators which are quantificable or in some way translatable to numerical ratings.
4. Select or construct a) the measuring method and b) measuring unit. For example, in measuring police field performance, the measuring method might include completion of a five-point ordinal-scaled questionnaire by supervisors of certain officers selected for evaluation for a preset time.

The measuring unit is the quantity or amount of the concerned concept contained in the observation. This unit is usually fixed arbitrarily and standardized for all observations. In the example of field performance after establishing a standard unit the investigator would estimate how many units of, say, "initiative" were "expressed" by each subject.

5. Apply the measuring unit to the concept to be measured according to the preset rules of correspondence. This step starts the main action of measurement in that it translates the observation to a number (the number of units).
6. Examine the data with appropriate methods of statistical analysis.
7. Evaluate effectiveness of the measurement process by assessing its contributions to the program's objectives.

Measurement prescribes certain processes involving an observer, an observation, and some form of measuring instrument, the combination of which produces a number (the measure) that stands for the observation. The overall process follows certain requirements decided initially and set forth in the research design.

Validity and Reliability

Data collection instruments used to define and measure the experimental variables should be selected or constructed to include all relevant factors pertaining to the variables.

The instruments should also be checked for the reliability, consistency, and dependability of their findings.

A central demand of the measurement process is the need to confirm the validity and reliability of the data. If these core criteria are unknown, little faith can attend results or conclusions.

Validity asks the question: Are we measuring what we think we are? Have we included the important properties in our measurements? Do our instruments satisfactorily get at the concept; that is, do they actually measure what they purport to measure. In the example of field performance, inclusion of such properties as mathematical or mechanical aptitude might be invalid in its operational definition.

Reliability deals with consistency and dependability of the measurement instrument. It asks the question: If the same phenomena were measured again and again with the same instrument, would the same or similar results be obtained?

Many factors influence measures: scoring errors, fatigue or emotional set, time, unclear questions, etc. The difference affected by these factors with successive administrations determines the reliability of the instrument.

Sampling Procedures

1. A sample is a smaller representation of a large whole. Sampling selects a portion of a certain population and treats it as representative of the population. Scientific sampling procedures focus on three primary criteria:

a. In defining the population from which to select the sample the evaluator should be guided by the purpose of the research and the relationships set forth in the experimental hypotheses.
b. The sample should be as large as possible. A general rule of thumb: the larger the sample, the better the representation, the more valid the results.
c. The sample should be representative of the large population. More specifically, it should represent, within known limits, the major characteristics of the experimental population. Whenever possible, the sample should be randomly selected. Only then will bias be controlled and representativeness assured.

2. The basis for deciding the sampling procedure should be to minimize the monetary costs and the cost of erroneous analysis of data that may accompany its use (bias and nonlikeness).

Program evaluators should possess a working knowledge of the three common sampling procedures: random, stratified, and purposive, defined as follows:

a. A random sample is one selected in such a way that every member enjoys an equal chance of being included and each selection is independent of any other selection. This procedure usually assigns a number to every member of the population from a table of random numbers. ("Member means any unit of research such as families, organizations, newspaper pages, time periods, individuals, and objects). The sample is then selected by taking numbers in succession from the table until attaining a preset size. A practical variation of partial randomness especially suited to card files and case studies is the systematic sample. This sample takes every nth item in the file, beginning with a randomly selected first card.
b. Stratified sampling is used when some segments of the population are known to be more homogeneous than others. This method identifies substrata of the sample containing the required characteristics and then randomly selects a certain percent from each strata. Examples of strata: age, religion, residence, ethnic background, and education.
c. Purposive sampling. When practicality precludes random sampling, a hand-picked sample is often substituted. Results apply to just the sample and should not be generalized to any population. Sampling errors and bias cannot be computed so this method should be used only when other methods are impossible.

Section 4. Analysis of Data

Statistics. Before data is collected, it should be decided on the basis of the design used, which statistical analyses are to be used on the data. Each type of statistic is limited by certain assumptions regarding the data which is analyzed. For example, analyses by t-tests or analyses of variance require random selection of the samples and interval or ratio data as opposed to

rankings (ordinal) or categorical (nominal) data. If the sampling is truly random, there shouldn't be any problems with another assumption that the data fall into a normal distribution. Data which involved classifying or ranking research subjects is limited in the types of statistics that can be used to analyze the data. Consequently, it is worthwhile to plan the entire research design in advance for several important reasons:

1) The assumptions of statistical inference are valid only for research for which the hypotheses is stated in advance.
2) It can be determined in advance whether the types of data available are able to provide the information desired.
3) Mistakes in design or analysis, etc, are less likely to occur.

Hypotheses are almost always stated in terms of a difference between 2 or more groups (one never predicts that two groups will be the same, since such a statement cannot be supported by statistical inference). If a t-test produces a result which is significant at the .05 level (written $p < .05$), that result means that only five times out of 100 could a difference between groups as large as that occur by chance. There is no guarantee that the difference between means is not in fact due to chance, although the probability is low. The .05 level of probability has been arbitrarily selected by most scientific researchers as the minimum cut-off point for "significance" of findings. Anything short of that, ie. $p < .10$, is not referred to as significant, although there is no magical transformation which occurs at the .05 probability level. It is merely a convention. Additional confidence in results can be gained by findings by oneself or other researchers which support the results obtained, eg. Several researchers evaluating probation programs found that effectiveness decreased as probation officer caseloads decreased.

Table 3 shows what types of statistics are most appropriate for standard experimental and non-experimental designs.

TABLE 3

ANALYSIS AND INTERPRETATION FOR SIMPLE DESIGNS

DESIGN	MEASUREMENT SCALE	STATISTIC	EXAMPLE	INTERPRETATION
2 Randomly selected groups for comparison Varying on only one dimension-eg. type of treatment program	Interval or ratio data	t-test for independent means	Classroom vs. programmed instruction on the measure of grade-level attainment.	If significant ($p < .05$) a difference between the groups is established, eg. programmed instruction produced greater grade-reason for difference otherwise. If significant, a difference between groups is established, eg. probation produced fewer recidivists proportionally.
	Category data	chi-square	Probation vs. group home on number succeeding and failing after 1 year period (recidivism)	
1 group randomly selected pretest and post test comparison, (quasi- experimental)	Interval or ratio data	Correlated t-test	Number of inappropriate behaviors reported during the first week of treatment.	If significant, the post test is different from the baseline (pretest) measure, eg. inappropriate behavior declined significantly in number during the course of treatment.
	Category data	chi-square	Assignment of group members to high- and low-risk categories before and after treatment.	If significant, the proportion of the group in each category differed from pretest to post test, eg. more low-risk people after treatment.

DESIGN	MEASUREMENT SCALE	STATISTIC	EXAMPLE	INTERPRETATION
2 groups randomly selected for comparison (varying on only one dimension, eg. type of treatment) Pretest and post test.	Interval or ratio data	Repeated measures or treatments-by-subjects (a type of analysis of variance)	Counseling vs. no treatment for runaway children, maturity scale pretest and post test for both groups (at same times)	If significant F, a difference between data sets is indicated. To determine the nature or locus of the difference, graphic representation or Neuman-Keuls test is required. (analysis of variance does not indicate locus of significant differences).
"	Interval or ratio data (alternative analysis)	t-tests (2) between 2 groups (pre 2. post)	"	If comparison between groups on pretest is significant, the sampling procedures were in error (start over). If t-test is significant between groups on post test, a difference between groups is indicated, eg. counselled runaway children scored higher on maturity scale than others.
1 group-time series design (this design is most appropriate for modifications in existing programs)	Interval or ratio data.	Graphic representation or average early vs. average late comparison.	Skill performance measures at regular intervals prior to and during modified training programs.	Graphic representation; if progress curve over time is regular, no change is evident. If curve changes near the point of program modification, change is indicated.
"	"	t-test	"	t-test - Procedures are essentially the same as for a one-group pretest, post test design.

CHAPTER 6 BEYOND PROGRAM EVALUATION TO AGENCY ACCOUNTABILITY

This report has stressed the "mechanics" of the evaluation process. It is evident, however, that this is not enough ; that there must be an individual as well as agency commitment to verifiable information, a disregard for mere assumptions, an insistence on measurement, a concern for awareness. Without these commitments, systems diagrams and formulas are inevitably relegated to mental and verbal gymnastics.

The emphasis on program evaluation underlines the need for an agency to be true to itself, accountable to itself. But this does not suppress or obviate the equal demand that an agency be accountable to others - to those who make the agency necessary as well as those who make it prosper. Obligation and often survival dictate that an agency justify its actions to society.

Thus, if this appendix were to be successful in assisting an agency to closely scrutinize its own activities, it would mark a beginning as well as an end, a means as well as a goal. For the next higher step would be for that same agency to open itself to the scrutiny of others, to the proper authority, to the government, to the people.

Is this being done already? Perhaps. It is suggested here, however, that agency accountability faces vicissitudes equal to or superior to those creating program evaluation.

Accountability is a holy word in public administration literature. The rhetoric is heavy and unmitigating. But responsibility for our own acts - being held to what we were committed to, elected to, appointed for - gives us the idea, and it does not take a soothsayer to tell us that the era of accountability for public agencies is hard upon us. The agency that sees this, sees the obvious. The agency that sees this and acts positively is still rare, and yet the pressures are vast. Local governments harassed by shrinking funds and the demands of growth are becoming weary of the cries of "wolf" that numerous agency heads substitute for assessment and justification of need. In addition, government and private agencies involved in grant funding are becoming more and more insistent that definable results emanate from the dispensation of funds. The past 3 or 4 years have spawned the expenditure of millions of dollars of criminal justice grant funds. Yet there is little valid information as to whether or not anything was accomplished by such expenditures. But perhaps most convincing of all is the attitude of the public - its insistence on learning the facts, its unwillingness to continue to accommodate mediocrity. There is no doubt, we will be held accountable.

As for solutions, there is no resource in sorcery nor does there have to be. If the typical police agency could objectvely evaluate its current attitude, it would readily see that significant improvement would not demand the immediate recruitment of magicians. To go before the heads of government and request ten, a hundred, or a thousand additional men to combat crime, returning each year to demand more men to fight more crime, constantly blaming the increases on public apathy and the collapse of the puritan ethic, does not sound convincing as the apex of any agency's effort to meet its responsibility or to account for it.

Nor is it convincing. If there is any wizardry in meeting the demands of accountability, it lies in the nonoccult devices of hard work.

What is your agency's reason for living? What is it going to do? How close is the agency to the completion of this or that objective? Why did your agency choose this objective? What is the cost vs. value? Where is the proof? These are the interrogations of accountability. Each agency must provide its own answers in its own way.

There are general guidelines, of course (and many tools and systems discussed in this appendix can be adapted to agency accountability).

Basically they consist of concepts similar to these:

1. Specific plans and the articulation and publication of those plans;
2. Analysis of the pros and cons of specific programs;
3. Analysis of cost vs. benefits;
4. Regular progress reporting;
5. Inclusion of funding agency into the planning process;
6. Budgeting for results not items;
7. Managing and staffing for effectiveness not "order"; and
8. Insistence by the funding agency to hold the funded agency accountable.

The remarks regarding magic solutions, however, were not spurious. Administrators and staffs must be aware of the black arts of the "in" term, the "name" system and the "one" way.

Such soundings as systems analysis, program evaluation review technique (PERT) and management information systems are often lauded, more frequently exaggerated, as organizational panaceas. Often the complexity of their structure tends to devalue their basic assets in order, creativity, planning, and a satisfied intellectual curiosity. The danger is that they evolve as organizational traditions rather than organizational tools. Far better to consider them as the emblems of both the simplicity and the complexity of the organizational process.

APPENDIX A

Information sources on treatment programs: A bibliography

In the preliminary steps of planning, a community is aware only that it has a juvenile problem and that the presently available resources provide too few alternatives for treatment. Deciding what kind of program is desirable and feasible is much harder. One approach to this problem is to become acquainted with some of the programs others have tried in similar circumstances. The annotated bibliography is intended as a source from which knowledge about a variety of programs can be obtained. The bibliography is not exhaustive; there are hundreds of other programs and sources which would also be applicable. Neither are the programs included intended to represent ideal programs. The sources do, however, offer information on a variety of kinds of programs and the problems they have faced. The bibliography includes sources on prevention programs, institutions ranging from foster homes to training school settings, probation-based programs, and community services using auxiliary community resources.

RESOURCE INDEX

I. PREVENTION

A. Work programs

Kansas City, Missouri, Work-Study Program. Youth Employment Programs in Perspective, Washington, D.C. Office of Juvenile Delinquency and Youth Development, U.S. Dept. of Health, Education, and Welfare, 1965.

Three stage program beginning in 8th grade when boys "vulnerable" to delinquency are identified by teachers and a socio-metric instrument. First year boys spend half-day in school, half-day in "socially useful work" around school (no pay). Boys are graded on work as well as classes. Stage two (usually 2nd year) boys work half-days away from school in service jobs for subminimum wages. If a boy wishes to leave school at this stage, school counselors will help him do so or he may elect to go into regular school program. Stage three boys leave school for full-time job while schoolwork-counselor continues to maintain contact and help if boys leave jobs or wish to acquire ones.

B. Police programs

Pursuit, Dan G.; Gerletti, John D.; Brown, Robert M.; Ward, Steven M. Police Programs for Preventing Crime and Delinquency. Springfield, Illinois, Charles C. Thomas, 1972. 490 p. $10.75.

Selected programs are described. Sections include: Role of Law Enforcement in Prevention; Community Relations; Prevention Programs for Specific Offenses; Educational and School Related Programs; Recreational Programs; Technological Programs; Funding; Requirements.

Cain, Thomas J.; Police Department, Pleasant Hill, California. Youth Services: A Police Alternative to the Juvenile Justice System. Part 2. Law and Order, 21 (1): 20-23, 33, 1973.

A program staffed by police officers (including female community aides) in which each officer is responsible for eight schools and spends ten to twenty hours per week becoming personally acquainted with students and staff. He provides service to alleviate problems before they begin. He also includes individual and group counseling. Children who might otherwise be cited to probation have been diverted. He improves rapport between police and young people.

Morrison, June. University of Arizona, Tucson, Arizona. The Controversial Police - School Liaison Programs, Police 13(2): 60-64, 1968.

Flint, Michigan program for police-school liaison efforts. Non-uniformed counselor assigned to a junior or senior high school; concerned with behavior problems around the school area. Preliminary results show a decrease in juvenile offenses.

Wattenberg, Wm. and Bufe, Noel. The Effectiveness of Police Youth Bureau Officers. Journal of Criminal Law, Criminology, and Police Science. 54 (4), December, 1963, pg. 470-475.

Detroit Police Department. Determined for each officer what proportion of boys for whom he was the first Youth Bureau contact became repeaters. Effectiveness defined as having a high proportion of non-repeaters. Comparisons confined to single precinct to reduce effects of rate variation in different neighborhoods. Qualities noted by supervisors used to form a record of each officer. From these a profile of an effective officer was drawn to aid in the selection of future Juvenile Officers. Most important result of study: evidence indicating the highly influential nature of first police contact on future delinquency.

C. Family Counseling

Teffesteller, Ruth. Delinquency Prevention through Revitalizing Parent-Child Relations. The Annals of the American Academy of Political and Social Science, 322 (March, 1959), 69-78.

Describes settlement house workers' attempts to redirect the gang type behavior of preadolescent (age 8-13) children by working with parents. Parents were led into groups and encouraged to support each other in developing a community which does not support gang delinquency.

Family Counseling (cont.)

Meeks, Tom, 1605 Eastlake Avenue, Los Angeles, California 90033. Project Summary: Family Treatment in Juvenile Halls. California Youth Authority; L.A. County Probation Department.
Short-term family treatment for children who have been detained by the juvenile court and are awaiting placement. Families come to special living units for short, intensive treatment focusing on parent-child communication system.

D. Schools

Scheier, I. Boulder County Juvenile Court, Boulder, Co. Volunteer Tutors in Court Probation Programs. 1968. 71 p.
Manual with guidelines for development of tutoring programs for juvenile probationers. Volunteers hope to reverse life-long trends of chronic under-achievers. Includes administration, qualifications, training, recruitment, methods, facilities.

E. Community action programs

Vaughan, Marion B. Flint, Michigan, Everybody Gets into the Act. American Journal of Correction. 32 (2): 6-8, 10-11, 1970.
Community-school philosophy based on keeping schools open day and evening, 6 days a week all year. Programs include regional counseling teams in the Junior High School; positive action program for boys on probation; rehabilitation including education, therapy, counseling; vocational guidance; police-cadet program; and community service officers for high school students.

Ikeda, Tsuguo. Seattle Atlantic Street Center, 2103 South Atlantic Street, Seattle, Washington 98144. Project Follow-up Summary: Effectiveness of Social Work with Acting-out Youth. NIMF: United Good Neighbor Fund, Seattle; Methodist Board of Missions. Begin September 1, 1962. Completion August, 1969.
Comparison study of impact of community treatment on a group of "high risk" seventh grade boys found less school discipline problems and less acting-out behavior in groups given intensive casework treatment.

II. INSTITUTIONS

A. Training schools

Catalino, Anthony. Bureau of Children's Institutions, Penn. Dept. of Public Welfare, Harrisburg, Pennsylvannia. A boys' committee as an instrument of communication Crime and Delinquency. 14 (4): 346-352, 1968.
Using a boys' committee to help broaden the channels of communication within the institution. Instrumental in improving regulations and procedures.

B. Foster care

Gater, L.; Hamm, R.; Osterberg, M. Boulder County, Juvenile Court, Boulder Co. A Home Away From Home: Community Volunteers Empty the Jail. Boulder Co., Boulder County Court, 1968, 46 p.
Volunteers serve as foster parents taking delinquents into their homes. Both group homes with paid parents and individual foster homes are used.

C. Camps

California, Los Angeles Co. Probation Department. 320 West Temple, L.A. Calif. 90012. Probation Camps: Information Series 8 L.A., L.A. County, Calif. Probation Dept.
Widely studied camp system for seriously delinquent boys. Junior camps (13-15) and senior camps (16-18) emphasize work training (forestry, road-building).

D. Short term

California Department of Youth Authority; Knight, Doug. The Marshall Program: Assessment of a Short-term Institutional Treatment Program. Part 1: Parole Outcome and background characteristics. Research report no. 56. Sacramento, Calif., Dept. of Youth Authority, 1969.
Therapeutic community model. Group interaction is principle treatment force. Paper presents evaluation comparing parole performances of Marshall boys to boys who were assigned to other institutions.

E. Detention

Simonson, Charles T. Detention Services, Summit County Juvenile Court Center, Akron, Ohio. Formulating Goals on Daily Program in Detention. Edwardsville, Illinois University, 6 p. 25 cents.
Prepared for the training of detention home personnel. Discusses movement toward goals, permitting decisions, interaction with group and with staff, sharing planning at all levels of staff. Evaluation is also considered.

F. Group homes

Keller, Oliver Jr.; Alper, Benedict. Florida Division of Youth Services, Tallahassee, Florida Halfway Houses: Community Centered Correction and Treatment. Lexington, Mass., D.B. Health, 1970. 203 p. $9.50.
General source on the community treatment center as a prerelease program for juvenile offenders and an alternative to institutional commitment. Describes professional and social concerns in the operation of such facilities. Discusses techniques, group intervention, guided group interaction as well as operation and administration.

Hargadine, John; Holmes, Horace B.; Scheir, Ivan H. Attention versus Detention. Boulder County Juvenile Court, Hall of Justice, Boulder Colorado, 80302.
Describes group foster homes for juveniles, directed by the juvenile court and supported by the local citizens to provide a homelike atmosphere as an alternative for children who would otherwise be in detention facilities, in inappropriate situations in their own homes, or in jail awaiting placement in foster homes or ranch-type "schools". Such homes are used as short-term placements.

G. Non-residential

Stephenson, Richard M.; Scarpitti, Frank R. Douglas College, New Brunswick, N.J. Essexfields: A non-residential experiment in group centered rehabilitation of delinquents. American Journal of Correction. 31 (1): 12-18, 1969.
A program patterned on Highfields (a residential group treatment center), which was successfully carried out in a non-residential setting. Includes careful experimental design and statistical analysis of results. Valuable both for the program description and the evaluation methods.

Henderson, Susan. Day Care for Juvenile Delinquents: An Alternative to Out-of-Home Placements. Judicature. 53 (1): 19-21, 1969.
Day care program dealing with the child in the family and community setting while she is attending school at the girls' detention facility at Juvenile Hall. Program used as a substitute both for institutionalization and probation. Has been successful in keeping girls in their homes with low recidivist rates and relatively low cost per girl.

III. PROBATION

A. Intensive

Los Angeles County Probation Dept. Reduction of Delinquency through Expansion of Opportunity (RODEO), L.A. County, Calif., office of research and standards, Probation Department, 1968, 40 p.

Intensive (cont.)

Substitution of intensive, in-community, supervision of 120 male wards of Juvenile Court. Experience with "hard-core" delinquents has been favorable. 1) reduce case loads 2) use indigenous aids 3) include family counseling.

Stark, Herman G. A Substitute for Institutionalization of Serious Delinquents: A California Youth Authority Experiment. Crime and Delinquency, 9: 242-248, 1963.
An intensive treatment-control program in the community as a substitute for institutionalization of delinquents committed by the juvenile courts. Agents work with caseloads of eight, employing a variety of programs.

Carpenter, H.A. 165 Grace Lane, Chicago Heights, Illinois. Project Follow-up Summary (final): Community Treatment Program for Delinquent and Neglected Children. Marquette Co. (Mich.) Probate Court; U.S. Office of Juvenile Delinquency and Youth Development.
Resocialization of delinquents by active environmental manipulation. Probation workers took active part in finding opportunities in social system for child on probation. Also used subprofessional area. Attempt to determine whether intervention services are successful in rural setting.

B. School

Maxwell, T. Mott Crime and Delinquency Prevention Program, Flint Board of Education. 923 E. Kearsley St., Flint, Mich. 48502. Project Summary: Positive Action for Youth (PAY) Mott Program, Flint Board of Education, Mich.; Juvenile Justice Division, Probate Court, Flint.
Fifty-five male probationers participate with peers, teachers, families. Program keeps boy in present environment (home and school) while offering group counseling, work experience, family counseling, supportive action, individual counseling. Includes evaluation of school improvement, measurement devices, other tools for use by interested schools and social agencies.

C. Paraprofessional volunteers

Rushen, R. and Hunter, F. Los Angeles County Probation Dept., Los Angeles, California. A Preliminary Report on the Indigenous Aid in Probation, Parole, and Correctional Association 5 (1): 10-16, 1968.
Using indigenous people from neighborhoods to provide services to help youths modify their behavior patterns. 120 male wards of juvenile court are receiving supervision in the community.

D. Family Counseling

Austin, Kenneth M.; Speidel, Fred R. San Bernardino County Probation Department, San Bernardino, California. Thunder: An Alternative to Juvenile Court Appearance. California Youth Authority Quarterly. 24 (4): 13-16, 1971.
Describes an intensive, short-term family group counseling program used in place of the judicial process in cases where the juvenile was not criminally oriented. Completion of program resulted in the case being settled out of court. Results suggest group techniques were as effective as individual casework.

IV. COMMUNITY SERVICES

A. Volunteer organization

Kirkwood, William. 44 E. 23rd St., New York, New York 10010. Volunteers in Corrections. Information Review on Crime and Delinquency, Vol. 1, No. 3. New York, National Council on Crime and Delinquency, 1968. 21 p.

Volunteer organization (cont.)

Deals with the use of volunteers to eliminate community uneasiness and facilitate re-entry of the offender into the community. Lists large number of successful programs and possible roles for volunteers from Big Brothers to employment programs for ex-convicts.

B. Training

Lee, Robert Edward; Piercy, Fred P. 708-202 S.W. 16th Avenue, Gainsville, Florida. 32601. Project Crest: Community-Based Treatment for Juvenile Delinquents. Gainsville, Florida, Florida Educational Research Development Council, 1972. 37 p.
Project Crest provides support for the Florida Division of Youth Services from teams of advanced graduate students in counselor education. The teams will provide a diagnostic program, crisis therapy, consulting service. Inservice training.

C. Consultation

Berlin, I.N. Mental Health Consultation with a Juvenile Probation Department. In: Szurek, S. The Antisocial Child; His Family And His Community. Palo Alto, Calif., Science and Behavior Books,1969. 224 p.
Deals with the use of mental health consultation consisting of lectures to the staff about child development and parent-child relationships, discussions of diagnostic categories of mental illness, and individual consultation with probation officers about their own conflicts about their jobs and clients. The consultation showed gratifying results in the functioning of the probation officer.

D. Planning

Office of Juvenile Delinquency and Youth Development. U.S. Dept. of Health, Education, and Welfare, Washington, D.C. Delinquency Today: A Guide for Community Action. Washington, U.S. Government Printing Office, 1969. 22 p. 20 cents.
A blueprint for community planning. Covers conflicting public attitudes, use of community based programs, personnel, legal services, federal role.

APPENDIX B

The following research report is based on fictitious data and is intended merely to serve as an example of a research design and report. No conclusions regarding the subject matter of the "experiment" can be drawn from any of the information included herein.

REVIEW OF LITERATURE

The effects of an "authoritarian" or a "democratic" group atmosphere on the behavior of children has been a topic of research since the study of group dynamics began. One of the early studies of leadership style compared the behavior of two groups of ten-and eleven-year old children; one a highly authoritarian group, the other a democratic group (Lewin and Lippitt, 1938). The authoritarian group was found to be more hostile, to produce a higher state of tension, and to turn aggression against a scapegoat. The democratic group was more cooperative, more constructive, more harmonious, and more stable.

Studies of leadership style have been conducted in many settings and the findings applied in educational and industrial institutions but they have been less widely adopted in correctional settings. However, with the wide recognition of the importance of peer influence on the behavior of delinquents, there have been programs which attempt to focus peer pressure in more pro-social directions. The Provo Experiemnt was one such program (Empey and Erickson, 1972), in which a system of self-government was developed wherein the boys worked out the solutions to group problems and the staff was committed to carrying out their solutions. Fixen, Phillips, and Wolf (1973) set about investigating experimentally the variables that effect youths' participation in self-government. They found that boys in a residential group home were capable of learning to participate effectively in a democratic system.

Maturation appears to involve an increasing capacity to relate constructively to others. This change has been conceptualized as seven successive levels of interpersonal maturity with subtypes within the level distinguished by the individual's characteristic manner of response (Sullivan, Grant, and Grant, 1957). Growing out of these categories, Jesness (1971) and Jesness and Wedge (1970) have developed a method of classifying delinquents as to their level of interpersonal maturity. This classification into one of nine I-level subtypes is used in their programs as a means of assigning delinquents to appropriate treatment programs.

Research on democratic and authoritarian groups to suggests that more constructive ways of relating to others can be taught in a democratic atmosphere. If interpersonal relationships are indicative of maturity level, then perhaps delinquent youths who have the experience of a democratic setting advance more rapidly to higher levels of interpersonal maturity.

METHOD

Subjects and Design. Six counties participated in the experiment. Each county has a boys' group home which houses 6 boys at a time. In three of the counties (county A, B, C), the rules of the homes were established by the house parents, and the structure was authoritarian. A token system was set up to enforce rules. In the other three counties (county D, E, and F), the homes were set up with democratic decision-making, where all rules and enforcement policies were arrived at by joint decisions between the boys and the house parents. Participants in the experiment were randomly selected from the pool of all boys in the 6 county region who were assigned by the county courts to enter a group home. Thirty-six boys in all were selected, ranging in age from 13-15 years. They boys were randomly assigned to one of the 6 homes, so that each boy had as much chance of being assigned to an authoritarian home as to a democratic home. Eighteen boys in all were assigned to the three authoritarian homes, and the other 18 were assigned to the three democratic homes. Upon entering the home, each boy was given the Jesness Inventory, and a score was computed for the immaturity scale. At the end of 6 months, all Subjects (Ss) were again tested on the Jesness Inventory and released to their own homes. The research design is a randomized 2 group design with pretest and post test in both groups (See Figure 1).

Figure 1. Design of Research.

	Pretest	Treatment	Post test
Authoritarian group homes	Jesness Immaturity Scale	Six months in home	Jesness Immaturity Scale Retest
Democratic group homes	Jesness Immaturity Scale	Six months in home	Jesness Immaturity Scale Retest

Instrument and Procedures. The immaturity scale of the Jesness Inventory is a paper and pencil test derived from an item analysis of criterion groups. It was designed to measure attitudes and perceptions of self and others which are inappropriate for the subject's age level. The reliability coefficients of the Jesness scales ranged from .62 to .88 for odd-even reliability and .40 to .79 for test-retest reliability. The immaturity scale was shown (Jesness, 1962, 1966) to relate behaviorally to such dimensions as low WISC IQ and achievement test scores, conforming, non-aggressive behavior, little insight, poor social poise, and low social status. Based on Jesness (1966), norms for the immaturity scale are as follows:

	Mean ⟶	Standard Deviation
Delinquent males	13.2	4.5
Non-delinquent males	11.6	3.7
Delinquent females	12.4	3.5
Non-delinquent females	11.9	3.6

These norms are based on populations ranging in size from 103 to 135 youth.

All six residents of each group home were tested together within a week of their first entrance to the home. Subjects were informed that the purpose of the testing was to provide data to be used to compare the effectiveness of their group home with that of other group homes.

RESULTS

Scores on the immaturity scale ranged from 8 to 19 for the democratic home pretest, 7-16 for the democratic home post test, 7-20 for the authoritarian home post test. Means for pretests and post tests are shown in Table 1.

Table 1. Means for Authoritarian and Democratic group homes--pretest and post test.

		Pretest Means	Post test Means
Authoritarian Homes	A1	13.17	12.50
	A2	13.50	13.00
	A3	13.50	12.67
Total Authoritarian		13.39	12.72
Democratic Homes	D1	13.33	11.00
	D2	13.33	11.33
	D3	13.5	10.67
Total Democratic		13.39	11.00

The immaturity scale is designed so that a lower score reflects less immaturity as defined by the scale items. There was no mean difference between combined authoritarian S scores and combined democratic S scores on the pretest measure, i.e. both means were 13.39. An independent means t-test between combined authoritarian S scores and combined democratic S scores on the post test measure was performed. A significant difference ($t = 2.049$, $p < .05$, $df = 34$) was obtained, in the direction of post test immaturity among democratic home Ss than among authoritarian home Ss. The present findings support the prediction that a democratic home structure is more conducive to a reduction of immaturity (as measured by the Jesness immaturity scale) than is an authoritarian home structure.

A t-test was computed comparing the democratic and authoritarian group home Ss with regard to the amount of change between pretest and post test measures, i.e. each S's score consisted of the post test minus the pretest score. The result was $t = +3.07$, $p < .01$, $df = 34$, in the direction of larger change scores for Ss from democratic homes than for Ss from authoritarian homes.

In order to demonstrate the weakness of non-experimental designs, a correlated t-test was performed on the pretest and post test data from the democratic homes, as though there were no authoritarian control group. The post test immaturity scale scores were significantly lower than the pretest scores, $t = 2.743$, $p < .05$. This significant result suggests that a change on the measured dimension occurred during the treatment period. However, there is no way of isolating the cause of the change, i.e. whether it was due to maturational processes or some unidentified aspect of the treatment program. The presence of a control group (the authoritarian homes) adds considerably to the amount of information gained. It can be observed that the direction of change is the same for both types of treatments. It would be instructive to have still another control group to help in determining whether similar changes also occur in a non-delinquent or untreated delinquent population.

SUMMARY

It was found that delinquent youth assigned to democratic group homes demonstrated a greater reduction in immaturity than comparable youth assigned to authoritarian group homes. While it has not been demonstrated that immaturity leads to or causes delinquent behavior, it has been previously shown (Jesness, 1966) that delinquent youth score higher on the immaturity scale than do non-delinquent youth. More research is needed to establish the nature of the relationship between immaturity and delinquent behavior.

REFERENCES

Empey, L. T.; Erickson, M.L. The Provo Experiment. Lexington, Mass.: Lexington Books, 1972.

Fixen, Dean L.; Phillips, Elery, L.; and Wolf, Montrose M. Achievement Place: Experiments in Self-Government with Predelinquents. Journal of Applied Behavior Analysis, 1973, (sp.) 6 (1), 31 - 47.

Jesness, C. F. The Jesness Inventory: Development and Validation. California Youth Authority Research Report No. 29, March, 1962.

Jesness, C. F. The Jesness Inventory Manual. Palo Alto, California: Consulting Psychologists Press, 1966.

Jesness, Carl F.; Wedge, R. F. Sequential I-Level Classification Manual. Sacramento, California Youth Authority, 1970.

Jesness, Carl F. The Preston Typology Study: An Experiment with Differential Treatment in an Institution. Journal of Research in Crime and Delinquency, 1971, 8 (1) 38 - 52.

Lewin, Jurt; Lippitt, Ronald. An Experimental Approach to the Study of Autocracy and Democracy: A Preliminary Note. Sociometry, 1938, 1, 292 - 300.

Sullivan, C. E.; Grant, M. Q.; and Grant, J. D. The Development of Interpersonal Maturity: Applications to Delinquency. Psychiatry, 1957, 20: 373 - 385.

APPENDIX C

For the present work to be maximally helpful, it should be used in conjunction with a source on statistics and design. The following are some of the sources available from psychology. There is a wide variety of other sources available in the social sciences which will serve adequately.

REFERENCES

1. Bruning, J. L. and Kintz, B1. Computational Handbook of Statistics. Glenview, Illinois: Scott, Foresman and Company, 1968.

2. Campbell, D. T. and Stanley, J. C. Experimental and Quasi-Experimental Designs for Research. Chicago, Illinois: Rand McNally and Company, 1963.

3. Matheson, D. W., Bruce, R. L. and Benchamp, K. L. Introduction to Experimental Psychology, New York: Holt, Rinehart, and Winston, 1970.

4. Hays, W. L. Statistics for Psychologists. New York: Holt, Rinehart, and Winston, 1963.

5. Koenker, R. H. Simplified Statistics. Totowa, N. J.: Littlefield, Adams and Co., 1971.

6. Maxwell, A.E. Basic Statistics in Behavioral Research. Baltimore, MD.: Penquin Books Ltd., 1970.

END