Goal attainment scaling as a clinical measurement technique in communication disorders: a critical review

Ralf W. Schlosser*

Department of Speech-language Pathology and Audiology, Northeastern University, 151B Forsyth, 360 Huntington Ave., Boston, MS 02115, USA

Received 18 April 2003; received in revised form 3 September 2003; accepted 23 September 2003

Abstract

Evaluation of client progress is an important topic in communicative disorders research and clinical literature. Goal attainment scaling (GAS) is a technique for evaluating individual progress toward goals. Despite recognition of GAS as a clinical-outcome assessment technique in other clinical professions, the current debate on measuring client progress and outcome measurement in communication disorders has largely ignored GAS. The purpose of this paper is threefold: (a) to introduce GAS to the field of communication disorders, (b) to offer a critical review, and (c) to explore directions for harnessing the value of GAS for the field. In addition to the ability of GAS to evaluate individualized longitudinal change, it offers the following positive attributes: (a) grading of goal attainment, (b) comparability across goals and clients through aggregation, (c) adaptability to any International Classification of Functioning, Disability, and Health levels and domains, (d) versatility across populations and interventions, (e) linkage tied to expected outcomes, (f) facilitator of goal attainment, and (g) a focal point for team energies. The unique value of GAS could render this technique as a welcomed addition to the present set of options available to clinicians interested in assessing progress and evaluating change. Reliability and validity of GAS will be discussed. Finally, directions for harnessing the potential of GAS for communication disorders are offered for clinical practice and clinical-outcome research.

Learning outcomes: (1) As a result of this activity, the participant will be able to delineate the steps involved in GAS. (2) As a result of this activity, the participant will be able to describe the positive attributes of GAS as a method for assessing client progress. (3) As a result of this activity, the participant will be able to identify issues that enhance the reliability and validity of GAS.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Goal attainment scaling; Client progress; Outcome measurement

*Tel.: +1-617-373-3785; fax: +1-617-373-2239.
E-mail address: r.schlosser@neu.edu (R.W. Schlosser).
Goal attainment scaling (GAS) is a technique for evaluating individual progress toward goals (Kiresuk & Sherman, 1968; Kiresuk, Smith, & Cardillo, 1994), and is a recognized form of clinical-outcome assessment technique in other clinical professions such as physical therapy, geriatrics, rehabilitation, early intervention, and mental health (Bailey & Simeonsson, 1988; Lewis, Spencer, Haas, & DiVittis, 1987; Rockwood, Graham, & Fay, 2002; Simeonsson, Bailey, Huntington, & Brandon, 1991; Stephens & Haley, 1991; Stolee, Stadnyk, Myers, & Rockwood, 1999). The current repertoire of methods for evaluating individual progress in communication disorders neither includes GAS nor does it exclude GAS based on a documented rationale (Flower, 1984; Golper, 1998; Hegde & Davis, 1999). Similarly, the current debate concerning outcomes measurement (e.g., Frattali, 1999) has not included the potential use of GAS as part of a repertoire of options. Thus, the threefold purpose of this paper are to introduce GAS to the field of communication disorders, to provide a critical review of the strengths and weaknesses of GAS, and to explore directions for harnessing the value of GAS for communication disorders.

1. The process

The process of using GAS is described in general terms before providing a specific illustration from the clinical practice of augmentative and alternative communication (AAC). The purpose of this illustration is to provide a necessary context for the subsequent critical review of the technique. GAS involves the following steps (Cardillo & Choate, 1994; Kiresuk & Sherman, 1968): (a) specify a set of goals; (b) assign a weight for each goal according to priority; (c) specify a continuum of possible outcomes (worst expected outcome (−2), less than expected outcome (−1), expected outcome (0), more than expected outcome (+1), and best expected outcome (+2)); (c) specify the criteria for scoring at each level; (d) determine current or initial performance; (e) intervene for a specified period; (f) determine performance attained on each objective; and (g) evaluate extent of attainment.

Specifying the continuum of possible outcomes is considered the most difficult task by most. Smith (1994) offers several considerations for the reference points of these scale levels. In determining the expected outcome (0) the clinician must make an accurate prediction of the status of the client at the end of treatment or for a pre-specified time of treatment application, based on the assumption that the treatment will be successful. The expected level of outcome should be what the clinician truly believes what would be clinically meaningful and what this client will most likely achieve at time of follow-up. After the expected level is set, the two mid-levels (i.e., more than expected outcome (+1), less than expected outcome (−1)) are determined. These outcomes are more or less likely to occur for this client. Still, they must be realistically attainable by this client. Finally, the extreme levels are set. These outcomes are much more (+2) and the much less (−2) favorable outcomes that can be realistically envisioned for a given client. Smith (1994) suggests that these extreme levels might be expected to occur in 5–10% of similar clients.

What constitutes evidence that a goal is attained? First of all, a goal can be attained at any of the following levels: expected, more than expected, and the best expected. GAS methodology mandates that the criteria for scoring at each of the five levels of the scale must be pre-specified rather than specified at the time of follow-up. The criteria may
include evidence from a variety of sources, including direct data from observations or indirect data such as interviews. At follow-up, performance will be evaluated against the required evidence. When should progress toward attainment be evaluated? In GAS applications this evaluation of progress is typically referred to as “follow-up,” which occurs after a predefined period of intervention. According to Smith (1994), the timing should be informed by the purpose of the application, the nature of the client’s problem, and the intervention provided. Depending on the purpose of evaluation or the outcomes attained, follow-up may take on the function of an interim-assessment observation (i.e., need for monitoring changes on a periodic basis), or an observation at termination of treatment. Treatment should be terminated after completion of a pre-specified maximum number of sessions or attainment of at least the expected level of performance. In clinical practice, should a client attain the expected level prior to the pre-specified maximum number of sessions, changes can be made to the follow-up guide by increasing the difficulty level or replacing the goal with another that fits with the treatment agenda. In clinical research, however, the follow-up date and measure must be strictly adhered to in order to perform treatment comparisons. In this way, “plus two” treatment outcomes, for example, can be compared among therapists or treatments, regardless of the rate of improvement.

The extent to which goals are attained is evaluated through visual analysis or through statistical analysis. This involves a comparison of the initial level of performance (i.e., baseline) with the attained level of performance. Visual analysis allows one to compare the initial performance with the level of attainment per goal and across goals for a client. Typically, however, follow-up assessment of goal attainment involves the generation of some kind of summary GAS score (Cardillo, 1994a). This summary score is then converted in to a standardized T-score (mean = 50, S.D. = 10) (or a weighted percentage improvement score) via the formula displayed in Table 1. When all goals are equally important, each \(w_i\) equals 1 and the equation simplifies accordingly. Table 1 offers sample computations for a mean scale score of 0 and a mean scale score of +2. The \(P\) value reflects the estimated average intercorrelation of attainment scores; Kiresuk and Sherman (1968) argued that it can be safely assumed that this value is 0.30 and safely used as a constant in this formula. MacKay, Somerville, and Lundie (1996), on the other hand, reasoned that the value of \(P\) could not be presumed based on intuition, but rather should be computed retrospectively on a case-by-case basis by adjusting it to achieve a desired range of scores. Alternatively, Cardillo and Smith (1994a) suggest calculating \(P\) by reserving one goal column (i.e., a column in the goal matrix) for the same type of goal. For example, in a school setting goal 1 might be reserved for targets in language, goal 2 for math, goal 3 for social behavior, etc.

Instead of using the formula displayed in Table 1 one can also use especially prepared tables that simplify the process tremendously (Cardillo, 1994b). Tables are provided for follow-up guides with one goal to follow-up guides up to eight goals. This works as follows: If a rater determined that performance for a client was \(-1\) for goal 1, 0 for goal 2, and \(+2\) for goal 3, then the sum of the individual scale scores would be: \[\text{sum} = (-1) + (0) + (+2) = +1.\] Using the appropriate table for three goals on a follow-up guide, this summed scale score is then located in the left hand column of the table, which leads to the converted T-score of 54.56 in the right hand column. These T-scores can be aggregated across individuals.

How are these T-scores to be interpreted? According to Sherman (1994), the mean of a series of T-scores would be expected to converge (more or less) to 50 as the size of the
series gets large with a standard deviation of 10. These assumptions have been confirmed by several analyses of GAS data. Sherman (1977) and Jacobs and Cytrynbaum (1977) reported actual mean T-scores of 51.8 and 47.1 and standard deviations of 11.4 and 9.9 for 698 GAS T-scores and scores on 113 clients, respectively. Cardillo and Smith (1994a) also noted that the distribution of the mean T-scores approaches normality. Returning to the earlier example with three goals, as the sum of the individual scale scores increases so does the T-score (see tables provided by Cardillo, 1994b). For example, if the sum were +2, resulting from individual scores of +2, −1, and +1, respectively,

\[ T = 50 + 4.56(+2); \ T = 59.12 \]

Example 2. When the attainment scores for three goals are +2, −1, and +1, respectively,

\[ x_1 = +2; \quad x_2 = −1; \quad x_3 = +1 \]

\[ T = 50 + 4.56(+2); \ T = 59.12 \]

2. An illustration

To offer the reader the necessary background for the subsequent critical review of the literature, an example from AAC shall be presented involving Josh as described by Jorgensen (1994). Josh is a 10-year-old boy who is included in a 5th grade class. A session
with Josh’s mother and sister, his current and future teachers, and selected peers, using the MacGill Action Planning System (M.A.P.S.) (Vandercook, York, & Forest, 1989), revealed some of the following about Josh. He uses facial expressions to communicate many of his needs, wants, and feelings. To be successfully included he needs a teacher who likes him and treats him like the other children, and he requires a communication system that helps people know more about what he is thinking, wanting, and feeling. A set of annual goals and short-term objectives was then developed for Josh that were consistent with what he needs to be included (Jorgensen, 1994).

To specify a set of goals for this illustration, some of these annual goals and short-term objectives developed for student Josh, were used (Table 2) to develop Josh’s goals. That is, a team consisting of Josh’s mother, his regular education teacher, the speech-language pathologist, derived the goals together through a consensus building process. All of the goals and the graded levels of attainment, as specified for each goal, were eventually entered into what is called a GAS follow-up guide (Kiresuk et al., 1994) (see Table 3 for an example). Subsequently, weights were assigned to the respective goals according to priority (Table 3). Processes described by Jorgensen (1994) such as M.A.P.S. (Vandercook et al., 1989), Personal Futures Planning (Mount, 1987), and Choosing Outcomes and Accommodations for Children (Giangreco, Cloninger, & Iverson, 1998) may be helpful in deriving the weights for each goal. In addition, the reader may also find it beneficial to first list those parameters, which are deemed variable and potentially pertinent as indicators of attainment levels. For the first goal one may vary the number of settings in which conversational turns should be performed along with how many times the skill should be demonstrated. For the second goal one may vary the number of settings, the intrusiveness of prompts used, and the latency for making a choice. The number of choice items available at a given time could be varied also, even though it may be more appropriate for Josh to keep his choices manageable with only three items. For the third goal, one may vary the accuracy to be attained, the number of materials, and the intrusiveness of prompts.

<table>
<thead>
<tr>
<th>Table 2 Goals and objectives for Josh</th>
</tr>
</thead>
</table>
| 1. Goal: When talking with a responsive peer, Josh will take three conversational turns using his communication book or natural gestures.  
**Objective:** During informal conversation or buddy-reading in the classroom, Josh will maintain his interest in a book or a picture from home by answering at least three questions posed by a peer. The SLP will interview Josh’s buddies and observe in the classroom once a week in different settings to evaluate progress.  
2. Goal: Within real-life situations such as choosing food in the lunchroom, selecting books in the library or picking friends to be on his team in gym, Josh will make a choice among three offerings.  
**Objective:** When presented with a natural cue and a gestural prompt across various settings (e.g., the server asking him what he wants for hot lunch and if he doesn’t choose, the server will point to the various choices). Josh will point to a choice (from among three) within the time limit given other students. This will be measured by observations across at least three settings and by interviewing peers and teachers.  
3. Goal: Josh will demonstrate an understanding of one-to-one correspondence by passing out materials such as books, milk cartons, or art supplies to peers in class.  
**Objective:** When accompanied by a peer who cues Josh by counting, “One, two, three, four,” Josh will pass out materials to peers in class, with accuracy up to 10. This objective will be evaluated by interviewing peers and observing Josh in class. |
Next, the levels of attainment were decided (Table 3). The expected level is best set first followed by the two mid-levels (i.e., more than expected outcome, less than expected outcome) and the extreme levels (i.e., best expected outcome, worst expected outcome). It makes sense to set the expected level first because it represents the center-point on the range of possible outcomes. Once the expected level is determined it becomes possible to envision outcomes that are somewhat better and somewhat worse than the expected level, followed by the best expected outcome and the worst expected outcome (Smith, 1994).

The next step involves the determination of current or initial performance (I = initial performance; see Table 3). For the goals presented here, as it would be for most goals, the initial performance will be at the level of the “worst expected outcome.” Exceptions are goals that target the maintenance rather than the acquisition of skills as might be the case with individuals who are presented with progressive disorders such as muscular dystrophy. Here, it may be best to set the initial performance at −1 rather than −2 in order to leave some room for a worsening in performance as might be predicted in individuals with progressive disorders.

Now, the team is ready to intervene for a specified period of time, which should be individualized for each goal. According to Smith (1994), the timing has to do with the

### Table 3

Example of Josh’s goals along with goal attainment levels displayed in a GAS follow-up guide

<table>
<thead>
<tr>
<th>Goal attainment levels</th>
<th>Goal 1: conversational turns (W = 1)</th>
<th>Goal 2: choice-making (W = 1)</th>
<th>Goal 3: one-to-one correspondence (W = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best expected outcome (+2)</td>
<td>Answers at least four questions during buddy-reading and informal conversation</td>
<td>Choose from among three choices within a typical time frame after natural prompts in 4 or more settings</td>
<td>Show correspondence with accuracy of at least 15 while handing out 3 materials based on gestural cues</td>
</tr>
<tr>
<td>More than expected outcome (+1)</td>
<td>Answers three questions during buddy-reading and informal conversation</td>
<td>Choose from among three choices within a typical time frame after natural prompts in 3 settings (A)</td>
<td>Show correspondence with accuracy up to 15 while handing out 3 materials based on spoken cues</td>
</tr>
<tr>
<td>Expected outcome (0)</td>
<td>Answers three questions during buddy-reading or informal conversation (A)</td>
<td>Choose from among three choices within a typical time frame after natural and gestural prompts in 3 settings</td>
<td>Show correspondence with accuracy up to 10 while handing out 3 materials based on spoken cues</td>
</tr>
<tr>
<td>Less than expected outcome (−1)</td>
<td>Answers two questions during buddy-reading or informal conversation</td>
<td>Choose from among three choices within a typical time frame after natural and gestural prompts in 2 or fewer settings</td>
<td>Show correspondence with accuracy up to 5 while handing out 3 materials based on spoken cues (A)</td>
</tr>
<tr>
<td>Worst expected outcome (−2)</td>
<td>Answers one or less questions during buddy-reading or informal conversation (I)</td>
<td>Choose from among three choices using more time than typical after modeling and/or physical prompts in 2 or fewer settings (I)</td>
<td>Show correspondence with accuracy up to 5 while handing out 2 or less materials based on spoken cues (I)</td>
</tr>
</tbody>
</table>

purpose of the evaluation of measurement, not with the technique of GAS itself. Hypothetically, we will intervene for 6, 5, and 3 months for goals 1, 2, and 3, respectively. Once the respective times have elapsed the goals will be revisited to evaluate the level of attainment (see Table 3). For a given individual the extent of attainment can be determined visually by examining scale levels for each of the goals. Alternatively, one could calculate a summed scale score for Josh by adding the individual scale scores together: \[ \text{sum} = (0) + (+1) + (-1) = 0. \] This summed scale score is then converted into a T-score of 50. An aggregation of attainment across individuals (as it may be the case for an entire AAC service) is possible as well (see the Section 3.2).

3. Positive attributes

Although the primary strength of GAS is the ability to evaluate individualized longitudinal change (Kiresuk & Sherman, 1968; Ottenbacher & Cusick, 1993), it has several other positive attributes. GAS offers (a) grading of goal attainment, (b) comparability across goals and clients through aggregation, (c) adaptability to any International Classification of Functioning, Disability, and Health (ICF) levels and domains, (d) versatility across populations, interventions, and fields, (e) linkage tied to expected outcomes, (f) facilitator of goal attainment, and (g) a focal point for team energies. Each of these positive attributes is described and substantiated with literature in detail in the following section. While some of these attributes apply to other forms of outcome measurement others are unique to GAS.

3.1. Grading of goal attainment

In clinical practice, the process of goal-setting has been an intricate part of what clinicians do. In order to determine whether a client improved, clinicians often rely on measures such as the “percentage of objectives attained.” One of the problems of using the percentage of objectives attained is that it does not account for goals that are partially attained, or exceeded. Yet, both scenarios are commonly found in clinical practice. GAS can account for both of these situations because it codifies grades of goal attainment.

3.2. Comparability across goals and individuals through aggregation

Another strength of GAS is that it permits legitimate comparisons across goals and individuals. The “percentage of objectives attained,” for example, does not lend itself to aggregation. Aggregation is crucial in clinical practice because outcomes can be evaluated for a given client across a variety of goals. The attainment levels for each of Josh’s goals (see Table 3), for example, were aggregated to get an overall sense of outcomes for Josh. Administrators may also be interested in aggregating GAS scores across individuals to evaluate their entire program or service. The AAC service that worked with Josh and many other clients, for example, may wish to know the overall attainment of goals across their clients. Alternatively, administrators may wish to assess outcomes across different
clinicians, or the responsiveness of different groups of clients to a given intervention or different interventions (Smith, 1994). For program evaluation, GAS has something to offer that standardized measures of status such as the Functional Assessment of Communication Skills (Frattali, Thompson, Holland, Wohl, & Ferketic, 1995) cannot. Smith (1994) points out that a client can change a great deal over the course of a treatment but still have a low level of functioning at the end due to a very poor initial status. On the other hand, a good post-treatment status may be due to high initial status, rather than indexing any accomplishment of the program. Because GAS is a measure of degree of change, Smith (1994) asserts that GAS can provide essential information in program evaluation.

What is the underlying logic that makes aggregation possible? Aggregation is possible because the scale levels of −2 to +2 assume the same attainment level across goals and individuals even though the criteria for scoring at a given level would differ for each goal and individual (Cardillo & Smith, 1994a). That is, −2 is always defined as the “worst expected outcome,” −1 is always defined as the “less than expected outcome,” 0 is always defined as the “expected outcome,” +1 is always defined as the “more than expected outcome,” and +2 is always defined as the “best expected outcome.” This assumption is based on a theoretical underpinning of GAS, and to be valid, requires that the −2 to +2 intervals of GAS are applied exclusively as prescribed. If applied as intended, the difference between the “more than expected outcome” and the “best expected outcome” should have the same meaning regardless of differences in individuals and goal content (Cardillo & Smith, 1994a). Aggregation is not advisable under all circumstances of goal-setting or goal-construction. To make comparisons across goals and individuals fair and equitable it is important to consider the manner in which the goals were set. It would not be fair, for example, to aggregate scores from individuals whose goals were clinician-constructed with those who were client-developed or those that were negotiated between client and clinician (Smith, 1994).

3.3. Adaptability to any ICF levels and domains

Although this applies to some other outcomes measurement techniques as well, GAS is flexible and adaptable to virtually any of the levels specified by ICF and any domain within these levels (World Health Organization, 2001). That is to say, GAS may be used to evaluate attainment of goals at the levels of impairment, activity limitation, and participation restriction, and domains within each of these levels such as voice and speech functions, community activities, and community participation. Such flexibility is important to the provision of AAC interventions, for example, where domains targeted may be drawn from any of the four aspects of communicative competence (i.e., operational competence, strategic competence, linguistic competence, and social competence) (Light & Binger, 1998), and from other areas such as participation, quality of life, and self-determination (Calculator & Jorgensen, 1994).

3.4. Versatility across populations, interventions, and subfields

GAS is also flexible in terms of being applicable to any population (Smith, 1994). GAS has been applied in various fields such as intellectual disabilities (Bailey &
Simeonsson, 1988), cognitive rehabilitation (Rockwood, Joyce, & Stolee, 1997), postacute brain injury rehabilitation (Malec, 1999), health promotion (Becker, Stuifbergen, Rogers, & Timmerman, 2000), early intervention (Simeonsson et al., 1991), geriatrics (Rockwood, 1995; Rockwood, Stolee, & Fox, 1993; Rockwood, Stolee, Howard, & Mallery, 1996; Stolee et al., 1999), and physical therapy and occupational therapy (Stephens & Haley, 1991). This attribute is important in AAC, for example, because clients with a variety of conditions such as developmental disabilities, acquired disabilities, progressive disorders, and temporary conditions are receiving services and interventions (Kangas & Lloyd, 1998). Further, GAS has the ability to express change brought about by any form of intervention (Smith, 1994). Because of this versatility across interventions and populations, it seems reasonable to hypothesize that GAS may be applicable to goals covering the range of populations or areas of clinical practice in communication disorders, including aphasiology, language disorders, stuttering, and so forth.

3.5. Linkage to expected outcomes

Another potential advantage of GAS is that the interpretation of progress occurs relative to expected outcomes because clinicians are used to viewing goals as expected outcomes. That is to say, the desired outcomes undoubtedly enter the definition of the scale attainment levels for each goal (Bailey & Simeonsson, 1988). Along a similar vein, Bailey and Simeonsson (1988) argued that GAS might provide a framework to examine a team’s ability to project client outcomes. In other words, should a team’s expectation of performance that is indicative of the “expected level” not be met over and over again, this team may use this information to revise their goal-writing accordingly. As such, GAS may be useful in program evaluation efforts. How successfully GAS is used in this function is yet to be examined.

3.6. Facilitator of goal attainment

Smith (1994) hypothesized that the goal-setting process through GAS itself may have a positive impact on goal attainment. Burgee (1996) examined this hypothesis. Specifically, her study was conducted to examine the effects on consultation outcome of incorporating GAS into an existing teacher-support consultation team model. GAS had a facilitative effect on teachers’ and consultants’ integrity with regard to monitoring/documenting student outcome, as well as with defining problems in behavioral terms and setting relevant student goals. Social validation data from interviews with the team members revealed that the consultants found GAS useful and beneficial and would continue using it in the future. In a related study, Parilis (1996) investigated the effects of GAS by comparing it to the typical setting of challenging goals in beginning college students. It was hypothesized that the use of GAS would improve the students’ attainment of goals in self-efficacy, motivation, and performance. Results supported this hypothesis, particularly for proximal goals rather than distal goals. In summary, these studies from other fields seem to suggest that GAS is not only a technique for measuring outcomes but perhaps also a facilitator of goal attainment. Whether or not this bears generality to treatments in communication disorders as well remains to be examined.
3.7. Focus of team energies

One of the possible reasons for this facilitative effect may be due to GAS serving a helping role in focusing the teams energies, which leads to the next attribute of GAS. Smith (1994) has postulated that clearly specified goals generated by a client and her treatment system can mobilize staff energies into a more coherent pursuit of relevant and feasible outcomes. This proposed underlying rationale is appealing yet difficult to measure. Perhaps, subjective evaluation data documenting the experiences of teams using GAS might help support this proposed underlying rationale (e.g., Evans, Oakey, Almdahl, & Davoren, 1999).

4. Potential issues and solutions

There are several issues that may need to be taken into account before using the technique, including: (a) validity, (b) reliability, (c) sensitivity, (d) goal scaling, (e) computation and statistical analyses, and (f) rater selection. These issues are not necessarily unique to GAS and may apply to other clinical-outcome measurement protocols. Each of these potential issues will be reviewed along with arguments and proposed solutions to these issues.

4.1. Validity

Although various studies have demonstrated that the procedure can be valid in terms of content, construct, and socially, these types of validity need to be established on a case-by-case basis.

4.1.1. Content validity

To date, a few studies have examined the content validity of GAS. Content validity addresses the question whether the content of the measure (i.e., here GAS) covers all aspects or elements of the attribute of interest being measured (Portney & Watkins, 2000). To demonstrate content validity there should be relative congruence between GAS content and expert opinion and/or a direct relation between the measure and theory (Tickle-Degnen, 2002). Stolee et al. (1999) examined the descriptive content validity of using GAS with a large group of geriatric patients through a content analysis of identified goal areas. Because the domains identified appeared to be relevant to those that would be identified in a comprehensive geriatric assessment according to experts, content validity was considered established for this study. In a study by Shefler, Canetti, and Wiseman (2001), descriptive content validity was demonstrated by comparing the similarity of GAS scales with the Target Complaints Scale as one of the standard measures; a major portion of the complaints indicated by the patients were reflected in the goal scales. Given that the specific tasks would necessarily differ across disciplines, disorders, and perhaps clinicians, it is important that the content validity be assessed on a case-by-case basis.

4.1.2. Construct validity and criterion validity

Construct validity addresses the question whether the score or conclusions drawn from GAS relate more to the validated measures of the same attribute than to validated measures
of a different attribute. An “adequate” degree of construct validity is achieved when there are larger correlations between similar measures compared to smaller correlation between different measures (Portney & Watkins, 2000; Tickle-Degnen, 2002). Criterion validity addresses the question whether the score or conclusions drawn from GAS relate to a score or conclusions drawn from a valid criterion? An “adequate” degree of criterion validity is demonstrated when there is a correlation between GAS and a theoretically linked measure or \( r \) greater than 0.60 (Tickle-Degnen, 2002).

The construct validity and criterion validity of GAS were evaluated via correlations with the following standardized outcome measures: Barthel Index, Brief Symptom Inventory, Health—Sickness Rating Scale, Independent Living Scale, Mayo-Portland Adaptability Inventory, Nottingham Health Profile, OARS Index of Instrumental Activities of Daily Living, Standardized Folstein Mini-Mental State Examination, Target Complaints Scale, and the Vocational Outcome Scale (Malec, 1999; Shefler et al., 2001; Stolee et al., 1999). GAS was hypothesized to correlate strongly with standardized measures that address clinically relevant domains, which are similar to the goal areas identified in the GAS follow-up guides. GAS was shown to correlate strongly with other measures that showed change, and it discriminated between lower and higher functional or QOL status. GAS scores, however, correlated poorly with the Nottingham Health Profile. Stolee et al. (1999) argued that GAS may or may not be a suitable QOL measure, depending on whether the GAS goals represent QOL goals. This suggestion, however, needs to be viewed cautiously. Ottenbacher and Cusick (1993) cautioned of the expectation that GAS scores should correlate highly with standardized functional status measures by pointing out the different purposes for which they were designed. Functional status measures were developed to determine the status of clients relative to a particular trait of interest such as activities of daily living or motor function; GAS, on the other hand, is a set of procedures designed to evaluate change not status. The study by Stolee et al. (1999) evaluating geriatric interventions with GAS and with the OARS Index of Instrumental Activities of Daily Living is based on the assumption that the two assessment methods are testing the same construct; that is, Activities of Daily Living. It needs to be acknowledged that the ability of GAS to test a construct, such as Activities of Daily Living will vary according to the selection of individual items and their relative weightings. Functional status measures, on the other hand, are fixed in terms of the scoring items and cannot vary across individuals. Thus, this author concurs with Ottenbacher and Cusick (1993) that the correlations between GAS (which are based on individualized goals that have been uniquely weighted) and standardized tests (using the same items across subjects) are generally expected to be low.

Low correlations between GAS and standardized measures have been found in a number of investigations across a variety of disciplines such as early intervention (Simeonnson et al., 1991), physical therapy (Stephens & Haley, 1991), and geriatrics (Stolee et al., 1999). Nonetheless, exceptions are possible even though they should not be expected. Shefler et al. (2001) found moderate to high correlations between GAS scores and the Health-Sickness Rating Scale (\( r = 0.70, P < 0.001 \)), the Target Complaints Scale (\( r = 0.50, P < 0.01 \)), the Brief Symptom Inventory (\( r = 0.38, P < 0.05 \)) and the Rosenberg Self-Esteem Scale (\( r = 0.34, P < 0.05 \)). Malec (1999) found moderate correlations between GAS \( T \)-scores and other outcome measures (e.g., Independent Living Scale, Mayo-Portland Adaptability Inventory—Sickness Rating Scale, Target Complaints Scale, and the Vocational Outcome Scale).
Inventory, and Vocational Outcome Scale) and noted that these correlations were similar to those among the functional outcome measures themselves. In explaining these correlations, Malec (1999) suggested that “…individual goals in rehabilitation tend to be less idiosyncratic and to relate to broader achievements of societal value” than other areas such as mental health (p. 270).

In summary, the construct validity and criterion validity of GAS are best tested on a project-by-project basis. Rather than relying on correlations with relevant standardized measures, which are expected to be relatively low, it is advisable to explore criterion validity through correlations of GAS scores with other measures of individual longitudinal change such as data obtained from single-subject experiments (Ottenbacher & Cusick, 1993). If researchers do rely on correlations with standardized measures to demonstrate construct validity it is important to select some measures that target the same or very similar attributes than those in the GAS follow-up guide and some measures that resemble this content less well.

4.1.3. Social validity

Perhaps the most important strength of GAS’s validity, however, rests with its apparent social validity. Social validity has been defined as the social significance and acceptability of goals, methods, and outcomes (e.g., Foster & Mash, 1999; Schlosser, 1999). This process may be implemented through the methods of subjective evaluation, which refers to the soliciting of opinions of persons who have a special position due to their expertise or their relationship to the client (Wolf, 1978). The social validity of GAS as a method has been demonstrated through subjective evaluations by clinicians. That is to say, GAS appears to be enthusiastically accepted by clinicians as a method (Bailey & Simeonsson, 1988; Lewis et al., 1987). The reason for this acceptance may be that it “…reproduces the typical clinician’s thinking as he or she judges actual outcome against what he or she would have predicted at the time treatment was started” (Lewis et al., 1987, p. 408).

4.2. Reliability

Information concerning the reliability of GAS is scarce, but what is available is encouraging. Stolee et al. (1999) determined inter-rater agreement on 61 GAS follow-up scores completed by a multidisciplinary team and an independent nurse involved in the patients’ care. Results indicated an intra-class correlation coefficient of 0.93. A secondary analysis, using the 255 individual goal scales as the unit of observation, found an intra-class coefficient of 0.89. These positive findings were attributed to the use of clear, objective, and measurable indicators for the goal scales. Shefler et al. (2001) determined inter-rater agreement on GAS scores prior to therapy, at follow-up, and after termination of therapy for 33 patients receiving psychotherapy. Mean inter-rater agreement scores between pairs of judges was $r = 0.88$. In their critical review of earlier work, Cytrynbaum, Birdwell, Birdwell, and Brandt (1979) found inter-rater agreement on GAS scores ranging from 0.51 to 0.95. GAS applications in rehabilitation, as reviewed by Malec (1999), have reported excellent inter-rater agreements with inter-class correlations of 0.90 or above. These results are indicative of the reliability of arriving at GAS.
scores. To be sure, however, it is best to assess the reliability of GAS scores on a case-by-case basis.

Critics of GAS have rightfully noted that the existing reliability data primarily speak to the accuracy of deriving at GAS scores, but not the reliability of the process of constructing the follow-up guides themselves (Cytrynbaum et al., 1979; Seaberg & Gillespie, 1977). One of the concerns with GAS is that the results may reflect the knowledge and skill (or lack thereof) and/or bias of those who construct the follow-up guides as much as they reflect client outcome (Cytrynbaum et al., 1979; Mintz & Kiesler, 1982; Simeonsson et al., 1991). Shefler et al. (2001) examined the similarity of GAS scales constructed by a pair of judges for the same client. Independent raters evaluated the degree of similarity on a scale ranging from 1 (no similarity) to 7 (scale content is identical). There was a relatively good level of agreement between judges in identifying content, ranging from a mean of 5.41–3.62. Kiresuk (1973) reported an agreement score of $r = 0.88$ between follow-up scales constructed for the same client by two different teams. These data are quite positive. However, there are other studies with less positive findings (e.g., Grygelko, Garwick, & Lampman, 1973). Together, these results are mixed. Thus, the larger question is whether content similarity is actually relevant to those who wish to apply GAS. Smith (1981, 1994) argued that different goals could reasonably emerge from the same problem area or domain by different goal setters. Nonetheless, content analysis would find these goals to be dissimilar and reliability would suffer. Thus, according to Smith (1994), and this author concurs, identical goal-construction should not be a requirement of reliability for GAS.

Adequate team training is also likely to reduce biases (Bailey & Simeonsson, 1988). In addition, bias may be minimized depending upon what is used as the underlying basis for arriving at the derivation of attainment levels (from $-2$ to $+2$). If, for example, the opinions of the team are used as the sole source of information in determining goal attainment, GAS scores are nothing more and nothing less than subjective evaluation data collected as part of typical social validation efforts (Schlosser, 1999). Subjective evaluation offers stakeholder perspectives and as such is particularly useful when it supplements more objective effectiveness data. According to Hegde (1994), objectivity is achieved when the methods and the results are publicly verifiable. Thus, in order to minimize bias it is essential that GAS scores be based not only on subjective data (e.g., interviews, progress notes in client files) but also on objective data such as direct observational data (see also Becker et al., 2000). Behaviors may be observed by independent observers and become publicly verifiable. It is for this reason that the goals outlined in Table 3 will be evaluated based on a combination of objective data (i.e., observations) and subjective data (i.e., interviews).

Some workers are starting to tackle the reliability issue of the process (e.g., Stolee, Rockwood, Fox, & Streiner, 1992). Others suggest, as previously discussed, that the selection and construction of goal guides is more a sampling issue than one directly linked to reliability (Cardillo & Smith, 1994b; Ottenbacher & Cusick, 1993). Rather, inter-observer agreement data need to be collected on the assessment of performance for each distinct goal similar to determining agreement within a single-subject experimental design. In terms of such inter-observer agreement, GAS fares fairly well as reported above.
Ottenbacher and Cusick (1993) and Becker et al. (2000) provide several other useful safeguards to minimize bias. First, they emphasize the importance of operational definitions of the outcome criteria (e.g., the operational definition of “choose within a typical time frame” and “natural prompts” in Table 3). Operational definitions are precursors to objectivity as described earlier. In addition, they point out that the examiner must be able to record accurately and reliably the levels of outcome before using GAS. To minimize bias further, they encourage the determination of attainment by an independent examiner without previous involvement in the treatment or the goal-setting process, or knowledge of the group to which the client was assigned (if groups are used in the study). Another avenue for minimizing that attainment reflects the bias of those who construct the guides is to ensure that the treatment was implemented as planned—with treatment integrity (Simeonsson et al., 1991). Here, the procedures for enhancing treatment integrity described by Schlosser (2002) would be useful.

In summary, ensuring the use of adequate safeguards such as those described above can minimize most of these concerns pertaining to reliability and bias. Obviously, bias cannot be (and probably should not be) removed entirely without compromising one of the strengths of GAS—that the interpretation of progress occurs relative to expected outcomes. Demonstrations of reliability through previous research cannot replace the need for reliability assessments on a case-by-case basis.

4.3. Sensitivity

Clinicians seek to evaluate progress in their clients. To do so, it is essential that the tools or techniques employed in evaluating progress are sensitive to sometimes subtle-but-important changes. Sensitivity, as used in this context, refers to the ability of a method to detect change when change did in fact take place. Standardized approaches may fall short on capturing subtle-but-important change in client-centered functional skills (Kiresuk et al., 1994; MacKay et al., 1996). Based on recent research in other clinical professions, GAS may be sufficiently sensitive to capture these changes (Malec, 1999; Rockwood et al., 1993, 1996, 1997; Stolee et al., 1999).

Stolee et al. (1999), for example, studied the sensitivity of GAS in geriatric care by using multiple methods and operationalizations of sensitivity, including effect size, relative efficiency, and analysis of variance. In comparison to standardized measures, each of the methods used determined GAS to be the most sensitive instrument. In keeping with the critique of Ottenbacher and Cusick (1993), individualized measures such as GAS are expected to be more sensitive than standardized measures. At any rate, although it can be assumed that GAS is more sensitive than standardized measures, it is best to evaluate the sensitivity of GAS on a project-by-project basis by comparing GAS with standardized measures and with other measures of individualized longitudinal change.

Specificity is a different concept from sensitivity yet closely linked. While sensitivity addresses the detection of changes when changes did occur, specificity is the opposite; that is, the measure displays an absence of change when no change occurred (Law, 2002). To date, there is no research available into the specificity of GAS. Because specificity and sensitivity interplay with each other, future research should explore tetrachoric analyses of specificity and sensitivity of GAS.
4.4. Goal scaling

Appropriate scaling is imperative to the validity and reliability of GAS. Some of the more frequently noted pitfalls in scaling include the development of overlapping levels, gaps between levels, multidimensional scales, and the setting of “easy” goals (Becker et al., 2000; Smith, 1994). The issue of overlapping scales occurs when goals can be scored at more than one level at a given time because of an overlapping indicator (Smith, 1994). For example, if the expected level states “requests preferred objects between 6 and 9 opportunities out of 15 opportunities” and the better than expected level says “requests preferred objects between 8 and 11 opportunities out of 15 opportunities, a client who emits requests in eight opportunities could be scored at either level. Thus, overlapping levels are avoidable and should be avoided. A scale with gaps between levels can be equally problematic for the rater who scores performance at follow-up (Smith, 1994). Returning to the earlier example, an expected level of “... 6 to 9 opportunities ...” and a better than expected level of “... 12 to 15 opportunities ...” would cause problems if the client requested 10 times. This is preventable at the time of scale construction. The third issue of multidimensional scales occurs when performance on one goal is scored on two or more dimensions (Smith, 1994). The client may perform at the best expected level on one of the dimensions and at a less than expected level on the other dimensions. This creates difficulty for the person who is scoring at follow-up. Smith (1994) suggests that this can be avoided by having each level differ on only one dimension from the next level. This strategy is modeled for goal 2 in Table 3 where choice-making performance is scored on multiple dimensions, including the number of settings, the type of prompts, and the latency for making a choice. The fourth issue raised by critics of GAS is the setting of smaller than actually expected goals (Kiresuk et al., 1994). What prevents a clinician who wants to show greater goal attainment from setting easier levels of attainments? Working in teams is likely going to provide adequate checks on such behavior. While it is conceivable that individual clinicians may be inclined to set “easy” goals, it is unlikely that an entire team would conspire to do so. Adequate team training is also likely to increase the accuracy with which clinicians predict attainment levels (Bailey & Simeonsson, 1988).

4.5. Computation and statistical analyses

Another area of caution pertains to the computation and statistical analyses of GAS scores. MacKay et al. (1996) argued that scale scores of $-2$ to $+2$ are treated as if they were interval data, when in fact they may be ordinal (rankings on perceived continuum of achievement) or even nominal (“used simply to classify an object, person or characteristic” (Siegel, 1956, p. 23)). Regardless of these properties, users of GAS typically transform them into standard scores (e.g., Kiresuk et al., 1994), which may cause distortions in the data and shed doubt about the conclusion drawn from any test. If used in this way, GAS may not fulfill its claim to be a parametric expression of non-parametric information. As noted by Ottenbacher and Cusick (1993), the use of parametric algorithms for non-parametric data is an age-old debate that is certainly not limited to GAS. Because this practice is widespread in GAS application studies (Kiresuk et al., 1994) solutions are needed.
As a solution, MacKay et al. (1996) advocate the use of non-parametric methods for inferential analyses involving GAS scores. To do so, they suggested ways by which \( -2 \) to \( +2 \) may become (a) stages in a sequence so that they may describe ordinal data or (b) categories in a particular sequence so that they may describe nominal data. This would be accomplished by (a) detailing how frequently each score occurs for each individual or by (b) categorizing the scores dichotomously (e.g., below goal, at or above goal), respectively. Either approach would then allow the interventionist to apply non-parametric statistics such as the median test (Siegel, 1956) to determine differences between the experimental group and the control group. According to MacKay et al. (1996), non-parametric methods make fewer assumptions about the nature of GAS data than do the original standard scores which are parametric. Cardillo and Smith (1994a) and Ottenbacher and Cusick (1993), on the other hand, argue that one can legitimately use parametric methods because the measure is approximately normally distributed; whether or not the measure is ordinal or interval will not make much difference. They further suggest that whether or not to use parametric or non-parametric methods should be based on statistical considerations such as sampling distributions, assumptions, sample size, etc. This debate is far from resolved at a theoretical level and is likely to persist in the foreseeable future (Cardillo & Smith, 1994a). On a practical level, several authors have demonstrated that the use of parametric procedures is likely to produce very similar results to non-parametric procedures (Cardillo & Smith, 1994a; Malec, 1999).

4.6. Rater selection

Who should do the follow-up and rate performance? Interestingly, there has been unanimous agreement among both proponents and critics of GAS to use independent raters at follow-up (Cardillo, 1994a; Cytrynbaum et al., 1979). Bias would be more likely if the follow-up were conducted by the very same clinician who has been involved in goal-setting or the implementation of treatment. Because of this involvement clinicians have a personal investment in the outcome score. Thus, at least for purposes of program evaluation and clinical-outcome research it is essential to have the rating conducted by someone who has not been directly involved in goal-setting or the client’s treatment (Smith, 1994). Ideally, an agency or unit that is not dependent on the one being evaluated should employ the raters. This, however, is not always followed (e.g., Becker et al., 2000). In clinical practice, such precautions may be cost-prohibitive and require a more reasonable approach such as the use of existing staff under certain restrictions. Cardillo (1994a) suggested the use of clinicians from the same unit as long as they are not directly involved in goal-setting or the provision of treatment. Regardless of who does the follow-up, however, the degree to which the raters are trained will bear on the reliability and validity of GAS. Thus, it is of utmost importance that raters receive adequate instruction in GAS.

5. Harnessing the potential of GAS in clinical practice

As a technique for measuring individualized progress toward unique goals, GAS is ready for application in clinical practice. Clinicians and students of communication disorders
need to receive training in the use of GAS. In order for this to occur, texts on clinical procedures (e.g., Hegde & Davis, 1999) may consider incorporating GAS as an option for measuring change in individual clients and professional preparation programs. Similarly, pre-professional training programs may consider curricular content on GAS in appropriate courses and in clinical placements. Although current training materials on GAS are available and are recommended for study (e.g., Kiresuk et al., 1994), additional training materials geared toward the needs of speech-language pathologists would be helpful. Given the importance of appropriate scaling for the validity and reliability of GAS (Smith, 1994), it is essential that clinicians and future clinicians get exposed to scaling examples and practice scaling across areas of clinical practice in communication disorders. Future research should also address the effects of rater instruction on the reliability and validity of GAS. Although the social validity of GAS has been demonstrated through assessments of clinician satisfaction, it would be prudent to study the link of the GAS results to the values of life-oriented changes as perceived by clients/participants and their caregivers and significant others.

6. Using GAS in clinical-outcome research

How might GAS be used in clinical-outcome research? This question will be addressed by discussing the use of GAS in conjunction with group designs and with single-subject experimental designs. Like standardized outcome measures, GAS by itself is not a research methodology that allows one to establish causal inferences between independent variables and dependent variables. In order to use GAS in clinical-outcome research it is necessary that GAS be folded into a (quasi-) experimental design to minimize threats to internal validity (for a rationale see Ottenbacher & Cusick, 1993). Commonly used design types include group designs and single-subject experimental designs.

How would GAS be operationalized with the group design strategy? The above discussed underestimation or overestimation of goals, for example, may constitute such a threat to internal validity associated with the use of GAS (e.g., Simeonsson et al., 1991). Using a control group (i.e., participants with goals that do not receive treatment) with random allocation of goals or participants to the respective group or using control goals (i.e., goals that are not targeted for treatment) can minimize this threat to internal validity. With random allocation, goals, for example, are randomly assigned to the control or the experimental condition. Those goals assigned to the control condition are not treated. After all, there is no reason to believe that overestimation or underestimation could occur differently for experimental goals/participants than for control goals/participants. Further, using control goals/groups provides a more suitable avenue for evaluating the robustness of the outcomes. None of the reviewed GAS studies from other fields has used control goals/groups.¹ One application, however, involved two experimental groups rather than a control

¹ Although Flowers and Booraem (1991) did use control goals and they improved not as much as the experimental goals, they employed a semantic differential scale (1 = much worse to 7 = greatly improved) without a priori definitions of each level. As such their use of the term GAS to describe their semantic differential scale is inappropriate due to extensive differences in procedures.
group: Bradshaw (1993) randomly assigned participants with schizophrenia to treatment in a coping-skills group or a problem-solving group. Participants in the coping-skills group had a mean change in GAS score of 38.6 points, compared with a change of 20.7 points for the patients in the problem-solving group. These already strong results in favor of the coping-skills group would have been even more compelling with a control group. When GAS is used with group designs, it is important that the above debates surrounding the level of measurement, the calculation of summary scores and the statistical analyses be taken into account. A body of group studies may be synthesized using meta-analytic techniques to calculate effect sizes for various treatments. Because meta-analyses serve an important function in clinical-outcome research (e.g., Robey & Schultz, 1998), the question arises whether the results from individual GAS studies could make a contribution to future meta-analyses. The answer appears affirmative as long as the calculation of effect sizes and their statistical analyses take into account the level of measurement of the original GAS data (e.g., Fleiss, 1994; Rosenthal, 1994).

Clinical-outcome research not only relies on group designs but also single-subject experimental designs (e.g., McReynolds & Kearns, 1983; Schlosser, 2003). So the question arises whether GAS has any utility with single-subject experimental designs? There appear to be several avenues for GAS with single-subject experimental designs. If the graded scale from −2 to +2 were used directly as the dependent measure within a suitable single-subject experimental design, performance on goals could be monitored during baseline and revisited at predefined intervals. The specific design, however, would need to be carefully selected. A multiple-baseline design across behaviors (here goals), for example, would only be appropriate if the treatment were the same across the goals. Thus, as long as only those goals are grouped within the same multiple-baseline design that are targeted with the same treatment, this would be a viable option. A multiple-baseline design across subjects may be another possibility as long as the treatment is the same across participants. Again, this would require researchers to group those participants together with similar goals. The use of multiple-baseline design across settings is a possibility only for goals such as goal 1 (Table 3) that do not stipulate performance in multiple settings as indicator of the graded attainment scale. In this situation, there may be little clinical reason to evaluate the attainment of this goal across multiple settings. Otherwise, it could be argued, performance in settings should be part of the graded scale in the first place. Unlike group designs, multiple-baseline designs do not require control participants. Nonetheless, the use of control goals (goals that are monitored without being targeted for treatment) may be warranted to eliminate threats to internal validity resulting from overestimation or underestimation of goals.

Decisions concerning the effectiveness of a treatment should not be based on one individual single-subject experimental study, but rather a synthesis of a body of studies. Analog to group designs, meta-analyses of single-subject research serve an important function in clinical-outcome research (Schlosser & Lee, 2000). GAS studies that are folded into single-subject experimental designs may contribute to future meta-analyses as long as the visual data generated and displayed conform to accepted norms in single-subject experimental research (McReynolds & Kearns, 1983). If so, the data generated should permit the calculation and statistical analysis of commonly used metrics in the synthesis of single-subject experiments (e.g., Scruggs & Mastropieri, 1998).
To evaluate whether GAS can live up to its potential in clinical-outcome research in communication disorders, studies need to be conducted that examine its validity, reliability, and sensitivity in various service delivery systems (e.g., center-based, private-practice, consultation, etc.) and a variety of populations (e.g., people with developmental disabilities, people with acquired disorders, etc.) in several areas of clinical practice within communication disorders. To obtain a sense of GAS’s sensitivity and specificity, control goals or a control group may be used (Ottenbacher & Cusick, 1993). Another avenue, however, may be a comparison of GAS results with other measures of longitudinal change such as single-subject experimental designs. Changes in measures from single-subject experimental designs should coincide with changes in GAS scores. In terms of reliability, GAS needs to be treated no different from other methods that evaluate intervention effectiveness, which require inter-observer agreement for dependent measures and treatment integrity. Specifically, inter-observer agreement data need to be collected on the assessment of performance for each distinct goal by individuals who were not involved in goal-setting or treatment application. Independent observers need to be trained up to a criterion before asking them to judge goal attainment of actual studies. Research into the effects of rater instruction on inter-rater and intra-rater reliability of rating goal attainment is warranted. Directing efforts into the reliability of goal-construction itself appears, as discussed earlier, largely counterproductive given the nature of the GAS process. In addition to the essential gathering of objective data on GAS, subjective evaluation data documenting the experiences of teams in using GAS are needed in order to streamline the process of goal-setting and the interpretation of change.

7. Conclusions

The purpose of this paper was to introduce GAS to the field of communication disorders, to offer a critical review of GAS, and to provide directions for harnessing this unique value of GAS for speech-language pathology. As demonstrated in this paper, GAS offers unique values such as an a priori expectation on change, a codified range of change, a sharpening of the focus of goals, a sharpening of the focus of treatment protocols, and capturing of subtle-but-important change in client-centered functional skills. This review also showed that GAS is not without issues that may delimit its utility when not taken seriously. Because of its unique values, however, GAS should be considered a welcomed addition (not a replacement) to the present set of options available to clinicians, administrators, and researchers interested in assessing change. Directions were provided for harnessing this potential for clinical practice and clinical-outcome research.

Appendix A. Self-study questions

1. Goal attainment scaling is a technique for
   a. measuring the percentage of objectives attained
   b. evaluating the client’s perception about the effects of therapy
   c. predicting the success of a therapy
d. evaluating individual progress toward goals  
e. estimating the length of therapy needed

2. Which of the following is not one of the steps involved in goal attainment scaling?
   a. Specify a set of goals  
b. Assign a weight for each goal  
c. Eliminate goals that are too difficult to attain  
d. Specify a continuum of possible outcomes  
e. Determine performance attained

3. The goal-setting should involve:
   a. a team of professionals  
b. individuals consistent with the philosophy of the service-delivery system  
c. a consensus-building process  
d. professionals and the client  
e. people who do not know the client

4. Which of the following are frequently noted pitfalls in scaling goals (mark all that apply)?
   a. Development of overlapping levels  
b. Inclusion of gaps between levels  
c. Development of multidimensional scales  
d. Setting of easy goals  
e. Setting of difficult goals

5. Which of the following statements most accurately represents what is known about the validity of GAS?
   a. Based on previous research, the validity of GAS can be assumed  
b. The validity of GAS is best demonstrated on a study-by-study basis  
c. GAS has strong content validity, but questionable construct validity  
d. GAS has strong content validity, but questionable criterion validity  
e. GAS has strong construct and criterion validity, but its content validity is questionable.

References


