Available online at www.sciencedirect.com
ScienceDirect
Research in Social and Administrative Pharmacy 8 (2012) 102–121
RESEARCH IN SOCIAL & ADMINISTRATIVE PHARMACY

Review Article

Goal-attainment scaling: A review and applications to pharmacy practice

Minhchau Vu, Pharm.D.[a], Anandi V. Law, B.Pharm., Ph.D.[b,*]

[a]*Inpatient Pharmacy, Kaiser Permanente Hospital, 10800 Magnolia Avenue, Riverside, CA 92879, USA*
[b]*Department of Pharmacy Practice and Administration, College of Pharmacy, Western University of Health Sciences, 309 E. Second Street, Pomona, CA 91766, USA*

Abstract

Background: Goal-attainment scaling (GAS) might represent a breakthrough, or at least hold significant utility in pharmacy practice for program evaluation, but it has not yet been widely applied in the discipline.
Objective: This study aimed to (1) review the psychometric properties of GAS—a measurement technique for the evaluation of outcomes, (2) explore the clinical utility of GAS in various settings, and (3) introduce its potential application in pharmacy practice.
Methods: This systematic review included identified published literature in Pubmed electronic database with keywords/search terms: GAS, goal attainment, goal scaling, goal-attainment procedure, goal-attainment method, GAS and health outcomes, GAS and Short-Form-36 (SF-36), and GAS and quality of life. The inclusion criteria were (1) articles pertaining to GAS method; (2) psychometric data of reliability, validity, and responsiveness were reported; (3) published in Pubmed from 1968 to July 2007; and (4) research on humans. The exclusion criteria were (1) articles published in languages other than English and (2) review articles.
Results: Of the 1055 articles screened, 26 articles from physical/occupational rehabilitation (17) and psychology (9) with psychometric properties evaluation met the inclusion criteria for review. Examination of the literature revealed that GAS demonstrated high reliability, variable validity, and excellent responsiveness.
Conclusion: Several reasons that make GAS a useful methodology include the capability for patient-specific and cooperative goal setting; incremental goal setting toward progress; versatility of clinical utility to cover medication therapy management; and indexing of individual scores for evaluation of program effectiveness.

Keywords: Goal-attainment scaling; Goal setting; Medication therapy management; Disease state management; Multiple patient and clinical goals

Measurement of effectiveness of clinical services

Assessment of direct patient care services in community pharmacy practice settings is imperative in judging the clinical and economic effectiveness of these services and of pharmacist impact on patient-reported outcomes. This review addresses gaps in the measurement of programs involving multidimensional indicators and reviews

* Corresponding author. Tel.: +1 909 469 5500; fax: +1 909 469 5428.
E-mail address: alaw@westernu.edu (A.V. Law).

1551-7411/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.sapharm.2011.01.003

goal-attainment scaling (GAS) as an alternative technique.

Patient care services in the community pharmacy setting, such as disease state management (DSM), involve the utilization of a systemic population-based approach to identify at-risk individuals with a specific disease state.[1] DSM programs are designed with an understanding of a disease progression and targeted interventions to provide coordinated and continuous quality of care to patients in a cost-effective manner; however, each disease state has its own specific treatment goals and guidelines. Medication therapy management or MTM, on the other hand, is a patient care service that involves the pharmacological therapy management of patients with several concurrent comorbidities. Each patient differs from the other in their disease conditions, course of illness, medication regimens, and personal needs in dealing with their illness. Additionally, patients increasingly have more than 1 chronic condition, related or otherwise, especially in the elderly population that is usually targeted for MTM and multiple DSM programs.

Current evaluations of DSM/MTM programs include assessment of clinical indicators and patient-reported outcomes such as health-related quality of life (HRQoL), usually for individual diseases. Given the multidimensional nature and complexity of patients with chronic conditions, a single disease-specific indicator or measure is often inadequate to incorporate the heterogeneity of the population and to capture the dynamics of concurrent disease states. For example, a patient with diabetes will have different treatment goals (eg, glycemic control, reduction in hemoglobin A1C (ie, glycosylated hemoglobin), or increase in energy level) from a patient with chronic heart failure (eg, reduction in symptoms, increase in exercise tolerance). These 2 patients also may differ in their treatment goals and responses to medication therapy. Further, patients with multiple conditions will have different treatment goals for each condition. These variations complicate measurement and comparison of outcomes.

There are several issues that the current measurements present: (1) Goals for an individual patient with multiple conditions (comorbidities) can only be assessed in a fragmented manner based on indicators for individual conditions; there is currently no instrument that provides a comprehensive evaluation of multidimensional DSM or MTM programs in patients with multiple conditions. For HRQoL, a generic instrument can be used in these cases but it lacks the specificity of examining disease-related impact when multiple conditions are present; (2) Comparison of differences between patients within a DSM program or between different DSM and MTM programs is challenging because of case-mix differences and other confounding variables; and (3) The current measurement tools standardized for a particular condition lack the versatility where pharmacists can develop individualized goals to meet the unique need of each patient with differing conditions and detect these individual changes.

What is needed then is a tool that can be individualized for each patient (with possible multiple conditions) to document progress but also be indexed or summarized to measure program effectiveness. GAS could be a possible method in this context. This systematic review and evaluation of the psychometric properties (validity, responsiveness, and reliability) defines, describes, and critiques published work in GAS before explaining its relevance to pharmacists engaged in DSM/MTM. GAS has been used extensively in psychology/mental health and rehabilitation settings but has found its way into the literature in pharmacy practice on only 2 occasions.[2,3]

What is GAS?

Developed by Kiresuk and Sherman in 1968 as a general method for evaluating community mental health programs, GAS has been since used widely in various clinical and educational settings.[4] GAS is a measurement technique that involves using individualized outcome indicators to construct detailed and comprehensive indexed measures that enable outcome evaluation of complex and multidimensional problems. It can be used to assess multiple individualized goals per subject over time and program efficacy as a whole, comparing performance across subjects within the same program.

The steps of the GAS method are as follows:

1. Identification of individualized goals. Setting objective goals is the first step in GAS. Goals are individualized to patient-specific needs and meet SMART (specific, measurable, attainable, realistic, and timely) criteria. The number of goals can be limitless; however, for feasibility, a range of 3-4 goals is suggested. Goal setting is usually conducted by the clinicians, patients, a multidisciplinary team, or a collaboration of the combination above.

2. Setting weight for each goal. Weight can be set for each goal based on its priority or importance; however, this step is optional. All goals are set at 1 if they are assumed to be equal.
3. Selection of follow-up time. Setting a reasonable time frame to follow-up after an intervention is also critical. This process usually occurs concurrently with the first step: goal setting.
4. Statement of expected outcome and development of other outcome levels. The statement of expected would be most likely to occur. Goals are assigned to a 5-point scale, with 0 as the expected outcome, −2 as least favorable or worse than expected, and +2 as most favorable or better than expected outcome. An example of goals for a patient with chronic heart failure (CHF) and diabetes mellitus type 2 (DM type 2) is illustrated in Table 1.
5. T-score conversion. Results from individual goals are summed up and converted to a standardized T-score with a mean of 50 and standard deviation (SD) of 10. One way of calculating the T-score is by using the formula shown below, where x_i = score of goal outcome (−2 to +2), w_i = weight of goal (optional), ρ = estimated average intercorrelation of attainment scores. According to Kiresuk and Sherman, the ρ values are assumed to be 0.3.

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

6. Standardized T-score. A score of 50 and above indicates that performance has met or exceeded the expected level of goal attainment. A score of less than 50 means individual progress was less or worse than the expected outcomes. The overall performance of a patient group with pre/postintervention scores can also be evaluated by the distribution of all reported T-scores, which usually approximates the normal distribution curve ($x = 50$, $SD = 10$). Individual patient performance can also be assessed and compared against the group performance (Fig. 1).

Methods

The aim of this review was to summarize studies in the literature that compared psychometric properties of GAS with other tools for a particular disease state or practice setting. This summary would help assess feasibility of using GAS as a valid technique for pharmacy practice settings.

Search strategy

This systematic review aimed to evaluate the psychometric properties of GAS based on published literature from 1968 to July 2007 in the Pubmed electronic database. Keywords/search terms used were GAS, goal attainment, goal scaling, goal-attainment procedure, goal-attainment method, GAS and health outcomes, GAS and SF-36, and GAS and quality of life. Limits were set for articles in English and on human subjects.

Inclusion/exclusion criteria

The inclusion criteria were (1) articles pertaining to GAS method; (2) psychometric data of reliability, validity, and responsiveness were reported; (3) published in Pubmed from 1968 to July 2007; and (4) research on humans. The exclusion criteria were (1) articles published in languages other than English; and (2) review articles. Articles published after July 2007 were not included in the review.

Selection of articles

Search results were screened by title of publication and abstracts to identify articles with reported psychometric data. If title or article abstract seemed to suggest that psychometric data might be reported, articles were obtained for further evaluation. Assessment of the screened articles was confirmed by the second author for eligibility of inclusion.

Results

A total of 1139 articles were displayed from various keyword searches; 84 articles were eliminated because of overlapping of articles from different searches (Table 2).

Of the 1055 articles screened, 26 articles with psychometric properties reported from various settings met the inclusion criteria for this review. The composition of these articles based on various settings was as follows: 17 from physical/occupational rehabilitation and 9 from psychology. Most of the articles on physical/occupational rehabilitation were focused on the disabled, geriatric, and pediatric patients. Articles from psychology/psychiatry settings mainly consisted of

Table 1
Examples of GAS goals and their outcome scales

Goal weights	Goal 1	Goal 2	Goal 3
	30	20	10
Patient A	Symptom management	Glycemic control	Diet modification, g/d
Most unfavorable treatment outcome thought likely (−2)	Symptoms worsened (↑ rales, edema [+2], orthopnea [3+ pillows], great limitation with daily activities)	• Increase HbA1C, A1C >4% from goal[a] (ie, A1C >10.5%) • Postprandial glucose and FBG not controlled	Na+>3
Less than expected success with treatment (−1)	Symptoms did not improve (↔ rales, edema [+1], orthopnea [1-2] pillows, marked limitation with daily activities)	• No change in HbA1C, A1C 2.1-4% from goal[a] (ie, A1C = 9.6-10.5%) • Postprandial glucose and FBG not controlled	Na+=2-3
Expected level of treatment success (0)	Symptoms improved slightly (↓ rales, slight edema, no orthopnea, some limitation with daily activities)	• HbA1C ↓ <2%, A1C 2-3% from goal (A1C = 8.5-9.5%) • Postprandial glucose not controlled • FBG controlled	Na+<2
More than expected success with treatment (+1)	Symptoms improved greatly (no symptoms, no limitation with daily activities, marked limitation with walking 1 block)	• HbA1C ↓ 2-3% • HbA1C 1-1.9% from goal (A1C = 7.5-8.4%) • FBG and postprandial glucose controlled	Na+<1.5
Best anticipated treatment success (+2)	Symptoms improved significantly (no symptoms and limitation with daily activities, no limitation with walking 1 block)	• HbA1C ↓ 3.1-4% • HbA1C is at goal (A1C = 6.5%) • FBG and postprandial glucose controlled	Na+<1

[a] HgbA1C goal = 6.5%.

patients with mental health and dementia. Excluding the review articles, the sample size in these studies ranged widely from 10 to 650.

Validity

Validity refers to the degree to which GAS assesses a specific concept it intends to measure. Types of validity found in these articles were content, construct, and concurrent validity. Each of these is described below with reference to the review.

Content validity

Content validity refers to a systemic examination of the test content to determine whether the behavior domain is representative of the sample. As seen in Table 1, 6 studies (5 rehabilitation, 1 psychology) included this psychometric parameter. Content validity was reported to be based on field experts and previous literature in 4 studies.[5-8] It was also demonstrated in 4 (2 rehab, 2 psychiatry) studies by the consensus (43.75-88%) in the identification and ranking between goal setters/raters or patients.[9-11] Content analysis was also used to explain this psychometric property of GAS in 1 rehabilitation study[12] (Appendix).

Construct, criterion, and concurrent validity

Construct and criterion validity denote the similarities between an abstract theoretical concept (eg, cognition, happiness) and the operationalization of that concept to which an instrument attempts to measure, whereas concurrent validity refers to the correlations between GAS outcomes compared with outcomes of other measures from

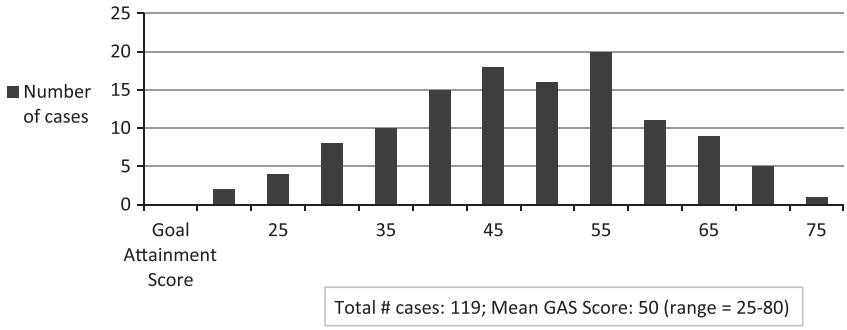

Fig. 1. Program distribution scores.

an existing standardized tool that has been validated. These 3 types of validity were not differentiated in several articles; therefore, they are categorized together in this review.

Seven rehabilitation studies explored the construct validity of GAS by comparing with a wide range of standard outcome measures.[6,10,12-16] High correlation was found between GAS and cognitive global impression (CGI). Moderate correlations were found with the following instruments: Barthel Index (BI), Locomotor Capabilities Index, Milwaukee evaluation of daily living, Katz Index of Older Americans Resource Scale for instrumental activities of daily living, Katz instrumental activities of daily living, Rappaport Disability Rating Scale, Functional independence measure, and Rappaport Disability Rating Scale (RDRS). Low correlations were found between GAS and the Mini-Mental State Examination (MMSE) Nottingham Health Profile, and the Peabody Gross Motor Scale (PGMS).

Table 2
Keyword searches

Keyword	Search resulted	No. of articles overlapped	No. of articles screened
GAS	112	NA	112
GAS and quality of life	6	5	1
Goal-attainment method	88	10	78
Goal-attainment technique	22	3	19
Goal-attainment procedure	16	2	14
Goal scaling	223	29	194
GAS and health outcomes	18	3	15
Goal attainment	654	32	622
Total	1139	84	1055

Seven rehabilitation studies reported criterion, convergent, and concurrent validity.[5,7,8,17-20] Similarly to the results for construct validity, CGI showed the highest correlation with GAS, compared with other evaluation scales. The following 11 instruments resulted in moderate correlation with GAS: Alzheimer's Disease Assessment Scale (ADAS-Cog), BI, Family Needs Survey, Global Deterioration Scale, Instrumental activities of daily living (IADL), Impact on Family Scale, Medical Psychological Questionnaire for lung patients, Oswestry Low Back Pain Disability Questionnaire, Lawton-Brody Physical Self-Maintenance Scale (PSMS), PGMS, and Self-perceived change. Six scales showed low correlations with GAS, especially the MMSE (Table 3).

Only 1 study in the psychology arena reported construct validity.[21] Correlations were high when comparing clinician scores between the Clinician's interview-based impressions of change-plus (CIBIC+) versus baseline GAS scores; however, only moderate correlations were found between GAS scores and CIBIC+ Scale when caretaker and patients scores were used. This same study also reported low and moderate correlations between GAS domains (cognition, function, and behavior) and other standard cognitive, functional, and behavioral measures (MMSE, ADAS-Cog, Lawton-Brody Physical Self-Maintenance Scale, IADL, Functional Activities Questionnaire, Cornell Depression Scale, and Center for Epidemiologic Studies Depression Scale). Validity was reported in 3 psychotherapy studies.[9,22,23] Two instruments showed high correlation with GAS: Fahrenberg Questionnaire (FLZ) and Health-Sickness Rating Scale (HSRS). Three instruments exhibited moderate correlations with GAS: CIBIC+, Target Complaint Scale (TCS), and Brief Symptom Inventory

Table 3
Validity comparison between GAS and other outcome measures

Type of validity	High correlation (r)	Moderate correlation (r)	Low correlation (r)
Construct validity	**Rehabilitation setting:** CGI (0.73, 0.80)	**Rehabilitation setting:** BI (0.41, 0.6) OARDS-IADL (0.41, 0.48) Katz IADL (0.41, 0.45) MEDLS (0.5, 0.51) RDR (−0.61, 0.47) FIM (0.45)	**Rehabilitation setting:** MMSE/sMMSE (0.028-0.2) NPH (−0.06 to 0.03) PGMS (0.25-0.33)
	Psychology setting: CIBIC+ (−0.82, −0.76)[a]	**Psychology setting:** CIBIC+ (−0.62, 0.51)[b] MMSE, ADAS-Cog, LBPSMS, IADL, FAQ, CDS, CES-D, PFMS, FNS, IFS (−0.56 to 0.58)	
Criterion, concurrent, and convergent validity	**Rehabilitation setting:** CGI (0.82, 0.85)	**Rehabilitation setting:** ADAS-Cog (0.52) BI (0.14, 0.86) FNS (0.43) GDS (0.63) IADL (−0.42) IFS (0.31) MPQL (0.42, 0.6) ODQ (−0.31) PSMS (−0.3) PGMS (0.44) SPC (0.14, 0.59)	**Rehabilitation setting:** BCRS (0.22) CIRS (−0.16) HABAM (0.169) MMSE (0.004) PFMS (0.18) QoL-RIQ (0.02, 0.13)
	Psychology setting: FLZ (0.61, 0.74) HSRS (0.7)	**Psychology setting:** CIBIC+ (−0.52, −0.31) TCS (0.5) BSI (0.38)	**Psychology setting:** IIP (0.03, 0.19) MMSE (0.09, 0.37) ADAS-Cog (0.1, 0.36)

[a] Clinicians' baseline scores.
[b] Patients/caregivers' baseline scores.

(BSI). MMSE, used in a psychology setting, also showed a low correlation with GAS, similarly to results found in other rehabilitation studies. Two other instruments, which also demonstrated low correlation with GAS are ADAS-Cog and Inventory of interpersonal problems (Table 3, Appendix).

Responsiveness

Responsiveness is a measure of how well an instrument can detect change had a substantial change occurred. All 16 (13 rehabilitation) and (3 psychology) articles reported high responsiveness associated with GAS compared with other outcome measures. The average effect size (ES) of GAS was 5.96, ranging from 0.61 to 18.96.[24] GAS also showed the largest relative efficiency when compared with other standard measures (BI, CGI, and RDRS). One study did not report exact values for ES and RE; however, high efficiency of GAS was demonstrated on a graph. The average standardized mean response from 3 studies was 0.98.[25] One study also reported responsiveness in terms of Norman's responsiveness statistic, which also indicated GAS as the most responsive instrument in detecting change compared with other measures such as PSMS, Comprehensive Geriatric Assessment, BI (GAS = 0.32-0.57), (others = −0.06 to 0.18). One study also reported higher percentage of change detected by GAS compared with behavioral objective format. Not only did GAS measure the goals achieved, but it was also able to determine goals that exceeded expected outcome and goals progressing toward outcome, a percentage of which was not detected by behavioral objective format. GAS detected 1 in 18 cases not detected

by BI, with sensitivity = 91% and specificity = 86%, reported in 1 study (Appendix).[26]

Reliability

Reliability refers to the consistency GAS yields with repeated measures. Thirteen studies reported high reliability with GAS. Correlation coefficient was greater than 85% in 9 studies.[27] One study demonstrated moderate reliability in scoring between 2 raters ($r = 67\%$). When compared with goal scoring between patients and therapists in 1 study, results only showed moderate reliability ($r = 0.46$, $P < .001$).[28]

Discussion

Psychometric parameters of GAS

In summary, the reviewed literature revealed that GAS had high reliability, variable validity, and excellent responsiveness. As the previous articles have indicated, because of the nature of GAS, content validity needs to be assessed on a case-by-case basis.[29] Construct validity of GAS varied widely depending on the standard measure in comparison and the selection of individual goals and their relative weighting. Although GAS consisted of a variety of goals, which were established by different goal setters/raters, high inter-rater reliability was observed between GAS scorers/raters. In terms of responsiveness, GAS has consistently shown high performance in detecting even the smallest change. The reason for this finding is the flexibility of goal setting with GAS; individualized items that could not be included in standardized measures could be incorporated as an outcome measure in GAS.

Although psychometric data were mainly reported in psychiatry and rehabilitation settings, GAS has been applied in a wide variety of settings (eg, pain management, sports medicine, long-term care, education programs) because of its clinical versatility.

Limitations of the review

Limitations of this review include (1) lack of differentiation between the various types of validity; (2) lack of specification whether GAS domain scores or GAS T-scores were used when comparing GAS to other instruments; and (3) exclusion of articles after July 2007.

Applications to pharmacy practice

There are several advantages of GAS that make GAS a useful methodology for DSM and MTM program evaluation (summarized in Table 4).

Cooperative goal setting

GAS involves a collaborative effort from patients, clinicians, and program administrators to establish attainable goals and realistic level of expected outcomes. Patient participation in this process enhances understanding of goal and expected level of performance, and grants patients with a sense of control. Group goal setting also provides clinicians the opportunity to identify their patient's individualized areas of improvement and educate their patients on goals and expected outcomes. GAS also allows integration of input from program administrators to set reasonable objectives to ensure the provision of quality and cost of program. All in all, increased agreement in mutual goals could potentially provide increased motivation for patient to comply, improve

Table 4
Summary of advantages and disadvantages of GAS

Advantages of GAS	Limitations of GAS
High reliability, "variable" variability, and excellent responsiveness.	There is an assumption that outcomes can be determined in advance.
Goals can be completely individualized for the client's needs.	Staff will need training in using the approach.
Goals can be changed or abandoned if circumstances change.	There is an additional time commitment involved in developing the outcome levels, although this is less of an impact if such discussion is part of the practice approach.
Research has shown that a maximum of 5 goals is likely to be manageable at any one time and that most people would be working on 2 goals in any one period of time.	Expected outcomes need to be set at a realistic level for the client's needs and circumstances and the time period set for review, or results will be distorted.

patient's satisfaction, clinical outcomes, and overall therapeutic benefits.[30,31]

Patient-specific goals

GAS allows patients to be involved in the goal-setting process, which helps them with understanding their disease states and therapy and outcomes of care. With that basic foundation, patients will be more likely to grasp the goal objectives and levels of goal expectation. Examples from the health psychology literature provide evidence that better knowledge and expectations of the condition help drive proactive behavior, compliance to therapy, and perception of improvement.[32-34] Furthermore, weighting of goals allow patients to rank or prioritize the importance of their objectives and tackle a few manageable objectives at a time. Additionally, information from goal and weight settings gathered from patients provides clinicians a better comprehension of patient's individualized goals instead of following generalized goals delineated in protocols. For example, goals for a patient with chronic heart failure (CHF)-NYHA class III and DM type 2 (patient A) will differ from goals set for a patient with CHF-NYHA class II and dyslipidemia (patient B). Patient A, with CHF-NYHA class III will experience marked limitation in daily activities, and therefore will more likely be concerned about alleviating his symptoms. Whereas patient B, with CHF-NYHA class II, will only experience slight limitation during vigorous physical activities, and therefore will most likely to want to work toward increasing his tolerance to exercise, instead of reducing daily symptoms like patient A. Because of their different severity of CHF and other comorbidities (patient A—DM, patient B—hyperlipidemia), health issues will be addressed with different sets of goals. For demonstration purposes, assume that clinicians and patients have agreed on the following important goals for 2 patients, A and B. Three goals for patient A are to improve symptoms, glycemic control, and diet modification; whereas goals for patient B are to improve exercise tolerance, cholesterol reduction, and medication compliance. The versatility of GAS allows the clinician to set individualized goals for these 2 patients, yet at the same time, compare their progress with other patients in the clinic/program using the standardized T-scores.

Combination of different types of goals

As seen above, multiple types of goals can be incorporated such as clinical- and patient-reported goals.

Incremental goals toward progress

GAS allows patients to set increments of progression that would correspond with their attainment levels. Involvement in a goal process, and being able to set smaller and more achievable goals, gives patient a sense of control. For example, patient B would like to improve on his medication compliance, which is also the key in any therapy management. With GAS, clinicians can collaborate with this patient to increase adherence by reducing the number of missed doses with tackling 2 or 3 drugs at a time. The sense of achievement of small goals will motivate patients further toward their next goals. Additionally, detection of small, important changes provides clinicians an idea of the direction and rate of progress of their patients' performance toward long-term goals. Using a systems approach, if anytime during this process where the patient regress in his or her performance, the clinician would be able to locate the point of lack of motivation or dissatisfaction, identify the problem, and address the issues immediately to get the patient back to the right direction. Frequent progress updates provide benefits in both clinicians and patients.

Versatility to cover MTM

The versatility of GAS clinical utility is a positive attribute in the evaluation of a multidisciplinary program such as MTM. Instead of multiple specific goals/scales needed for each disease state in 1 patient, GAS only requires 1 scale to evaluate the longitudinal change of a patient with multiple diseases. Goals in terms of medication therapy, specific to drug response, could be used to differentiate progress attributed to appropriate medication use and added effects of lifestyle modifications.

Evaluation of program effectiveness

Individual scores can also be standardized to T-scores, which can be evaluated for progress or change of a group as a whole. For example, after a defined time point (eg, 3 months), patients A and B along with other patients participating in this evaluative GAS program obtained T-scores of mostly 50 and above. This indicated that patients were performing better than expected, thus suggesting an overall good performance of the program. Plotting these T-scores on a graph will provide the evaluator with a distribution scale showing the patient individualized progress contributing to the program's overall performance (Fig. 1). Managers/administrators can take

advantage of this aspect of GAS to set program objectives, evaluate patient/clinicians performance, and devise clinical inventions after incremental progress update to steer program goals or train clinicians toward the right direction. Administrators also could use clinical indicators to correlate with pharmacy services for demonstration of program effectiveness or benefits. This attribute is also extremely important especially at this time, where clinician/program performance demonstration is essential for the reimbursement of the provision of services.

Limitations of GAS

The challenges of GAS should also be taken into consideration.

Requirement of GAS training

The skill levels of clinicians to set pertinent concrete goals from theoretical concepts and to accurately predict levels of outcomes determine the validity of GAS; thus, clinical expertize and thorough training in the usage of GAS are pertinent requirements of clinicians conducting this process. Skills in the rating process of patient outcomes are also pertinent to the reliability of GAS. To maximize the consistency of the instrument, training of evaluators in understanding the operational detail of the scales is also essential.

Lack of psychometric data in other settings

Besides the psychotherapy and rehabilitation settings, literature containing GAS psychometric data is limited. This may be due in part to the varied applications of GAS. Psychometric data may not be reported or as relevant when GAS is used to evaluate program effectiveness; however, psychometric data are relevant and reported in a discriminative assessment, where GAS is used to compare differences between patients at a particular point in time or to compare between GAS and other normative tests. Although clinical utility of GAS has been demonstrated in various settings, comparison of its application across these different settings is challenging.

Requirement of time and resources

Training of clinicians and evaluators before program implementation demands time and commitment from staff and administrators. The goal-setting process, which involves the discussion between clinicians and patients, also requires time and effort from parties involved. Resources (such as multiple raters) and time (multiple-rating periods) are encouraged during the evaluation process to ensure high reliability of GAS, which can also contribute to the time and costs of implementation. Frequent discussion and feedback on the usage of GAS between evaluators are also an essential component for the sustenance of GAS as a program evaluative measurement. Other factors critical to the survival of GAS in the program include integration of GAS into host organization, a supportive administration team of GAS, utility in that particular organization, and a perception of high benefits associated with GAS by the majority of all concerned.[35]

Conclusion

This review examined the psychometric properties and clinical utility of GAS in various settings and to introduce its potential use in pharmacy. With its focus in patient individualized goals, versatility in outcome measurement, and high responsiveness to subtle, but important clinical longitudinal changes, GAS might be a potentially valuable instrument in evaluating client/pharmacist/program in pharmacy practice, particularly in an MTM/DSM setting. The implementation of GAS in a pharmacist-provided MTM or DSM clinical program and evaluation of GAS psychometric properties in this setting would be an appropriate next step in future GAS research.

Acknowledgments

The authors would like to thank the library staff at Western University for their prompt service in the retrieval of their numerous journal requests.

References

1. Disease Management Association of America. Definition of disease management. Available at: http://www.dmaa.org/definition.html. Accessed 22.03.05.
2. Kiresuk TJ, Lund SH, Larsen NE. Measurement of goal attainment in clinical and health care programs. *Drug Intell Clin Pharm* 1982;16:145–153.
3. Nieuwstraten C, Dolovich L, Chase H. Goal attainment scaling in the provision of pharmaceutical care to hospitalized patients with cardiovascular disease: a pilot study. *Can J Hosp Pharm* 2002;55(4):262–271.
4. Kiresuk T, Sherman R. Goal attainment scaling: a general method for evaluating comprehensive community mental health programs. *Community Ment Health J* 1968;4(6):443–453.
5. Gordon JE, Powell C, Rockwood K. Goal attainment scaling as a measure of clinically important change in nursing-home patients. *Age Ageing* 1999; 28:275–281.

6. Yip AM, Gorman MC, Stadnyk K, Mills WG, MacPherson KM, Rockwood K. A standardized menu for goal attainment scaling in the care of frail elders. *Gerontologist* 1998;38:735–742.
7. Rockwood K, Stolee P, Howard K, Mallery L. Use of goal attainment scaling to measure treatment effects in an anti-dementia drug trial. *Neuroepidemiology* 1996;15:330–338.
8. Stolee P, Rockwood K, Fox RA, Streiner DL. The use of goal attainment scaling in a geriatric care setting. *J Am Geriatr Soc* 1992;40:574–578.
9. Shefler G, Canetti L, Wiseman H. Psychometric properties of goal-attainment scaling in the assessment of Mann's time-limited psychotherapy. *J Clin Psychol* 2001;57:971–979.
10. Palisano RJ. Validity of goal attainment scaling in infants with motor delays. *Phys Ther* 1993;73:651–658.
11. Woodward CA, Santa-Barbara J, Levin S, Epstein NB. The role of goal attainment scaling in evaluating family therapy outcome. *Am J Orthopsychiatry* 1978;48:464–476.
12. Stolee P, Stadnyk K, Myers AM, Rockwood K. An individualized approach to outcome measurement in geriatric rehabilitation. *J Gerontol A Biol Sci Med Sci* 1999;54:M641–M647.
13. Rushton PW, Miller WC. Goal attainment scaling in the rehabilitation of patients with lower-extremity amputations: a pilot study. *Arch Phys Med Rehabil* 2002;83:771–775.
14. Rockwood K, Joyce B, Stolee P. Use of goal attainment scaling in measuring clinically important change in cognitive rehabilitation patients. *J Clin Epidemiol* 1997;50:581–588.
15. Joyce BM, Rockwood KJ, Mate-Kole CC. Use of goal attainment scaling in brain injury in a rehabilitation hospital. *Am J Phys Med Rehabil* 1994;73:10–14.
16. Rockwood K, Stolee P, Fox RA. Use of goal attainment scaling in measuring clinically important change in the frail elderly. *J Clin Epidemiol* 1993;46:1113–1118.
17. Fisher K, Hardie RJ. Goal attainment scaling in evaluating a multidisciplinary pain management programme. *Clin Rehabil* 2002;16:871–877.
18. Palisano RJ, Haley SM, Brown DA. Goal attainment scaling as a measure of change in infants with motor delays. *Phys Ther* 1992;72:432–437.
19. Simeonsson RJ, Bailey DB Jr, Huntington GS, Brandon L. Scaling and attainment of goals in family-focused early intervention. *Community Ment Health J* 1991;27:77–83.
20. van Stel HF, Colland VT, Heins NL, Rijssenbeek-Nouwens LH, Everaerd W. Assessing inpatient pulmonary rehabilitation using the patient's view of outcome. *J Cardiopulm Rehabil* 2002;22:201–210.
21. Rockwood K, Graham JE, Fay S. Goal-setting and attainment in Alzheimer's disease patients treated with donepezil. *J Neurol Neurosurg Psychiatry* 2002;73:500–507.
22. Leichsenring F, Biskup J, Kreische R, Staats H. The Göttingen study of psychoanalytic therapy: first results. *Int J Psychoanal* 2005;86:433–455.
23. Asp E, Cloutier F, Fay S, et al. Verbal repetition in patients with Alzheimer's disease who receive donepezil. *Int J Geriatr Psychiatry* 2006;21:426–431.
24. Rockwood K, Howlett S, Stadnyk K, Carver D, Powell C, Stolee P. Responsiveness of goal attainment scaling in a randomized controlled trial of comprehensive geriatric assessment. *J Clin Epidemiol* 2003;56:736–743.
25. Rockwood K, Fay S, Song X, MacKnight C, Gorman MVideo-Imaging Synthesis of Treating Alzheimer's Disease (VISTA) Investigators. Attainment of treatment goals by people with Alzheimer's disease receiving galantamine: a randomized controlled trial. *CMAJ* 2006;174:1099–1105.
26. Ashford S, Turner-Stokes L. Goal attainment for spasticity management using botulinum toxin. *Physiother Res Int* 2006;11:24–34.
27. Jones MC, Walley RM, Leech A, Paterson M, Common S, Metcalf C. Using goal attainment scaling to evaluate a needs-led exercise programme for people with severe and profound intellectual disabilities. *J Intellect Disabil* 2006;10:317–335.
28. Fiester AR. Goal attainment and satisfaction scores for CMHC clients. *Am J Community Psychol* 1979;7:181–188.
29. Schlosser RW. Goal attainment scaling as a clinical measurement technique in communication disorders: a critical review. *J Commun Disord* 2004;37:217–239.
30. Naik A, Kallen M, Walder A, Street RL Jr. Improving hypertension control in diabetes mellitus: the effects of collaborative and proactive health communication. *Circulation* 2008;117:1361–1368.
31. MacGregor K, Handley M, Wong S, et al. Behavior-change action plans in primary care: a feasibility study of clinicians. *J Am Board Fam Med* 2006;19:215–223.
32. Koller M, Lorenz W, Wagner K, et al. Expectations and quality of life of cancer patients undergoing radiotherapy. *J R Soc Med* 2000;93(12):621–628.
33. Litt MD. Self-efficacy and perceived control: cognitive mediators of pain tolerance. *J Pers Soc Psychol* 1988;54(1):149–160.
34. Llewellyn-Thomas HA, Thiel EC, McGreal MJ. Cancer patients' evaluations of their current health states: the influence of expectations, actual health states and mood. *Med Decis Making* 1992;12:115–122.
35. Glaser EM, Backer TE. Durability of innovations: how goal attainment scaling programs fare over time. *Community Ment Health J* 1980;16:130–143.
36. Willer B, Miller G. On the validity of goal attainment scaling as an outcome measure in mental health. *AJPH* 1976;66:12.
37. Kaplan JM, Smith WG. The use of attainment scaling in the evaluation of a regional mental health program. *Community Ment Health J* 1977 Summer;13(2):188–193.

Appendix

Author, year published	Sample size	Study design	Outcome measured	Validity	Reliability	Sensitivity/ responsiveness
Studies in the clinical rehabilitation settings						
Jones (2006)[27]	N = 22 Age = 34.8-60.8 yr Setting A and B consisted of patients with intellectual and multiple physical disabilities	Pre-experimental/ pre- postevaluation design for a 16-wk individualized exercise program	Change in total GAS scores and GCI at baseline vs weeks 4, 8, 12, and 16. Change in Aberrant Behavior Checklist scores at baseline vs week 16	Cannot be determined	Inter-rater reliability: $r > 0.80$	ES (GAS) = 0.23-0.33
Ashford (2006)[26]	N = 18 Age X = 44.4 yr, SD = 13.4 Disabled patients with brain injury	Retrospective study	Outcome GAS score Change in GAS and BI scores (positive (≥ 10) vs negative (< 10) overall clinical outcome (ICF classification, system, &and activity function)	—	—	Sensitivity = 91% Specificity = 86% Change detected in 1 of 18 cases by BI
Rockwood (2003)[24]	N = 165 (control = 80, intervention = 85) Age = 81.8 yr, SD = 7.2 Disabled, geriatric patients in rural community dwellings	Randomized, controlled trial	Change in GAS treatment status score compared with baseline scores in functional improvement Change in scores compared with baseline scores in other secondary measures (BI, IADL, PSMS, SQLI)	—	—	ES: GAS > other measures (2.93-4.43) RE: GAS > BI (29.92-100.0+) SMR: GAS > other measures (0.78-1.07) Norman's Responsiveness Statistics: GAS > other measures (0.32-0.57)

Fisher (2002)[17]	N = 149 Age = 42.5 yr, range = 16-65 Patients with pain >1 yr	Prospective, observational study	GAS and other physical mobility measures (MPQ, NRS, ODQ, GHQ, PAIRS).	**Concurrent validity:** Pearson's $r = -0.31$ (GAS vs ODQ)	—	—
Rushton (2002)[13]	N = 10 (6 female, 4 male) Age = 72.3 yr, SD = 10.7 Unilateral lower-extremity amputations	Pilot study comparing GAS with BI and LCI	Change scores of GAS, BI, LCI of the Prosthetic Profile of the Amputee	**Construct validity** GAS vs BI, Pearson's $r = 0.44$ ($P = .1$) GAS vs LCI, $r = 0.35$ ($P = .2$)	Inter-rater reliability = 0.67 63% (24/38) goals in common between 2 independent investigators	**GAS is most responsive** ES (GAS = 6.5; BI = 1.04; LCI = 3.7) RE (GAS vs BI = 3.1; GAS vs LCI = 4.7)
Stolee (1999)[12]	N = 173 Age = 81 yr, SD = 7, range = 61-96	Prospective, descriptive study	Change scores and follow-up scores (BI, OARS of IADL, self-rated health question, MMSE, a quality-of-life measure, NHP, global clinical assessment)	**Content validity:** Goals reviewed and modified by clinicians as appropriate **Construct validity:** Change GAS and other scores (BI = 0.6; OARS IADL = 0.48; MMSE = 0.2; NHP = [−0.06 to 0.03]) GAS follow-up vs other scores (BI = 0.66; OARS IADL = 0.54; MMSE = 0.30; Global rating = 0.67, NHP = [−0.31 to −0.16])	Inter-rater reliability (team conference vs nurse): guides = 61, goals = 255 GAS follow-up scores, $r = 0.93$ GAS individual goals, $r = 0.89$	ES (GAS = 3.52; BI = 1.14, OARS IADL = 1.32, Other measures <0.5) SRM (GAS = 1.73, BI = 0.97, OARS IADL = 0.80, other measures <0.55) RE (GAS = 3.14, BI = standard, OARS IADL = 0.69, other measures <0.25)
Gordon (1999)[5]	N = 53 Age = 81 ± 8 yr	Prospective, descriptive study	ES and relative efficiency of the BI, HABAM, GDS,	**Content validity:**	—	GAS = highest ES: 1.29 GAS = highest relative efficiency: 53.7

(*Continued*)

(*Continued*)

Author, year published	Sample size	Study design	Outcome measured	Validity	Reliability	Sensitivity/ responsiveness
	Disabled nursing home patients		CIRS, axis 8 of BCRS.	Inferred from other geriatric settings **Convergent validity:** GAS vs other standard measures (axis 8 of BCRS = −0.2158, CIRS = −0.1649, BI = 0.1484, HABAM = 0.169)		
Yip (1998)[6]	N = 143 Age = 77 yr, SD = 8 Geriatric patients admitted to rehabilitation unit for therapy	Prospective, descriptive study	GAS total (outcome and personal) goal scores at discharge and change scores on the BI, Katz, ARS IADL, and SMMSE	**Content validity:** Based on literature and team evaluation **Construct validity:** GAS total change scores vs BI, Katz, and OARS IADL (Spearman's r = 0.41-0.45), SMMSE (r = 0.11) GAS outcome scores vs others (r = 0.43-0.45); GAS process scores vs others (r = 0.24-0.31)	—	SRM (GAS = 1.56, BI = 0.89, Katz = 0.82, OARS IADL = 0.72, SMMSE = 0.54) RE (GAS = 3.16, BI = 1, Katz = 0.86, OARS IADL = 0.58, SMMSE = 0.32)
Rockwood (1997)[14]	N = 44 Age = 29.2 yr, range = 22-61 Patients with at least moderate functional impairment on admission	Prospective, descriptive study	Scores at admission and discharge, and change scores of the different measuring instruments (IADL, KELS, MEDLS, Spitzer QLI, RDR, ADLELIM, ADLMOB, CGI)	**Construct validity:** GAS change scores vs other measures (Spearman's r = CGI = 0.73, MEDLS = 0.5176, RDR = −0.4722, all other measures <0.5)	0.95 Discharge scoring, ICC = 0.95 Change scores, ICC = 0.93	RE (GAS = 7.8, SQLI = 1.95, MEDL = 1.41, RDRS = 1, other measures <1) ES (GAS = 5.11, SQLI = 1.4, RDRS = 0.48, KELS, MEDLS = 0.59, other measures <0.5)

Rockwood (1996)[7]	N = 15; Age X = 72.8 yr Alzheimer's patients with mild-moderate dementia	Randomized, placebo-controlled, double blind, parallel group study	Mean change in GAS outcome and total, and other scores (ADAS-cog, GDS, CGI, MMSE, PSMS, IADL) at pre/post intervention	**Content validity:** Assessed by panel of experts **Concurrent validity:** GAS vs CGI (Pearson's $r = 0.85$) GAS vs PSMS, IADL, ADAS-cog, GDS ($r = -0.3, -0.4255, 0.52, 0.63$), respectively GAS vs MMSE ($r = 0.004$)	—	ES = 0.61 RE = 0.47 Largest ES and RE compared with standard (CGI)
Joyce (1994)[15]	N = 16 Age = 27 yr, range = 17-49 years old Patients with severe traumatic brain injury	Prospective, descriptive study	Scores at admission and discharge of GAS vs other outcome measures (IADL, KELS, MEDLS, Spitzer QLI, RDR, ADLELIM, ADLMOB and CGI—discharge score only)	**Construct validity:** GAS vs other measures (Spearman's r, CGI = 0.8061, MEDLS = 0.5, RDR = −0.6162, all other measures <0.5)	**Inter-rater reliability:** Admission scoring, ICC = 0.92 Discharge scoring, ICC = 0.94 Change scores, ICC = 0.87	—
Rockwood (1993)[16]	N = 45 Age = 81 yr, SD = 8 Frail elderly with polypharmacy and comorbidities, or medically stable patients admitted for rehabilitation after surgical procedures	Prospective, descriptive study	Scores at admission and discharge of GAS vs other outcome measures (BI, FIM, MMSE, Katz IADL, Spitzer QLI, Physical Self-Maintenance Scale)	**Construct validity:** GAS change scores vs others (Spearman's r, BI = 0.5889, FIM = 0.4525, MMSE = 0.02806)	Inter-rater reliability: ICC=0.91	No exact values reported for ES and RE Graph showed that GAS is more efficient than any other measure

(*Continued*)

(*Continued*)

Author, year published	Sample size	Study design	Outcome measured	Validity	Reliability	Sensitivity/ responsiveness
Palisano (1993)[10]	N = 21 Infants (4-24 mo) Infants with developmental delay and regular atten dance of weekly 2-h early intervention sessions	Prospective, descriptive, intervention study	Scores and change scores of GAS and PGMS	**Content validity:** 79-88% agreement between GAS scores with questions 1-3 by scorers (Q1: $X^2 = 13.7$, $df = 9$, $P = .13$; Q2: $X^2 = 15.8$, $df = 9$, $P = .07$; Q3: $X^2 = 12.8$, $df = 9$, $P = .17$) **Construct validity:** Change scores of GAS vs PGMS (correlation $r = 0.25$; $P = .14$ (first 3 mo); $r = 0.33$, $P = .08$ (second 3 mo)	—	GAS and behavioral objectives detected 39% of achievable goals, of which GAS detected 79% of goals that exceeded criteria GAS detected change in 61% of goals not detected by behavioral objectives
Palisano (1992)[18]	N = 65 Age = 19.4 mo, SD = 7.1 Infants with motor delay receiving physical therapy, occupational therapy, or both as part of their early intervention program	Prospective, descriptive, intervention study	GAS T-scores and PGMS age-equivalent change scores	**Criterion validity:** GAS vs PGMS age-equivalent change scores: $r = 0.44$ ($P < .001$); explained 19% of variance with gross motor GAS T-scores GAS vs PFMS age-equivalent change scores: $r = 0.18$ ($P < .12$); explained 3% of variance associated with fine motor GAS T-scores	**Inter-rater reliability:** Kappa coefficient = 1 (10 goals)	Mean GAS T-score=55.4, SD=14.6 (X=50, Z=2.48, $P < .001$)

Stolee (1992)[8]	N = 15 Age X = 79 yr, range = 65-94	Prospective, descriptive study	Scores on admission and discharge of GAS, BI, global rating of outcome (subjective 10-pt scale)	**Content validity:** 82% concensus in identifying and ranking of goals between 2 geriatricians 12 of 13 problem areas (goals) are in agreement with literature from the field Concurrent validity: GAS vs CGI ($r = 0.82$) Change scores of GAS vs BI ($r = 0.86$)	**Inter-rater reliability:** 59% in agreement in expected outcome levels of goal areas between 2 geriatricians ICC (nurse vs physician) -discharge score, $r = 0.88$ -change score, $r = 0.87$ GAS discharge score, $r = 0.91$ GAS change score, $r = 0.90$	—
Simeonsson (1991)[19]	N = 23 families Age (infants) = 13.8 mo Infants with delayed development receiving community home-based interventions	Prospective, intervention study	GAS T-scores, FNS, IFS (initial, follow-up, and change scores)	**Criterion validity** GAS vs FNS, $r = 0.43$, $P = .08$ GAS vs IFS, $r = 0.31$, $P = .2$	—	Mean T-scores (beginning) = 23.8, range = 16.95-42.31 Mean T-score (follow-up) = 51.6, range = 21.0-76.5 Mean T-score change=27.8 (60% > 50)
Van Stel (2002)[20]	N = 79 Age = NA Patients with moderate-severe asthma or COPD	Prospective, intervention study	GAS scores and change scores from QOL-RIQ, MPQ-LP, FET, and global-rating of change questionnaire	**Concurrent validity:** GAS vs MPQL (Spearman's r = 0.42-0.64) GAS vs QoL-RIQ ($r = 0.02$-0.13) GAS vs SPC ($r = 0.14$-0.59) GAS vs 6=min walking distance ($r = 0.58$)	**Inter-rater reliability:** Correlation between therapist/co therapist/ patients (kappa $r = 0.27$- 0.49)	SRM (GAS = 3.57, QOL-RIQ = 1.01, other outcome measures < 1.0) RE (GAS/ QOL-RIQ)=10.2

(*Continued*)

(*Continued*)

Author, year published	Sample size	Study design	Outcome measured	Validity	Reliability	Sensitivity/ responsiveness
Studies in psychology settings						
Wilier (1976)[36]	N=72	Prospective, descriptive study	GAS scores at admission, discharge, and 3mo after discharge Other outcome measures: client satisfaction, length of stay, goal specific adjustment, community adjustment, and recidivism	**Concurrent validity:** ICC between therapists' GAS scores and client's GAS scores: $r = 0.47$; $P < .01$, n = 59 ICC between therapists' GAS scores and other outcome measures, $r = -0.5$ to 0.14 ICC between clients' GAS scores and other outcome measures, $r = 0.20$-0.49; except $r = 0.04$ for recidivism	—	—
Leichsenring (2005)[22]	N = 36 Age = 35 yr, SD = 8.63 Patients with chronic psychiatric disorders of symptoms lasting >2 yr before start of psychoanalytical therapy	Prospective, intervention study	GAS scores, SCL-90R, IIP, VEV Questionnaire, Fahrenberg Questionnaire, at 5 different time points (before therapy, after 50 sessions, after 160 sessions, immediately after psychoanalysis, and 1yr postpsychoanalysis).	**Concurrent validity:** GAS change scores vs others (Pearson's *r*, GAS = 0, 55-0.56, IIP total score = 0.03-0.19, FLZ total score = 0.61-0.74)	—	ES, week 50-1yr postintervention (GAS=1.33-2.67, SCL-90-R-GSI=0.57-1.38, IIP total score 0.38-1.85, FLZ total score=0.18-1.43)
Fiester (1979)[28]	N=650	Prospective, descriptive study	Mean goal ratings of clients and therapists	—	GAS scores between client vs therapists, Pearson's r=0.46, P<.001	—

Rockwood (2006)[25]	N = 130 Age = 77 yr, SD = 7.7 Patients with mild to moderate Alzheimer's disease treated with either galantamine or placebo for 4 mo, followed by a 4-mo open-label extension during which all patients received galantamine	Randomized, controlled trial	GAS, ADAS-cog scale, CIBIC-plus, DAD, and CBS scores by clinicians and patients/caregivers	—	—	**Change from baseline to end: SRM (GAS):** Drug vs placebo (clinician), $x = 4.8$, SD = 9.6 vs × 0.9, SD = 9.5; SRM = 0.41, $P = .02$) Drug vs placebo (patients), $x = 4.2$, SD = 9.6 vs $x = 2.3$, SD = 9.0; SRM = 0.20, $P = .27$) SRM, others: ADAS-cog −0.36, P=.04; CIBIC+ scores, −0.40, P=.03; DAD scores, 0.28, P=.13; CBS scores −0.17, P=.38
Woodward (1978)[11]	N = 279 families Age = 6-16 yo Families with children who were referred for academic or behavioral problems at school	Prospective, intervention study	GAS T-scores at initial at follow-up	**Content validity:** 87% high confidence of scale scored 86% relevant to measuring change	Correlation of individual scales = 0.84 ($P < .001$) Inter-rater reliability (scores of follow-up workers) = 33-95% GAS scores as a function of follow-up worker ($F = 1.25$; $df = 6268$; $P > .10$) Correlation of post-GAS scores between 2 therapists = 0.8 ($P < .1$) No. of guides on therapist's goals	—

(*Continued*)

(*Continued*)

Author, year published	Sample size	Study design	Outcome measured	Validity	Reliability	Sensitivity/ responsiveness
					($F = 1.18$; $df = 1269$; $P > .10$) Therapist's discipline and GAS postscores ($F = 0.47$, $df = 3267$; $P > .10$).	
Asp (2006)[23]	N = 100 Age = 76 yo (SD = 7.7) Community-dwelling people	Multicenter, open-label 52-wk trial of donepezil in the treatment of mild-moderate AD	Changes in repetition, and other GAS domains (cognition, behavior, leisure, social, function), vs other outcome measures (eg, MMSE, ADAS-Cog, and CIBIC-plus) at months 3, 6, 9, and 12	**Criterion validity:** Correlation coefficients between changes in repetition and other outcome measures GAS domains (0.09-0.69) Total GAS scores (0.71-0.74) MMSE (0.09-0.37) ADAS-Cog (−0.36 to −0.10) CIBIC-Plus (−0.52 to −0.31)	—	—
Rockwood (2002)[21]	N = 100; Age X = 76 ± 8 yr; range = 51-96 Alzheimer's patients with mild-moderate dementia;	Open-label, prospective multi center study	Donepezil efficacy results on ADAS-cog Mean change from baseline to weeks 12, 24, 32, and 52 for GAS domains (global goals, cognition, function, behavior, social, leisure) of patient/carer and/or clinicians goals	**Construct validity:** Clinician global GAS score vs CIBIC+ from baseline to week 52 (Spearman's r = −0.76 to −0.82) Patient/carer global GAS scores vs CIBIC+ from baseline to week 52 ($r = -0.51$ to -0.62) GAS scores and other second efficacy measures (MMSE, ADAS-Cog, PSMS, IADL, FAQ, CSD, CES-D) ($r = 0.01$-0.56)	—	—

Shefler et al (2001)[9]	N=33 Age=24-45yo Patients suitable for Mann's time-limited psychotherapy (TLP, a short-term psychotherapy model)	Controlled, randomized outcome study based on Mann's TLP	Three individual outcome measures: GAS, TCS, and GAS Three universal outcome measures (BSI, RSE, and HSRS) Five goals: severity of symptoms, self- esteem, romantic relationship, same-sex friendships, and work performance	**Convergent validity:** Health-Sickness Rating Scale (HSRS) $r = 0.7$, $P < .001$ Target Complaints Scale $r = 0.5$, $P < .01$ BSI $r = 0.38$, $P < .05$ **Internal consistency:** GAS composite scores (pre- vs posttext), $r = 0.61$-0.78, $P < .001$ **GAS subscale scores (before vs 6-mo follow-up)** 3 subscales: $r = 0.73$-0.84 ($P < .001$) 2 subscales: $r = 0.32$-0.36 ($P < .05$)	Mean inter-rater reliability, r=0.76-0.92 (mean=0.86)	—
Kaplan (1977)[37]	N=96	General evaluation of performance of mental hospital	GAS scores at 6, 12, and 18mo after admission	Validity not established	ICC pairs of raters (follow-up scores): $r = 0.87$ ICC between 3 raters (goal guides): $r = 0.71$ (mean scores): $r = 0.91$	—

MMSE, Mini-Mental State Examination; PSMS, Lawton-Brody Physical Self-Maintenance Scale; IADL, Instrumental activities of daily living; FAQ, Functional Activities Questionnaire; CES-D, Center for Epidemiologic Studies Depression Scale; ADAS-cog, Alzheimer's Disease Assessment Scale-Cognitive; CIBIC-plus, Clinician's interview-based impressions of change-plus; GDS, Global Deterioration Scale, BCRS, Brief Cognitive Rating Scale; Bl, Barthel Index; SQLI, Spitzer Quality-of-Life Index; HABAM, Hierarchical assessment of balance and mobility; CIRS, Cumulative Illness Rating Scale; OT-SI, Occupational Therapy applying in sensory integration approach; MPQ, McGill Pain Questionnaire; ODQ, Oswestry low back pain Disability Questionnaire; NRS, Pain Intensity Numerical rating Scale; GHQ, General Health Questionnaire; PAIRS, Pain and Impairment Relationship Scale; KELS, Kohlman evaluation of living skills; MEDLS, Milwaukee evaluation of daily living; RDRS, Rappaport Disability Rating scale; ADLE-LIM, Klein-Bell activity of daily living scale, elimination; ADLMOB, Klein-Bell activity of daily living scale, mobility; FIM, Functional independence measure; PGMS, Peabody Gross Motor Scale; PFMS, Peabody Fine Motor Scale; PDMS, Peabody Developmental Motor Scale; FNS, Family Needs Survey; IFS, the Impact on Family Scale; SCL-90-R, Symptom Checklist; IIP, Inventory of interpersonal problems; FLZ, Fahrenberg Questionnaire; QOL-RIQ, Quality of Life for Respiratory Illness Questionnaire; MPQ-LP, The Medical Psychological Questionnaire for lung patients; FET, Functional exercise tolerance; OARS IADL, Katz Index of Older Americans Resource Scale for instrumental activities of daily living; Katz IADL, Katz instrumental activities of daily living; NPH, Nottingham Health Profile; SPC, self-perceived change; X, average age; COPD, chronic obstructive pulmonary disease; ICC, inter-rater correlation coefficient; AD, Alzheimer's Disease.