Goal Attainment Scaling
A Responsive Measure of Client Outcomes

Dorothy A. Forbes, RN, PhD

ABSTRACT
Goal Attainment Scaling (GAS) was developed by Kiresuk and Sherman in 1968. Although highly criticized in the early years of its conception, GAS is beginning to be recognized as a reliable, valid, and responsive outcome measurement approach. This article describes the application of GAS, using elderly home health care clients as an example. Reliability and validity issues are examined. Criticism of GAS is often based on traditional psychometric assumptions associated with normative assessment and does not reflect the evaluative nature of the GAS methodology. The advantages and limitations of GAS are discussed. Goal Attainment Scaling has the potential to demonstrate the important contribution home health care programs make to clients by measuring the change that has occurred during their admission to the program. This kind of information is required by policy makers, case managers, and home health care providers to develop policy, allocate limited resources, and offer appropriate and effective services.

With the high cost of health care, interest in outcomes research has been generated at regional, national, and international levels. Health care providers are being held accountable for the efficiency and effectiveness of their interventions. However, standardized outcome instruments often fail to reflect all of the client problems addressed and ignore the individuality of the client (Grenville & Lyne, 1995; Rockwood, Joyce, & Stolec, 1997). The purpose of this article is to present a measurement approach, known as Goal Attainment Scaling (GAS), which accommodates multiple, individualized goals (Ottenbacher & Cusick, 1990). The application of GAS is described for elderly clients who are registered in a home health care program which offers a variety of services such as nursing, occupational therapy, physical therapy, social work, personal care assistance, and homemaking.

Goal Attainment Scaling was developed by Kiresuk and Sherman in 1968 at the Program Evaluation Resource Center in Minneapolis, Minnesota as a method for evaluating community mental health programs. Funding through the National Institute of Mental Health was received from 1969 to 1973 to implement and evaluate the methodology (Saltmarche, 1987). By 1973, more than 70 programs had adopted GAS or adaptations of it. However, the procedure was criticized conceptually and operationally, and reliability and validity problems were identified (Calsyn & Davidson, 1978; Cytrynbaum, Ginath, Birdwell, & Brandt, 1979; Seaberg & Gillespie, 1977). The effect of these critiques was a sharp decline in the use of GAS in evaluating the effectiveness of health care interventions. In recent years, however, GAS has received renewed attention in mental health (Bradshaw, 1993; de Beurs, Lange, Blonk, & Koele, 1993; Lloyd, 1986); family therapy (Fleuridas, Rosenthal, Leigh, & Leigh, 1990; Simeonsson, Bailey, Huntington, & Brandon, 1991); brain injury rehabilitation (Joyce, Rockwood, & Mate-Kole, 1994; Malec, Smigieliski, & DePompolo, 1991; Zweber & Malec, 1990); stroke rehabilitation (Grenville & Lyne, 1995); following orthopedic surgery (Saltmarche, 1987); and with frail, cognitively impaired older adults in nursing homes (Blair, 1991; Scherder, Bouma, Steen, 1995; Tappen, 1994) and the frail elderly in acute care settings (Rockwood, Stolee, & Fox 1993). Indeed, GAS has potential for use...
APPLICATION OF GAS

Goal Attainment Scaling has been described in detail elsewhere (Daley, 1987; Kiresuk & Sherman, 1968; Kiresuk, Smith, & Cardillo, 1994; Stolee, Rockwood, Fox, & Streiner, 1992). For the purpose of this article, GAS is briefly presented as a seven-step process to be used with elderly home health care clients.

Step 1: Selection of Goals

In the home health care setting, elderly individuals are assessed by professional case managers, who are primarily nurses. This involves completing a detailed assessment instrument and may include observing the elderly individual performing certain tasks, using standardized instruments to evaluate skill areas; assessing their environment; and identifying their support network (Law et al., 1990). The individual’s strengths and problem areas are jointly identified. The focus of treatment and the desired goals are established through negotiation and agreement with the individual. This process may take considerable time because more than one home visit may be necessary to identify the goals. However, this process helps clients and families better understand the possible outcomes of care and their roles in achieving these outcomes (Rockwood et al., 1997).

Some clients may not be able to identify their areas of difficulty because of cognitive impairment. In these situations, the informal caregiver, usually a family member, acts as the client's proxy and participates on behalf of the client. However, it is recognized there is a low level of agreement between family caregiver-client pairs although a trend to higher agreement is seen when the caregiver lives in the same household (McCusker & Stoddard, 1984; Epstein, Hall, Tognetti, Son, & Conant, 1989). The establishment of mutual goals is essential because without the client or caregiver's active participation, the expected outcomes most likely will not be achieved.

Step 2: Weighting

Once identified the relative importance of each goal may be assessed and differential numerical weights may be assigned to each goal scale. However, the author concurs with Cardillo and Smith (1994a) and does not recommend the use of differential numerical weights because the criteria for judging importance have not been specified (Clark & Caudrey, 1983; Seaberg & Gillespie, 1977). Importance may reflect:

- Availability of effective interventions.
- Social desirability responses (i.e., the goal may reflect prevailing social mores rather than client need).
- Length of time the problem has persisted.
- Importance of the goal to the clients' significant others.

Additionally, applying numerical weights to the scales increases the complexity of the goal-setting task and greatly complicates the calculation of summary GAS scores. Rockwood et al. (1997) computed a comparison between scores with differential weighting and scores with all weights set to one. No clinically important or statistically significant differences in GAS scores or in responsiveness scores were revealed. This study provides support for omitting the use of weighting.

Step 3: Follow-Up Time Selection

The follow-up interview is usually conducted at the end of the intervention period or at a prespecified time set by the client and case manager when the goals are established. For example, if the intervention is expected to be short-term, the follow up would occur at the client’s discharge from the home health care program. If the intervention is long-term, follow up would occur every 3 months.

Step 4: Title and Indicator Selection for Each Goal

The case manager selects a briefly worded label as the title. For example, in Table 1 the title, Mobility, captures the area of interest. Indicators then are selected. An indicator is a behavior, affective state, skill, or process reflecting the goal and can be used to indicate progress in meeting the goal (Smith, 1994). Examples of indicators in Table 1 are ability to transfer and walk.

Step 5: Statement of Expected Outcomes

The most realistic outcome expected following effective intervention is recorded in the middle level ("0" category) for each scale. Achievement of this result would constitute a successful outcome. Using the indicator, the goal setter specifies the nature and parameters of the expected outcome as precisely as possible. For example, in Table 1 the expected level for Mobility is "independent transfer" and "walks up to 10 meters with a walker." Because the goal in this situation is to maintain the client’s mobility, the expected outcome in 3 months’ time is the same as the client’s ability to mobilize on admission to the program.

Step 6: Completion of the Other Scale Levels

Other scale levels are identified for each goal indicating possible outcomes that are much less than expected (−2), somewhat less than expected (−1), somewhat more than expected (+1), and much more than expected (+2). The scale levels should be realistically attainable, with the extreme levels representing the outcomes that may be expected to occur in 5% to 10% of similar clients (Smith, 1994). The case man-
<table>
<thead>
<tr>
<th>Level of Attainment</th>
<th>Scale 1: Mobility</th>
<th>Scale 2: Foot Care</th>
<th>Scale 3: Relief for Spouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 (much less than expected)</td>
<td>Good bed mobility. Two-person transfer.</td>
<td>Six or more suggestions required.*</td>
<td>No arrangements made for relief.*</td>
</tr>
<tr>
<td>-1 (somewhat less than expected)</td>
<td>One-person transfer. Walks with walker up to 5 meters indoors. Uses wheelchair outdoors.</td>
<td>Four to five suggestions required.</td>
<td>1 day per week of relief.</td>
</tr>
<tr>
<td>0 (expected level)</td>
<td>Independent transfer. Walks up to 10 meters with walker.*</td>
<td>Client demonstrates foot care, and RN inspects feet, RN has two to three suggestions.</td>
<td>2 days per week of relief.*</td>
</tr>
<tr>
<td>+1 (somewhat better than expected)</td>
<td>Walks up to 15 meters with cane.†</td>
<td>One suggestion required.†</td>
<td>3 days per week of relief.</td>
</tr>
<tr>
<td>+2 (much better than expected)</td>
<td>No aid, no distance limitation.</td>
<td>No suggestions required.</td>
<td>4 or more days per week of relief.</td>
</tr>
</tbody>
</table>

* Admission status. † Follow-up status.

ager also should ensure there are no gaps or overlap among levels. To enhance reliability, the use of 5-point scales, rather than 3-point scales, with a minimum of three scales per client are recommended (Cardillo & Smith, 1994b).

**Step 7: Follow Up**

The GAS score is determined at the follow-up interview. For program evaluation purposes, the follow-up interview should not be conducted by an individual who has a personal investment in the outcome score or who has been involved in setting the goals or in providing the treatment (Smith, 1994). The interviewer assigns a value of +2, +1, 0, -1, or -2 depending on the level achieved for each goal. Then the scores for the goals are added and converted into a GAS score using a conversion table developed by Kiresuk et al. (1994) (Table 2). The elderly client assessed in Table 1 has a score on admission to the home health care program of -4. When converted to a t score from Table 2, the value was found to be 31.74. Similarly, the goal attainment score after 3 months was +2 and when converted to a t score, the value was found to be 59.13, with a goal attainment change of 27.39 between the two. The statistical significance of this difference can then be determined.

Alternatively, if the conversion tables are not available, the goal attainment score may be determined by using the formula developed by Kiresuk and Sherman (1968). If the scales are weighted equally, as is recommended, the formula is as shown in Figure 1. See Figure 2 for score determination based on the elderly client example in Table 1. These results are the same as those obtained by using the conversion key in Table 2.

A score of 50 represents an outcome at the expected level. Kiresuk (1973) reported that on the basis of the evaluation of more than 900 GAS assessments with an average of four goals, a normal distribution had been found with a mean of 50 and a standard deviation of 10.

It is important to note the t score formula as proposed by Kiresuk and Sherman (1968) represents a within-person standardization. The distribution reflected by the t score is not a characteristic being assessed against a normative standard, rather it is a change in performance or characteristic within the client (Ottenbacher & Cusick, 1993). This distinction has important implications in terms of reliability and validity.

**ISSUES OF RELIABILITY AND VALIDITY**

Critics of GAS as an outcome measurement approach have focused on problems of reliability and validity and contrasted GAS unfavorably with standardized measurement instruments (Cytrynbaum et al., 1979; Lewis, Spencer, Haas, & DiVittis, 1987; Seaberg & Gillespie, 1977). Reliability is the consistency...
\[ t = 50 + \frac{10 \sum x_i}{\sqrt{(n - np) + n^2p}} \]

Where: \( \Sigma = \text{sum, } x_i = \text{attainment score ranging from -2 to +2}, n = \text{number of scales, } p = .3 \) (which is the expected overall intercorrelation among goal scales).

Figure 1. Formula to determine GAS score (Kiresuk & Sherman, 1968).

with which an instrument measures the concept of interest, and validity is the extent to which the instrument actually reflects what it is intended to measure (Politi & Hungler, 1995). Critics have raised questions such as, “Will two case managers identify the same strengths and problems, the same areas to be addressed, and the same goals for a given client?” This is a reliability issue, but also affects the validity of GAS. Indeed, critiques of the process of GAS have indicated the validity of GAS is dependent on the predictions of the goalsetters and the results may reflect the bias of those who construct the guides as much as they measure client outcomes (Cytrynbaum et al., 1995).

These criticisms are based on traditional psychometric assumptions associated with normative (discriminative) assessment and do not similarly pertain to the evaluative nature of GAS methodology. A normative assessment is used to distinguish between individuals or groups based on whether or not predetermined characteristics exist (Cole, Finch, Gowland, & Mayo, 1994). An evaluative assessment is used to measure the magnitude of change after treatment and over time in an individual or group on a dimension of interest such as mobility (May, 1996). From the perspective of normative assessments, the construction of identical goals for all clients is an important psychometric requirement. However, from the evaluation perspective, the formation of identical goals among clients is not of primary importance. Identical goals would eliminate the ability of GAS to establish goals individualized to the client’s unique situation and needs (Ottenbacher & Cusick, 1993).

Likewise, the issue of interrater reliability in goal construction is not a relevant issue for GAS because different case managers will most likely conceptualize a client’s problems in different ways (Cardillo & Smith, 1994b). However, Stolee et al. (1992) measured the extent of agreement of two geriatricians in identifying and ranking goal areas. They found 71 of the 87 (82%) goal areas and the same primary goal in 14 of the 15 (93%) clients were identified independently by both geriatricians. Of greater importance is the measurement of the interrater reliability of the raters who evaluate the client outcomes at the follow-up interviews (Cardillo & Smith, 1994b). Training programs for the raters, adequate operational detail of the scales, multiple evaluation periods, and use of multiple raters have been found to improve the interrater reliability (Ottenbacher & Cusick, 1993). For example, when the training program for the follow-up interviewers was minimal (42 minutes), the interrater reliability was found to be .72 (Saltmarche, 1987), whereas if the follow-up interviewers received adequate formal training in interviewing and scoring, the interrater reliability scores ranged from .91 to .99 (Cardillo & Smith, 1994b; Joyce et al., 1994; Rockwood et al., 1993; Stolee et al., 1992).

Other forms of reliability include:

- Internal consistency (i.e., the degree to which the subparts of an instrument are all measuring the same attribute or dimension).
- Test-retest (i.e., the stability of the instrument over repeated administrations) (Politi & Hungler, 1995).

Neither of these forms of reliability apply to GAS because different goals for each client, that may change over time, are being measured (Saltmarche, 1987). Having extensively examined the issues of reliability, Cardillo and Smith conclude GAS scores with at least three 5-point scales per follow-up guide are “at least as reliable as scores on conventional rating scales, but much more research is needed” (1994b, p. 241). Other authors such as Joyce et al. (1994) and Stolee et al. (1992) concur.

Few studies have examined and
reported on the validity of GAS. Although content validity is usually considered in terms of the degree to which the items of an instrument adequately represent the universe of content (Polot & Hungler, 1995), the content validity of GAS has been assessed in terms of clinical relevance at the time of goal setting, content appropriate for the client, and realism of goals (Cardillo & Smith, 1994b). For example, the content validity of GAS in assessing infants with motor delays was examined by 10 physical therapists. Ten randomly selected GAS-formatted goals were rated on three dimensions:
- Importance of the goal for motor development and function.
- Whether the expected progress was achievable.
- Whether each of the four levels of change was clinically important.

Between 77% and 88% of the therapists’ ratings for each dimension met the criteria for content validity (Palisano, 1993).

Although content validity may be necessary, it is not sufficient for the validity of a test score. Essential to establishing the validity of a test is convergent validity (i.e., the degree to which two methods of measuring a construct are similar) (Polot & Hungler, 1995). Depending on the goals identified, GAS scores were found to be modestly correlated with other outcome measures such as the Brief Psychiatric Rating Scale (r = .64, p not reported) (Smith & Cardillo, 1994), the Portland Adaptability Inventory (r = .62, p > .01) (Malec et al., 1991), the Rappaport Disability Rating Scale (r = -.62, p not reported) (Joyce et al., 1994), and weakly correlated with others such as the Instrumental Activities of Daily Living (r = .16, p not reported) (Joyce et al., 1994) and the Spitzer Quality of Life Index (r = .22, p not reported) (Joyce et al., 1994). When the content of the treatment goals was better represented in the comparison outcome measures, the correlations between GAS and the other instruments improved. This demonstrates that the GAS measures some of the same features assessed by other outcome instruments.

However, GAS should not be considered a general measure of outcome or a measure of posttreatment status. Rather, GAS is a sensitive measure of treatment-induced change. The construct measured by the GAS score for elderly clients receiving home health care is the average change on a variety of variables that have occurred over a specified time frame relative to the amount of change predicted. The degree to which the instrument measures the construct under investigation is referred to as construct validity (Polot & Hungler, 1995). For example, GAS was compared with other outcome measures such as the Barthel Index, the Katz Activities of Daily Living Index, and the Spitzer Quality of Life Index to assess their sensitivity to change in the health status of frail elderly patients admitted to two geriatric medicine wards. The GAS was found to be more responsive to change than these measures commonly used in evaluating the effectiveness of specialized interventions (Rockwood et al., 1993). Similarly, GAS was found to be more responsive to change in cognitive rehabilitation patients than standard measures such as the Rappaport Disability Rating Scale, the Milwaukee Evaluation of Daily Living, the Instrumental Activities of Daily Living Scale, and the Spitzer Quality of Life Index (Rockwood et al., 1997). Indeed, GAS is able to measure change as small as the goalsetters are capable of noting (Smith & Cardillo, 1994)

### Table 2

<table>
<thead>
<tr>
<th>Sum of Scale Scores</th>
<th>Average Scale Score</th>
<th>t Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>-6</td>
<td>-2.00</td>
<td>22.62</td>
</tr>
<tr>
<td>-5</td>
<td>-1.67</td>
<td>27.18</td>
</tr>
<tr>
<td>-4</td>
<td>-1.33</td>
<td>31.74</td>
</tr>
<tr>
<td>-3</td>
<td>-1.00</td>
<td>36.31</td>
</tr>
<tr>
<td>-2</td>
<td>-.67</td>
<td>40.87</td>
</tr>
<tr>
<td>-1</td>
<td>-.33</td>
<td>45.44</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>50.00</td>
</tr>
<tr>
<td>+1</td>
<td>.33</td>
<td>54.56</td>
</tr>
<tr>
<td>+2</td>
<td>.67</td>
<td>59.13</td>
</tr>
<tr>
<td>+3</td>
<td>1.00</td>
<td>63.69</td>
</tr>
<tr>
<td>+4</td>
<td>1.33</td>
<td>68.26</td>
</tr>
<tr>
<td>+5</td>
<td>1.67</td>
<td>72.82</td>
</tr>
<tr>
<td>+6</td>
<td>2.00</td>
<td>77.38</td>
</tr>
</tbody>
</table>

**ADVANTAGES AND LIMITATIONS OF GAS**

The advantages and limitations of GAS alluded to throughout this article are highlighted in this section. As indicated above, a major advantage of GAS is its responsive-
ness as an outcome measure. It can accurately detect clinically meaningful change and detect a difference when one is present (Cole et al., 1994; Guyatt, Walter, & Norman, 1987). Research usually draws conclusions on the basis of aggregate data and statistically significant treatment effects. In contrast, case managers are concerned with individual clients, their unique situations, and with clinically significant treatment effects (Lewis et al., 1987). The GAS meets these concerns by reflecting the client's and case manager's perspectives as multiple goals are mutually set and indicators are established that measure small increments of change salient to the client and case manager.

Additionally, GAS expedites the essential process of establishing mutual goals. This process promotes several potential therapeutic benefits:

- Facilitation of the clients' problem-solving efforts.
- Increased interdisciplinary staff and client agreement and contribution.
- Establishment of realistic client and staff expectations of therapy.
- Improved clarity of treatment goals for the clients, case managers, and other professional and support staff involved with client care.
- Increased participation of clients toward meeting their goal(s).
- Increased client satisfaction.

Within home health care programs, goal attainment has traditionally been a dichotomous measure (Stoelle et al., 1992) and seldom has a time frame been specified with the goal. For example, two outcomes have been possible: the goal was either achieved or unmet. The magnitude of success or deficiency has not been determined. By defining a time frame and placing the target goal in the center of the 5-point range of possible outcomes goals, GAS enhances understanding of the meaning of the outcome measurement (Salatmarche, 1987). The numeric summary score provides a quantitative measure allowing for comparison among clients or programs with similar identified problem areas.

However, several reasons for resistance to use GAS have been identified. Case managers may view GAS as a means of evaluating their ability to accurately predict expected outcomes. Additionally, the time required to implement GAS and the extra paperwork involved have been concerns expressed. Demonstrating the clinical benefits of GAS and conducting an individualized approach to training may allow these issues to be addressed (Choate, Smith, Cardillo, & Thompson, 1981). Concern also has been raised that case managers will set goals too easy to attain. A peer review process monitoring for consistently high scores, supplemental use of standard scales, and conducting the follow-up scoring of goals by individuals who were not involved in setting the goals may ensure the goals are appropriately established (Rockwood et al., 1997).

An additional limitation is GAS's inability to compare the absolute level of adjustment or disability of clients in relation to other individuals because GAS is criteria-referenced rather than norm-referenced. The GAS conveys only one aspect of outcome. If information on comparative status is warranted, standardized instruments should be used. Indeed, several researchers have encouraged the use of standardized instruments to complement GAS (Smith & Cardillo, 1994; Studenski & Woods Duncan, 1993). However, if interest is only in the change produced during treatment rather than the level of adjustment after treatment, GAS is recommended because it provides the most accurate estimate of treatment-induced change (Palisano, 1993; Rockwood et al., 1993, 1997; Smith & Cardillo, 1994).

IMPLICATIONS FOR CASE MANAGERS

Goal Attainment Scaling is an appropriate outcome measurement approach for case managers, who are primarily nurses. It captures the uniqueness of the client situation, and the multiplicity and complexity of problems encountered in home health care. The GAS provides powerful insights into the expectations of clients and their caregivers in relation to their treatment and reveals more precise understanding of the clients' relevant experiences during their home health care admissions (Rockwood, Stoelle, Howard, & Mallery, 1996). The GAS has the potential of demonstrating the important contribution home health care programs make to clients by measuring the change that has occurred during their admission to the program. To achieve a comprehensive assessment of outcomes, GAS should be supplemented by standardized instruments that report defensible estimates of posttreatment status. Policy makers and case managers require both kinds of information to develop policy, make decisions regarding the allocation of limited resources, and enhance the appropriateness and effectiveness of home health care services.

REFERENCES


Cardillo, J.E., & Smith, A. (1994a). Psychometric issues. In T.J. Kiresuk, A. Smith, & J.E. Cardillo (Eds.), Goal attain-