



HAL
open science

Goal Attainment Scaling in rehabilitation: A literature-based update

A. Krasny-Pacini, J. Hiebel, F. Pauly, S. Godon, M. Chevignard

► To cite this version:

A. Krasny-Pacini, J. Hiebel, F. Pauly, S. Godon, M. Chevignard. Goal Attainment Scaling in rehabilitation: A literature-based update. *Annals of Physical and Rehabilitation Medicine*, 2013, 56 (3), pp.212-230. 10.1016/j.rehab.2013.02.002 . hal-01513662

HAL Id: hal-01513662

<https://hal.sorbonne-universite.fr/hal-01513662>

Submitted on 25 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Update article / Mise au point

Goal Attainment Scaling in rehabilitation: A literature-based update

Goal Attainment Scaling en rééducation : une mise au point à partir des données de la littérature

A. Krasny-Pacini^{a,*}, J. Hiebel^{a,b}, F. Pauly^{b,d}, S. Godon^b, M. Chevignard^{c,e}

^a Institut universitaire de réadaptation Clemenceau-Strasbourg, 45, boulevard Clemenceau, 67082 Strasbourg, France

^b Pôle de pédiatrie, hôpital de Hautepierre, CHU de Strasbourg, 67098 Strasbourg cedex, France

^c Laboratoire ER 6-UPMC, Eq 4 « Recherche clinique en neuropsychologie », groupe hospitalier Pitié, Paris, 47, boulevard Hôpital, 75013 Paris, France

^d Service de MPR et pédiatrie, centre hospitalier du Luxembourg, 1210 Luxembourg, France

^e Service de rééducation des pathologies neurologiques acquises de l'enfant, hôpitaux de Saint-Maurice, 14, rue du Val-d'Osne, 94410 Saint-Maurice, France

Received 26 September 2012; accepted 2 February 2013

Abstract

Goal Attainment Scaling (GAS) is a method for quantifying progress on personal goals. Turner-Stokes's guide to GAS is a method for quantifying progress towards personal goals. Turner-Stokes's guide and the use of Kiresuk's T-score are the most widely used GAS-based approaches in rehabilitation. However, the literature describes a number of other approaches and emphasizes the need for caution when using the T-score. This article presents the literature debates on GAS, variations of GAS (in terms of the score level assigned to the patient's initial status and description of the scale's different levels), the precautions to be taken to produce valid GAS scales and the various ways of analyzing GAS results. Our objective is to (i) provide clinical teams with a critical view of GAS (the application of which is not limited to a single research group's practices) and (ii) present the most useful resources and guidelines on writing GAS scales. According to the literature, it appears to be preferable to set the patient's initial level to -2 (even when worsening is a possible outcome) and to describe all five GAS levels in detail. The use of medians and rank tests appears to be appropriate, given the ordinal nature of GAS.

© 2013 Elsevier Masson SAS. Open access under [CC BY-NC-ND license](#).

Keywords: Goal Attainment Scaling; Scale; Goal; Psychometric properties; Rehabilitation

Résumé

La Goal Attainment Scaling (GAS) est une méthode qui permet d'écrire des échelles d'évaluation personnalisées. La version de GAS de Turner-Stokes et l'utilisation du T-score de Kiresuk sont les approches les plus utilisées en rééducation, alors que les données de la littérature recommandent des approches différentes et remettent en cause l'utilisation du T-score. Cet article présente ces débats de la littérature, les variantes dans l'écriture des GAS (en termes de chiffre alloué à l'état initial du patient et de description des niveaux de l'échelle), les précautions à prendre pour que les GAS rédigées soient valides et les différentes manières d'analyser les résultats des GAS. L'objectif est de procurer aux équipes une vue critique de la GAS ne se limitant pas à la pratique d'une équipe et de présenter les articles les plus utiles pour se familiariser avec la méthode. Au vu de la littérature, il paraît plus judicieux de fixer le niveau initial d'un patient à -2 (même lorsqu'une aggravation est possible), et de décrire précisément les cinq niveaux de GAS. Analyser les résultats GAS en utilisant des médianes et des tests de rang, permet de respecter le caractère ordinal des GAS.

© 2013 Elsevier Masson SAS. Cet article est publié en Open Access sous licence [CC BY-NC-ND](#).

Mots clés : Goal Attainment Scaling ; Objectif ; Échelle ; Qualité psychométriques ; Rééducation

* Corresponding author.

E-mail address: agatakrasny@yahoo.com (A. Krasny-Pacini).

1. English version

1.1. Introduction

Goal Attainment Scaling (GAS) is a method for writing personalized evaluation scales [27] in order to quantify progress toward defined goals. This approach is attracting growing interest in clinical practice because it enables assessment of a treatment's efficacy in terms of goals set by the patient him/herself (rather than on generic scales, which may not always include the problem that most severely bothers the patient). GAS is used in many fields, including medicine and especially in psychiatry, geriatrics and physical and rehabilitation medicine (PRM) – fields in which setting precise goals is a fundamental part of treatment planning. In fact, GAS can be used to cover all the fields of the International Classification of Functioning, Disability and Health (ICF) [82] by choosing goals that cover activity, participation, quality of life and environmental factors [33]. Involving the patient and his/her family and carers in the choice of treatment goals may enable better integration of these goals into activities of daily living by transforming goals related to ICF activity domain into participation goals in the patient's usual context [44]. Patients undergoing rehabilitation are more motivated when their goals are clearly defined and consistent with their life project [19,58]. Rehabilitation outcomes are better when the patient is involved in setting his/her goals [80]. In PRM departments, GAS helps to:

- plan rehabilitation programmes by setting priorities;
- structure team meetings and multidisciplinary consultations around precise objectives;
- better quantify a patient's progress;
- better communicate with the patient, his/her family and rehabilitation funding bodies [36].

Lastly, GAS can also be used to address ethical issues (resuscitation orders) [18] or to assess health care system functioning [26].

Furthermore, GAS is being increasingly used as an outcome measure in research on rehabilitation programmes or treatments for disabled people (e.g. prosthetics [49,68], occupational therapy [15], neuro-orthopaedics [2–4,14,73,78], paediatric rehabilitation [38,47,79], locomotor rehabilitation [40] and special education [43]).

Several literature reviews on GAS have been published [5,20,35,55,62]. The most recent of these (by Vu and Law [77]) cited 17 articles in the field of rehabilitation. These various reviews have covered the psychometric qualities of GAS but do not provide any concrete guidance on its application in clinical practice or research. Furthermore, they do not show how methodological differences may influence the scientific validity of GAS and they barely address the interpretation of GAS results. Hence, guides on using GAS tend to reflect the practices adopted by a small number of research groups. Indeed, Turner-Stokes' work [72] has attracted so much attention that some researchers may even gain the impression that it is the only

guide to GAS. In fact, Turner-Stokes's guide does not feature a number of important aspects of GAS published by other groups (particularly those by Tennant [69] and Steenbeek [63–65]).

The objective of the present article is to review the literature on GAS in a pragmatic way, so that interested practitioners may use the method in their clinical practice and/or research. In particular, we shall review methodological variations, the latter's influence on interpretation of the scale's psychometric qualities, the T-score's properties and debates prompted by this method.

1.2. Writing Goal Attainment Scales

Here, we shall not provide a detail description of the procedure for using GAS in patients in rehabilitation departments described by Turner-Stokes et al. [71,72] because it has been widely disseminated in recent years (including a French-language version presented at the Ipsen symposium during the French PRM Society's annual congress [9]).

Overall, GAS methodology consists in:

- defining a rehabilitation goal;
- choosing an observable behaviour that reflects the degree of goal attainment;
- defining the patient's initial (i.e. pretreatment) level with respect to the goal;
- defining five goal attainment levels (ranging from a “no change” to a “much better than expected outcome”);
- setting a time interval for patient evaluation;
- evaluating the patient after the defined time interval;
- calculating the overall attainment score for all the rehabilitation goals.

Optional extensions of this method consist in dividing long-term goals into short-term sub-goals with corresponding GAS sub-scales and giving more weight to some goals than to others.

A five-point scale is generally used: “–2” is the initial pretreatment (baseline) level, “–1” represents progression towards the goal without goal attainment, “0” is the expected level after treatment, (and therefore, the “most likely” level after treatment), “+1” represents a better outcome than expected, and “+2” is the best possible outcome that could have been expected for this goal. Since there may be several rehabilitation goals for a given patient, each goal will have its own GAS scale. Determining the rehabilitation goal is relatively easy in routine PRM practice, inasmuch as GAS is a formalization of the therapeutic objectives discussed on a daily basis with patients and their families. However, it is more difficult to draft a full goal attainment scale, i.e. to precisely describe the five attainment levels. Bovend'Eerdt's and Steenbeek's groups have focussed on how to choose the GAS levels.

Bovend'Eerdt's group [7] developed a method for easily determining the various GAS levels once the main goal has been defined. The first step consists in identifying the patient's expectations and the environmental factors influencing the performance of the activity in question (e.g. the patient's house

Table 1
Examples of GAS scales written for a child with traumatic brain injury, presenting a dysexecutive syndrome, left-side hemiparesis and impairment of the right arm (significant ulnar deviation and spasticity of the elbow) that complicates eating.

Goal attainment scale	The child's main goal: walking around at home more easily, including the staircase Observable target behaviour: walking down 10 steps of the stairs Weighting: w = 4	The child's main goal: eating on his/her own more easily Observable target behaviour: eating a bowl of mashed potatoes with a spoon, unaided Weighting: w = 2	The family's main goal for the child: being able to prepare his/her school bag Observable target behaviour: preparing the school bag Weighting: w = 1
-2	Walks down the stairs without alternating steps with one hand on the stair rail and the other held by a carer	Starts to eat a bowl of mashed potatoes unaided but cannot finish it	The school bag is prepared by the parents or the teacher. The child is unable to prepare the bag alone
-1	Walks down the stairs without alternating steps and holding the stair rail only, while supervised by a carer	Manages to eat a bowl of mashed potatoes unaided but takes more than 15 minutes	Prepares the school bag but requires constant verbal guidance from the parents or teacher
0	Walks down the stairs with alternating steps, with one hand on the stair rail and the other held by a carer	Eats a bowl of mashed potatoes in 11 to 15 minutes	The child manages to prepare the school bag thanks to a check-list of the necessary steps and under supervision, so that steps are not forgotten
+1	Walks down the stairs with alternating steps and holding the stair rail only, while supervised by a carer	Eats a bowl of mashed potatoes in 7 to 11 minutes	The child manages to prepare the school bag alone, thanks to a check-list of the necessary steps. No supervision is required and the child only occasionally forgets items
+2	Walks down the stairs unaided, with alternating steps, holding the stair rail and not supervised by a carer	Eats at an essentially normal speed, like the family's other children	The child manages to prepare the school bag alone and without a check-list and only occasionally forgets items
Level attained after 4 months	After rehabilitation with physiotherapy: walks down the stairs with alternating steps, with one hand on the stair rail and the other held by a carer (Score = 0)	After botulinum toxin injection, occupational therapy and use of a splint against ulnar deviation: is able to eat alone and finish a bowl of mashed potatoes but never in less than 15 minutes. (Score = -1)	After cognitive rehabilitation with scripts, sequencing exercises and training in following step-wise instructions for tasks: Manages to prepare the school bag alone, thanks to a check-list of the necessary steps (Score = +1)

The weighting corresponds to the child's priorities (goals 1 and 2) and the parents' priority (goal 3) at the time of the evaluation; more weight is given to the child's motor goals than to the parents' cognitive goal.

has two floors and thus the patient needs to walk up and down stairs: Table 1). The second step consists in determining the observable target behaviour corresponding to the target activity (e.g. walking down 10 steps of the stairs). In the third step, the rehabilitation team works with the patient to identify the assistance required to perform this activity: human assistance, technical aids, assistive devices, verbal guidance, cognitive assistance, etc. The fourth step consists in quantifying the initial performance at the target activity in terms of the time required, quantity (e.g. the number of steps) and frequency (e.g. frequency of falls) of the target behaviour. The five attainment levels are then written by adding or changing the "assistance required" and/or "performance quantification" categories. It is important to modify only one characteristic at a time.

This type of formulation appears to be preferable to quantitative evaluation (e.g. use of cognitive compensation less than 10% of the time, 10 to 25% of the time, 26 to 40% of the time and so on [35]). Firstly, it is rarely possible to measure activity on this type of numerical scale. Secondly, patients are not

observable all of the time (even by the main carer). Likewise, the use of visual analogue scales (VASs) for quantifying the difficulty of dressing (for instance) is not recommended, since we are not aware of literature reports in which the correlation between the difficulty of dressing rated on a VAS on one hand and according to GAS on the other has been confirmed.

Steenbeek's group sought to identify for objective, observable measurements with a view to writing scales that were as accurate as possible. Their GAS cover activities in which performance is measurable and is assumed to reflect attainment of the goal. For instance, the ability to walk on uneven ground is evaluated by a timed walk between the rungs of a ladder that simulates an uneven surface [65]; the ability to handle a joystick is evaluated in terms of the number of spaces coloured during a given time interval in a computer drawing programme [63].

GAS must meet a series of criteria that have been defined as research in this field has progressed [7,24,26,27]:

- each GAS level must be described accurately enough to allow a person not involved in the GAS-writing process to easily classify the patient at one of the GAS levels described therein;
- each scale must represent a single dimension of change;
- the levels must be measurable and thus defined in terms of observable behaviours;
- the scales must correspond to goals that are important to the patient;
- all the levels must be realistic and attainable. In particular, the +2 level must not correspond to an unexpected or miraculous goal;
- the time scale within which goals must be attained and scales must be scored should be defined in advance;
- the interlevel differences in difficulty must all be the same, i.e. it must be as difficult to go from –2 to –1, as from –1 to 0 or from 0 to +1, etc.

These criteria are based broadly on the idea that regardless of the GAS scale, all rehabilitation goals must be “SMART” [24,56]: specific, measurable, acceptable, realistic and defined in time.

Consequently, the most frequent mistakes in writing goal attainment scales are as follows:

- attainment levels that overlap or, in contrast, are not covered by any of the goals;
- unequal gaps between levels (although this problem can never be completely eliminated);
- the use of multidimensional scales (e.g. standing up and walking);
- over-simple goals, the attainment of which does not correspond to a significant clinical difference;
- subjective criteria for goal attainment (i.e. based on opinions and interviews, rather than objective, quantifiable observations).

Some GAS training methods have been proposed [63]; they show that well-trained rehabilitation staff are able to draft realistic, pertinent GAS for their patients. One of the best ways of writing a goal attainment scale is to use existing scales, such as those published as illustrative examples by experienced research groups [6,23,24,46,52,53] (see [63] in particular, for examples of initially erroneous scales that were corrected after training).

1.3. Use of Goal Attainment Scaling in research

It has been suggested that goal attainment scales designed for demonstrating the efficacy of a treatment should follow stricter rules, in order to diminish the level of subjectivity. The most reasonable proposals (which are not necessarily applied in the literature) include:

- the revision of the goals and of the GAS scales by an independent third party [27];
- attainment level scoring by a person who is not part of the team that set the goals at the outset [65];

- the use of “control goals” that are not targeted by rehabilitation [55];
- evaluation of the patients on two different attainment scales developed by independent research groups (i.e. treatment success must be independent of how the goals were formulated) [27];
- goal-setting by a group (rather than a single therapist or the patient alone), in order to avoid overly simple or unrealistic goals [26].

1.4. Expressing Goal Attainment Scaling results

Four different ways of expressing the results can be found in the literature:

- scoring each goal between –2 and +2, resulting in as many raw scores as there are scales [65] and giving a direct result for each goal, which is easily understood by the patient and easy to use in clinical practice;
- a T-score [27], which is supposed to enable GAS scores to be normalized and then analysed with parametric statistics (please refer to the section on this topic below);
- the mean of the raw scores [79], giving an overall score between –2 to +2 for the goals as a whole;
- the sum of the differences between the initial level and the attained level for each of the patient’s goals [37].

The respective advantages and drawbacks of these four methods are summarized in Table 2 and will be discussed in the last part of the article. It is important to bear in mind that the complex calculation of the T-score is not the only method that can be used with GAS: a simple –2 to +2 scale is sufficient in clinical practice because the aim is to see where the patient is with respect to the agreed goal.

The T-score is the most frequently used method [27] and expresses all the patient’s scales results as a single standardized value. Although the T-Score and the raw scores are considered to be highly correlated, this statement is based on studies with small sample sizes [23,52]. It is possible to weight the T-score by giving more weight to certain goals and thus to the scores on the corresponding scales. The T-score is calculated by applying an equation that transforms the raw scores from the individual scales into a single number.

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

where: X_i = the GAS score, W_i = the weighting of each goal attainment scale, ρ = the correlation coefficient between the various scales.

If all the goals have the same weight, this equation simplifies to [28]:

$$T = 50 + C_x \sum x_i$$

where C is a coefficient that depends on the patient’s number of scales (and thus the number of scores). C is 10 for one scale, 6.2

Table 2
Advantages and drawbacks of the different ways of expressing Goal Attainment Scaling results and the corresponding scores for the example given in Table 1.

	Advantages	Drawbacks	Expression of the results for the GAS example in Table 1
Raw scores ranging from -2 to +2 for each scale [65]	Easily understood by the patient A direct, rapid result for each goal Does not affect the ordinal nature of GAS data	Parametric statistics are not applicable. No overall score for the efficacy of the treatment	GAS #1: score = 0 GAS #2: score = -1 GAS #3: score = +1
Sum of the differences between the initial level and level attained [37]	Enables different initial levels (-1 or -2) to be taken into account Since the score is dependent on the number of goals, it is only meaningful when divided by the number of goals	Not applicable in group studies if the patients in the study group sample have different numbers of goals and thus GAS scores	Sum of the differences: $0 - (-2) + -1 - (-2) + 1 - (-2) = 6$ NB. if one divides this score by the number of scales (three, in this example), one can see that the child has improved by two attainment levels, on average
Mean of the raw scores [79]	Rapid and easy to perform during a consultation Easily understandable for the patient Independent of the number of scales	The performance of arithmetic operations on ordinal data is problematic	Mean = $[0 + (-1) + 1]/3 = 0$ On average, the goals have been attained as expected (mean 0)
T-score [27]	Most frequently used in the literature. Supposedly enables GAS scores to be normalized	The performance of arithmetic operations on ordinal data is problematic	T-score = 48.2 with the weightings chosen by the parents T-score = 50 if the same weight is allocated to all the goals

for two scales, 4.56 for three scales; 3.63 for four scales and 3.01 for five scales [28].

1.4.1. Weighting and properties of the T-score

Various weighting methods have been suggested in the literature: weighting as a function of the importance and difficulty of the goal [72] and according to the probability of attaining the goal in question [23]. Although the weighting is supposed to influence the T-score [27], weighted scores and scores in which all the goals have the same weight are correlated [52].

If one decides to use the T-score, one must bear in mind that four of its characteristics will influence the final result. Firstly, in the T-score equation, ρ represents the coefficient for the correlation between a patient’s various scales. Kiresuk suggested using a ρ value of 0.3 [27]. In principle, this value produces a standard deviation of 10 for the T-score values [27]. In reality, ρ is often lower than 0.3 because the goals and thus the goal attainment scales can belong to as many

fields as the patient wishes and are thus poorly interrelated. In contrast, some GAS from the same field can have higher ρ values. In the field of cognitive rehabilitation for example, where one expects use of a memory book to be correlated with better organizational abilities and perhaps better job-seeking abilities, Malec [35] found a value of $\rho = 0.44$. Although ρ should, in practice, be adjusted on a case-by-case basis [34], it hardly changes the value of the T-score (Table 3).

Secondly, Kiresuk and Sherman [27] postulated that the T-scores should be distributed around a value of 50 with a standard deviation of 10; this was confirmed by several studies in the 1970s [55]. Since that time, many authors have considered that use of the T-score is equivalent to normalization – however, the use of T-score does not guarantee that the score data are normally distributed [29].

Thirdly, the T-score equation is built so that the initial T-score varies according to the number of goals and scales set at the outset – even though the patients all start from the -2 level

Table 3
Illustration of the slight variations in the T-score caused by changing (i) the weight allocated to each scale and (ii) the correlation coefficient ρ , using the example from Table 1 where GAS #1 score = 0; GAS #2: score = -1; GAS #3: score = +1.

	T-score with $\rho = 0.3$ [27]	T-score with $\rho = 0.44$ [35]
No weighting i.e. with equal weights: GAS #1 w = 1; GAS #2 w = 1; GAS #3 w = 1	50.0	50.0
With weighting GAS #1 w = 4; GAS #2 w = 2; GAS #3 w = 1	48.2	48.1
Simulation with other weightings If GAS #1 w = 40; GAS #2 w = 20 and GAS #3 w = 10	48.2	48.1
If GAS #1 w = 10, GAS #2 w = 9 and GAS #3 w = 2	45.7	45.4

for all the goals (e.g. an initial T-score of 23 for a patient with three goals and a value of 30 for a patient with just one goal).

Fourthly and lastly, use of different multiples for T-score weighting will not alter the result; for example, weightings of 1, 2 and 3 will result in the same T-score as weightings of 10, 20 and 30 [27]. However, the score does not vary greatly as a function of the weighting (Table 3).

Consequently, these four characteristics imply that:

- since the true value of ρ is unknown, one can use other values [34];
- use of a T-score does not dispense with the need to check for a normal data distribution before applying parametric statistics;
- to compare two groups before and after treatment using T-scores, the practitioner must check that the groups have a similar number of goals;
- since weighting is subjective, it is rarely used in rehabilitation [35] and so the T-score can be calculated using the simplified equation:

$$\text{T-score} = 50 + Cx \sum xi$$

1.5. Variants of the Goal Attainment Scaling methodology

1.5.1. Who sets the goal and writes the Goal Attainment Scaling scales?

All scenarios have been described in the literature, with scales written by the patient [23], by the therapist [46], by the rehabilitation team [53] (with [72] or without [14] the patient's involvement), by an independent "goal selector" [27] or even an external "goal selection committee" [26]. However, it appears that goals are more likely to be attained if the patient is involved in selecting them [24,25,80].

Setting a goal before the start of treatment, defining the goal precisely and agreeing on the different attainment levels helps to transfer information and to negotiate realistic goals [35,72]. In Turner-Stokes' work [72], patients are encouraged to set their goals themselves. However, to help them do this, Turner-Stokes's group have developed a "menu" of pre-stated goals in the most common fields in rehabilitation (walking, pain, dressing, etc.), which can help the patient and the care team to formulate their own objectives.

In anosognosic patients or a patient who are not greatly aware of their difficulties, goal-setting is indeed more complicated but becomes a therapeutic process per se [35].

In paediatric rehabilitation, it appears to be essential to involve the family in goal-setting because the literature data show that children, parents, therapists and physicians have differing concerns and priorities [12,16,75,76]. Hence, there is a growing body of literature in favour of rehabilitation based on goals set by the family and the child [1,22,59].

1.5.2. What is the initial level? How can worsening be expressed?

In Turner-Stokes's method, the value corresponding to the starting level is chosen according to whether or not worsening is

possible [72]. The patient's initial status will be -2 if worsening is not possible. If worsening after treatment is plausible, the initial level will be set to -1 so that any worsening can be rated as -2 . Although this method has the advantage of enabling aggravation to be scored, it has several disadvantages. By setting the initial level at -1 , three different levels of goal attainment are defined but none corresponds to progress without attaining the goal – a situation that is frequently encountered in clinical practice. A patient who has not progressed and a patient who has progressed (but not enough to attain the expected goal at level 0) will both scored as -1 , even though their respective clinical responses to treatment differed. Furthermore, progress is measured on three or four levels depending on whether the initial level is set to -1 or -2 , respectively; this makes it difficult to compare scales.

A growing number of researchers set the initial level to -2 for all patients, in order to obtain comparable scales. This usefully enables one to measure improvement in the absence of goal attainment, although the floor effect associated with this method makes it impossible to score aggravation. Steenbeek [65] suggested adding a -3 level for expressing aggravation. However, this would prevent the calculation of a T-score with a Gaussian distribution centred on 0. In contrast, this approach may be appropriate if the T-score is not used.

1.5.3. How many levels should be described?

According to Turner-Stokes et al. [72] and other clinicians and researchers, only two levels need to be described in detail: the initial level (the patient's current status) and the expected level (the goal). The other levels are set afterwards and can be expressed as follows: the goal has been attained, as expected: 0; the patient's status has not changed: -2 ; the patient has improved and progressed towards the goal but has not attained it: -1 ; the patient has marginally surpassed ($+1$) or greatly surpassed ($+2$) the expected outcome. The disadvantage of this scoring method is subjectivity, notably when deciding between $+1$ and $+2$. Turner-Stokes et al. recommended reserving this method for clinical practice but wrote that all five levels should be precisely defined if GAS is used as an efficacy criterion in research [72]. However, this latter recommendation is not always applied [2,73]. The advantage of this method relates to the fact that describing two levels is less time-consuming.

Nevertheless, most authors [6,24,46,52,53,61,63] consider that it is essential to describe each of the five levels with a great degree of precision and to define on which task goal attainment will be evaluated and in which context (see Steenbeek et al. [63] for examples of GAS scales with detailed descriptions of these aspects [63]). The major disadvantage of this method relates to the time needed to accurately describe the five attainment levels: an average of 45 minutes, according to Steenbeek et al. [63], but just 10 to 12 minutes per scale according to other authors [6,13]. All the literature data suggest that the time required to define rehabilitation goals decreases as the rehabilitation team gains experience, including situations in which the goals have to be defined with the family [22].

The "three-milestone GAS" [29,30] falls between Turner-Stokes's method and that of Steenbeek. The objective is less

subjective (notably when differentiating between +1 and +2) but ensures that a scale can be written relatively quickly. Although this variant is still a conventional, five-level goal attainment scale, only the –2, 0 and +2 levels are precisely described prior to treatment; these levels then act as “milestones” within a continuum of possible scores from –2 to +2. If the patient’s status corresponds exactly to the description of the –2, 0 or +2 level, scoring is easy. If not, it is easy to score the patient between two “milestones”; for example, if the patient’s status is better the description for 0 but not good enough for +2, it will be scored as +1.

1.6. Other personalised goal attainment measures

It is important to differentiate between conventional GAS (which, despite the above-mentioned variations, remains a precise, codified methodology), and other personalized goal attainment measures (notably Cusick’s seven-level GAS [13], Weigl’s Modified GAS [81], Treatment Goal Attainment [60] and the Global Clinical Impression [23]) and approaches that have been referred to as GAS but use neither Kiresuk’s methodology nor a scale from –2 to +2 [42,54]. To the best of our knowledge, these methods had not been studied in terms of their psychometric qualities. In contrast, the Canadian Occupational Performance Measure [32] (COPM) is a well-studied method that is more structured than GAS. It is based on a semi-structured interview with three sections: self-care (personal care, functional mobility and community management), productivity (paid/unpaid work, household management, play/school) and leisure (quiet recreation, active recreation and socialization), during which the patient identifies his/her problems and sets treatment goals accordingly.

1.7. Precautions regarding the psychometric qualities of GAS methods described in the literature

Given that GAS variants do not all use the same initial level or do not describe their levels with the same degree of precision, the scales’ psychometric qualities cannot be compared in a valid way.

1.7.1. Interrater reliability

The interrater reliability (IRR) is described as good in literature reviews [77] but does appear to vary according to:

- the precision with which the levels are described;
- the person writing the scale [64];
- the person scoring the scale [63,65];
- the field in question [64].

Goal attainment scales written by speech therapists may have greater IRR ($\kappa = 0.92$) than those written by physiotherapists ($\kappa = 0.73$) [64]. Greater IRR is obtained when GAS scales are written by the therapist treating the patient than when the GAS is written by an outside therapist writing the scale after examining the child just for an hour [64]. This is even more true for cognitive domains ($\kappa = 0.85$ versus 0.63, respectively) than

Table 4
Interrater reliability in various rehabilitation studies.

Study	Field of rehabilitation	ICC/kappa
Rockwood [52]	Cognitive	0.97
Palisano [46]	Paediatric	0.89 (prestudy) 0.75 (study)
Steenbeek [65]	Paediatric	0.63
Bovend’Eerd [8]	Neurological	0.48
Rushton [53]	Amputees	0.67
Joyce [21]	Neurological	0.92–0.94
Steenbeek [64]	Paediatric	0.65–0.92

Interpretation of the IRR (expressed as the kappa/intraclass correlation coefficient (ICC): $\kappa \geq 0.9$: excellent; 0.9–0.71: good; 0.7–0.51: moderate; 0.5–0.31: poor; ≤ 0.3 very poor.

motor domains ($\kappa = 0.76$ versus 0.65, respectively) [64]. The IRR is moderate when one rater observes the patient directly and the other views video recordings ($\kappa = 0.61$ –0.66) [63]. Table 3 summarizes the interrater reliability in various rehabilitation areas when all five GAS levels are described in detail. To the best of our knowledge, the IRR of Turner–Stokes’s method [72] (in which only two levels are described and the others are “deduced”) has never been reported (Table 4).

1.7.2. Content validity

Palisano [46] showed that it was feasible to write GAS that had good content validity as long as the team writing the GAS had sufficient experience and had thought the attainment levels through carefully (notably in terms of the need for a single dimension per scale). However, a GAS validity depends on the way it is written and one cannot extrapolate literature data on other scales. The goal attainment scale’s validity will depend on the setter’s objectivity and ability to anticipate the range of possible outcomes [39] based on their knowledge of the pathology, of the patient’s potential and of the available therapeutic resources. Consequently, GAS results may sometimes reflect the setter’s knowledge and ability, rather than the treatment’s true efficacy [55].

1.7.3. Criterion concurrent validity

The GAS scores are poorly or not at all correlated with standard scales used in routine practice in rehabilitation in the fields of geriatrics [18], cognition [52], neurological disease [23], orthoses [53] and paediatrics [46,61] (such as the Barthel Index, for instance [18,23,53]).

GAS and the Global Clinical Impression are strongly correlated [23,52]. In contrast, Cusick et al. found that GAS was poorly correlated with the COPM [13], which nevertheless also measures the attainment of personalized objectives. The researchers suggested that this was due to the fact that the GAS took account of two parameters that were not addressed by the COPM in their study (specific arm function and behaviour).

1.7.4. Sensitivity to change

GAS scales present excellent sensitivity to change, as has been demonstrated in different populations and contexts [18,52,53]. In rehabilitation, GAS is more sensitive to change than the Barthel Index and Functional Independence Measure

are [23,51,74]. In some studies, GAS is the only method capable of detecting a change after treatment [51,78]. Standard scales sometimes fail to detect a change when the goal has been attained [61]. The main reason for this is that often the goals and attainment levels do not correspond to any of the standard scales' items [61].

1.8. Unresolved issues concerning GAS methodology: the problem of considering Goal Attainment Scaling ordinal scales as interval data

One of the criticisms of GAS scores is based on the non-interval nature of data generated [34,69]. Despite the care taken to ensure that extent of progression from one level to the next is regular, goal attainment scales are ordinal scales [35], i.e. the distances between levels are not equal. Each level to which a number is assigned (–2; –1, 0, etc.) can be considered as “more than” or “less than” the following level. In fact, each “level” represents a category of possible values; one solely knows that the “–2” level is below the “–1” level, which in turn is below the “0” level. Consequently, arithmetic operations (e.g. the calculation of means and T-scores) are not applicable to this type of data [41] and so non-parametric statistics (such as rank tests) must be used [34]. Debates concerning the respective processing of interval versus ordinal data are not limited to the field of GAS [45]: many commonly used ordinal scales assign numerical values to categories [31] and interpretation of the results can be significantly flawed, due to irregular intervals between categories [66]. Although this issue has been emphasized in literature reviews [55,62], the T-score continues to be used in both clinical practice and research.

The problem applies to all ordinal measures to which arithmetic operations are applied but is more significant when calculating GAS T-score, for several reasons. Firstly, several ordinal variables are multiplied together in the Kiresuk equation (score x weighting, etc.) [69]. Secondly, the clinical difference between the different goal attainment levels is not constant. For example, when walking down stairs, is letting go of the carer's hand as difficult as letting go of the handrail? Thirdly, the scales used for the T-score may not be from the same dimension (i.e. one scale concerns walking and another concerns sleep). When one combines these sources of variability with the T-score's dependence on the number of goals, the weightings and the choice of ρ (see above), the T-score appears to be a particularly poor metric for assessing outcome.

Recently, use of the Rasch model [50] has improved the psychometric qualities of scales and questionnaires used in rehabilitation [10,11,17]. The model linearizes ordinal scales and thus enables the use of parametric statistics. This type of process is only applicable to a well-defined scale and cannot be used with personalized scales. However, Tennant [69] simulated GAS scales in order to test them in the Rasch model. The T-scores of 300 subjects (tested on ten GAS scales and with scores weighted by the goals' importance and difficulty) were compared with T-scores in which the scores

from the same GAS had been linearized for three variables: the raw score on each scale, the goal's importance and the goal's difficulty. Tennant found that 14.7% of the T-scores differed by over 10 points when comparing the non-linearized scales (wrongly considered as linear scores for calculation of the T-score, despite being ordinal) and the scales truly linearized with the Rasch model. Importantly, a 10-point difference is clinically significant. Hence, Tennant's study demonstrated that GAS results are inaccurate when the ordinal nature of the data is not taken into account.

In contrast, Malec [35] showed that the use of parametric vs. non-parametric statistics gave similar results, but his statement is based on a less rigorous methodology: he compared correlation coefficients between GAS T-scores and the scores of other ordinal scales (used in brain injury rehabilitation) obtained by either a parametric method (yielding Pearson's correlation coefficient) or by a non-parametric method (yielding Spearman's correlation coefficient). He showed that Pearson's and Spearman's correlation coefficients were indeed quite similar. Although Malec's study is cited as justification of the use of parametric statistics [55], it does not enable one to affirm that the T-score is a valid measure.

Given that GAS is not linear, what other options are there for analyzing GAS results? Tennant [69] suggested establishing “item banks” of precalibrated goal attainment scales as a single-dimensional measure via Rasch's Differential Item Functioning [70], which enables use of the T-score. However, the disadvantage of this method relates to the loss of the truly personalized nature of GAS.

The best solution is may be that initially suggested by MacKay [34] and applied by Steenbeek's group [65]: the median [57] of the raw scores (–2 to +2) is analyzed with rank tests and non-parametric statistics. The ordinal nature of the scales is therefore taken into account. For example, in a study on botulinum toxin [65], Steenbeek used Wilcoxon's two-tailed signed-rank test to compare two groups in terms of the median score before and after treatment.

1.9. Conclusion and recommendations

Setting precise goals, describing the patient's initial status, defining the possible attainment levels and agreeing on how that goal can be attained: these steps themselves constitute a pedagogic process, that enables:

- to negotiate realistic goals;
- to discuss what is the most important for the patient and the patient's family;
- to obtain a truly informed consent for the rehabilitation plan proposed;
- to actively involve the patient and his/her family in the rehabilitation project.

In this sense, GAS is above all a tool for dialogue, patient education and formalization of the patient-caregiver contract, rather than just another dataset in our patient's already voluminous medical records.

Most studies of the psychometric qualities of GAS have analyzed attainment scales within which the five levels are described very precisely (i.e. an approach that differs from that proposed by Turner-Stokes [72]). Accordingly, we propose that the term GAS should be applied solely to precisely described, five-level scales. The validity and IRR of the Turner-Stokes method must be explored further before it can be applied. It appears more judicious to set the initial level to –2, in order to obtain comparable scales and to detect progression towards a goal in the absence of attainment (despite the issue of the floor effect). In view of the debates on the ordinal character of the goal attainment scales and the erroneous use of arithmetic calculations (such as the T-score) to interpret GAS ordinal data, it appears more sensible to perform non-parametric analyses of the data (using Steenbeek's method [65]) and to abandon the T-score. GAS must assess observable behaviours lest the ambiguity of the ratings add to the imprecision of the scales, whose construct and content validity can never be definitively evaluated due to their idiosyncratic nature.

GAS scales used in research may sometimes prompt erroneous conclusions [69]. Literature proposals of additional rules for GAS used in research should be studied in more detail and applied more rigorously. In particular, the use of control GAS (i.e. concerning goals that are not targeted by treatment and which are unlikely to be attained through generalization) can be useful in multiple baseline protocols [48,67] by monitoring the respective changes over time in GAS scores for control and target goals. Bearing in mind the wide range of IRR values reported in the literature, clinical trials using GAS as an outcome measure should evaluate their own IRR, explain which variant of GAS is being used (i.e. who sets the goals and writes the GAS scale, whether or not the scale is proofread/corrected by a different person, the number of attainment levels described, the choice of the initial level, etc.). Rehabilitation teams wishing to learn more about GAS can follow published training modules [63] and the guides issued by Bovend'Eerd et al. [7] and King et al. [24].

Disclosure of interest

The authors declare that they have no conflicts of interest concerning this article.

This work forms part of a broader project that won the SOFMER-Allergan Prize in October 2012.

2. Version française

2.1. Introduction

La Goal Attainment Scaling (GAS) est une méthode qui permet d'écrire des échelles d'évaluation personnalisées [27], afin de quantifier la réussite aux objectifs fixés. La GAS connaît un intérêt grandissant en pratique clinique car elle permet d'apprécier l'efficacité d'une prise en charge sur les objectifs choisis par le patient et non sur des échelles génériques, qui peuvent parfois omettre le problème qui est le plus important pour le patient. Elle est utilisée dans de

nombreux domaines dont en médecine, en particulier en psychiatrie, gériatrie et MPR où fixer des objectifs précis est fondamental pour planifier la prise en charge. Les échelles écrites par méthode GAS permettent de couvrir tous les champs de la Classification internationale du fonctionnement (CIF) [82] en choisissant des objectifs portant aussi bien sur l'activité que la participation, la qualité de vie ou l'environnement du patient [33]. Impliquer le patient et la famille dans le choix des objectifs, permettrait de mieux implanter les objectifs dans la vie quotidienne, en transformant les objectifs concernant le domaine activité de la CIF en objectifs de participation dans le contexte habituel du patient [44]. La motivation des patients en rééducation est accrue si les objectifs en sont clairement définis et s'ils coïncident avec leur projet de vie [19,58]. Les résultats de la rééducation sont meilleurs si les patients participent au choix des objectifs [80]. Dans les services de médecine physique et de réadaptation, la rédaction de GAS permet de:

- planifier la rééducation en fixant des priorités ;
- de structurer les réunions d'équipes (synthèses) et les consultations multidisciplinaires autour d'objectifs précis ;
- de mieux quantifier les progrès des patients ;
- de mieux communiquer avec le patient, sa famille mais aussi les organismes finançant la rééducation [36].

Enfin des GAS peuvent aussi être écrites pour les enjeux éthiques (ex. désir d'être réanimé) [18] ou pour évaluer le fonctionnement d'un service de soins [26].

La GAS est également de plus en plus utilisée en recherche comme critère de jugement dans les études visant à démontrer l'efficacité d'une rééducation ou d'une thérapeutique pour des personnes en situation de handicap: en appareillage [49,68], en ergothérapie [15], en neuro-orthopédie [2–4,14,73,78], en rééducation pédiatrique [38,47,79], en rééducation de l'appareil locomoteur [40], en éducation spécialisée [43].

Plusieurs revues de la littérature ont été publiées sur la méthodologie GAS [5,20,35,55,62], dont la plus récente écrite par de Vu et Law [77] a inclus 17 articles dans le domaine de la rééducation. Ces revues de la littérature portent sur les qualités psychométriques de la méthodologie GAS mais ne donnent pas de recommandations concrètes pour son utilisation en pratique clinique ou en recherche, ne montrent pas comment les variantes de la méthodologie influent sur sa validité scientifique et insistent peu sur les manières d'interpréter les résultats des échelles GAS. Les guides d'utilisation de la GAS, quant à eux reflètent les pratiques de certaines équipes, en particulier celle de Turner-Stokes [72], si bien que la GAS est souvent assimilée à cette publication. Pourtant le guide de Turner-Stokes n'intègre pas d'importants aspects de la GAS publiés par d'autres équipes, et en particulier celles de Tennant [69] ou Steenbeek [63–65].

Le but de cet article est de présenter les données de la littérature concernant le GAS de façon pragmatique, afin que les équipes qui le souhaitent puissent l'utiliser en clinique et en recherche. En particulier, une analyse des variantes de la méthodologie sera présentée, leur influence sur l'interprétation

des qualités psychométriques des GAS, les propriétés du T-score et les débats que cette méthodologie suscite.

2.2. Écriture des échelles GAS

La procédure d'utilisation de la GAS pour les patients hospitalisés en service de réadaptation décrite par Turner-Stokes et al. [71,72] ne sera pas rappelée ici car elle a été largement diffusée ces dernières années, notamment en version française lors des symposium IPSEN des congrès de la SOFMER [9].

D'une manière générale, la méthodologie GAS consiste à :

- définir un objectif de rééducation ;
- choisir un comportement observable témoignant du degré d'atteinte de cet objectif ;
- définir le niveau initial (avant traitement) du patient vis-à-vis de cet objectif ;
- définir cinq niveaux d'atteinte de cet objectif, correspondants à une progression de « pas de changement » à « meilleur résultat espéré » ;
- fixer un délai pour évaluer le patient sur cet objectif ;
- évaluer le patient après le délai fixé et calculer un score global d'atteinte des objectifs de rééducation.

Les optionnels de la méthodologie consistent à :

- diviser les objectifs de long terme en sous-objectifs atteignables à court-terme (avec des sous-échelles GAS correspondantes aux sous-objectifs) ;
- à pondérer les objectifs en accordant plus de poids à certains d'entre eux.

L'échelle est classiquement écrite en cinq points : « -2 » est le niveau initial (avant traitement), « -1 » représente la progression vers l'objectif sans que celui-ci ne soit atteint. « 0 » est le niveau attendu après traitement donc le niveau « le plus probable » après traitement. « +1 » représente un objectif réussi mieux que prévu et « +2 » le meilleur résultat que l'on pouvait espérer par rapport à cet objectif. Une échelle GAS mesure l'atteinte d'un objectif. Comme les objectifs de rééducation peuvent être multiples, autant d'échelles GAS que d'objectifs seront rédigés pour chaque patient. Déterminer les objectifs de rééducation est relativement aisé en pratique courante en rééducation. En ce sens l'écriture de GAS est en fait la modélisation des contrats discutés au quotidien avec les patients et leurs familles lors de l'élaboration d'un projet thérapeutique. Cependant il est plus difficile de rédiger une échelle GAS complète c'est-à-dire de décrire avec précision les cinq niveaux d'atteinte de cet objectif. Les équipes de Bovend'Eerdt et de Steenbeek se sont particulièrement penchées sur la manière de choisir les niveaux des GAS :

L'équipe de Bovend'Eerdt [7] a développé une méthode pour trouver facilement les différents niveaux des échelles GAS une fois que l'objectif principal est défini. La première étape consiste à identifier les attentes du patient ainsi que les

facteurs environnementaux influençant la réalisation de l'activité sur laquelle porte l'objectif (ex. domicile du patient étant sur deux niveaux, d'où nécessité de pouvoir descendre les marches, [Tableau 1](#)). La deuxième étape consiste à déterminer le comportement cible observable correspondant à l'activité cible (ex. descendre dix marches). Dans la troisième étape l'équipe identifie avec le patient les aides nécessaires pour réaliser cette activité : moyens humains, aides techniques, guidance verbale, aide cognitive, compensations (ex. utilisation d'une rampe, aide d'un adulte)... La quatrième étape consiste à quantifier la performance initiale lors de l'activité cible en termes de temps nécessaire, « quantité » (ex. nombre de marches), fréquence (ex. fréquences de chutes) du comportement cible. Les cinq niveaux de l'échelle sont alors rédigés en supprimant ou modifiant les catégories « aides nécessaires » et/ou « quantification de performance ». Il est important de n'en modifier qu'une caractéristique à la fois.

Ce type de formulation semble préférable à celles s'efforçant d'être quantitatives (ex. utilisation d'une compensation cognitive moins de 10 % du temps, entre 10 et 25 % du temps, entre > 25 à 40 % du temps... [35]). En effet les activités sont rarement mesurables avec une telle précision numérique et les patients ne sont pas observable 100 % du temps, même par l'aidant principal. De même l'utilisation d'échelles visuelles analogiques (EVA), pour quantifier par exemple la difficulté de l'habillage, est à déconseiller car la réelle difficulté de l'habillage sur l'EVA et la GAS n'ont pas de correspondances fiables établies dans la littérature à notre connaissance.

L'équipe de Steenbeek a recherché des mesures objectives et observables pour rédiger des échelles les plus précises possibles. Leurs GAS portent sur des activités où la performance est mesurable et dont on présume qu'elles reflètent l'atteinte de l'objectif. Par exemple, la difficulté de marche en terrain irrégulier est évaluée par un parcours chronométré entre les barreaux d'une échelle qui simulent le terrain irrégulier [65] ; les capacités à manier un joystick sont évaluées sur un programme informatique de dessin par le nombre d'espaces coloriés en un temps donné [63].

La rédaction d'une échelle GAS doit répondre à une série de règles construites au fur et à mesure de la recherche utilisant la GAS [7,24,26,27] :

- chaque niveau doit être décrit suffisamment précisément pour qu'une personne n'ayant pas participé à la rédaction de l'échelle puisse sans difficulté classer le patient dans l'un des niveaux décrits ;
- chaque échelle doit représenter une seule dimension du changement ;
- les niveaux doivent être mesurables, donc définis en termes de comportements observables ;
- les échelles doivent correspondre à des objectifs importants aux yeux du patient ;
- tous les niveaux doivent être réalistes et atteignables et notamment le niveau +2 ne doit pas correspondre à un but inespéré ou miraculeux ;

Tableau 1

Exemples de GAS, écrites pour un enfant traumatisé crânien présentant un syndrome dysexécutif, une hémiparésie gauche avec une atteinte du membre supérieur droit (inclinaison ulnaire importante et une spasticité du coude) le gênant pour manger.

Niveau GAS	Objectif principal de l'enfant : circuler plus facilement dans son domicile, y compris dans les escaliers Comportement cible observable : descendre 10 marches d'escaliers Pondération : w = 4	Objectif principal de l'enfant : manger seul plus facilement Comportement cible observable : manger un bol de purée avec une cuillère sans aide Pondération : w = 2	Objectif principal de la famille : avoir toutes les affaires nécessaires à l'école sans devoir constamment préparer les affaires pour lui Comportement cible observable : préparation du cartable Pondération : w = 1
-2	Descend marche par marche sans alternance, une main sur la rampe, une main tenue par un aidant	Commence à manger seul un bol de purée mais n'arrive pas à finir	Le cartable est préparé par les parents ou la maîtresse. Il est incapable de le préparer seul
-1	Descend marche par marche sans alternance, en utilisant la rampe seule, sous la supervision d'un aidant	Arrive à manger seul et finir un bol de purée mais nécessite plus de 15 minutes	Prépare son cartable lui-même mais en étant constamment guidé verbalement par ses parents ou sa maîtresse
0	Descend marche par marche avec alternance, une main sur la rampe, une main tenue par un aidant	Mange le bol de purée en 11 à 15 minutes	Arrive à préparer son cartable grâce à une check-list des étapes nécessaires et sous surveillance pour ne pas omettre d'étapes
+1	Descend marche par marche avec alternance, en utilisant la rampe seule, sous la supervision d'un aidant	Mange le bol de purée en 7 à 11 minutes	Arrive à préparer son cartable seul grâce à une check-list des étapes nécessaires. Pas de surveillance nécessaire et il ne manque des affaires que occasionnellement
+2	Descend de façon autonome avec alternance, en utilisant la rampe seule, sans la supervision d'un aidant	Mange à un rythme quasi normal, comme les autres enfants de la fratrie	Autonome, prépare seul son cartable sans check-list et il n'y manque des affaires que occasionnellement
Niveau atteint au bout de 4 mois	Après rééducation en kinésithérapie : Descend marche par marche avec alternance, une main sur la rampe, une main tenue par un aidant (Score = 0)	Après injection de toxine botulinique, rééducation en ergothérapie et appareillage par attelle d'alignement ulnaire : Arrive à manger seul et finir un bol de purée mais nécessite plus de 15 minutes (Score = -1)	Après rééducation cognitive avec exercices de scripts, séquençage et suivi d'instruction en étapes sur différentes tâches : arrive à préparer son cartable seul grâce à une check-list des étapes nécessaires (Score = +1)

Les pondérations correspondent aux priorités de l'enfant (objectif 1 et 2) et des parents (objectif 3) au moment de l'évaluation avec une pondération priorisant les objectifs moteurs l'enfant plutôt que l'objectif cognitif des parents.

- le délai dans lequel les objectifs doivent être atteints et donc les échelles cotées doit être défini à l'avance ;
- l'intervalle entre les niveaux doit être le même entre les différents niveaux, c'est-à-dire qu'il doit être aussi difficile de passer de -2 à -1, que de -1 à 0, que de 0 à +1...

Ces critères s'appuient largement sur l'idée qu'indépendamment de l'échelle GAS, tout objectif de rééducation doit être « SMART » [24,56] : spécifique, Mesurable, Acceptable, Réaliste, définis dans le Temps.

Par conséquent, les erreurs les plus fréquentes dans la rédaction des GAS sont :

- des niveaux se chevauchant entre eux ou au contraire une performance n'étant couverte par aucun des niveaux ;
- une difficulté inégale entre les niveaux (cette erreur ne pouvant jamais être éliminée totalement) ;
- des échelles multidimensionnelles (ex. se lever et marcher) ;
- des objectifs trop simples, dont l'atteinte ne représente pas une différence clinique significative ;
- des critères d'atteinte de l'objectif subjectifs, basés sur des impressions, des entretiens plutôt que sur des mesures objectives et observables.

Des méthodes d'entraînement à la GAS ont été publiées [63] montrant que les paramédicaux bien entraînés sont capables d'écrire des GAS réalistes et pertinentes pour les patients qu'ils prennent en charge. L'une des meilleures manières d'écrire des GAS est de s'inspirer de celles déjà existantes et publiées à titre d'exemples par les équipes en ayant l'expérience : [6,23,24,46,52,53] (voir en particulier [63] pour des exemples de GAS initialement erronés et corrigés après formation).

2.3. Goal Attainment Scaling en recherche

Il a été proposé que les échelles GAS élaborées dans le but de démontrer l'efficacité d'un traitement obéissent à des règles plus strictes afin de diminuer leur subjectivité. Les propositions les plus raisonnables, mais pas forcément appliqués dans la littérature, incluent :

- la révision des objectifs et des échelles GAS par une personne indépendante [27] ;
- la cotation des échelles par un cotateur indépendant de l'équipe qui a fixée les objectifs au départ [65] ;
- l'utilisation d'« objectifs contrôlés » non visés par la rééducation [55] ;

- l'évaluation des patients sur deux groupes de GAS élaborés par des groupes indépendants (la réussite d'un traitement devant être indépendante de la manière dont ont été formulés les objectifs) [27] ;
- le choix des objectifs par une équipe plutôt qu'un seul thérapeute ou le patient seul afin d'éviter de fixer des objectifs trop faciles ou au contraire irréalistes [26].

2.4. Expression des résultats

Quatre moyens d'exprimer les résultats sont retrouvés dans la littérature :

- cotation entre -2 et +2 de chaque échelle séparément aboutissant à autant de scores bruts que d'échelles [65] et donnant un résultat direct de chaque objectif, facilement compréhensible pour le patient et simple d'utilisation en pratique clinique ;
- T-score [27] permettant en théorie de normaliser les scores pour y appliquer des statistiques paramétriques (cf. paragraphe y dédié) ;
- moyenne des scores bruts [79] donnant une note globale de -2 à +2 pour l'ensemble des objectifs d'un patient ;
- somme des différences entre niveau initial et niveau atteint de chacune des échelles d'un patient [37].

Les avantages et inconvénients de ces méthodes sont résumés dans le **Tableau 2** et seront discutés dans la dernière partie de cet article. Il est important de souligner que le calcul complexe du T-score n'est pas obligatoire pour utiliser la méthode GAS et la cotation simple de -2 à +2 de chaque échelle est suffisante en pratique clinique, puisque l'objectif est de voir où en est le patient par rapport à l'objectif fixé.

Le calcul du T-score représente la méthode la plus utilisée. Le T-score [27] regroupe en un chiffre standardisé les résultats à toutes les échelles écrites pour un patient. Il est admis que le T-Score et les scores bruts sont hautement corrélés mais cette affirmation repose sur des études à petits effectifs [23,52]. Le T-score peut être éventuellement pondéré en accordant plus de poids à certains objectifs et donc au résultat de l'échelle qui s'y rapporte. Ce T-score est généré par une formule qui transforme les scores bruts des échelles individuelles en une seule valeur numérique.

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

avec : X_i = score GAS, W_i = pondération de chaque échelle GAS, ρ = coefficient de corrélation entre les scores des différentes GAS ;

Si tous les objectifs ont le même poids cette formule peut être simplifiée [28] :

$$T = 50 + Cx \sum x_i$$

où C est un coefficient qui dépend du nombre d'échelles (et donc de scores) pour le patient. C = 10 si une seule échelle est écrite, C = 6,2 si deux échelles, C = 4,56 si trois échelles ; C = 3,63 si quatre échelles ; C = 3,01 si cinq échelles [28].

2.4.1. Pondération et propriétés du T-score

Différents schémas de pondération ont été proposés dans la littérature : pondération selon l'importance et la difficulté de l'objectif [72] ; selon la probabilité d'atteindre l'objectif [23]. La pondération est supposée influencer le T-score [27] mais en

Tableau 2

Avantages et inconvénients des différentes manières d'exprimer les résultats aux GAS et calcul par chacune des méthodes des scores GAS du **Tableau 1**.

	Avantages	Inconvénients	Expression des résultats pour l'exemple du Tableau 1
Scores bruts de -2 à +2 pour chaque échelle [65]	Facilement compréhensible pour le patient Résultat direct et rapide pour chaque objectif Préserve la nature ordinale des données GAS	Statistiques paramétriques non applicables Pas de score global pour le traitement	GAS n° 1 : score = 0 GAS n° 2 : score = -1 GAS n° 3 : score = +1
Somme des différences entre niveau initial et niveau atteint [37]	Permet d'intégrer différents niveaux initiaux (-1 ou -2) Score dépendant du nombre d'objectifs, ne permettant pas de se représenter la signification du chiffre obtenu à moins de le diviser par le nombre d'objectifs	Non applicable dans des études de groupe si dans l'échantillon tous les patients n'ont pas le même nombre d'objectifs et de GAS correspondantes	Somme des différences = $0 - (-2) + -1 - (-2) + 1 - (-2) = 6$ NB. si on divise ce score par le nombre de GAS (ici 3), on voit que l'enfant a en moyenne progressé de 2 niveaux GAS
Moyenne des scores bruts [79]	Rapide et facile à réaliser en consultation Facilement compréhensible pour le patient Indépendant du nombre d'échelles	Problématique d'utiliser des opérations arithmétiques sur des données ordinales	Moyenne = $[0 + (-1) + 1]/3 = 0$ En moyenne les objectifs ont été atteints comme prévu (moyenne 0)
T-score [27]	Le plus utilisé dans la littérature Permet en théorie de normaliser les scores	Problématique d'utiliser des opérations arithmétiques sur des données ordinales	T-score = 48,2 avec les pondérations choisies par les parents T-score = 50 si le même poids est alloué à tous les objectifs

Tableau 3

Illustration des faibles variations du T-score en changeant le poids alloué à chaque échelle et le coefficient de corrélation ρ , à partir de l'exemple du Tableau 1 où, GAS n° 1 : score = 0 ; GAS n° 2 : score = -1 ; GAS n° 3 : score = +1.

	T-score si $\rho = 0,3$ [27]	T-score si $\rho = 0,44$ [35]
<i>Sans pondération</i>	50,0	50,0
c.-à-d. si poids égaux : GAS n° 1 w = 1 ; GAS n° 2 w = 1 ; GAS n° 3 w = 1		
<i>Avec pondérations</i>		
GAS n° 1 w = 4 ; GAS n° 2 w = 2 ; GAS n° 3 w = 1	48,2	48,1
<i>Simulation avec d'autres pondérations</i>		
Si GAS n° 1 w = 40 ; GAS n° 2 w = 20 ; GAS n° 3 w = 10	48,2	48,1
Si GAS n° 1 w = 10 ; GAS n° 2 w = 9 ; GAS n° 3 w = 2	45,7	45,4

pratique les T-scores pondérés et non pondérés (c'est-à-dire où tous les objectifs ont le même poids) sont corrélés [52].

Si l'on décide d'utiliser le T-score, il faut bien avoir à l'esprit quatre caractéristiques de la formule du T-score, qui vont influencer le résultat final. Premièrement, dans la formule du T-score, le ρ représente le coefficient de corrélation des scores des différentes échelles du patient entre elles. Kiresuk propose $\rho = 0,3$ comme valeur pour l'intercorrélation des échelles GAS entre elles pour le calcul du T-score [27]. Cela est une valeur présumée permettant d'obtenir un écart-type de dix aux valeurs de T-score [27]. En réalité le ρ est souvent plus bas car les objectifs et donc les échelles GAS pouvant appartenir à autant de domaines que le patient le souhaite, elles sont très peu corrélées entre elles. À l'inverse, des GAS appartenant à un même domaine peuvent avoir des ρ plus élevés (ex. : en rééducation cognitive, où il est espéré que l'utilisation d'un carnet-mémoire sera corrélée à de meilleures capacités d'organisation et éventuellement de recherche d'emploi, Malec [35] avait trouvé $\rho = 0,44$). En pratique le ρ devrait être ajusté au cas par cas [34] mais il change peu la valeur du T-score (Tableau 3).

Deuxièmement, Kiresuk et Sherman [27] avaient postulé que les T-scores seraient normalement distribué autour de 50 avec un écart-type de 10, ce qui a été confirmé dans plusieurs études des années 1970 [55]. Depuis, pour beaucoup d'auteurs l'utilisation du T-score équivaut à une normalisation alors que l'utilisation de T-score ne garantit pas une distribution normale des scores.

Troisièmement, la formule du T-score est construite de telle manière que selon le nombre d'objectifs et d'échelles GAS fixées au départ, le T-score initial varie alors même que les patients partent tous du niveau -2 pour tous les objectifs (ex. T-score initial = 23 pour un patient ayant trois objectifs, versus T-score = 30 pour celui qui n'en a qu'un).

Enfin quatrièmement, les T-scores seront identiques, quel que soit le multiple utilisé pour les pondérations (ex. des pondérations à 1, 2, 3 ou à 10, 20, 30 aboutiront au même T-score) [27] mais dans tous les cas le score varie peu en fonction des pondérations (Tableau 3).

Par conséquent ces quatre caractéristiques impliquent que :

- la valeur vraie du ρ étant inconnue, on peut utiliser d'autres valeurs [34] ;

- l'utilisation de T-score ne dispense pas d'en vérifier la distribution normale avant d'y appliquer des statistiques paramétriques ;
- pour comparer deux groupes avant et après traitement en utilisant les T-scores, il faut s'assurer que les groupes sont comparables en nombre d'objectifs ;
- la pondération étant subjective, elle est peu utilisée en rééducation [35] et par conséquent le T-score peut être calculé par la formule simplifiée :

$$\text{T-score} = 50 + Cx \sum x_i$$

2.5. Variantes possibles dans la méthodologie Goal Attainment Scaling

2.5.1. Qui choisit les objectifs et écrit les échelles Goal Attainment Scaling ?

Tous les cas de figure sont décrits dans la littérature : GAS écrites par le patient [23], par son thérapeute [46], par l'équipe de rééducation [53] avec [72] ou sans [14] participation du patient, par un « goal selector » indépendant du thérapeute [27] ou même un « comité de sélection des objectifs » extérieur au service [26]. Il semble que les objectifs ont plus de probabilité d'être atteints si les patients sont impliqués dans leur sélection [24,25,80]. Fixer un objectif avant le début d'une intervention, le définir précisément et se mettre d'accord sur les différents niveaux que l'on peut espérer atteindre aide au transfert d'informations et à la négociation d'objectifs réalistes [35,72]. Chez Turner-Stokes [72], les patients sont encouragés à identifier eux-mêmes leurs objectifs, mais pour les aider, l'équipe de Turner-Stokes a développé un « menu » d'objectifs préverbalisés dans les domaines les plus fréquents en rééducation (marche, douleur, habillage...) qui peuvent aider le patient et l'équipe à formuler leurs propres objectifs.

Chez les patients anosognosiques ou du moins peu conscients de leur difficultés, la rédaction de GAS est certes difficile mais elle devient un processus thérapeutique en soi [35]. En rééducation pédiatrique, il paraît crucial d'impliquer la famille dans le choix des objectifs car la littérature montre que les préoccupations et domaines d'importance ne sont pas les mêmes chez les enfants, leurs parents, leurs thérapeutes et leurs

médecins [12,16,75,76] et une littérature croissante appuie la rééducation basée sur les objectifs de la famille et de l'enfant [1,22,59].

2.5.2. *Quel est le niveau initial ? Comment exprimer une aggravation ?*

La méthode de Turner-Stokes consiste à choisir le chiffre correspondant au niveau de départ selon qu'une aggravation est possible ou non [72]. L'état initial du patient correspondra au niveau -2 s'il n'existe pas d'aggravation possible. En revanche, pour tous les cas où une aggravation est possible, le niveau initial sera fixé à -1 afin de permettre de chiffrer une aggravation, cotée alors à -2. L'avantage de la méthode est certes de permettre d'intégrer dans les scores finaux les aggravations possibles, cependant elle pose plusieurs problèmes. En fixant le niveau initial à -1, trois différents niveaux de réussite d'objectif sont définis mais aucun niveau ne correspond au progrès sans atteinte de l'objectif, qui est une situation clinique pourtant fréquente. Un patient n'ayant pas progressé et un patient ayant progressé mais pas au point de réussir l'objectif attendu (niveau 0) auront tout deux un score de -1 alors que leur réponse au traitement a bien été différente cliniquement. Par ailleurs, selon que le niveau initial soit fixé à -1 ou -2, le progrès est mesuré sur trois ou quatre niveaux, rendant les différentes échelles peu comparables entre elles.

De plus en plus d'études fixent le niveau initial à -2 pour tous les patients afin d'obtenir des échelles comparables. L'avantage est de détecter une amélioration quand l'objectif n'a pas été atteint mais l'effet plancher de cette méthode empêche d'exprimer une aggravation. Steenbeek [65] a proposé de rajouter un niveau -3 exprimant l'aggravation. Mais dans ce cas le T-score, basé sur une répartition gaussienne centrée par 0 n'est plus applicable. En revanche, cette approche convient si une autre méthode d'expression des résultats autre que le T-score est utilisée.

2.5.3. *Combien de niveaux faut-il décrire ?*

Pour l'équipe de Turner-Stokes [72], et pour beaucoup d'auteurs aussi bien en pratique clinique qu'en recherche, il suffit de décrire deux niveaux : le niveau initial et le niveau attendu. En d'autres termes, il faut décrire précisément l'état actuel du patient et l'objectif. Les autres niveaux sont retrouvés à posteriori, et exprimés sous la forme suivante : l'objectif a été atteint comme attendu : 0 ; l'état du patient n'a pas changé : -2 ; le patient s'est amélioré et a progressé par rapport à l'objectif fixé mais sans l'atteindre : -1 ; le patient a réussi mieux que l'objectif attendu : un peu mieux (+1) voire beaucoup mieux (+2). L'inconvénient de ce mode de cotation est la trop grande subjectivité dans la cotation, notamment entre les niveaux +1 et +2. Les auteurs préconisaient que cette méthode soit réservée à la pratique clinique mais que les cinq niveaux soient précisément définis si les GAS sont utilisés comme critères d'efficacité en recherche [72]. Cependant ce postulat n'est pas toujours appliqué en recherche [2,73]. L'avantage est bien entendu le temps puisque décrire deux niveaux est moins chronophage.

Cependant pour la plupart des auteurs, [6,24,46,52,53,61,63] il faut impérativement décrire avec grande précision chacun des

5 niveaux, mais aussi fixer sur quelle tâche exacte et dans quel contexte particulier sera évalué le patient au moment de la cotation (voir Steenbeek pour des exemples de GAS décrites avec ces précisions [63]). L'inconvénient majeur de cette méthode est le temps nécessaire pour rédiger une échelle : 45 minutes en moyenne chez Steenbeek [63], mais d'autres auteurs arrivent à décrire les cinq niveaux en dix à 12 minutes par échelle [6,13]. L'ensemble de la littérature suggère que le temps nécessaire pour définir les objectifs de rééducation décroît au fur et à mesure que l'équipe gagne en expérience, y compris lorsque les objectifs doivent être choisis avec la famille [22].

La « GAS 3 bornes » [29,30] est intermédiaire entre la méthode de Turner-Stokes et celle de Steenbeek. Le but est d'arriver à être moins subjectif (notamment dans la différenciation entre niveau +1 et +2) tout en rendant la méthode faisable en termes de temps nécessaire pour écrire une échelle. La GAS 3 bornes reste une GAS classique en 5 niveaux mais seuls les trois niveaux : -2, 0 et +2, sont précisément décrits avant le traitement et agissent comme des « bornes » au sein d'un continuum de résultats possibles de -2 à +2. Si l'état du patient correspond à la description des niveaux -2, 0 ou +2 la cotation est aisée. Sinon il est facile de situer le patient entre deux « bornes » (par exemple si son état est meilleur que la description du 0 mais ne correspond pas au +2, son score sera de +1).

2.6. *Autres mesures personnalisées d'atteinte des objectifs*

Il est important de différencier la GAS classique, qui malgré ces variations, reste une méthodologie précise et codifiée, d'autres mesures personnalisées quantifiant la réussite d'un objectif, notamment la GAS en sept niveaux de Cusick [13], la Modified GAS de Weigl [81], le Treatment Goal Attainment [60], l'Impression Clinique Globale [23] ainsi que les échelles se faisant appeler « Goal Attainment Scale » mais qui n'utilisent ni la méthodologie de Kiresuk ni la gradation de -2 à +2 [42,54]. À notre connaissance, ces méthodes n'ont pas fait l'objet d'études concernant leur qualité psychométriques, contrairement à la Canadian Occupational Performance Measure [32] (COPM) : la COPM est une mesure bien étudiée dans la littérature, qui est plus structurée que la GAS. Elle repose sur un entretien semi-structuré comprenant trois sections : hygiène personnelle et mobilité, productivité (activité salariée, gestion du lieu de vie, école et jeux pour les enfants) et loisirs pour lesquels les patients identifient leurs problèmes et aboutissent à des objectifs de traitement.

2.7. *Précautions quant aux qualités psychométriques des échelles GAS dans la littérature*

En sachant que les variantes de GAS n'utilisent pas toutes le même niveau initial et qu'elles ne décrivent pas toutes avec la même précision les niveaux des échelles, leurs qualités psychométriques sont peu comparables.

2.7.1. *Fidélité interjuge*

En particulier la fidélité interjuges (FIJ) décrite comme bonne dans les revues de la littérature [77], varie avec la précision de

Tableau 4
Fidélités interjuges dans les différentes études en rééducation.

Étude	Domaine de rééducation	ICC ou κ
Rockwood [52]	Cognitive	0,97
Palisano [46]	Pédiatrique	0,89 (préétude)
		0,75 (étude)
Steenbeek [65]	Pédiatrique	0,63
Bovend'Eerdts [8]	Neurologique	0,48
Rushton [53]	Amputés	0,67
Joyce [21]	Neurologique	0,92–0,94
Steenbeek [64]	Pédiatrique	0,65–0,92

Interprétation des FIJ par intraclass corrélation (ICC) et coefficient kappa (κ) : $\geq 0,9$: très bonne ; 0,9–0,71 : bonne ; 0,7–0,51 : modérée ; 0,5–0,31 : médiocre ; $\leq 0,3$ très mauvaise.

description des niveaux, en fonction de qui rédige les échelles [64], qui les cote [63,65] et en fonction du domaine mesuré [64]. Les GAS des orthophonistes auraient de meilleures FIJ ($\kappa = 0,92$) que celles des kinésithérapeutes ($\kappa = 0,73$) [64]. De meilleures FIJ sont obtenues quand les GAS sont rédigées par le thérapeute traitant le patient que par un paramédical extérieur rédigeant les GAS en se basant sur une heure d'examen de l'enfant [64], et cela est d'autant plus vrai dans les domaines cognitifs ($\kappa = 0,85$ versus 0,63) que moteurs ($\kappa = 0,76$ versus 0,65) [64]. La FIJ est modérée lorsqu'une des cotation est réalisée à partir d'enregistrements vidéos et l'autre à partir de l'observation directe du patient ($\kappa = 0,61$ –0,66) [63]. Le Tableau 3 résume les FIJ dans différents domaines de la rééducation dans le cas où les cinq niveaux de GAS sont précisément écrits. À notre connaissance aucune étude n'a testé la FIJ de la méthode de Turner-Stokes [72] pour des échelles où seuls deux niveaux sont décrits et les autres « déduits » (Tableau 4).

2.7.2. Validité de contenu

Palisano [46] a montré la faisabilité de rédiger des GAS ayant une bonne validité de contenu, si l'équipe les fixe à une expérience suffisante et si les GAS sont réfléchies rigoureusement, notamment en termes d'unidimensionnalité. Cependant, la validité des GAS dépend de la manière dont elles sont écrites et on ne peut assumer leur validité sur les données publiées d'autres GAS. La validité de l'échelle GAS va dépendre de l'objectivité de celui qui la rédige et de son habileté à anticiper un panel de résultats possibles [39], selon les connaissances qu'il a de la pathologie, des possibilités du patient, et des moyens thérapeutiques à disposition. Par conséquent, les résultats aux échelles GAS peuvent refléter autant, voire davantage, les connaissances et les capacités de ceux qui les rédigent que la réelle efficacité d'une prise en charge [55].

2.7.3. Validité Concurrente contre critère

Les résultats aux GAS sont peu ou pas corrélés avec les échelles standard, utilisées en pratique courante en rééducation gériatrique [18], en rééducation cognitive [52], en rééducation neurologique [23], en appareillage [53] et en rééducation pédiatrique [46,61], tel que le Barthel [18,23,53] par exemple.

GAS et Impression clinique globale sont fortement corrélés [23,52]. En revanche, dans l'étude de Cusick, la GAS était peu corrélée avec la COPM [13], qui mesure pourtant aussi l'atteinte d'objectifs personnalisés. Les auteurs l'expliquent par le fait que les GAS de cette étude prenaient en compte deux mesures non testées par la COPM : la fonction spécifique du membre supérieur et le comportement.

2.7.4. Sensibilité au changement

La GAS présente une excellente sensibilité au changement, démontrée dans différentes populations et contextes [18,52,53]. En réadaptation, la GAS est plus sensible au changement que l'index de Barthel et que la Mesure d'Indépendance Fonctionnelle [23,51,74]. Dans certaines études, la GAS est la seule échelle capable de détecter un changement après traitement [51,78]. Les échelles standard ne montrent parfois aucun changement alors que l'objectif GAS est atteint [61]. La raison principale est que les objectifs et GAS fixés ne correspondent souvent à aucun des items des échelles standard [61].

2.8. Débats non résolus de la méthodologie Goal Attainment Scaling : la problématique des échelles ordinales considérées comme intervallaires

Une des critiques des scores GAS concerne leur caractère non intervallaire [34,69]. Malgré toute la rigueur employée pour que la progression d'un niveau à l'autre soit stable, les échelles GAS sont des échelles ordinales [35], c'est-à-dire que la distance entre chaque niveau n'est pas égale. Chaque niveau, auquel est assigné un chiffre (–2 ; –1 ; 0 ; . . .), peut être considéré comme « plus que » ou « moins que » le niveau suivant. Chaque « niveau » représente en fait plutôt une catégorie de valeurs possibles dont on sait juste que le niveau –2 est inférieur au niveau –1, qui est inférieur au niveau 0 . . . Par conséquent les calculs arithmétiques (calcul de moyennes, T-scores . . .) ne sont pas applicables à ce type de données [41] et il faut alors utiliser des statistiques non paramétriques tels des tests de rang [34]. Le débat concernant le traitement des données intervallaires versus ordinales ne se limite pas à la GAS [45] : beaucoup d'échelles ordinales communément utilisées donnent des valeurs numériques à des catégories [31] et l'interprétation des résultats peut être largement faussée par l'inégalité de difficulté pour passer d'une catégorie à l'autre [66]. Cette problématique a été soulignée dans les revues critiques de la littérature [55,62] mais les T-score continuent à être employés en clinique comme en recherche.

Toutes les mesures ordinales auxquelles on applique des opérations arithmétiques présentent ce problème mais le problème est plus important dans le calcul du T-score GAS pour plusieurs raisons :

- plusieurs variables ordinales sont multipliées entre elles dans la formule de Kiresuk (score x pondération . . .) [69] ;
- la différence clinique n'est pas constante entre les différents niveaux de la GAS (ex. est-il aussi difficile de lâcher la main de l'aidant pour descendre un escalier que de lâcher une rampe ?) ;

- les échelles utilisées pour le T-score n'appartiennent pas à une même dimension (une échelle va concerner la marche, une autre le sommeil. . .).

Si l'on rajoute à ces problèmes, la variabilité des T-scores (cf paragraphe « pondération et propriétés du T-score ») selon le nombre d'objectifs fixés, selon la gradation des pondérations et selon le choix du ρ , le T-score semble présenter bien des faiblesses pour être utilisé comme critère de jugement.

Récemment le modèle Rasch [50] a permis d'améliorer les qualités psychométriques d'échelles et questionnaires utilisés en rééducation [10,11,17] : en linéarisant des échelles ordinales, le modèle Rasch permet de les rendre intervallaires et permet l'utilisation de statistiques paramétriques. Ce genre de processus n'est applicable que pour une échelle bien définie et non à des échelles personnalisées. Cependant, Tennant [69] a utilisé une simulation d'échelles GAS pour les tester par le modèle Rasch. Dans son étude, les T-scores de 300 sujets (testés sur dix GAS, pondérées par importance et par difficulté de l'objectif) étaient comparés aux T-scores de ces mêmes GAS, linéarisés par analyse Rasch pour les trois variables : score brut à chaque échelle, importance de l'objectif, difficulté de l'objectif. 14,7 % des T-scores différaient de plus dix points entre les GAS non linéarisées (c'est-à-dire considérées à tort comme intervallaires pour le calcul du T-score mais en réalité ordinales) et les GAS linéarisés (c'est-à-dire rendues réellement intervallaires par Rasch), dix étant une différence cliniquement significative. Cette étude met en évidence l'imprécision des résultats lorsque la nature ordinaire des données des GAS n'est pas prise en compte.

À l'inverse, Malec [35] a montré que l'utilisation de statistiques paramétriques ou non paramétriques donne des résultats similaires, mais sa méthodologie est moins rigoureuse. En effet, il a comparé les coefficients de corrélation entre T-scores des GAS et scores d'autres échelles ordinales (utilisées en rééducation de patients cérébrolésés) obtenus par méthode paramétrique (Pearson) et par méthode non paramétrique (Spearman). Il a montré que les coefficients de Pearson et de Spearman étaient effectivement proches. Bien que cette étude soit citée pour justifier l'emploi de statistique paramétrique [55], elle ne permet pas d'affirmer que le T-score est une mesure valide.

Si la GAS n'est pas intervallaire, quelles alternatives pour analyser les résultats ? Tennant [69] a proposé d'établir une banque d'items correspondant à des échelles GAS pré-calibrées en mesure unidimensionnelle par la « Differential Item Functioning » du Rasch [70] ce qui permettrait d'utiliser le T-score. L'inconvénient est de perdre le caractère réellement personnalisé de la GAS. La meilleure solution est peut-être celle proposée initialement par MacKay [34] et utilisée par l'équipe de Steenbeek [65] : utiliser les médianes [57] des scores bruts (-2 à +2) en les analysant par des tests de rang et des statistiques non paramétriques. La nature ordinaire des échelles est alors prise en compte. Par exemple dans son étude sur la toxine botulique [65], Steenbeek compare deux groupes en utilisant les médianes des scores avant et après toxine par le two-tailed Wilcoxon signed-rank test.

2.9. Conclusion et recommandations

Fixer des objectifs précis, décrire le niveau dont part le patient, définir les différents niveaux d'atteinte possible de cet objectif et se mettre d'accord sur les moyens de l'atteindre est en soi un processus pédagogique, qui permet de négocier des objectifs réalistes, d'échanger sur ce qui est important pour le patient et sa famille, et qui permet d'obtenir un consentement réellement éclairé et une implication du patient et de sa famille dans le projet de rééducation. En ce sens la GAS est avant tout un outil de dialogue, d'éducation thérapeutique et de modélisation de contrat et non une échelle supplémentaire dans les bilans déjà souvent longs de nos patients.

La plupart des études ayant évalué les qualités psychométriques de la méthodologie GAS analysent des échelles GAS où les cinq niveaux sont décrits avec précision, c'est-à-dire une méthodologie différente de celle du guide de Turner-Stokes [72]. Par conséquent nous suggérons que le terme de GAS soit réservé à ces échelles à cinq niveaux précisément décrits. La validité et la FIJ de la méthodologie de Turner-Stokes doivent être explorées davantage avant d'être applicables. Il semble plus judicieux de fixer le niveau initial à -2 pour obtenir des échelles comparables et afin de pouvoir détecter la progression vers un objectif sans que celui-ci ne soit atteint, malgré la problématique de l'effet plancher. Au vu des débats quant au caractère ordinal des échelles GAS et à l'utilisation erronée de calculs arithmétiques tel le T-score pour interpréter les résultats des GAS, il semble plus prudent de privilégier une analyse non paramétrique des données, suivant la méthode de Steenbeek [65] et de ne pas calculer de T-score. Il est crucial que les GAS évaluent des comportements observables, sans quoi l'imprécision de cotation s'ajoute aux imprécisions des échelles dont les validités de contenu et de construit ne pourront jamais être définitivement évaluées du fait de leur aspect idiosyncrasique.

Les GAS utilisées en recherche peuvent potentiellement mener à des conclusions erronées [69], et les propositions de la littérature de règles additionnelles pour la GAS en recherche devraient être plus étudiées et appliquées pour gagner en rigueur scientifique. En particulier, l'utilisation de GAS contrôles, concernant des objectifs non visés par le traitement et pour lesquelles une généralisation n'est pas attendue, peuvent s'appliquer aux protocoles en ligne de bases multiples [48,67], en suivant l'évolution au cours du temps des scores GAS des objectifs contrôles et des objectifs cibles. Vu la variabilité des FIJ dans la littérature, les essais cliniques utilisant la GAS comme critère de jugement, devraient évaluer leur propre FIJ, expliquer quelles variantes de la méthodologie ils utilisent (qui rédige, l'échelle, qui la cote, relecture/correction de l'échelle par un tiers, nombre de niveaux décrits, choix du niveau initial). Les équipes voulant se former à la GAS peuvent s'inspirer des formations publiées [63] et des guides de Bovend'Eerd et al. [7] et King et al. [24].

Déclaration d'intérêts

Les auteurs déclarent ne pas avoir de conflits d'intérêts en relation avec cet article.

Cet article a été rédigé dans le cadre d'un travail plus large qui a été récompensé par le prix SOFMER-Allergan en octobre 2012.

References

- [1] Ahl LE, Johansson E, Granat T, Carlberg EB. Functional therapy for children with cerebral palsy: an ecological approach. *Dev Med Child Neurol* 2005;47:613–9. <http://dx.doi.org/10.1111/j.1469-8749.2005.tb01213.x>.
- [2] Ashford S, Turner-Stokes L. Goal attainment for spasticity management using botulinum toxin. *Physiother Res Int* 2006;11:24–34.
- [3] Bjornson K, Hays R, Graubert C, Price R, Won F, McLaughlin JF, et al. Botulinum toxin for spasticity in children with cerebral palsy: a comprehensive evaluation. *Pediatrics* 2007;120:49–58. <http://dx.doi.org/10.1542/peds.2007-0016>.
- [4] Borg J, Ward AB, Wissel J, Kulkarni J, Sakel M, Ertzgaard P, et al. Rationale and design of a multicentre, double-blind, prospective, randomized, European and Canadian study: evaluating patient outcomes and costs of managing adults with post-stroke focal spasticity. *J Rehabil Med* 2011;43:15–22. <http://dx.doi.org/10.2340/16501977-16502663>.
- [5] Bouwens SFM, Van Heugten CM, Verhey FRJ. Review of goal attainment scaling as a useful outcome measure in psychogeriatric patients with cognitive disorders. *Dement Geriatr Cogn Disord* 2008;26:528–40. <http://dx.doi.org/10.1159/000178757>.
- [6] Bouwens SFM, Van Heugten CM, Verhey FRJ. The practical use of goal attainment scaling for people with acquired brain injury who receive cognitive rehabilitation. *Clin Rehabil* 2009;23:310–20. <http://dx.doi.org/10.1177/0269215508101744>.
- [7] Bovend'Eerd TJ, Botell RE, Wade DT. Writing SMART rehabilitation goals and achieving goal attainment scaling: a practical guide. *Clin Rehabil* 2009;23:352–61. <http://dx.doi.org/10.1177/0269215508101741>.
- [8] Bovend'Eerd TJH, Dawes H, Izadi H, Wade DT. Agreement between two different scoring procedures for goal attainment scaling is low. *J Rehabil Med* 2011;43:46–9. <http://dx.doi.org/10.2340/16501977-16502624>.
- [9] Boyer F. GAS: décollage immédiat pour une réponse aux difficultés d'évaluation et de prise en compte des attentes des patients. Lyon: Symposium IPSEN; 2009.
- [10] Burger H, Franchignoni F, Kotnik S, Giordano A. A Rasch-based validation of a short version of ABILHAND as a measure of manual ability in adults with unilateral upper limb amputation. *Disabil Rehabil* 2009;31:2023–30. <http://dx.doi.org/10.3109/09638280902887420>.
- [11] Caty GD, Arnould C, Stoquart GG, Thonnard J-L, Lejeune TM. ABIL-OCO: a Rasch-Built 13-Item Questionnaire to assess locomotion ability in stroke patients. *Arch Phys Med Rehabil* 2008;89:284–90. <http://dx.doi.org/10.1016/j.apmr.2007.08.155>.
- [12] Chiarello LA, Palisano RJ, Maggs JM, Orlin MN, Almasri N, Kang L-J, et al. Family priorities for activity and participation of children and youth with cerebral palsy. *Phys Ther* 2010;90:1254–64. <http://dx.doi.org/10.2522/ptj.20090388>.
- [13] Cusick A, McIntyre S, Novak I, Lannin N, Lowe K. A comparison of goal attainment scaling and the Canadian occupational performance measure for paediatric rehabilitation research. *Dev Neurorehabil* 2006;9:149–57. <http://dx.doi.org/10.1080/13638490500235581>.
- [14] Desloovere K, Schörkhuber V, Fagard K, Van Campenhout A, De Cat J, Pauwels P, et al. Botulinum toxin type A treatment in children with cerebral palsy: evaluation of treatment success or failure by means of goal attainment scaling. *Eur J Paediatr Neurol* 2012;16:229–36. <http://dx.doi.org/10.1016/j.ejpn.2010.09.006>.
- [15] Doig E, Fleming J, Kuipers P, Cornwell P, Khan A. Goal-directed outpatient rehabilitation following TBI: a pilot study of programme effectiveness and comparison of outcomes in home and day hospital settings. *Brain Inj* 2011;25:1114–25. <http://dx.doi.org/10.3109/02699052.2011.607788>.
- [16] Dunn N, Shields N, Taylor NF, Dodd KJ. Comparing the self concept of children with cerebral palsy to the perceptions of their parents. *Disabil Rehabil* 2009;31:387–93. <http://dx.doi.org/10.1080/13682820802052125>.
- [17] Franchignoni F, Giordano A, Ferriero G, Muñoz S, Orlandini D, Amor-esano A. Rasch analysis of the locomotor capabilities Index-5 in people with lower limb amputation. *Prosthet Orthot Int* 2007;31:394–404. <http://dx.doi.org/10.1080/03093640701253952>.
- [18] Gordon JE, Powell C, Rockwood K. Goal attainment scaling as a measure of clinically important change in nursing-home patients. *Age Ageing* 1999;28:275–81.
- [19] Hart T, Evans J. Self-regulation and goal theories in brain injury rehabilitation. *J Head Trauma Rehabil* 2006;21:142–55.
- [20] Hurn J, Kneebone I, Cropley M. Goal setting as an outcome measure: a systematic review. *Clin Rehabil* 2006;20:756–72. <http://dx.doi.org/10.1177/0269215506070793>.
- [21] Joyce B, Rockwood K, Mate-Cole C. Use of Goal Attainment Scaling in Brain Injury in A Rehabil. . . : Am J Phys Med Rehabil [Internet]. [cité 2012 sept 12]. Consulté à l'adresse: http://journals.lww.com/ajpmr/Full-text/1994/02000/Use_of_Goal_Attainment_Scaling_in_Brain_Injury_in.3.aspx.
- [22] Ketelaar M, Vermeer A, Hart H, Beek E, van P, Helder PJ. Effects of a functional therapy program on motor abilities of children with cerebral palsy. *Phys Ther* 2001;81:1534–45.
- [23] Khan F, Pallant JF, Turner-Stokes L. Use of goal attainment scaling in inpatient rehabilitation for persons with multiple sclerosis. *Arch Phys Med Rehabil* 2008;89:652–9. <http://dx.doi.org/10.1016/j.apmr.2007.09.049>.
- [24] King GA, McDougall J, Palisano RJ, Gritzan J, Tucker MA. Goal Attainment Scaling: its use in evaluating pediatric therapy programs. *Phys Occup Ther Pediatr* 2000;19:31–52. http://dx.doi.org/10.1080/1006v19n02_03.
- [25] King GA, Rosenbaum PL, King SM. Evaluating family-centred service using a measure of parents' perceptions. *Child Care Health Dev* 1997;23:47–62. <http://dx.doi.org/10.1046/j.1365-2214.1997.840840.x>.
- [26] Kiresuk TJ, Lund SH, Larsen NE. Measurement of goal attainment in clinical and health care programs. *Drug Intell Clin Pharm* 1982;16:145–53.
- [27] Kiresuk TJ, Sherman R. Goal Attainment Scaling: a general method for evaluating comprehensive community mental health programs. *Comm Mental Health J* 1968;4:443–53.
- [28] Kiresuk TJ, Smith A, Cardillo JE. Goal attainment scaling: applications, theory, and measurement. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc; 1994.
- [29] Krasny-Pacini A. Goal Attainment Scaling 3 bornes dans l'évaluation du traitement par la toxine botulinique chez l'enfant handicapé neuro-moteur. [Thèse de médecine]. Strasbourg; 2012.
- [30] Krasny-Pacini A. Utilisation du Goal Attainment scaling pour l'évaluation des résultats de la toxine botulinique chez les enfants paralysés cérébraux en pratique clinique. Marseille: SOFMER; 2012.
- [31] Küçükdeveci AA, Tennant A, Grimby G, Franchignoni F. Strategies for assessment and outcome measurement in physical and rehabilitation medicine: an educational review. *J Rehabil Med* 2011;43:661–72. <http://dx.doi.org/10.2340/16501977-0844>.
- [32] Law MC, Baptiste S, Carswell A, McColl MA, Polatajko H, Pollock N. Canadian Occupational Performance Measure. Thorofare (USA): SLACK Incorporated; 1998.
- [33] Löwing K, Hamer EG, Bexelius A, Carlberg EB. Exploring the relationship of family goals and scores on standardized measures in children with cerebral palsy, using the ICF-CY. *Neuropsychol Rehabil* 2011;14:79–86. <http://dx.doi.org/10.3109/17518423.2011.552088>.
- [34] MacKay G, Somerville W, Lundie J. Reflections on goal attainment scaling (GAS): cautionary notes and proposals for development. *Educ Res* 1996;38:161–72. <http://dx.doi.org/10.1080/0013188960380204>.
- [35] Malec JF. Goal Attainment Scaling in Rehabilitation. *Neuropsychol Rehabil* 1999;9:253–75. <http://dx.doi.org/10.1080/096020199389365>.
- [36] Malec JF, Smigielski JS, DePompolo RW. Goal attainment scaling and outcome measurement in postacute brain injury rehabilitation. *Arch Phys Med Rehabil* 1991;72:138–43.
- [37] Maloney FP. Goal Attainment Scaling. *Phys Ther* 1993;73:123.
- [38] Mayson TA, Ward VJ, Harris SR. Parent goals as outcome measures for children receiving treadmill training: a series of case reports. *Dev Neurorehabil* 2012;15:219–22. <http://dx.doi.org/10.3109/17518423.2011.654282>.

- [39] McLaren C, Rodger S. Goal attainment scaling: clinical implications for paediatric occupational therapy practice. *Aust Occup Ther J* 2003;50:216–24. <http://dx.doi.org/10.1046/j.1440-1630.2003.00379.x>.
- [40] Mendelson AD, McCullough C, Chan A. Integrating self-management and exercise for people living with arthritis. *Health Educ Res* 2011;26:167–77. <http://dx.doi.org/10.1093/her/cyq077>.
- [41] Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. *Arch Phys Med Rehabil* 1989;70:308–12.
- [42] Mullis R, Lewis M, Hay EM. What does minimal important change mean to patients? Associations between individualized goal attainment scores and disability, general health status and global change in condition. *J Eval Clin Pract* 2011;17:244–50. <http://dx.doi.org/10.1111/j.1365-2753.2010.01429.x>.
- [43] O'Connor C, Stagnitti K. Play, behaviour, language and social skills: the comparison of a play and a non-play intervention within a specialist school setting. *Res Dev Disabil* 2011;32:1205–11. <http://dx.doi.org/10.1016/j.ridd.2010.12.037>.
- [44] Ostensj  S, Oien I, Fallang B. Goal-oriented rehabilitation of preschoolers with cerebral palsy – a multi-case study of combined use of the Canadian Occupational Performance Measure (COPM) and the Goal Attainment Scaling (GAS). *Dev Neurorehabil* 2008;11:252–9. <http://dx.doi.org/10.1080/17518420802525500>.
- [45] Ottenbacher KJ, Cusick A. Discriminative Versus Evaluative Assessment: some observations on Goal Attainment Scaling. *Am J Occup Ther* 1993;47:349–54. <http://dx.doi.org/10.5014/ajot.47.4.349>.
- [46] Palisano RJ. Validity of goal attainment scaling in infants with motor delays. *Phys Ther* 1993;73:651–8. discussion 658–60.
- [47] Palisano RJ, Haley SM, Brown DA. Goal attainment scaling as a measure of change in infants with motor delays. *Phys Ther* 1992;72:432–7.
- [48] Perdices M, Tate RL. Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognised and undervalued? *Neuropsychol Rehabil* 2009;19:904–27. <http://dx.doi.org/10.1080/09602010903040691>.
- [49] Phillips MF, Robertson Z, Killen B, White B. A pilot study of a crossover trial with randomized use of ankle-foot orthoses for people with Charcot-Marie-tooth disease. *Clin Rehabil* 2012;26:534–44. <http://dx.doi.org/10.1177/0269215511426802>.
- [50] Rasch G. *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Oxford, England: Nielsen & Lydiche; 1960.
- [51] 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;56:736–43.
- [52] Rockwood K, Joyce B, Stolee P. Use of goal attainment scaling in measuring clinically important change in cognitive rehabilitation patients. *J Clin Epidemiol* 1997;50:581–8.
- [53] Rushton PW, Miller RE. Goal attainment scaling in the rehabilitation of patients with lower-extremity amputations: a pilot study. *Arch Phys Med Rehabil* 2002;83:771–5. <http://dx.doi.org/10.1053/apmr.2002.32636>.
- [54] S tila H, Iisalo T, Pietikainen T, Seppanen R, Salo M, Koivikko M. Botulinum Toxin Treatment of Spastic Equinus in Cerebral Pal...: *Am J Phys Med Rehabil* [Internet]. [cit  2012 sept 10]. Consult    l'adresse: http://journals.lww.com/ajpmr/Fulltext/2005/05000/Botulinum_Toxin_Treatment_of_Spastic_Equinus_in.6.aspx.
- [55] Schlosser RW. Goal attainment scaling as a clinical measurement technique in communication disorders: a critical review. *J Commun Disord* 2004;37:217–39. <http://dx.doi.org/10.1016/j.jcomdis.2003.09.003>.
- [56] Schut H, Stam H. Goals in rehabilitation teamwork, disability and rehabilitation *Informa Healthcare*. *Disabil Rehabil* 1994;16:223–6.
- [57] Siegel S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill; 1956.
- [58] Sivaraman Nair KP. Life goals: the concept and its relevance to rehabilitation. *Clin Rehabil* 2003;17:192–202.
- [59] Sorsdahl AB, Moe-Nilssen R, Kaale HK, Rieber J, Strand LI. Change in basic motor abilities, quality of movement and everyday activities following intensive, goal-directed, activity-focused physiotherapy in a group setting for children with cerebral palsy. *BMC Pediatrics* 2010;10:26. <http://dx.doi.org/10.1186/1471-2431-10-26>.
- [60] Spikman JM, Boelen DHE, Lamberts KF, Brouwer WH, Fasotti L. Effects of a multifaceted treatment program for executive dysfunction after acquired brain injury on indications of executive functioning in daily life. *J Int Neuropsychol Soc* 2010;16:118–29. <http://dx.doi.org/10.1017/S1355617709991020>.
- [61] Steenbeek D, Gorter JW, Ketelaar M, Galama K, Lindeman E. Responsiveness of Goal Attainment Scaling in comparison to two standardized measures in outcome evaluation of children with cerebral palsy. *Clin Rehabil* 2011;25:1128–39. <http://dx.doi.org/10.1177/0269215511407220>.
- [62] Steenbeek D, Ketelaar M, Galama K, Gorter JW. Goal attainment scaling in paediatric rehabilitation: a critical review of the literature. *Dev Med Child Neurol* 2007;49:550–6. <http://dx.doi.org/10.1111/j.1469-8749.2007.00550.x>.
- [63] Steenbeek D, Ketelaar M, Galama K, Gorter JW. Goal Attainment Scaling in paediatric rehabilitation: a report on the clinical training of an interdisciplinary team. *Child Care Health Dev* 2008;34:521–9. <http://dx.doi.org/10.1111/j.1365-2214.2008.00841.x>.
- [64] Steenbeek D, Ketelaar M, Lindeman E, Galama K, Gorter JW. Interrater reliability of goal attainment scaling in rehabilitation of children with cerebral palsy. *Arch Phys Med Rehabil* 2010;91:429–35. <http://dx.doi.org/10.1016/j.apmr.2009.10.013>.
- [65] Steenbeek D, Meester-Delver A, Becher JG, Lankhorst GJ. The effect of botulinum toxin type A treatment of the lower extremity on the level of functional abilities in children with cerebral palsy: evaluation with goal attainment scaling. *Clin Rehabil* 2005;19:274–82. <http://dx.doi.org/10.1191/0269215505scr859oa>.
- [66] Stucki G, Daltroy L, Katz JN, Johannesson M, Liang MH. Interpretation of change scores in ordinal clinical scales and health status measures: The whole may not equal the sum of the parts. *J Clin Epidemiol* 1996;49:711–7. [http://dx.doi.org/10.1016/0895-4356\(96\)00016-9](http://dx.doi.org/10.1016/0895-4356(96)00016-9).
- [67] Tate RL, McDonald S, Perdices M, Togher L, Schultz R, Savage S. Rating the methodological quality of single-subject designs and n-of-1 trials: introducing the Single-Case Experimental Design (SCED) Scale. *Neuropsychol Rehabil* 2008;18:385–401. <http://dx.doi.org/10.1080/09602010802009201>.
- [68] Ten Berge SR, Boonstra AM, Dijkstra PU, Hadders-Algra M, Haga N, Maathuis CGB. A systematic evaluation of the effect of thumb opponens splints on hand function in children with unilateral spastic cerebral palsy. *Clin Rehabil* 2012;26:362–71. <http://dx.doi.org/10.1177/0269215511411936>.
- [69] Tennant A. Goal attainment scaling: current methodological challenges. *Disabil Rehabil* 2007;29:1583–8. <http://dx.doi.org/10.1080/0963280701618828>.
- [70] Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *J Rehabil Med* 2003;35:105–15.
- [71] Turner-Stokes L. Goal Attainment Scaling: a practical guide. <http://www.bsrm.co.uk/eventdiaries/LeamingtonSpa2010/Docs/Goal%20Attainment%20Scaling%20in%20Rehabilitation%20a%20practical%20guide.pdf>.
- [72] Turner-Stokes L. Goal attainment scaling (GAS) in rehabilitation: a practical guide. *Clin Rehabil* 2009;23:362–70. <http://dx.doi.org/10.1177/0269215508101742>.
- [73] Turner-Stokes L, Baguley JJ, De Graaff S, Katrak P, Davies L, McCrory P, et al. Goal attainment scaling in the evaluation of treatment of upper limb spasticity with botulinum toxin: a secondary analysis from a double-blind placebo-controlled randomized clinical trial. *J Rehabil Med* 2010;42:81–9. <http://dx.doi.org/10.2340/16501977-0474>.
- [74] Turner-Stokes L, Williams H, Johnson J. Goal attainment scaling: does it provide added value as a person-centred measure for evaluation of outcome in neurorehabilitation following acquired brain injury? *J Rehabil Med* 2009;41:528–35. <http://dx.doi.org/10.2340/16501977-0383>.
- [75] Vargus-Adams JN, Martin LK. Domains of importance for parents, medical professionals and youth with cerebral palsy considering treatment outcomes. *Child Care Health Dev* 2011;37:276–81. <http://dx.doi.org/10.1111/j.1365-2214.2010.01121.x>.
- [76] Vargus-Adams JN, Martin LK. Measuring what matters in cerebral palsy: a breadth of important domains and outcome measures. *Arch Phys Med Rehabil* 2009;90:2089–95. <http://dx.doi.org/10.1016/j.apmr.2009.06.018>.

- [77] Vu M, Law AV. Goal-attainment scaling: a review and applications to pharmacy practice. *Res Social Adm Pharm* 2012;8:102–21. <http://dx.doi.org/10.1016/j.sapharm.2011.01.003>.
- [78] Wallen M, O'Flaherty SJ, Waugh MCA. Functional outcomes of intramuscular botulinum toxin type a and occupational therapy in the upper limbs of children with cerebral palsy: A Randomized Controlled Trial. *Arch Phys Med Rehabil* 2007;88:1–10. <http://dx.doi.org/10.1016/j.apmr.2006.10.017>.
- [79] Wallen M, Ziviani J, Naylor O, Evans R, Novak I, Herbert RD. Modified constraint-induced therapy for children with hemiplegic cerebral palsy: a randomized trial. *Dev Med Child Neurol* 2011;53:1091–9. <http://dx.doi.org/10.1111/j.1469-8749.2011.04086.x>.
- [80] Webb PM, Glueckauf RL. The effects of direct involvement in goal setting on rehabilitation outcome for persons with traumatic brain injuries. *Rehabil Psychol* 1994;39:179–88. <http://dx.doi.org/10.1037/h0080321>.
- [81] Weigl DM, Arbel N, Katz K, Becker T, Bar-On E. Botulinum toxin for the treatment of spasticity in children: attainment of treatment goals. *J Pediatr Orthop B* 2007;16:293–6. <http://dx.doi.org/10.1097/BPB.0b013e328092562b>.
- [82] WHO International Classification of Functioning, Disability and Health (ICF) [Internet]. WHO. [cité 2012 juill 27]. Consulté à l'adresse: <http://www.who.int/classifications/icf/en/>.