

Goal Attainment Scaling: A General Method for Evaluating Comprehensive Community Mental Health Programs

Thomas J. Kiresuk, Ph.D.

Robert E. Sherman, M.S.

ABSTRACT: A mental health enterprise may be described by either (a) rather general philosophical total mental health goals, or (b) highly diverse and individualized patient-therapist goals. Goals a. have not provided a workable framework for program evaluation. This paper proposes that evaluation be done in the framework of goals b. by setting up, before treatment, a measurable scale for each patient-therapist goal, and specifying, for each patient, a transformation of his overall goal attainment into a standardized T-score. This method, together with random assignment of patients to treatment modes, was devised to permit comparison of treatment modes within a program, but it also provides a good basis for a judgmental evaluation of the total program.

Evaluation of mental health services has become especially difficult because two major lines of development in the mental health enterprise have proven to be nearly incompatible. On the one hand, several different models of mental illness and treatment are considered useful (e.g., medical, sociological, learning theory) and a corresponding diversity of techniques are employed in a variety of settings, using a full range of professional and lay manpower. The counterpoint to this diversification is an increasing demand for program evaluation aimed at justifying the commitment of economic and manpower resources and demonstrating the effectiveness of the particular techniques utilized. A program evaluation unit, therefore, is faced with all the known conceptual and procedural difficulties that occur in research on the traditional therapies as well as an increasing number of comparable problems associated with the proliferation of nontraditional techniques of very broadly defined mental health intervention.

Mental health centers attempting to devise evaluation systems can turn to extensive bibliographies (Abt & Riess, 1966; Bloom, 1965; Dent, 1966; Evaluation in Mental Health, 1955) or recognized reviews and reports (Gruenberg, 1966; Williams & Ozarin, 1968) to find a research design and

Dr. Kiresuk and Mr. Sherman are members of the staff of the Hennepin County Mental Health Center, Hennepin County General Hospital, Minneapolis, Minnesota. The authors would like to express appreciation to William W. Jepson, M.D., Program Director of the Hennepin County Mental Health Center, for his considerable contribution to the earlier formulation of the method reported here.

method of measurement that would be applicable to their programs. Such a review, however, would show that the problems of definition and measurement have not been resolved and that each mental health center has proceeded with its own working or operational definitions and measures. These working solutions have permitted description and analysis of programs at the same time that they have helped avoid the philosophical paralysis that occurs during the search for a comprehensive research definition of mental health. The operational definitions have had an unspecified, informal, or presumptive relationship to the higher order definition, a relationship that has depended on the common sense and goodwill of fellow evaluators for its validity. While these procedures have led to the creation of a great diversity of criteria and measures, together with their rationales, they have not encouraged the development of a standard method of evaluation that would have applicability to a variety of mental health programs.

The tendency for mental health programs to develop new criteria and instruments that are difficult to extend to other programs also has existed in evaluation studies conducted on the program elements themselves. Thus, one would select from among the numerous consultation studies listed by Bindman (Abt & Riess, 1966, pp. 78-106) for evaluation of a consultation program but would probably utilize rating scales or other factor analytically derived *Q* sorts such as those described in Lyerly and Abbott (1966) in order to evaluate the outpatient psychotherapy program. Comparison between the two program elements even with regard to patients and treatment objectives that are similar would not be possible. The treatment objectives would not be formulated in the same way, and the measures chosen would be sufficiently different to prevent efficacy comparisons.

Within the program element itself, however, the measurement problem seemed to reverse itself. Instead of a large number of different techniques, at this level there has been a tendency to use a fixed battery of evaluation measures regardless of the individual patient characteristics or problems. In our outpatient therapy unit, for instance, there are patients of both sexes, of varying marital status, ages ranging from 16 to 80, education level ranging from less than eighth grade to Ph.D. candidates. The use of the same psychometric or rating device for all patients inevitably led to evaluating some patients on variables that were irrelevant to their particular dilemma or circumstances. The clinicians justifiably objected, stating that the evaluation process did not truly represent their therapy aims—"I'm not trying to cure his Multiphasic!" To rate all patients with regard to their anxiety level, sexual problems, and thought disorder, whether or not these variables had anything to do with the patient's reason for coming to the clinic, appears to be unreasonable and wasteful.

Turning from the problems of measurement to that of determination and specification of goals, it is at the patient-therapist level, or at the time of the establishment of a contract between the patient and the organization, that the aims, intentions or goals of the program element can be described

in concrete terms. Statements of general program goals that are not tangibly related to the activities of the program elements, and therefore the goals of the patients and therapists in that element, are basically platitudes or statements of good intentions. The recent federal and state encouragement of welfare and mental health organizations to relate their activities to clearly stated goals, and then demonstrate their progress toward them, has served to stimulate interest in the formal definition and measurement of goals. The central problem of mental health in this area has been to bridge the gap between the broadly stated goals of intervention and rehabilitation and the particular objectives of the activities of the program.

Of the possible ways of specifying goals of a mental health program, the use of community-wide epidemiological data as a means of justifying and evaluating a program has been considered too indirect (Carstairs in Williams & Ozarin, 1968). The number of factors influencing such demographic and public health measures is too great, and the number of these factors under the control of any given program are too few to expect significant variation in these measures to be attributable to the program effort. On the other hand, the particular goals of the patient-therapist contract within a program element may seem trivial, and often are not recorded in a systematic way.

The measurement procedure described here is a method of goal definition and goal measurement that permits both description of the total program and evaluation of the program elements. The procedure includes a means of formally and routinely specifying the actual goals undertaken by the staff of any program element. When used in the design recommended, it should be possible to determine the extent to which these goals have been attained and compare the relative effectiveness of the treatments used to attain them.

The proposed solution includes a scaling technique and a basic evaluation design common to all therapy research. There are three steps: (1) goal selection and scaling; (2) random assignment of the patient to one of the treatment modalities; and (3) a follow-up of each patient with regard to the goals and scale values chosen at intake. The following sections of this paper describe these steps further in terms particularly applicable to an outpatient mental health program.

PROCEDURE, GOAL SELECTION AND SCALING

After the screening process, but before random assignment to a treatment mode, the goal selector (or goal selection committee) decides upon a realistic set of mental health goals for the patient, and, for each goal specified, a scale composed of a graded series of likely treatment outcomes, ranging from least to most favorable. At least two points on each scale should have sufficiently precise and objective descriptions to enable an unfamiliar observer to determine whether the patient lies above or below that point. These points are then assigned numerical values, -2 for a least favorable outcome, $+2$ for a most favorable outcome, with the value zero assigned to the treatment outcome considered most likely. This scale is

TABLE 1
Sample Goal Attainment Scaling

Outcome Value	Goals				MMPI — 78
	Fear of Sex Involvement	Dependency on Mother	Decision-making	Social Functioning	
Goal Weights	20	30	20	30	10
Most unfavorable treatment outcome thought likely (-2)	Avoidant No dating No sex	Lives at home Does nothing without mother's approval	No new decisions made, still weighing same alternatives (job, vocation)	Institutionalized prison or hospital	Up at all over previous score
Less than expected success with treatment (-1)			Complains of being unable to make up mind	On probation Further arrests	Remains in double prime range
Expected level of treatment success (0)	Dating Petting	Chooses own friends, activities without checking with mother	Makes up mind on vocation, other major items	On probation No further arrests for peeping	Mid 60's T-score
More than expected success with treatment (1)	Some satisfactory intercourse	Returns to school		No contact with police, states peeping no longer a problem	
Best anticipated treatment success (2)	Regular dating Regular satisfactory intercourse Marriage	Establishes own way of life Chooses when to consult mother			40-60 T-score

thus a judgmental transformation of the treatment outcome, into, approximately, a random variable with mean zero and variance one.

These "outcome scales" are specific to individual patients and the defined points may be only indirectly related to the mental health goals. For example, (see Table 1) a patient may have a mental health goal "less dependency on mother" with a scale point "return to school." Only two scale points need be specified. However, in pre-testing the procedure it was found that three or four points were readily determined. It is important that the scale points be stated in terms of events the presence or absence of which can be easily judged by a follow-up worker who has had no contact with the clinical or therapeutic procedures of the treatment unit.

In Table 1 a variety of scale points were chosen in order to demonstrate the capacity of the scaling technique to utilize any form of objectively determinable event. The change specified in the MMPI scales is an example of a psychometric event; probation status and arrests are readily available public records; "complains of being unable to make up mind" is self-rating; dating and petting are autobiographical reports. Other more abstract concepts such as self-actualization can be included as long as verifiable stages of this process are specified in advance.

The goal selector may also specify a set of weights for the goals, reflecting his appraisal of the relative value of each goal as an indicator of successful treatment. Thus in the opinion of the goal selection committee that set up the goals for the patient referred to in Table 1, the MMPI changes were considered of low importance and received a weight of 10, while "dependency on mother" was judged to be a key psychodynamic problem area and was assigned a value of 30. Weights do not need to sum to a fixed total, e.g., weights 1, 2, 3, perform the same as weights 2, 4, 6, or 10, 20, 30. In the absence of a decided preference for certain goals, equal weighting would lose little information.

Each patient would have to have at least one goal, together with specified scale values. One goal might be sufficient for patients with clearly circumscribed problems such as an environmental dilemma, vocational or financial problem, etc. An example would be an elderly citizen with no family seeking a way to visit his friends downtown after bus service had been discontinued to his area. On the other hand, there is no limitation on the number of goals that can be chosen for any given patient. This provision follows from the desire to have this scaling process reflect the clinical realities of the treatment unit.

After the goal selection and weighting, the patient is randomly assigned to one of the treatment modes. The treatment is administered, and then, after a predetermined interval (for example, one year after assignment), the case is called to the attention of the follow-up unit. The unit then contacts the patient and reviews with him progress toward the goals specified prior to treatment. This also provides an excellent opportunity to determine whether further services are required, or whether disposition planning has

been carried out successfully, and thereby contributes in an informal way to total program evaluation. Most important, however, is the determination of the patient's position on the goal attainment scales so that a composite goal attainment score can be computed.

DERIVATION OF THE STANDARDIZED COMPOSITE GOAL ATTAINMENT SCORE

Let x_i be the numerical value for the outcome on the i th goal attainment scale. If the scale has been properly constructed, the expected value of x_i is zero, and its standard deviation is about one. Assume a patient has n scales, each with its relative weight, w_i . We now wish to take the composite score, $\sum w_i x_i$ and transform it to a standard variable with mean 50 and standard deviation 10. That is, we define the standardized goal attainment score:

$$(1) \quad T = 50 + \frac{10[\sum w_i x_i - E(\sum w_i x_i)]}{\sqrt{\text{VAR}(\sum w_i x_i)}}$$

Where the mean and variance of $\sum w_i x_i$ are yet to be determined. We have the mean immediately:

$$(2) \quad E(\sum w_i x_i) = \sum E(w_i x_i) = 0$$

So that

$$\text{VAR}(\sum w_i x_i) = E[(\sum w_i x_i)^2]$$

or

$$\begin{aligned} \text{VAR}(\sum w_i x_i) &= E\left(\sum_{i=1}^n w_i x_i \sum_{j=1}^n w_j x_j\right) \\ &= \sum_{i=1}^n w_i^2 E(x_i^2) + \sum_{i \neq j} w_i w_j E(x_i x_j) \end{aligned}$$

But

$$E(x_i^2) = \text{VAR}(x_i) = 1, \text{ and } E(x_i x_j) = \rho_{ij},$$

where ρ_{ij} denotes the correlation between x_i and x_j . So

$$(3) \quad \text{VAR}(\sum w_i x_i) = \sum_{i=1}^n w_i^2 + \sum_{i \neq j} w_i w_j \rho_{ij}.$$

Now even if the goal selector could specify the $n(n-1)/2$ different values ρ_{ij} for each patient, the consequent complexity of (3) would cripple the usefulness of the standardizing procedure. Suppose, however, that the goal selector could specify a single value ρ such that

$$(4) \quad \rho \sum_{i \neq j} w_i w_j = \sum_{i \neq j} w_i w_j \rho_{ij}.$$

ρ bears an intuitive meaning of a kind of weighted average correlation among the x_i , but need bear only the formal meaning of formula (4). Equation (3) may then be written:

$$\begin{aligned} \text{VAR} \left(\sum_{i=1}^n w_i x_i \right) &= \sum_{i=1}^n w_i^2 + \rho \sum_{i \neq j} \sum w_i w_j \\ &= (1-\rho) \sum_{i=1}^n w_i^2 + \rho \sum_{i \neq j} \sum w_i w_j + \rho \sum_{i=1}^n w_i^2 \\ &= (1-\rho) \sum_{i=1}^n w_i^2 + \rho \sum_{i=1}^n \sum_{j=1}^n w_i w_j \end{aligned}$$

and finally

$$(5) \quad \text{VAR} (\sum w_i x_i) = (1-\rho) \sum_{i=1}^n w_i^2 + \rho \left(\sum_{i=1}^n w_i \right)^2$$

a relatively useable form. Applying (5) and (2) to (1), the composite goal attainment score becomes

$$(6) \quad T = 50 + \frac{10 \sum w_i x_i}{\sqrt{(1-\rho) \sum w_i^2 + \rho (\sum w_i)^2}}$$

Effect of Unrealistic Scaling

If the various goal scales are not scaled realistically, i.e., so that the expected value and standard deviation of each x_i are close to zero and one respectively, then the mean and standard deviation of T may be considerably different from the anticipated values. This will not affect comparison of T values in a randomized comparison of treatment modes, but it does suggest caution in comparing T values longitudinally in time for a given program, or across programs using different goal scaling procedures.

Effect of Erroneous Choice of ρ

A value for ρ , the expected overall intercorrelation among goal scores, must be assumed. In most circumstances it will be sufficient to assume a value of, say, $\rho=.3$ without formal justification. A good choice of ρ results in accurate standardization of the composite goal attainment score.

If, however, ρ is badly chosen, then T will be scaled to a standard deviation other than 10. From the point of view of analysis it is of no consequence that the variable T have a standard deviation of 10, but it is desirable that all T scores be scaled to a common standard deviation. These T scores will be averaged over patients in each treatment mode, and common standard deviation is a device to give equal weighting to the observation on each patient. A mischoice of ρ , then, would result in differing weights for some of the T scores. While not desirable, this is clearly not critical. In any event, random assignment of patients to treatment modes remains an effective insurance of the validity of treatment comparisons.

DISCUSSION

This scaling procedure has been tested in consultation and outpatient program elements of the Hennepin County Mental Health Center. Through discussion and experience obtained in this preliminary study several possible areas of difficulty were encountered. The most frequently occurring objections or problems of application will be discussed here.

First of all, is the technique feasible? Will the screening clinicians be able or willing to formulate the therapy goals in the clear, objective, and quantifiable manner required? Our experience has indicated that a definite change in the reporting habits of clinicians is required. However, it also appears that the actual therapy goals occurring in our Center are frequently conceived of in terms that are nearly measurable, and that with only minor modification these goals can be converted into the objective events necessary for adequate follow-up.

Another frequent concern relates to the bias in the goal selection procedure. It is general knowledge that therapists are biased in that they have preferences for particular modes of therapy, prefer to deal with certain kinds of patients and problems, tend to conceive of their role and purpose in ways that will emphasize certain problem areas and exclude others. What kind of error does this bias introduce into the evaluation score of the therapist or program element?

The scaling technique makes such biases public, recording them in a systematic manner that permits review. The goals can be inspected by anyone interested in or responsible for program direction to determine whether they are frivolous, inappropriate or unrelated to the mental health goals of the organization. If the goals selected are too "easy" or too "hard," follow-up of these goals will indicate 100 percent success or failure across all therapy modes and therapists. Random assignment of the patient to the different therapists would tend to keep the goal selector or selection committee "honest" since all therapists would be equally likely to receive any particular case. The review of goals and their outcome would eventually lead to a more realistic choice of goals.

During the initial period of use of this procedure it is recommended that two independently operating goal selectors or committees select goals and scale points for the same patients. Follow-up of these independently derived sets of goals should indicate whether the goal attainment score is reliable or whether the score is tied only to the particular goals chosen. It is our expectation that favorable or unfavorable therapy outcome is a general phenomenon. One goal selector may perceive a patient's problems in terms of intrapsychic symptoms and psychodynamics while another may see them in terms of his relationship to others. The first goal selector might specify anxiety level, psychiatric symptoms, transference, and other psychodynamic insights as outcome events, while the other goal selector might choose reports of social interactions, management of social interaction or alterations in the

way the patient understands his group membership and participation. It is our belief that success or failure with the one set of goals would imply a comparable success or failure in the other set. This, however, is an empirical matter that will be settled through experience with the procedure.

Another concern has been that the staff will tend to select goals that it knows are possible at the expense of the patient's desires. This eventuality of course is not inherent in the technique but may occur in any setting where the goals of the organization are incompatible with the goals of the patient. Once again, this kind of incompatibility can be made public since the goals of the patient can be written or checked on a symptom and problem reporting sheet. In the case of delusions or inappropriate vocational aspirations, one would expect that the reasonable therapist would have as a goal the negotiation of a more workable method of dealing with social realities. If the organization tends to impose middle-class values where they are inappropriate, this will be evidenced in the frequency with which such goals are chosen, as compared with the social class of the patient and his stated incompatible goals. Just who would win in this kind of contest between organization and patient would be revealed in the follow-up.

A procedural question that has been frequently asked is whether or not the therapist should be aware of the goals chosen by the goal selector. In some cases, of course, where the goal selector and the therapist turn out to be the same staff member, the possibility exists that the goals would be remembered. In our particular setting, however, processing about 4,000 patients per year, this would probably be a negligible event. It is our belief that it would be best if the therapist did not know the goals selected by the committee. He would proceed in the usual manner, choosing his own goals based on his contacts with the patient and review of the same written materials available to the goal selectors. One would expect that if the goal selectors and the therapists are all well trained, reasonable professionals, the goals that would be chosen would be reasonably comparable. The separation of the two processes would minimize the possibility that the therapist would focus only on the particular goals chosen, tending to spend less effort on the more general therapeutic improvement. Even so, if we are correct in our assumption of a general therapeutic effect, improvement on a particular set of goals would imply comparable improvement on other possible choices of goals or scale values.

The assignment of numerical values indicating the relative importance of the goals chosen for any particular patient is acknowledged to be a very subjective procedure. That clinicians and therapists do in fact hold some goals to be more important than others is general knowledge. It is in keeping with our intent to develop a procedure that reflects the clinical realities of the mental health operation that such a provision for value judgments be included. The weight assignment is very flexible, accommodating any number of different goals for a given patient, and any preference of relative numerical weighting. This provision reflects the belief that success or failure

in highly critical goal areas should not be equated with success or failure in relatively trivial areas.

Another frequently occurring concern of the potential users of this procedure is that they may be unable to conceive of scale points for a goal that they believe is very important. This concern arises because of a feeling that five scale points must be specified for all goals. Actually a minimum of two is required. During the early experience with the procedure a great deal of resourcefulness will be necessary in order to invent new scales or choose from among existing measures. Preferred methods of measuring frequently chosen goals will probably accumulate over time and could be catalogued. The goal selector or committee will have to work closely with members of the program evaluation unit—the latter bringing their familiarity with a variety of measures to the goal selection procedure.

Other reservations and areas of concern regarding the application of this procedure were related to the random assignment of patients to treatment modes and therapists, and the follow-up. A full discussion of the merits and hazards of these will not be presented here. There are some practical considerations related to the use of the goal attainment scaling procedure, however. In order to assure the welfare of the patient, the random assignment to treatment modes would be pre-empted by medical priority if it is believed that a particular treatment is clearly the treatment of choice. Such medical priority cases could still have goals and scale values chosen and followed up. The resulting information would still be useful for purposes of program description, checking on therapy progress and outcome, and for a review of disposition planning and implementation.

The definition of therapy and the research problems related to therapy dropouts can be managed by arbitrary working solutions. Our preference was to define therapy as one or more visits kept by the patient following assignment to the particular therapist and therapy mode. Because of the random assignment of patients one would expect dropouts and reassignments to occur equally in all therapy modes. The analysis of goal attainment scores would not be affected unless there were an unexpectedly high dropout rate for certain therapies.

With regard to follow-up, the only special problem arising from the use of the goal attainment scaling derives from the highly particular and therefore confidential nature of the scale values. Consequently the mails and telephone cannot be used except in the effort to locate the former patient and obtain his cooperation.

The advantages of the proposed scaling procedure are many. It can be used in a wide variety of settings. The flexibility of the open grid (Table 1) avoids measuring patients on irrelevant variables, yet places no restriction on possible goals, and gives freedom to assign relative weights appropriate for each patient. In addition, results have a very direct interpretation, as compared with, say, a factor analytically derived score, or a score reflecting MMPI "improvement." For all of this, computational difficulty is small,

and the mathematical model is simple. And, of course, the resultant composite goal attainment score permits comparison of treatment outcomes for widely varying patient types and widely differing treatment modes.

Finally, the inspection of the goals actually chosen for each patient may be the only route to a full appreciation of the purposes and operational character of the entire program.

REFERENCES

- Abt, L. E., & Riess, B. F. *Progress in clinical psychology*. Vol. 7. New York: Grune and Stratton, 1966.
- Bloom, B. L. *Mental health program evaluation, 1955-1964*. NIMH, U.S. Public Health Service, 1965, mimeographed.
- Dent, J. K. *A bibliographic index of evaluation in mental health*. Bethesda, Md.: Public Health Service, National Institutes of Health, Department of Health, Education and Welfare, 1966.
- Gruenberg, E. M. (Ed.) Evaluating the effectiveness of mental health services. *Milbank Memorial Fund Quarterly*, 1966, 45.
- Lyerly, S. B., & Abbott, P. S. *Handbook of psychiatric rating scales*. Bethesda, Md.: National Institute of Mental Health, National Institutes of Health, 1966.
- Williams, R. H., & Ozarin, L. D. (Eds.) *Community mental health, an international perspective*. San Francisco: Jossey-Bass Inc., 1968.
- NIMH. *Evaluation in mental health*. Public Health Service Publication No. 413. Washington, D.C.: U.S. Government Printing Office, 1955.