Goal Attainment Scaling as a Method of Clinical Service Evaluation

Kenneth J. Ottenbacher, Anne Cusick

Key Words: assessment technics • documentation • professional practice • quality assurance, health care

The demand for clinical accountability and the documentation of therapeutic effectiveness continues to increase in health-related settings. Therapists are attempting to address this increasing demand by adapting methods based on traditional experimental models of research to evaluate their clinical practice. Experimental and quasi-experimental designs, however, are often of limited usefulness in clinical environments for a variety of practical and ethical reasons. This paper presents a method of evaluating the effectiveness of a therapeutic intervention called goal attainment scaling, which involves goal setting procedures and assessment techniques that are practice-based and practitioner-oriented. The procedures are presented and the argument made that goal attainment scaling is a viable method by which one can document therapeutic change and demonstrate clinical accountability.

Kenneth J. Ottenbacher, PhD, OTR, is Professor of Occupational Therapy and Associate Dean, School of Health Related Professions, State University of New York at Buffalo, 435 Stockton Kimball Tower, Buffalo, New York 14214.

Anne Cusick, BSc(OT), MS, is Senior Lecturer, School of Occupational Therapy, University of Sydney, Cumberland College of Health Sciences, Sydney, Australia.

This article was accepted for publication September 15, 1989.

Program evaluation was defined by Anderson and Ball (1978) as a set of procedures with which to appraise a program's merit and provide information about its goals, activities, outcomes, impact, and costs. A formal hierarchy of evaluation activity has been proposed by Rossi and Freeman (1982), who identified four types of program evaluation: (a) that which is concerned with the gathering and developing of information for program planning, (b) that which includes strategies for the monitoring of program implementation, (c) that which deals with the assessment of a program's impact or the determination of whether the program worked, and (d) that which is designed to determine program efficiency. This last type of evaluation addresses the determination of cost–benefit ratios and the evaluation of how efficiently program resources are used.

Ideally, a comprehensive evaluation of any occupational therapy service should include all four types of assessment activity: program planning, program implementation, program impact, and program efficiency. Rossi and Freeman (1982) noted, however, that for some programs, particularly those that are established and operating, not all four types of evaluation procedures are necessary or even possible. Information on program impact evaluation is the most frequent type of assessment requested by administrators, policymakers, and other groups concerned with the provision of health care services (Datta & Perloff, 1979). Therapists in a wide variety of practice environments are increasingly required to demonstrate and document the outcomes of their interventions. The need for accountability in occupational therapy was accurately described by Gillette (1982), who said that

the practice claims of the profession must be established in order to provide ample evidence of the value of occupational therapy to consumers of the service and to other health care providers as well. In the absence of careful and thorough documentation, members of a profession such as occupational therapy will not receive appropriate recognition nor adequate reimbursement for their services. (p. 499)

The program impact evaluation represents a set of procedures intended to address the concerns raised by Gillette (1982) and others (e.g., Rich, 1985). It is designed to gauge the extent to which treatment produces change in the desired direction (Rossi & Freeman, 1982). It implies a set of predetermined, specifically defined therapeutic objectives. An intervention that produces an impact is one that achieves some movement or change toward the therapeutically desirable objectives.

To conduct an impact evaluation, one must have a plan for collecting, organizing, and analyzing information. The plan or design allows the evaluator to demonstrate that changes in the anticipated direction
have occurred. Specific impact assessment methods may vary considerably, depending on the nature of the intervention and the clinical environment. One can occasionally use traditional experimental (or quasi-experimental) designs for program impact evaluation. In such designs, the investigator manipulates an independent variable and randomly assigns subjects to groups or conditions, and the outcome measure is blindly recorded. In most clinical settings, however, it is more productive to use alternative procedures rather than classic true or quasi-experimental methods. Single-subject designs are often advocated as an alternative method of program impact evaluation for individual clients (Bloom & Fischer, 1982; Kazdin, 1982; Ottenbacher, 1986). Single-subject methods, however, are not applicable to all evaluation situations or questions. For example, Kazdin (1982) observed that single-subject designs are limited in their ability to explore treatment by subject interactions within a single patient. A focus on 1 subject does not allow for a comparison of different treatments among multiple subjects who differ in various characteristics, at least within a single evaluation. Alternative models of program impact evaluation are needed to address the increasing demand for accountability and documentation encountered in clinical environments.

**Goal Attainment Scaling**

Goal attainment scaling is one method of program impact evaluation that has potential for applied fields such as occupational therapy. Goal attainment scaling, which was originally developed for the mental health field, was first described by Kiresuk and Sherman (1968) in a study of mental health practitioners. The procedure they described involved two basic steps. The first step included the construction of a plan, or goal attainment guide, for each client based on the desired or expected level of performance. The second step was to use this guide to rate the client's performance at a predetermined time after intervention. Kiresuk and Sherman developed a goal attainment guide, illustrated its use in a clinical setting, and presented a formula for computing a goal attainment score.

Since the original description, goal attainment scaling has been successfully used to evaluate program impact in psychotherapy and mental health (Greenspan & Starfstein, 1981; Grey & Moore, 1982; Lewis, Spencer, Haas, & DiVittis, 1987), education (Garwick, 1974), mental retardation (Bailey & Simeonsson, 1988), and general rehabilitation (Clark & Caudrey, 1986; Goodyear & Bitter, 1974). The method, however, is not generally known or used by occupational therapists despite its presentation in an article by Maloney, Mirrett, Brooks, and Johannes (1978). The purpose of the present paper is to describe and illustrate the use of goal attainment scaling as a method of program evaluation with relevance to the assessment of occupational therapy.

**Establishing Goals**

The basic idea of goal attainment scaling is not a new one. The practice of setting measurable goals for intervention is well established. Occupational therapy goals may, however, be vague or global, thus making the evaluation and measurement of goal attainment difficult. A therapist may state, for example, that a goal for a client who has suffered a stroke is "to improve fine motor function." This goal is both vague and global, which makes it difficult for the therapist, the client, and others to recognize improvement. Goals may also look too far into the future, with no anticipated date of accomplishment. Thus, "improve fine motor function" is a never-ending goal, because no time limit or performance parameter is identified. As a result, occupational therapy goals may be irrelevant, unmeasurable, or unattainable. Goals are also sometimes determined by one professional, with little input from the client, family, or other team members. This raises the issue of whose goal is being pursued. Finally, goals should have social and functional validity, that is, they should be established in relation to the client's home, work, and community environments.

Goal attainment scaling provides the framework for the development of program goals that are measurable; attainable; desired by all; and socially, functionally, and contextually relevant. In addition, goal attainment scaling produces a quantitative index of a client's progress that can be used to compare the performances of one client over time or to compare performances across clients in the same program.

A major component of goal attainment scaling is the establishment of the desired or expected level of a client's performance. This level of performance is then used to construct a goal attainment guide or plan. Information related to the expected level of the client's performance is best collected from multiple sources. The accuracy of this information, and hence the accuracy of the Goal Attainment Scale score, is enhanced when the data are collected from the client, family members, and other service providers.

The first criterion for a goal is that it consist of observable and recordable behavior. In addition, the goal should be time limited—it should provide the therapist, client, and family with an identified time line for expected performance. At the end of this specified time period, the scale is scored and the goals may be changed to meet a new set of expectations or a revised intervention program.

The process of establishing program impact goals...
can be accomplished in eight interrelated steps. To illustrate these steps, we have developed a Goal Attainment Scale for a participant in a program that teaches competitive skills to persons with severe developmental disabilities.

**Step 1. Identify an overall (general) objective.** In the present example, the group-home providers, the teacher, the therapist, and the client may all agree that part-time employment in a competitive (i.e., non-sheltered workshop) environment is the overall program goal.

**Step 2. Identify specific problem areas that should be addressed.** This requires the therapist to prioritize the problem areas and then reduce them to observable, reportable components. For example, problem areas for the client might include social interactions and weight reduction. The important component in this step is to identify problem areas in which a measurable indication of performance can be obtained.

**Step 3. Specify what behaviors or events will indicate improvement in each of the areas selected in Step 2.** This step provides the operational detail required to make the scale a useful instrument for the evaluation of performance. For example, if social interactions has been identified as a problem area (see Step 2), then the number of verbal contacts between the client and other participants in the training classes might be selected as one measurable indicator of improvement.

**Step 4. Determine the methodology that will be used to collect the desired information.** In this step, a plan is developed that will determine how the information will be collected, who will collect it, and in what setting the evaluation data will be gathered.

**Step 5. Select the expected level of performance.** This step, which relies on the therapist's professional judgment and realistic appraisal of the client, is both a strength and a weakness of the evaluation method. If the expected levels differ from the actual performance, then the Goal Attainment Scale score will reflect this inaccuracy. This step is based on the assumption that experienced clinical practitioners will be able to predict treatment outcomes with information from the client, family members, and other health care providers. For example, if the therapist and others judge that it is realistic for the client to lose 5 lb in 1 month, then this is a satisfactory operational criterion for the goal of weight loss. This operational criterion relies heavily on the therapist's and others' knowledge of the client, the weight-loss program, and the environment.

**Step 6. Identify the most favorable outcome, the least favorable outcome, and intermediate levels of the client's performance.** Kiresuk and Sherman (1968) suggested that the evaluator develop five levels of performance, ranging from the most favorable to the least favorable outcome. The activities identified at the five levels represent a continuum of behaviors (or events) with no gaps or obvious overlaps in behavior. These five levels of performance are used in the scoring of the Goal Attainment Scale. Again, the focus is on realistic outcomes. The most favorable outcome, for example, should reflect what the client can accomplish if everything in the program goes smoothly. In the case of weight loss, for example, the operational criterion of this goal may indicate that if all goes smoothly in the program the client could lose 15 lb during 1 month of intervention. Conversely, the least desirable outcome would be that the client gains 5 lb during 1 month of intervention.

Table 1 describes the behaviors and levels of attainment for the two problem areas identified in Step 1 (i.e., social interactions and weight loss). Each of the five levels is operationally defined and is assigned a numeric value, with 0 indicating the expected level of performance and -2 and +2 indicating the least and most favorable outcomes, respectively.

<table>
<thead>
<tr>
<th>Predicted Attainment</th>
<th>Score</th>
<th>Social Interactions</th>
<th>Weight Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most unfavorable outcome</td>
<td>-2</td>
<td>Speaks to no one except therapist during 3-hr session</td>
<td>Gains 5 lb within 1 month</td>
</tr>
<tr>
<td>Less than expected outcome</td>
<td>-1</td>
<td>Says &quot;hello&quot; or other greeting to fellow workers during 3-hr training session</td>
<td>Maintains weight over 1-month period</td>
</tr>
<tr>
<td>Expected level of outcome</td>
<td>0</td>
<td>Holds sustained, interactive conversation of 200 words (or 10 min) with one other worker during 3-hr session</td>
<td>Loses 5 lb within 1 month</td>
</tr>
<tr>
<td>Greater than expected outcome</td>
<td>+1</td>
<td>Holds interactive conversation of more than 200 words (or 10 min) with two or more workers (independently or simultaneously) during 3-hr session</td>
<td>Loses 10 lb within 1 month</td>
</tr>
<tr>
<td>Most favorable outcome likely</td>
<td>+2</td>
<td>Holds interactive conversation of 500 words (or 20 min) with three or more workers during 3-hr session</td>
<td>Loses 15 lb within 1 month</td>
</tr>
</tbody>
</table>
Step 7. Once the goal attainment guide has been completed, ascertain whether there are overlapping levels, gaps between levels, or more than one indicator in a problem area. If a client exhibits a stereotypic behavior that interferes with his or her ability to work productively, for example, then the therapist can measure progress in this area by using the amount of time spent engaging in the stereotypic behavior or the percentage of productivity. Both measures could be used, but not in the same column. The scale should also be checked to make sure that the definitions of behaviors are clear and that the instructions on how to collect data are not ambiguous.

Step 8. Ascertain the client’s current status and determine future evaluations to document progress. The selection of the time period between evaluations depends on several factors, including the type of intervention provided; the expected level of performance; and external criteria imposed by third-party payers, accrediting agencies, and others. In the example, the client could be weighed before program implementation, and this weight recorded. The therapist and others would then determine when the client would again be weighed.

Scoring the Goal Attainment Scale

The Goal Attainment Scale can be scored in several ways. The option selected, however, must be used consistently across all protocols so that a standardized comparison can be made across clients. The most commonly used scoring system, proposed by Kiresuk and Sherman (1968), involves a 5-point performance scale (see Step 6) ranging from -2 to +2, in which -2 = least favorable outcome; -1 = less than expected; 0 = expected level of performance; +1 = greater than expected; and +2 = most favorable outcome. These scores are used for each of the levels across all goals (see Table 1).

In scoring the Goal Attainment Scale, the therapist assigns relative weights to each of the goals identified for the client. There is no standard procedure for determining how each goal is weighted. The weighting, ideally, is achieved by a consensus of the client, therapist, teacher, family members, and other persons concerned with the client’s performance. Generally, the weighting simply reflects a prioritizing or ranking of the goals. If four primary goals are prioritized, the most important goal is given a weight of +4 and the least important goal a weight of +1. In the example, the goal of reducing stereotypic behaviors might be identified as very important and thus given a weight of +3, the goal of social relations might be identified as important and thus given a weight of +2, and the goal of weight loss might be considered least important and thus given a weight of +1. The weights must be determined in the goal planning stage rather than in the evaluation phase. If the goals are weighted during the evaluation, the weights might reflect the patient’s or therapist’s priorities as they look back on the outcome of the program and its strengths and weaknesses. This would introduce the possibility of systematic error or bias into the evaluation process.

After the intervention has been administered, the weights for the goals and the rating of each level of performance are used to compute a Goal Attainment Scale score. This score represents a numeric index of the client’s improvement or lack thereof. The formula used to compute the goal attainment score is

\[ T = 50 + \frac{(10 \sum W_i X_i)}{V(1 - r) \sum W_i^2 + r \sum W_i^2)} \]

where \( W_i \) represents the weighting for a particular goal and \( X_i \) represents the outcome score for each behavior (i.e., a value from -2 to +2). The \( r \) value in the formula reflects the estimated average intercorrelation for the outcome scores. Kiresuk and Sherman (1968) argued that an \( r \) value of .30 can be safely assumed and used as a constant in the formula. Finally, the \( T \) value is a standardized score with a mean of 50 and a standard deviation of 10.

The function of the formula can be demonstrated through an example. Table 2 presents information on a participant with a developmental disability who is enrolled in the program described in Step 1. Four goals were identified and weighted from +1 to +4, as shown in Table 2. The table also includes the outcome score for each of the goals obtained at the final evaluation as well as other information needed to generate the \( T \) score from the formula presented above.

When the information from Table 2 is included in the formula, the following results are obtained:

\[ T = \frac{50 + (10 \times 7)}{\sqrt{70 \times 30) + (30 \times 100)}} \]

\[ T = \frac{50 + 70}{\sqrt{51}} \]

\[ T = 50 + 9.80 \]

\[ T = 59.80 \]

Because \( T \) is a standardized score, it can be compared to a hypothetical normal distribution. A \( T \) of 50 corresponds to the 0 point on the original profile, that is, the expected level of performance ranging from -2 to +2. A \( T \) of greater than 50 reflects performance above the expected level and a \( T \) of less than 50 reflects performance below the expected level. In the example given, it means that the client’s overall performance following program involvement is above average.

The \( T \) score is a better reflection of the client’s performance than the simple raw score because it
Table 2
Data for Four Goals for a Developmentally Disabled Participant

<table>
<thead>
<tr>
<th>Goal</th>
<th>W_i</th>
<th>X_i</th>
<th>W_iX_i</th>
<th>W_i^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of productivity</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>Stereotypic behaviors</td>
<td>3</td>
<td>+2</td>
<td>+6</td>
<td>9</td>
</tr>
<tr>
<td>Social interactions</td>
<td>2</td>
<td>+1</td>
<td>+2</td>
<td>1</td>
</tr>
<tr>
<td>Weight loss</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
<td>-7</td>
<td>30</td>
</tr>
</tbody>
</table>

Note. W_i = weight. X_i = outcome. Σ = total.

combines the outcome scores for all the goals, thus providing an overall measure of the client’s improvement or lack thereof (Clark & Caudrey, 1986). The T score also allows the weighting that has been given to the individual goals to be incorporated into the final outcome.

Discussion

Advantages

The goal attainment scaling system is a flexible set of procedures for the evaluation of an individual’s or group’s performance in a variety of areas. The system is advantageous in that it is not bound to any theoretical orientation or particular type of treatment or outcome measure. In addition, goals can be individualized and are specifically designed to represent realistic expectations concerning the client’s performance. Along this line, the Goal Attainment Scale strategy actively encourages cooperative goal setting. The thoughts and opinions of clients, their families, and other health service providers are important when one is establishing and prioritizing (weighting) the goals and determining realistic levels of expected performance.

The Goal Attainment Scale provides a numeric index that reflects the client’s performance over time. Comparisons may be made across clients or with the unit normal distribution. The procedure is easily computerized so that data can be quickly and efficiently processed and summarized.

The goal attainment scaling method is also flexible; it can be used with selected individuals or with groups. It can be used in a relatively informal evaluation context or in the context of experimental design, in which the therapist randomly assigns clients to groups and carefully manipulates interventions.

Limitations

Despite the positive aspects of goal attainment scaling, it is not a panacea for the difficulties encountered in clinical documentation and accountability. The procedures may be misused or misinterpreted, and therapists using goal attainment scaling must be aware of its limitations.

The Goal Attainment Scale is not a research tool. Rather, it is a set of procedures designed to assist service providers in assessing client change. If the Goal Attainment Scale is used in a research context, then the therapist must ensure that the clients’ rights are protected and that they are assigned to groups, preferably at random, after goal setting has been completed. In a research investigation, the person who sets the goals should not be the person who provides treatment. The final follow-up assessments should also be performed independently (i.e., recorded blindly). In addition, care must be taken to ensure the reliability of the outcome measures and the integrity of the treatment. Lewis and colleagues (1987) have provided some valuable suggestions and guidelines to those interested in using goal attainment scaling procedures in a research context.

A comprehensive review of the limitations associated with goal attainment scaling was provided by Cytrynbaum, Ginath, Birdwell, and Brandt (1979). Their major criticism was that the T score generated as a final product in the process is influenced by the reliability of the predictions and assessments made by evaluators at both the initial and final evaluations. They suggested that extensive training may be necessary to achieve adequate reliability in some areas of goal setting and assessment. For example, the social interactions goal presented in the earlier example (see Table 1) may present reliability problems if the raters are not properly trained. If enough operational detail is not provided, or if only one evaluation period or only one rater is used.

The problem of reliability in the assessment of a client’s performance is not unique to goal attainment scaling. Whenever judgments concerning human performance are made, care should be taken to ensure that these judgments are as reliable and as accurate as possible. One can improve reliability by including multiple measurement periods, a defined training program for raters, and more explicit definitions or examples of the client’s performance. Procedures to determine the reliability of observations and assessments are beyond the scope of this paper but, along with methods for the enhancement of observer reliability, are widely available in the research literature on applied behavioral science (e.g., Hartman, 1982; Webb, Campbell, Schwartz, & Sechrest, 1966).

A second area of concern identified by Cytrynbaum et al. (1979) is the possibility of floor effects in the evaluation process of the Goal Attainment Scale. If the expected level of performance after intervention is assigned a 0 value on a scale ranging from −2 to +2, then one can probably safely assume that many clients will begin the program at a lower level, that is, −2.
When this happens, there may be no room for regression. The worst possible outcome for the client is that he or she remains at the -2 level during the final evaluation. Such a finding would suggest that the client's status has remained unchanged during the period of intervention. No allowance is made for clients who may have regressed during the time the program was in effect. In the example provided earlier, the client may gain 10 lb over a 1-month period, but the attainment rating scale does not permit such regression to be noted (see Table 1). Although this is not a major problem associated with the Goal Attainment Scale, it is a limitation that therapists should be aware of, particularly if they are working with a client population in which deterioration in performance is expected.

Clark and Caudrey (1986) also expressed concern that the weighing of $T$ is part of the formula itself, that is, the weighting is internal rather than external. They believed that the internal weighting of $T$ reduces its sensitivity and masks interactions between the difficulty and importance of a particular goal. Clark and Caudrey (1986) proposed that each goal be rated according to both its difficulty and its importance and that this rating be done with the same numeric rating scale. The product of this rating (Difficulty X Importance) is then used as the internal weight in the computation of the $T$ scores. Clark and Caudrey further suggested that, after the $T$ value is obtained, the evaluator externally weight the $T$ score by multiplying the $T$ value by the overall mean importance rating. They argued that this approach allows for a more sensitive comparison of performance. The disadvantage is that with an external weighting, the final $T$ value no longer represents a standardized score, because the weighting will differ for each client.

Implications and Conclusion

Despite the limitations described above, goal attainment scaling represents an improvement over the subjective and anecdotal evaluations that typify clinical assessments. Goal attainment scaling is a flexible evaluation methodology that can address the documentation and accountability concerns facing health care providers. Perhaps most important, goal attainment scaling is a method that is practice-based and practitioner-oriented, which gives it an important advantage over traditional evaluation procedures modeled on experimental and quasi-experimental designs.

Lipsey, Crosse, Dunkle, Pollard, and Stobart (1985) observed that the dominant methodological approach to program evaluation in most applied fields is based on the use of the “experimental paradigm.” They noted that the experimental approach emphasizes the quantitative measurement of dependent variables and the manipulation of independent variables to establish cause-and-effect relationships. They conducted a detailed review and analysis of program evaluation studies in several applied fields and concluded that the studies were characterized by weak designs, low statistical power, ad hoc measurements, and neglect of treatment implementation and program theory.

The lesson we draw from our analysis is that it is time to acknowledge that, despite its current wide spread use, the experimental paradigm is not an all purpose program evaluation methodology. (Lipsey et al., 1985, p. 25)

One could argue that it is unrealistic to expect clinically based occupational therapists to complete complex and time-consuming evaluation studies based on traditional experimental procedures. Clinical therapists do, however, have the opportunity and the obligation to assess clients' performance, document change, and report results. Goal attainment scaling can assist clinicians in achieving these important therapeutic objectives.

References


---

**Coming in July:**

Special Issue on the Sensory Integration and Praxis Tests

- An overview of the SIPT
- Relations between design copying and other tests of sensory integrative dysfunction
- Measuring change with the SIPT
- Praxis on verbal command and imitation
- Sensory integrative processing in adolescents

Turn to *AJOT* for the latest information on occupational therapy treatment modalities, aids and equipment, legal and social issues, education, and research.