

LRI REPORT No. 18

RESEARCH, EVALUATION, AND REPRESENTATION

Peter G. Ossorio

© Copyright 1978

Linguistic Research Institute
Boulder, Colorado

The purpose of this report is to delineate (a) two general research paradigms, (b) the relation of these paradigms to evaluation research methodology, and (c) the derivation of these paradigms from certain Descriptive Psychological schemas for representing real world phenomena.

I. The Actor-Observer-Critic Schema

This Descriptive schema refers to three methodologically fundamental behavioral statuses, or ways of functioning. These three are characterized briefly as follows:

- A. In functioning as Actor, the individual acts in response to his circumstances in accordance with his nature. This form of functioning is, therefore, characterized as (1) before-the-fact, (2) spontaneous, (3) creative, and (4) value-giving.
- B. In functioning as Observer-describer the individual takes note of (discriminates, describes) the course of events, particularly those which are relevant to the behavior of the Actor. This form of functioning is, therefore, characterized as (1) after-the-fact, (2) passive, (3) reflective, and (4) either value-neutral or value-finding.
- C. In functioning as Critic-appraiser the individual begins with the results of Observer-describer functioning and makes an appraisal of whether the course of events is satisfactory or unsatisfactory.
 - (1) If satisfactory, the Critic takes no positive action but appreciates/enjoys the success.
 - (2) If unsatisfactory, the Critic generates (a) a diagnosis, i.e., an account of what it is that has gone wrong and/or (b) a prescription,

i.e., a practical guide in regard to what to do differently so as to improve matters.

This form of functioning is, accordingly, characterized as (1) after-the-fact, (2) reflective, and (3) judgmental, verdict-giving, or value-finding.

In the functional form of the AOC schema, the three statuses are related in the form of feedback loop. As shown in Figure 1, what the Critic judges to be the case is an essential part of the circumstances to which the Actor then responds, which then begins a new feedback cycle.

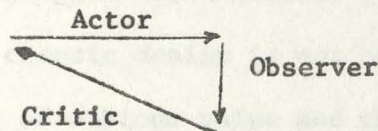


Figure 1. Functional AOC Schema

In this connection, it should be noted that Observer-describer functioning is a special case of Actor functioning and that Critic-appraiser functioning is a special case of Observer-describer functioning, whereas Actor functioning is simply the general case of a person behaving.

One of the primary values of the Actor-Observer-Critic schema is that it provides a systematic representation (in conjunction with other notational Descriptive devices) of a person as being self-regulating while nevertheless being responsive to both personal and circumstantial factors. It is this feature which finds a fairly direct exemplification in one of the research schemas, the Precaution-Assurance Paradigm.

II. The Precaution-Assurance Paradigm

Some version of the classical experimental designs involving a control group and experimenter manipulation of independent variables is commonly con-

sidered to be the standard of rigor in establishing general conclusions on an empirical basis. Correspondingly, "applied research," including validation or evaluational research, is generally seen in either one of two lights. (a) In the first case, it is seen as the application of general "knowledge" which was itself acquired in this foolproof (rigorous) way. That is, it is seen as a mere demonstration that some general principles arrived at via 'rigorous' methodology do apply to some phenomenon of interest in the real world. (b) In the second case the classic experimental procedures are held up as the ideal to be striven for in the empirical examination of the real world phenomenon itself. It follows in this case that to the extent that the classic design is not adhered to (and usually, it cannot) the results are of dubious value and the conclusions drawn are unsound (though possibly true).

It should be noted however, that classical experimental methods are designed to support general conclusions. The primary interest is never on the actual subjects as such. In contrast, much 'applied' research, and certainly most evaluation research, is designed to support particular conclusions. That is, what is of primary interest is not whether a certain kind of program, procedure, instrument, etc. is generally successful in this or that way or degree, but rather whether a historically particular program (or set of programs), procedures, etc. is in fact successful in this or that way or degree over a particular span of time.

Even when more general conclusions are desirable in evaluation research the emphasis is properly on a case by case approach because ordinarily the relative influence of historical-situational contexts is so marked that simple generalizations about procedures (programs, etc.) across contexts would be rash or foolhardy (even if it were a case of 'applying rigorously

established findings'). One can always sum across replications of a fine-grained analysis; one cannot recover individual patterns from group data.

Clearly then, the specifications for a viable evaluation methodology would include (a) being capable of rigorous implementation in the real world setting, (b) providing fine-grained or single case conclusions, and (c) lending itself to systematic variation and replication for supporting more general conclusions. The Precaution Paradigm is responsive to these requirements.

Let us first examine the fine-grained module. In this case we presuppose a real world setting where there is a procedure (instrument, program, etc.) which is engaged in and which is a candidate for evaluation. Our first move is to construct a representational formula (cf the Process Description, below) for the phenomena in question. This will have the general form (1) A uses/does B with C in ways D (e.g., to accomplish G), which can go wrong in ways E_i , as indicated by observations F_{ij} .

For example, a therapist (A) may use a projective test (B) with a client (c) by interpreting M responses (D) to arrive at a psychodiagnosis and differential treatment (G). This might go wrong in that the client was misdiagnosed and the treatment would be ineffective (E_i), as indicated by the continued presence of the initial symptoms (F_{ij}). Or, again, an agency worker (A) may use a "mastery of English" entry on an application form (B) with an unemployed refugee (C) to direct him to a laborer's job and a vocational training class (D). This might go wrong in that the refugee would perceive himself as having lost face and subsequently become depressed (E_i), as indicated by self-report of feelings and symptoms and failure to hold the job or progress in training (F_{ij}).

Given the representational formula, we next focus on the ways in which

A could go wrong by using /doing B with C in those ways. At this point there arises the issue of real, or practical, doubt versus idle skepticism. In this regard, we introduce two test questions. The first is, which of these ways of going wrong am I most worried about? Given an answer to that, the second question is, is that possibility important enough so that it's worth taking precautions against being wrong in that way? If the answer here is "yes," the next question is, which of the indicators of its going wrong that way are both convincing enough and accessible enough to make them worth getting (via the relevant observations)? If there are such indicators, the observations are made. These observations provide either a reassurance that the project is not going wrong that way or a warning that it probably is going wrong in that way.

The foregoing constitutes the required methodological unit or module. In this regard the following may be noted.

- (a) The sequence, and the logic of the sequence, is that of Actor-Observer-Critic, and the sequence is one of an error-detecting feedback cycle.
- (b) What qualifies as "going wrong" (the "diagnosis") is context-bound insofar as it depends on the specific purpose(s) for which B is used/done and the specific setting, persons, etcetera which are involved.
- (c) Correspondingly, the indicated corrective measures (the "prescription") will be context bound. (In the context of program or treatment evaluation, an overall evaluation will correspond to a single feedback cycle whereas a pragmatic evaluation will involve repeated feedback cycles and modification of the program during its progress if the feedback information so indicates.)
- (d) The functional AOC schema is both an error-detecting and a success-detecting feedback loop. Although one might define success logically

in this context as simply the absence of failure, in fact there may be, and usually are, positive indicators of success. Thus, there is a corresponding module based on a modified formula in which we refer to ". . . can go right in ways E_1 , as indicated by observations F_{ij} ." And, of course, we may combine the two, i.e., ". . . can go right or wrong in ways E_1 etc."

- (e) What qualifies as success will in general be no less specific and context-bound than what qualifies as failure.
- (f) If there are no observational indicators of B having gone wrong in a given way in the real world settings, we may conclude that there is no real problem of its going wrong in that way.

The methodological module lends itself to systematic variation and replication because the representational formula amounts to a parametric analysis of the kind of use (instrument, program, etc.) which is to be evaluated. That is, A, B, C, D, E, and F will be parameters of that kind of use. Thus, one may replicate across A, i.e., persons or groups engaging in the procedure; across B, i.e., different forms, instances, or variations of the procedure; across C, i.e., different recipients or groups of recipients of the procedure; across D, i.e., the various ways of using/doing B, with special reference to the purposes, circumstances, settings, etc.; across E, i.e., ways of going right or wrong; and across F, i.e., different indicators of success or failure. And, of course, one may replicate across any or all combinations of these various possibilities.

By way of elaboration we may note the following.

- (a) Although the paradigm presented above involves the representation of a procedure in a historical setting, the procedure may be either actual or hypothetical, e.g., merely intended or planned. Correspond-

ingly, and particularly in the latter case, the ways in which the procedure might go right or wrong might be exemplified in some other setting, e.g., an experimental setting. For example, if the failure envisioned in the historical setting were the misdiagnosis of the client, that failure might be equally well exemplified in an experimental setting with clients selected specifically for the purpose of conducting the test. (But it also might not be. The problem of the "external validity" of experiments is a real one and there are no general solutions to it.)

- (b) The modular approach makes relatively clear something which is often glossed over in the classic experimental tradition, namely that it is impossible to take all logically possible precautions against being wrong or to obtain all logically possible assurances of being right.
 - (c) Classic experimental design can be seen as a case of (1) taking certain standard precautions, whether relevant and important or not, and (2) taking them in advance of serious real world practice ('application'). Such a strategy might have some real value, but it might also be counterproductive.
 - (d) Experimental procedures, like any procedures, will exemplify B in Formula (1). We may represent the real world phenomenon by saying that A uses experimental procedures (B) with subjects and problems (C) in certain ways and with certain purposes in particular settings (D) such that A can go wrong in ways E_i as indicated by observations F_{ij} . This holds equally for the use of traditional experimental paradigms and for the newer paradigms such as the Precaution Paradigm.
- The use of experimental procedures is thus not something which per se carries any guarantee of any kind of success or avoidance of failure.

And we cannot, with respect to a given kind of use of experimental procedures take all possible precautions etc. Experimental procedures provide a framework for exercising human judgment and competence, not a way of doing without it.

III. Process Representation

The systematization of reality concepts (object, process, event, state of affairs, relationship) via a transformational calculational system and the systematic explicit representation of objects, processes, et cetera are presented in "What Actually Happens"(Ossorio, 1975, 1978). The following is a brief summary and review.

The two most relevant transformational rules ("Transition Rules") dealing with the general concept of a process are as follows.

Rule 4. A process is a sequential change from one state of affairs to another.

Rule 5. A process is a state of affairs having other, related, processes as immediate constituents.

Because these recursive rules deal with the general concept of a process they apply to all processes and will therefore not distinguish one process from another. In order to have a general method for giving explicit representation to a particular process or kind of process we require (a) a parametric analysis of "process" and (b) a systematic notation or format for giving process representations. Both of these requirements are met by the "Basic Process Unit" (BPU) shown in Table 1.

The Basic Process Unit has a gross structure of "Name" and "Description." The former identifies the process and the latter gives the explicit representation.

In the explicit "Description" we may distinguish a "gross structure"

TABLE 1

<i>Basic Process Unit (BPU)</i>	
P-NameA:	The process "Name" of process A.
P-DescriptionA:	The "Description" of A. It specifies:
I. P-Paradigms:	The major varieties of P-NameA. This is a technical option. If only one paradigm exists, it will be the same as P-NameA. For each paradigm, the following is specified:
(a) Stages I-K:	These are "Names" of subprocesses within A. They are systematically specified, e.g., as P-NameA11, P-NameA12, . . . , P-NameA1K for Paradigm 1. For each stage, specify:
(1) Options 1-N:	These are the various exemplars of the process (stage) in question. That is, these are the various ways in which that process could happen. Each Option is systematically indexed as P-NameA111, P-NameA112, . . . , P-NameA11N. Each of these can now be expanded (decomposed) on the model of P-NameA.
(b) Individuals	
(c) Elements	
(d) Eligibilities	
(e) Contingencies	
(f) Versions	

(Stages, Options) and a "fine structure," or "State of Affairs Structure," (Elements, Eligibilities, Individuals, Contingencies, and Versions).

Rules 4 and 5 are most clearly expressed in the fact that the process is represented as a sequence of Stages each of which is itself a process. The fact that the latter processes (stages) can each have some number of distinguishable exemplars is expressed in the association of some number of Options (exemplars) to each Stage. It is the Stage-Option structure which codifies the recursiveness of Transition Rules 4 and 5, for each Option is itself a process and can therefore be represented by a Basic Process unit involving a new set of Stages and Options, and the latter can in turn be so represented, and so on ad infinitum. This enables us to represent processes of any magnitude in any degree of detail, down to continuous processes.

Processes generally involve object constituents as well as process constituents. These objects have certain relationships (their having these relationships is a state of affairs) which change over time and the changing of these relationships over time is (the same state of affairs as) the occurring of the process. "Element" refers to the formal ingredients (objects) of the process. "Individuals" refers to historical individuals in the abstract, and "Eligibilities" assigns Individuals to Elements. For example in the play "Hamlet," Hamlet, Polonius, the skull, and the castle are all Elements. Person X, Person Y, prop A, and prop B are abstract historical individuals. Either Person X or Person Y may be Eligible to play Hamlet or Polonius (but not the skull or the castle), whereas Prop A is eligible to play the part of the skull and Prop B is eligible to play the part of the castle. Both Individuals and Elements are needed in the formal specification because there need not be a one to one relation between them. For example, Person Y may

be eligible to play Hamlet in Scene I and a spearbearer in Scene II. Actual performances of "Hamlet" require actual individuals in place of the abstract Individual even though "Hamlet" as a kind of process involves no reference to particular persons.

The occurrence of a process involves the occurrence of one of the Options for Stage 1 followed by the occurrence of one of the Options for Stage 2, and so on. In the process representation restrictions on the occurrence of particular Options are given by Contingency statements, Contingencies may be Attributional or Co-occurrence contingencies.

In the case of a Co-occurrence contingency the occurrence of a given Option in a given Stage is incompatible with (or necessitated by) the occurrence (or non-occurrence) of some other Option in some other Stage (or a combination of such options). For example, in a chess game, White's possible fifth moves are highly restricted by which of the possible first, second, third, and fourth moves have actually occurred, and many moves which are in principle possible in chess as fifth moves (are Options in Stage 5 of that process) are not possible in this game (this Version (see below) of that process).

In the case of an Attributional contingency the availability of a given Option in a given stage is contingent on some attribute of the individual who is serving as a given Element. For example, throwing a 90-yard pass in football is one of the formal possibilities, but it could only occur if the passer had an exceptionally strong arm.

Statistical Contingencies represent empirical correspondences rather than conceptual requirements. They may be of either a Co-occurrence or Attributional sort. For example, a 90-yard pass would be unlikely if the line of scrimmage were more than 30 yards from the offensive team's goal

line (which reduces to a Co-occurrence Contingency). Similarly, a checkmate in four moves would be unlikely to be available to a player who did not see it ahead of time. And a refugee for whom attending class represented an extreme loss of face would be unlikely to attend class.

When a process takes place it takes place in one of the ways in which it can take place. The specification of Stages, Options, Elements, Individuals, Eligibilities, and Contingencies is a way of specifying the conceptual restrictions (optionally, augmented by empirical restrictions) on what sort and sequence of happenings would qualify as an occurrence of (an exemplar of) the process in question. Each such distinguishable exemplar is a Version of the process. Each of the ways in which a given process can take place is a Version of that process. The occurrence of the process on a given occasion is (the same thing as) the occurrence of one of its Versions on that occasion. In contrast, occurrences of the same process on different occasions will generally involve the occurrence of different Versions on the different occasions. And different Versions need not resemble each other in any other way than in being Versions of the same process.

The problem of "generalizing" or "applying" the results of laboratory or 'analogue' research is in part the problem of trying to draw conclusions about one process (Version A of Process X) on the basis of observing another process (Version B of Process X) which is not unlikely to be different in many crucial respects.

The requirements for describing a given process as such are in principle not different from the requirements for a parametric analysis of "process," i.e., for specifying the ways in which one process (or kind of process) can be the same as another process (or kind) as such or different from it. Thus, the reference to Stages, Options, Elements, Individuals,

Eligibilities, Contingencies, and Versions constitutes a parametric analysis of "process."

One important feature of the process representation provided by the BPU format is that it exemplifies a holistic, or "top down," approach. Description consists of specifying various facts about the process in question. Each fact serves to further distinguish that process from other processes of the same general sort, and there is no definitive 'complete' description. Thus, the BPU form of representation begins to be informative as soon as any information is available concerning the process in question.

Systematically incomplete process descriptions are codified by Means-Ends Descriptions or Task Analyses ("What Actually Happens," Chapter III). In a Means-Ends Description we merely specify the Elements of the process (or of each of the Stages) which contribute to the outcome of the process (including desired changes, if any). In a Task Analysis we merely specify what sets of states of affairs would qualify as successful Outcomes (in effect, we specify different 'Versions' of a successful outcome).

IV. The Simulation Paradigm

This research paradigm is designed to enable us to (a) formulate our understanding of a phenomena (and/or our guesses about it) by generating a process representation of it and (b) test that understanding by predicting certain facts about the phenomenon on the basis of other facts. What makes these predictions possible is the structure of the process description including, importantly, Co-occurrence and Attributional Contingencies. Because of the predictive implications of the contingencies which are stated in "if-then" form, the obvious technical implementation for this research paradigm is computer simulation, hence the designation "Simulation Paradigm."

In the context of the BPU process representation, the Simulation

Paradigm involves the following features.

- (a) We begin with the general notion of a process and ask, "What is the phenomenon?" The answer here is given by specifying it as a process, e.g., "providing (certain) mental health services to children," and distinguishing various possible exemplifications (Versions) as categorized by the investigator. Classes of exemplifications correspond to the "outcome variables" of the classic evaluation design.
- (b) We ask, "What makes a difference in how the process goes?" The answer is given in terms of (1) the parameters of the process or any of its elements, and (2) the values of those parameters. The first of these two is a set that corresponds to the interventions or other predictive (causal or noncausal) variables in the traditional outcome study. There is no substantive limitation on the type of parameter which may be involved. Thus, aspects of institutions, programs, persons, and situations or relationships may all enter the picture as "what makes a difference."
- (c) We ask, "Where does it make a difference?" In answer, we merely specify a locus in the process representation. Whatever makes a difference has to make a difference somewhere.
- (d) Finally, we ask, "What difference does it make there?" The answer will have the general form, "Depending on whether the value of parameter Q is x or y, the exemplification (of, e.g., "providing (these) mental health services to children") will belong to classes A, B, C," There is no restriction on the kind of functional relation which may appear here. (The statement of these relations is an integral aspect of the Process Description; formally, it consists of specifying attributional and co-occurrence contingencies.) For example, it may be a

simple linear function, but it may be a non-numerical decision table or a logical "either . . . or . . . and . . . unless . . . if" kind of function. The latter is of particular interest in connection with the problem of strongly interacting variables (it is the potential for non-numerical values which leads us to speak generally of "parameters" rather than "variables"). In this case, the outcome is expressed as an explicit joint function of the values of the several "interacting variables," and that function will commonly have "or," "unless," and "if" components. In some of the very simple special cases, the joint function will be identical to a representation within the familiar linear additive model or other models used in multivariate analysis. (Any type of multivariate analysis can be accomplished within the Simulation Model.)

The simulation model has the following features of interest.

- (a) It permits (indeed, requires) explicit statements of functional relationships between predictor and outcome parameters.
- (b) There is no limit to the form which these functional relationships may take or to the nature of the parameters involved (e.g., personal, institutional, physiological, etc.).
- (c) It permits a clean separation between the conceptual model of the phenomenon and a mathematical model of the statistics.
- (d) It permits the detailed examination of the effects of any subset of predictor parameters, since the remaining parameters may be held constant, e.g., by using group means for individuals, hence allowing no variation in that parameter. (Such analysis would normally call for computer simulation implementation.)
- (e) Because it is, in effect, a predictive test of the entire set of rela-

tionships simultaneously, the sample size required for testing does not increase exponentially as it does in purely empirical multivariate designs.

- (f) There are at least two significant restrictions on the use of simulation as against measurement. The first is that the actual Process Description must reach an appreciable level of accuracy before it is of practical value. The second is that there is no analytical procedure for establishing the "best fit" to a set of sample data. Inspection methods or systematic variation of the hypothesized relationships are available, however, and the post hoc revision with cross validation is a standard procedure. Because the limitations of covariation and simulation methods are relatively complementary, a general approach to impact analysis is to use both methods.

IV. Evaluation Research

Evaluation research and routine program evaluation occur within the context of some basic facts which impose strong limits on the manner and ease with which evaluation can be effectively accomplished and the uses to which a given evaluation can legitimately be put. Among these brute facts are the following.

- A. Improvement does not occur in "pure form." Just as an actual success is never merely a success but is also a particular accomplishment such as winning a race, so an actual improvement on the part of some particular individual is always also a more specific change in personal characteristics, behavior, relationships, achievements, etc. But it is the fact of improvement that we are interested in, not the particular changes per se.
- B. There is no specific change which is per se necessarily or intrinsically an improvement. The decision as to whether a given change will qualify

as an improvement requires evaluative human judgment. This judgment will, in part, reflect the target individual's personal characteristics and his life setting, as well as the purposes or norms with respect to which the issue of improvement arises for the person making the judgment.

- C. When improvement is exhibited by different individuals, in general, it is exhibited in different specific ways, even when it is "the same" improvement. (The differences will correspond to the differences between different Versions of the same process.) Thus, in principle, there is no way to specify rigorously and in objective terms the specific changes which would qualify as improvement for an entire group of individuals. At its worst, such a stipulative attempt would approximate the irrelevance of flipping a coin to decide the question of improvement. At best, it might be good enough for some purposes for some people.
- D. In principle, there is no change of a specific sort which all observers would agree qualified as an improvement for a given individual. In general, a given change in an individual will affect different significant persons differently and they will, properly, make different judgments of "improvement." Judgments of this sort will, therefore, carry weight only among persons who are in agreement in this respect. The likelihood of agreement may be increased in several ways:
1. By moving towards "lowest common denominator" phenomena: "If he's learned to dress himself (or "if he can now hold a job," "if she can now sit still in class") surely that's an improvement."
 2. By specifying norms or frameworks: "From the point of view of reality contact, his being more aggressive at work is an improvement."
 3. By restricting the implications of the judgment or uses to which it

is to be put: "For purposes of deciding whether to terminate."

"From the standpoint of how to advise the parents," etc.

4. By stipulating that one party, e.g., the client, the parent, the school or agency representative makes the decision.
- E. In a real world setting the effects of a given influence, e.g., mental health treatment, are inevitably confounded with an unknown number of important events and other influences. Changes from before to after cannot simply be attributed to a given influence. The experimental devices of random assignment to groups or extensive control of individual cases are palliatives, not solutions, and they are almost never available in field settings, since therapeutic, administrative, political, or other considerations will properly take precedence.
 - F. Evaluative phenomena, being historical rather than universal, will, in general, be different across time and place. A treatment procedure which is effective here now is likely not to be effective somewhere else (different clients, different problems, different milieu) or here ten years later (different staff, changing times leading to changing problems, etc.). Thus, the utility of a given evaluation is limited. In an organization, evaluation is likely to be required on a continuing or periodic basis.
 - G. All of the foregoing were phrased in terms of the evaluation of "improvement." Corresponding statements could be made in respect to other evaluative concepts such as "adequate functioning," "psychopathology," "mental health," and "life problem." There is no set of behaviors, objective indicators, et cetera which is logically equivalent to any of these evaluative phenomena.
 - H. There is no set of procedures which intrinsically or necessarily qualifies as performing an adequate program evaluation. The decision as to

whether a given set of procedures does so qualify on a given occasion requires human evaluative judgment. The adequacy of a program evaluation procedure provides the same general problem of evaluation as does the degree of improvement of clients under treatment.

Consider the case of new programs for children and elderly persons initiated by an urban mental health center. The systematic evaluation described below is designed to be responsive to the general considerations noted above and to be of practical value for a variety of purposes, including decisions regarding changes in program procedures and decisions regarding allocation of resources to different aspects of the Child and Elderly programs.

A. Evaluation Structure

The overall evaluation has three major components, which are here designated as predictive, procedural, and retrospective.

1. The predictive component resembles the classic outcome of treatment design in which indicators of improvement are designated initially and are assessed before and after treatment.
2. The procedural component is related to traditional "process" studies of psychotherapy in that it involves a process description of the "treatment" or "program" and an evaluation of its appropriateness independent of outcome.
3. The retrospective component is a closure achieving procedure. It involves:
 - a. An evaluative review of the original assessment of the treatment planning;
 - b. An evaluation of improvement unrestricted by the predictively

designated indicators of improvement; and

- c. An evaluation of the extent to which improvement or lack of improvement could be attributed to factors other than the treatment program.

B. Evaluation Procedures

The following is a narrative outline of the procedures involved in the evaluation. The integration of treatment and evaluation is such that all but the last section (the retrospective components) is also essentially an outline of the treatment process.

1. Intake

In this phase, an intake worker obtains standard background information and a statement of the problem, and makes a routine assessment of personal characteristics, including behavioral tendencies and personal resources and deficiencies.

2. Treatment Planning

An ad hoc treatment team is formed for the client. The treatment team will include at least (a) the primary therapist or caregiver, (b) another clinically knowledgeable person, (c) the continuity of care person, and (d) for designated cases, a member of the evaluation team. The treatment team reviews the assessment information and does the following.

- a. Requests further assessment, if needed
- b. Judges the degree of need for treatment ("severity")
- c. Decides on strategy and type of treatment
- d. Specifies the nature of the problem
- e. Specifies the particular ways the problem is manifest ("ad hoc indicators")

- f. Specifies prima facie indicators for improvement. It is primarily the "severity" rating and these indicators which appear in the predictive "before and after" analysis. Among the indices which are likely to be used are the following:
- (1) Specific achievements, e.g., expresses affection toward her son, establishes a friendship with someone, learns class lessons without disrupting class.
 - (2) Changes in personal characteristics, e.g., becomes more responsible, less passive-aggressive, more tolerant of other people's shortcomings, less anxious, etc. Along with the specification of such changes is a specification of the preferred way of establishing these changes, e.g., therapist judgment, self-report, standardized test, special interview.
 - (3) Judgments by two significant figures in the client's life in regard to either characteristics or changes. Among such possible significant figures are the client, the therapist, a family member, an employer or teacher, a friend, a spouse or spouse-equivalent, an agency representative. Part of the specification here is what judgments on the part of those persons would confirm the evaluation of improvement. (In light of the introductory comments, it may be noted that sometimes the confirming indicator would be a negative judgment by a significant figure.)
 - (4) The length of time required for effective treatment.
- g. Specifies, if possible, a set of criteria for classifying treatment as successful and for terminating treatment accordingly.

3. Treatment Review

At least once during treatment (perhaps at an interval of one-third the time estimated for effective treatment) the treatment team reviews the treatment procedures. The review is based on the treatment plan and the therapist's progress notes together with any supplementary information which may be available from the evaluation team. At this time, the treatment team decides (a) the degree to which treatment is in accordance with the plan, (b) the degree to which the treatment now appears to be appropriate, and, if indicated, (c) how to proceed with a modified treatment. The evaluation, together with a parallel evaluation by the evaluation team after termination constitutes the procedure component of the evaluation.

4. Termination Evaluation

This evaluation is conducted by the treatment team and may be initiated by either the primary therapist or the continuity of care person. The evaluation is based on the treatment planning and review data, summaries by the primary therapist and the continuity of care person, and assessment (of improvement indicators) data provided by the evaluation team member. The treatment team makes a new rating of need for treatment, reviews its previously formulated criteria for success, and recommends termination or continuation. In the latter case the procedure is repeated when termination is again in question.

5. Follow-up

At a suitable interval after termination, normally six, nine, or twelve months, the indicators of improvement, or a subset thereof, are again assessed. A comparison of this data with the termination assessment provides a descriptive characterization of the stability

of improvement.

6. Retrospective Review

The retrospective review is conducted by an evaluation team which is formed ad hoc for each client. Only the continuity of care person will be common to the treatment team and evaluation team. The evaluation team will include at least (a) a member of the evaluation staff, (b) the continuity of care person, and (c) a mental health professional who would be competent to serve as the primary therapist for the client in question. The evaluation team will have access to all the information available for the given client. The evaluation team makes judgments in regard to the following.

- a. The appropriateness of the assessment procedures and of the diagnostic conclusion or problem formulation.
- b. The adequacy of the treatment plan.
- c. The validity of the improvement indicators.
- d. The degree of need for treatment at termination (independently of the treatment team decision and the predictive criteria of of success and indicators of improvement).
- e. The degree of improvement shown.
- f. The extent to which improvement or lack of improvement is attributable to treatment or other influences.
- g. A post hoc reformulation of problem formulation and treatment plan. ("If we'd known then what we know now, . . .") for reference and future practice.

C. Integration of data

Each of the before and after indicators of improvement will be obtained directly in quantitative form or will be transformed into quantitative

form, so that the before and after differences may be evaluated as to their statistical significance by means of, e.g., t-tests. Since these are only indicators rather than genuine criteria, these results will be only suggestive rather than decisive (fortunately, since they are likely to be contradictory).

The pattern of results which would most clearly support an overall evaluation of "improved" or "successful" would be the following.

1. All the improvement indicators show a statistically significant positive change.
2. Severity ratings show a change from "needs treatment" (to some degree) to "doesn't need treatment."
3. Either all significant figures agree that improvement has taken place or else the pattern of disagreement among these judges is predicted and explained adequately.
4. The treatment plan was implemented and appropriate, as rated in the treatment review procedures.
5. Retrospective review specifies that
 - a. Assessment and treatment were appropriate.
 - b. The client improved and was not in need of treatment at termination.
 - c. The client's improvement is attributable primarily to treatment rather than to other influences.

The major likelihood, of course, is that the results will be more or less equivocal rather than conforming to the "ideal" case above. Thus, the final evaluation problem will be how to count different patterns of less than completely unequivocal results. No rigorous general principle or procedure for accomplishing this evaluation is possible. It is possible,

however, for the evaluation committee to formulate ad hoc procedures for generating composite "degree of success" indices for particular purposes. (This is in accordance with the principle of limiting use of results as a way of increasing the likelihood of agreement.) It may also be possible to specify general categories of data which would have utility across different evaluations. Such categories would be useful for summarizing the results of several evaluations or for comparing them.

D. Modifications

The general structure and procedure described above are primarily designed for the evaluation of improvement of individual clients undergoing comprehensively planned treatment for a substantial period of time and under conditions ideal for evaluation. Certain modifications would be called for under other conditions. Among the modifications of present interest are the following.

1. Conditions for evaluation are almost never ideal. Because of the amount of professional time involved in the evaluation procedure it almost certainly will not be feasible to carry out all of the procedures with every client. In that case it would be appropriate to select every nth client at random for the complete evaluation procedure. A ratio of one in five or one in ten would seem to be reasonable in this regard.
2. Some telescoping of the procedure would be required in cases where treatment was not continued for a substantial period of time (e.g., on site treatment or early termination).
3. The evaluation of the Child and Elderly programs has both qualitative and quantitative aspects. The evaluation design described above is primarily responsive to the qualitative aspect; if clients

improve, then, qualitatively, the program may be judged to be a success.

For the Child and Elderly programs the relevant quantitative question is whether the programs are reaching enough people. The specifics of this problem are found in some fictitious statistics. For example, suppose that in 1970 children under eighteen comprised 31% of the catchment area population, yet in 1976 only 3% of the MHC caseload consisted of persons under 18 years of age; and in 1970 persons over 65 comprised 11% of the catchment area population, yet in 1976 only 1% of the MHC caseload consisted of persons over 65. Although there is no conclusive evidence to indicate that the prevalence of mental health problems is as great among children as it is among adults, these figures would strongly suggest that both children and the elderly are currently underserved. Thus, one criterion for the success of these programs is that the MHC caseload should more nearly approximate the population proportions in the service area for these groups. As a quantitative indicator of success it would seem reasonable to propose that by the end of the first year of the Child and Elderly programs, the proportion of admissions should be between one-fourth and one-half of the population proportions for these groups.

