ACRM
AMERICAN CONGRESS OF REHABILITATION MEDICINE

Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

Archives of Physical Medicine and Rehabilitation 2016;97:157-70

SPECIAL COMMUNICATION

Proposed Criteria for Appraising Goal Attainment Scales Used as Outcome Measures in Rehabilitation Research

Agata Krasny-Pacini, MD,[a,b,c] Jonathan Evans, PhD,[d] McKay Moore Sohlberg, PhD,[e] Mathilde Chevignard, PhD[b,f]

From the [a]Outreach Department for Patients With Acquired Brain Injury, University Rehabilitation Institute Clemenceau Strasbourg, Strasbourg, France; [b]Sorbonne Universités, Univerisité Pierre et Marie Curie Paris 06 CNRS. INSERM, laboratoire d'imagerie biomédicale, Paris, France; [c]Pediatric Orthopedics Department, Hautepierre Hospital, Strasbourg University Hospitals, Strasbourg, France; [d]Institute of Health and Wellbeing, University of Glasgow, The Academic Center, Gartnavel Royal Hospital, Glasgow, United Kingdom; [e]Communication Disorders and Sciences, University of Oregon, Eugene, OR; and [f]Rehabilitation Department for Children with Acquired Brain Injury, Saint Maurice Hospitals, Saint Maurice, France.

Abstract

Goal Attainment Scaling (GAS) is a method for writing personalized evaluation scales to quantify progress toward defined rehabilitation goals. In the published literature, GAS methodology is used with different levels of rigor, ranging from precisely written GAS scales that ensure minimal bias and explicitly describe 5 levels of goal attainment to subjective ratings of goal attainment by adjectives (eg, worse/better than expected), which are transformed into a T score, wrongly giving the reader the impression of a truly standardized, interval scale. A drawback of GAS methodology is that it is highly dependent on the ability of the GAS setting team/person to generate valid, reliable, and meaningful scales; therefore, reliability and validity of GAS scales are idiosyncratic to each study. The aims of this article were to (1) increase awareness of potential sources of bias in GAS processes; (2) propose GAS quality appraisal criteria, allowing judgment of the quality of GAS methodology in individual rehabilitation studies; and (3) propose directions to improve GAS implementation to increase its reliability and validity as a research measurement tool. Our proposed quality appraisal criteria are based on critical appraisal of GAS literature and published GAS validity studies that have demonstrated that precision, validity, and reliability can be obtained when using GAS as an outcome measure in clinical trials. We recommend that authors using GAS report accurately how GAS methodology was used based on these criteria.

Archives of Physical Medicine and Rehabilitation 2016;97:157-70

Goal Attainment Scaling (GAS)[1] is a method for writing personalized evaluation scales to quantify progress toward defined goals (both practical guidelines[2-5] and literature reviews on GAS are available in the literature[6-10]). GAS produces an individualized, criterion-referenced measure of a client's goal achievement. Scores can be aggregated to quantify the extent to which a group of clients who are receiving the same type of intervention achieve their personalized rehabilitation goals. One GAS scale is written for each identified rehabilitation goal, with an emphasis on the client's participation in goal selection when possible. Success of the intervention is then quantified on an ordinal scale, typically ranging from −2 (or −3) to 2.

GAS has therefore 2 intertwined components: (1) GAS methodology is a person-centered approach in rehabilitation that emphasizes collaborative goal setting with the establishment of goals and levels of progress that are meaningful to the client; and (2) GAS is an outcome measure that can be used both in clinical work and research to assess the effectiveness of an intervention based on personally relevant goals. This article focuses on the use of GAS as an outcome measure specifically for rehabilitation efficacy research. The reader is referred elsewhere to reviews of

Supported in part by the French Society of Physical Medicine and Rehabilitation-Allergan prize.

Disclosures: none.

0003-9993/15/$36 - see front matter © 2016 by the American Congress of Rehabilitation Medicine
http://dx.doi.org/10.1016/j.apmr.2015.08.424

the literature on the clinical aspects of collaborative goal setting.[10-12]

Writing personalized scales through GAS methodology is useful in measuring rehabilitation outcomes, and use of GAS methodology is expanding in research settings, especially in areas where standard scales do not adequately capture a study participant's progress or when a standardized assessment does not exist to measure the construct. GAS methodology offers benefit in the provision of individualized, dependent variables, a critical characteristic for measuring rehabilitation effects. GAS allows use of the same 5-point scale method for all clients and therefore aggregation of results independent of goal type. Further, the goal of rehabilitation is to improve clients' activity and participation in natural contexts, but very few measures are designed to ecologically assess performance. By contrast, GAS allows the transformation of goals related to the *International Classification of Functioning, Disability and Health* (ICF) activity domains into participation goals in defined contexts where the activities occur.[13,14] Feasibility of GAS has been shown across a variety of rehabilitation fields.[15-19] GAS scales are sensitive to change when testing an intervention in rehabilitation.[15,16,20-23] GAS characteristics in terms of safety, utility, and responsiveness are therefore encouraging.

However, in the published literature, GAS methodology is used with different levels of rigor, ranging from precisely written GAS scales that ensure minimal bias and explicitly describe 5 levels of goal attainment to subjective ratings of goal attainment by adjectives (eg, worse/better than expected), which are transformed into a T score, wrongly giving the reader the impression of a truly standardized, interval scale. Although the less rigorous form of GAS methodology can be convenient, useful, fast, and practical to use in clinical practice, there is growing concern for its use as an outcome measure in clinical trials[24,25] and mixed findings as to the reliability[24,26] and validity of GAS as an outcome measure.[25]

The aims of this article are to (1) increase awareness of potential sources of bias in GAS processes; (2) propose GAS quality appraisal criteria, allowing judgment of the quality of GAS methodology in individual rehabilitation studies; and (3) propose directions to improve GAS implementation to increase its reliability and validity as a research measurement tool. This article is not addressing use of GAS in clinical setting outwith research.

Methods

A literature search using the PubMed database was conducted to ensure that our critical appraisal of the research was inclusive. The keywords *goal attainment scaling* AND *rehabilitation OR therapy* were used to identify articles published between 1990 and 2014. The search returned 179 articles. Twelve articles were excluded because an abstract was not available or because the article was not written in English. A title and abstract review was conducted to identify those articles that evaluated GAS methodology as an outcome measure. Included articles were literature reviews on GAS, GAS clinical guidelines, articles relating to GAS validity and reliability, and articles relating to training in GAS. We purposefully included articles referring to fields outside physical medicine and rehabilitation that face the same challenges in evaluating treatment efficacy as rehabilitation (especially cognitive interventions from the field of psychiatry and developmental disorders). Articles were excluded if they assessed only GAS feasibility or sensitivity to change/responsiveness, without references to its validity and reliability as an outcome measure. This yielded 36 relevant full-text articles that were reviewed to identify bias in GAS and generate the quality appraisal criteria.

List of abbreviations:

GAS	**Goal Attainment Scaling**
ICF	**International Classification of Functioning, Disability and Health**
IRR	**interrater reliability**

Results

Potential sources of bias in GAS processes and published recommendations for constructing goal attainment scales

Usual criticisms of how GAS methodology has been used include: (1) unknown clinimetric qualities of GAS scales used in a given study because of their idiosyncratic nature[25]; (2) subjective scoring, especially if not all levels of the scale are formulated or if descriptions are not precise enough; (3) risk of choosing goals that are not clinically relevant or too easy/too difficult to attain[27] and therefore do not represent a meaningful or realistic change in function; (4) ordinal (rather than interval) nature of GAS scales[28] and the lack of equidistance between GAS levels which cannot be controlled for[24]; and (5) the use of a T score that uses subjective values, especially a subjective weighting of GAS scores and a ρ coefficient assumed to be 0.3, which has not been confirmed in the literature.[6,9,28]

A major drawback of GAS methodology is that it is highly dependent on the ability of the GAS setting team/person to generate valid, reliable, and meaningful scales. It has even been proposed that GAS is more a measure of how adequately a therapist can foresee outcome than an outcome measure itself.[6,29-32] A group of clients may show progress on their GAS scale because of a measurement error, on a GAS scale that is not reliable because of poor interrater reliability (IRR), of too easy goals, of unequal distances between GAS levels, or of use of subjective criteria for goal attainment. This issue has been raised by Ruble et al[27(p3)]:

"If GAS scores are higher in the experimental conditions…one could argue that the targeted outcomes as scaled using GAS were less difficult and easier for [clients] in the experimental group to achieve compared to the control group; that skills were written in more measureable terms and thus easier to be observed and coded in the experimental groups; or that the intervals between each scaled description were unequal and favored the experimental group."

Because these potential biases can threaten reliability of results obtained through GAS, Kiresuk et al[1,33] recommended the review of GAS scales by an independent third party, and they even suggested that clients should be evaluated on 2 different sets of GAS scales developed by 2 independent research groups[1] (ie, treatment success should be independent of how the goals were formulated)[1,34] to minimize bias. Although few publications address this demanding recommendation,[35,36] it seems crucial that authors using GAS as a research outcome measure provide the reader with information on how the scales were generated and verified (and/or compared between groups on items that may impact on GAS scoring as suggested by Ruble[27]) to provide information on