Goal Attainment Scaling in paediatric rehabilitation practice

a useful outcome measure

Duco Steenbeek

Address of correspondence
Duco Steenbeek
Revant Rehabilitation Centre Breda
Brabantlaan 1
4817 JW Breda
The Netherlands
E-mail: d.steenbeek@revant.nl
Telephone: 0031-76-5331753

The following organisations participated in the research projects of this dissertation:
Revant Rehabilitation Centre Breda, The Netherlands
Rehabilitation Centre De Hoogstraat, Utrecht, The Netherlands
Rehabilitation Centre De Trappenberg, Huizen, The Netherlands
University Medical Center Utrecht, The Netherlands
Rudolf Magnus Institute of Neuroscience, Utrecht, The Netherlands
NetChild, Network for Childhood Disability Research in The Netherlands

Cover *Zwaan* (mirrored), Minne de Groot, Amsterdam
Please refer to http://minnedegroot.ontheweb.nl/
Cover design & layout Renate Siebes, Proefschrift.nu
Printed by Labor Grafimedia BV, Utrecht
ISBN 978-94-90791-01-8

Goal
Attainment
Scaling
in paediatric rehabilitation practice

a useful outcome measure

Goal Attainment Scaling
in de kinderrevalidatie

een waardevol instrument voor uitkomstmeting

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht op gezag van de rector magnificus,
prof.dr. J.C. Stoof, ingevolge het besluit van het college
voor promoties in het openbaar te verdedigen
op dinsdag 15 juni 2010 des ochtends te 10.30 uur

door

Duco Steenbeek

geboren op 30 juni 1967
te Den Dolder

Promotor: Prof.dr. E. Lindeman

Co-promotoren: Dr. J.W. Gorter
Dr. M. Ketelaar

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van:
Stichting Wetenschappelijk Fonds Revalidatiecentrum De Hoogstraat, Utrecht
Revant Revalidatiecentrum Breda
Adams, Orthopedische Schoentechniek, Breda en Tilburg
Allergan B.V., Eindhoven
O.I.M. Brabant, Breda

CONTENTS

1

General introduction

INTRODUCTION

The aim of this thesis is to contribute to the knowledge in paediatric rehabilitation on the use and several clinimetric issues of Goal Attainment Scaling (GAS). GAS, an individualized outcome measurement tool rating the extent to which goals are attained, was first introduced in mental health care in 1968.[1] The method was introduced in rehabilitation in 1983 by Clark et al.,[2] followed by introduction in paediatric rehabilitation in 1992 by Palisano et al.[3]

Many rehabilitation interventions that are provided for children and families are believed to improve daily function, but there is often little evidence of their effectiveness. Many authors purport that it is deficiencies in measurement that have impeded both scientific research and clinical practice in rehabilitation.[4] GAS has been used since the nineties in several outcome studies in rehabilitation, as its responsiveness was assumed to be better than that of other measures. Although the GAS method, as any other outcome measure, should not have been introduced without standardization and sufficient evidence of its reliability and validity, it was after its introduction in paediatric rehabilitation that many clinimetric questions arose. The aim of this thesis is to explore the use of GAS in paediatric rehabilitation practice and answer the question if the presumed positive attributes of its use are true. The lessons learned from this thesis will contribute to clinimetrics in paediatric rehabilitation. This thesis aims to improve evidence based practice as knowledge of clinimetrics is one of the basic requirements for outcome measurement. Improving the quality of treatment evaluation is necessary in order to improve interventions and ultimately serve the needs and special interests of the children with disabilities and their families.

Research in childhood disability

Because children with disabilities often develop and change differently than typically developing children and the study of childhood disability is multidisciplinary and multifaceted, the ability to demonstrate clinical effectiveness of interventions is perhaps more complex and difficult than in any other fields of health care.[5] Another complex variable is the fact that the needs and goals of children and families change over time as children mature.

Outcome in rehabilitation pre-eminently concerns a person's activities and participation in life. The WHO International Classification of Functioning, Disability and Health (ICF)[6] (**Figure 1.1**) was endorsed in 2001 and provides a common framework for terminology, rehabilitation care policy and focus for research. The ICF provides a detailed list of codes

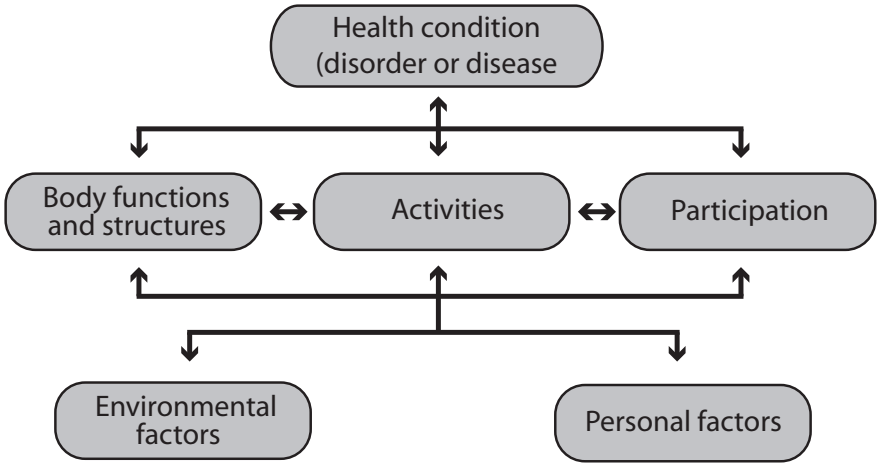


Figure 1.1 International Classification of Functioning, Disability and Health (ICF).[6]

to describe the integrity of body functions and structures, the ability and performance of activities of daily life, and the scope of individual social participation. Moreover, the framework can serve as a useful tool to analyse and to understand the interactive relationship between health conditions and personal and contextual factors. Simeonsson[7] discusses the ICF as framework for measuring the biopsychosocial dimensions of the UN Convention on the Rights for Children, which state that a mentally or physically disabled child should enjoy a full and decent life in conditions which ensure dignity, promotes self reliance and facilitates the child's active participation in the community. A Children and Youth version of the ICF (ICF-CY) was introduced in 2007. Measurement of change over time at the ICF activities and participation level is highly relevant in contemporary interdisciplinary paediatric rehabilitation care.[8] New outcome measures should be compatible with this ICF level,[9] providing one of our research questions on GAS.

Outcome measurement in paediatric rehabilitation in The Netherlands

The need for reliable and valid outcome instruments in paediatric rehabilitation in The Netherlands was recognized during the past decade. This has led to a national research programme, called Pediatric Rehabilitation Research in The Netherlands (PERRIN). This programme has resulted in a wide range of outcome instruments focusing on daily activities and participation of children

with Cerebral Palsy (CP). For details of the measures developed, translated and/or adapted by PERRIN, please refer to the PERRIN website: http://www.perrin.nl.

A Dutch translation of the Pediatric Evaluation of Disability Inventory (PEDI) was performed first. The PEDI was developed to provide a standardized clinical assessment of key functional capabilities and performance in children. Then secondly, a Dutch translation of the Gross Motor Function Measure (GMFM) was implemented. The GMFM is a clinical tool designed to evaluate change in gross motor function in children with CP. In Chapter 6 of this thesis GAS is compared to the GMFM and PEDI. Two functional classification systems for children with CP were also translated into Dutch and implemented: the Gross Motor Function Classification System (GMFCS) for self mobility and Manual Ability Classification System (MACS) for manual abilities. These classification systems are used to describe patient characteristics in this thesis. In addition to the extensive work within PERRIN, several other instruments have recently been developed and made available, for example the Motor Function Measure for children with a neuromuscular disease.

Our research group consists of members of *NetChild*. *NetChild* is an Academic Network for Childhood Disability Research in The Netherlands, launched in January 2003, aiming to bring colleagues from rehabilitation medicine and special education together. Outcome measurement in paediatric rehabilitation is one of the focuses of *NetChild*. *NetChild* is modelled after *CanChild*, centre for childhood disability research in Canada, which has a long tradition of multidisciplinary collaboration in this field. Many outcome measures at the activity and participation level for children have been developed by *CanChild*.

Paediatric rehabilitation practice in The Netherlands

The interdisciplinary rehabilitation services in The Netherlands are directed primarily at children with motor disturbances. In The Netherlands, paediatric teams in rehabilitation centres consist of a permanent group of professionals, who treat patients and their families with complex care needs and who benefit from a comprehensive interdisciplinary treatment. After an initial observation period in the rehabilitation centres, the options for intervention generally depend on issues like the request for help, complexity of the problems, the child's age and the predicted influence of the involvement of the rehabilitation team on the problem.

In 1996, the Rehabilitation Activities Profile for Children (Children's

RAP) was introduced in The Netherlands.[10] This instrument was developed for communication in paediatric rehabilitation during team conferences and provides a common language with the ICF terminology for therapists and parents. The Children's RAP system provides basic information about the child and its proxies, the present situation, the needs of the child and its proxies, and conclusions of the team conference. If the child is eligible for an interdisciplinary program, specific, measurable, acceptable, relevant and time-related (SMART) goals should be set after the observation period. The conclusion of 2 recent studies on the use of the Children's RAP by Nijhuis et al.[11, 12] was that integration of needs and problems on the one hand and rehabilitation goals on the other was not yet optimal in The Netherlands. Moreover goals were poorly documented. GAS may help to improve team conferences and the integration of needs in rehabilitation programs, by constructing the scales after the observation period. GAS scales can be constructed for each discipline or for the main interdisciplinary goal. The use of GAS in team conferences will be discussed in Chapter 4.

The study locations

The first project of this thesis was performed at rehabilitation centre "De Trappenberg" in Huizen, The Netherlands and the second at Rehabilitation Centre Breda, The Netherlands. The centres are two of the 23 specialized rehabilitation centres in The Netherlands with similar paediatric departments. During the projects, the participating children were treated by a team consisting of physicians in paediatric rehabilitation medicine, paediatric physical therapists, occupational therapists, speech-language therapists, child psychologists and social workers. Both rehabilitation centres cooperate directly with a special school for children with physical disabilities. About half of the children involved in the projects were in mainstream education and the other half attended the special school.

Purposes of measures

Measurement of functioning and disability generally has one of three aims: diagnosis, prognosis or evaluation.[13] In diagnosis the aim of measuring is to discriminate between subjects. An example is to identify children who are behind their peers in motor development. In prognosis, the aim is to discriminate between subjects on a longitudinal basis. Measuring prognosis requires predictive measures, providing information about future status. An example is the use of the GMFCS for children with CP at different young ages

in order to predict gross motor function later on.[14] In evaluation, the aim of measurement is to evaluate changes of functioning and disability over time. Evaluative measures need to be responsive to clinically important changes and are used to track development or the effect of interventions. This can be illustrated by monitoring walking ability or the level to which treatment goals are attained, as an indicator of progress during rehabilitation.

A discriminative measure is usually *norm*-referenced, in order to demonstrate that the difference between those with and without the problem of interest can be discriminated. A *criterion*-referenced measure is one in which the items that are chosen reflect the domain of interest and the criterion is explicitly described regarding whether or not the person can accomplish the activity. This type of measure is more appropriate when assessing the achievement of therapy goals and does not require comparisons with a general population or "normal" peers.

Measurement issues in general

In order to place this thesis in the proper context, four other measurement issues must be considered.

The first issue is to define *generic* instruments as opposed to *disease specific* instruments. Generic measures intend to measure the same construct (for example activity limitations) across different patient groups. Specific measures are developed for application in only one diagnostic group. An important advantage of generic rehabilitation measures is that the level of functioning in different diseases can be compared.

The second issue is that measures may focus on *individual* patients versus *a group of* patients. Measurement in medical research typically concerns a group of patients. An important purpose for measuring groups is to reduce measurement error by taking the average of the measurements in the group of patients:[15] the larger the group, the lower the error; the smaller the group, the greater the requirements of the instruments. In clinical practice one is interested in measuring an individual patient. The reliability of many instruments, although adequate in the group setting, does not meet the high standards of the individual setting: measuring individual patients is generally associated with a relatively high degree of error.

The third issue is to distinguish *standardized* versus *individualized* (*idiosyncratic*) instruments. The standardized measures at the ICF activity and participation level describe a selection of potential problems, using common wording. In patient-oriented rehabilitation, it is emphasized that the patients define the problems that need to be addressed. For this

approach, individualized measures can be used in which the nature of problems are described in the patient's words and in the context of his own daily experiences.[15]

The final issue is distinguishing 4 possible levels of measurement, including nominal, ordinal, interval and ratio.[16] *Nominal* data are unordered (e.g. job classification). An *ordinal* measure presents a number of response options that order the characteristic of interest quantitatively (e.g. educational level or GMFCS level as described above). In contrast, variables in which the interval between responses is known, are called *interval* variables (e.g. temperature in Celsius). Finally, variables where the zero point is meaningful, so that the ratio of two responses has some meaning, are called *ratio* variables (e.g. blood pressure in mmHg). Interval/ratio scales can be summarized using powerful parametric statistics, as in contrast ordinal and nominal data require different analysis.

GOAL ATTAINMENT SCALING

In accordance with the terminology that has been introduced, GAS is an example of a generic individualized evaluative criterion-referenced instrument, which can be used for measurement of changes in individual patients and in groups of patients. GAS is a method to measure progress towards individual goals. GAS was originally developed to generate a 5-point ordinal scale, defined as follows:[1]

–2	much less than expected goal attainment
–1	less than expected goal attainment
0	complete goal attainment
+1	more than expected goal attainment
+2	much more than expected goal attainment

Different versions of Goal Attainment Scaling

Since the introduction of GAS, authors have experimented with scales different from the original 5-point scale, for example Cusick introduced 7-point (Likert) scale rating.[17] In the original version, after constructing the GAS scales, a baseline value was scored. That means that deterioration could only be scored if the start was coincidentally scored as –1 or 0. This may introduce a bias for the calculation of group effects, as only 3 values remain to score progress, the first of which is the expected goal (0). Most users score

'no change' by definition as –2, but this means that deterioration cannot be scored, which results in a potential bias of so-called *bottoming effects*. For this thesis, we wanted to introduce the possibility of scoring deterioration for all participants, as several interventions or progression of disease can cause decline. On the other hand, we wanted to retain all original goal attainment levels in order to facilitate comparison to other studies described. We decided to add a '–3' score for deterioration and to define '–2' as the level equal to the start, as a solution for this dilemma, resulting in a 6-point scale. Stipulating that the starting point be higher than the lowest level, helps prevent bottoming effects. The projects described in this thesis use this 6-point scale version of GAS, including the '–3' score.

The selection of a measurement construct using Goal Attainment Scaling

Current debate about the clinimetrics of GAS includes issues such as (1) sensitivity to subtle changes; (2) the ability to detect important change (responsiveness); (3) reliability of the scale construction (content reliability) as well as of the scores (inter-rater reliability); (4) whether the content of the scales cover all aspects or elements of interest that are to be measured (content validity) and (5) whether the scale measures the construct that it purports to measure (construct validity). All properties of individual measures depend on agreements about the construct that is to be measured. Consequently, we developed criteria for GAS scale construction and scoring. This thesis discusses the selection of constructs, the feasibility of our criteria and consequences for the clinimetric properties.

Different methods for data computation

Originally, the attainment levels for each goal are combined in a single aggregated '*T* score' by applying the following formula:

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

where ω_i is the weight assigned to the ith goal and x_i is the score of the ith goal. ρ = 'average correlation'.

This accounts for variable numbers of goals, intercorrelation of goal areas and variable weighting, to transform goal attainment into a standardized measure or *T* score per patient with a mean of 50 and standard deviation of 10. Although

this formula is typical for GAS and still widely used, from the start of our studies we were convinced that it had mathematical pitfalls. An alternative is analyzing raw GAS scores instead of using this formula. We avoided the *T* sum formula for three reasons. First, we were convinced that the *T* sum score with several decimal places would give false impression of precision and a false sensitivity to change. A fundamental problem is that in this score data are treated as interval data, using mean scores for groups, whereas they represent at best an ordinal scale of value judgements. Second, according to the originators, if goal setting is unbiased so that results exceed or fall short of expectations in roughly equal proportions, one would expect a normal distribution of scores. However, our studies in daily rehabilitation practice were expected to result in non-normal distribution, not allowing the application of parametric statistics. Finally, although optional in the formula, we believed that weighting different goals per patient would threaten the reproducibility. The issue of the *T* score is discussed in more detail in Chapter 3.

CHILDREN WITH CEREBRAL PALSY

Chapter 2, 4, 5 and 6 focus on outcome measurement for children with CP. CP is one of the most common childhood disabilities and may present the most severe disabilities in terms of making great demands on the health, educational and social service systems.[18] Nearly two out of every thousand children born in The Netherlands have some type of CP.[19] In April 2006 a revised consensus report was generated on the Definition and Classification of CP.[20] CP can be defined as a group of permanent (but not unchanging) disorders of the development of movement and posture, as the result of one or more non-progressive abnormalities in the brain, before its growth and development are complete. It is caused by damage to one or more specific areas of the brain, usually occurring during fetal development or infancy. It can occur before, during or shortly following birth. Several causal pathways of CP have been identified, such as kidney or urinary tract infections, metabolic disturbances, toxaemia, jaundice, rhesus incompatibility, rubella, direct trauma and hypoxia. CP is characterized by an inability to fully control motor function, particularly muscle control and coordination. This can result in involuntary movement, disturbance in gait and mobility, abnormal sensation and perception. The motor disorders of CP are often accompanied by disturbances in sensation, perception, cognition (learning difficulties), hearing, vision, speech and swallowing difficulties, communication, behaviour and secondary musculoskeletal problems.[20] Children with CP may also be affected by other

associated conditions, such as epilepsy. The condition commonly appears in infancy, and children with CP are frequently slow to reach developmental milestones such as learning to roll over, sit, crawl, smile, or walk.

As CP is a heterogeneous condition, further classification of its types and severity is fundamental in research and clinical practice. In terms of body functions, classification is recommended using the system advocated by the Surveillance of Cerebral Palsy in Europe.[18] There are three main types of CP: spastic, dyskinetic and ataxic CP. Spastic CP is the most common type (about 90% of prevalence). The spastic type can be classified further according its limb distribution, as either unilateral or bilateral. The severity can be classified by the GMFCS and MACS.

As the type as well as the severity of CP is widely varied, a broad range of needs of children and families exist, resulting in a diversity of therapy goals. In The Netherlands, the Dutch Institute for Health Care Improvement CBO implemented a medical guideline for diagnosis and treatment of children with spastic CP in 2006, guiding several treatment strategies.[21] Physicians in paediatric rehabilitation medicine, paediatric physical therapists, occupational therapists, speech-language therapists, child psychologists, social workers and other disciplines are generally involved in all of the ICF areas in the activity and participation domain, including learning and applying knowledge, general tasks and demands, communication, mobility, self-care, domestic life, interpersonal interactions and relationships, major life areas (e.g. education) community and social and civic life. Interdisciplinary rehabilitation intervention and exercise programs are usually combined with adaptive equipment, like splints, braces, communication and mobility devices.

Spasticity is one of many child, environmental and family factors that can be related to gross motor development[22] and may cause serious secondary musculoskeletal problems including scoliosis, hip displacement, deformities and pain.[23] In addition to the rehabilitation team interventions, the children described in this thesis were treated with a variety of supplementary interventions to treat spasticity. Spasticity management strategies can be general versus focal and reversible versus irreversible. Drug therapy, generally and reversibly reducing spasticity, includes Baclofen and Dantrolene. If this is insufficient, additional spasticity-reducing therapy may be necessary, for example Botulinum toxin type A (BTX-A). BTX-A is a focal reversible method reducing muscle disbalance. Chapter 2 of this thesis describes one of the effect studies using GAS to evaluate the effect of multi-level BTX-A. When other treatments fail to prevent complications, surgery

(focal irreversible) may be necessary. Several surgical techniques are used to prevent hip luxation, improve the child's ability to walk, improve balance, or prevent further deformity. Selective Dorsal Rhizotomy (focal irreversible) involves identifying specific sensory nerve fibres behind the spinal cord, and then selectively cutting those nerve fibres to reduce spasticity. Intrathecal Baclofen Therapy (ITB) is an option for management of severe spasticity, using an implantable infusion system to deliver precise amounts of baclofen directly to the intrathecal space via a surgically implanted pump and catheter.[21]

THE MOTIVATION FOR THIS THESIS

In addition to the standardized measures developed, each with advantages and disadvantages for different applications in research and practice, the members of *NetChild* strongly felt the necessity to improve the uniformity of GAS and answer the questions on its use and the clinimetric properties. We felt the effect of some rehabilitation applications might have failed to be proven convincingly unless patients were satisfied with the outcome. Failure to demonstrate improvement for an individual patient may be due to a lack of improvement, but it may also be the result of using instruments with poor responsiveness. For future research as well as for measurement in rehabilitation practice, the further development of responsive instruments that reliably demonstrate relevant outcome, is very important. The motivation for this thesis is based on several crucial questions raised about its use in our own team practice. We decided to weight the positive expectations of GAS against these questions. The administration and scoring procedures in the paediatric rehabilitation setting differ from those in the field of mental health or geriatrics. The clinimetric properties can not be compared with those found in other fields. The first project at "de Trappenberg", Huizen (Chapter 2) had a double function: to evaluate the application of multi-level BTX treatment and to explore the use of the GAS application. The reason to continue our research at the centre in Breda after this first project was further exploration of the use, reliability and added value to previously developed measures for the ICF activity and participation domain and to examine team experiences (Chapter 3, 4, 5, 6).

The project in Breda was also intended to improve goal-oriented evidence based inter-disciplinary paediatric rehabilitation care locally, as the involvement of all team members could be realized.

OUTLINE OF THIS THESIS

The first study was performed to measure the effect of lower extremity treatment with Botulinum Toxin A using GAS (Chapter 2). As this study confirmed the positive expectations of GAS, a critical review of the available literature on content reliability, inter-rater reliability, validity, sensitivity to change and use in paediatric rehabilitation research of GAS was performed (Chapter 3). The findings resulted in the performance of projects on the clinimetric properties of GAS. We started by developing a practical clinical training program because training in the GAS method was found to be necessary. The aim of Chapter 4 is to share our experiences with this training program and introduction of GAS in the team. The training method and the team experiences after the training period in GAS are described. The next step was to enrol 23 children with CP and 20 therapists of 3 disciplines in a reliability study, as described in Chapter 5. The aims of this project were: (I) to determine the inter-rater reliability of GAS in the routine practice of interdisciplinary paediatric rehabilitation and to explore the reasons for discrepancies between scores and (II) to examine the difference in the inter-rater reliability of the scores between GAS scales constructed by the children's own therapists and that of scores on scales constructed by independent therapists. In the final study, the GAS outcome was compared to the outcome on the PEDI-NL and GMFM-66. Based on these data, we explored the added value of GAS (Chapter 6). The added value was examined in regard to construct differences and differences in responsiveness between GAS and both standardized measures. In Chapter 6, several issues on the responsiveness of GAS as discussed in Chapter 2 have been substantiated further. Finally, Chapter 7 is the general discussion summarizing the lessons we learned, the positive attributes of GAS and the remaining issues.

REFERENCES

1. Kiresuk T, Sherman R. (1968) Goal attainment scaling: a general method of evaluating comprehensive community mental health programs. *Community Ment Health J* **4**: 443–453.
2. Clark MS, Caudrey DJ. (1983) Evaluation of rehabilitation services: the use of goal attainment scaling. *Int Rehabil Med* **5**: 41–45.
3. Palisano RJ, Haley SM, Brown DA. (1992) Goal attainment scaling as a measure of change in infants with motor delays. *Phys Ther* **72**: 432–437.

4. Johnston MV, Keith RA. (1993) Measurement Standards for medical rehabilitation and clinical applications. In Granger CV, Gresham GE (eds): *New developments in functional assessment. Physical Medicine and Rehabilitation Clinics of North America*. WB Saunders Company, Philadelphia, PA, 425–449.

5. McConachie HR. (1999) Conceptual frameworks in evaluation of multidisciplinary services for children with disabilities. *Child Care Health Dev* **25**: 101–113.

6. World Health Organization. (2001) *International classification of functioning, disability and health (ICF)*. Geneva: World Health Organization.

7. Simeonsson RJ, Leonardi M, Lollar D, Bjorck-Akesson E, Hollenweger J, Martinuzzi A. (2003) Applying the International Classification of Functioning, Disability and Health (ICF) to measure childhood disability. *Disabil Rehabil* **25**: 602–610. Review.

8. Rosenbaum P, Stewart D. (2004) The World Health Organization International Classification of Functioning, Disability and Health: A model to guide clinical thinking, practice and research in the field of Cerebral Palsy. *Semin Pediatr Neurol* **11**: 5–10.

9. McDougall J, Wright V. (2009) The ICF-CY and Goal Attainment Scaling: Benefits of their combined use for pediatric practice. *Disabil Rehabil* **26**: 1–11.

10. Roelofsen EE, The BA, Beckerman H, Lankhorst GJ, Bouter LM. (2002) Development and implementation of the Rehabilitation Activities Profile for children: impact on the rehabilitation team. *Clin Rehabil* **16**: 441–453.

11. Nijhuis BJ, Reinders-Messelink HA, de Blécourt AC, Ties JG, Boonstra AM, Groothoff JW, Nakken H, Postema K. (2008) Needs, problems and rehabilitation goals of young children with cerebral palsy as formulated in the rehabilitation activities profile for children. *J Rehabil Med* **40**: 347–354.

12. Nijhuis BJ, Reinders-Messelink HA, de Blécourt AC, Boonstra AM, Calamé EH, Groothoff JW, Nakken H, Postema K. (2008) Goal setting in Dutch paediatric rehabilitation. Are the needs and principal problems of children with cerebral palsy integrated into their rehabilitation goals? *Clin Rehabil* **22**: 348–363.

13. Guyatt G, Kirschner B, Jaeschke R. (1992) Measuring health status: what are the necessary measurement properties. *J Clin Epidem* **45**: 1341–1345.

14. Gorter JW, Ketelaar M, Rosenbaum P, Helders PJ, Palisano R. (2009) Use of the GMFCS in infants with CP: the need for reclassification at age 2 years or older. *Dev Med Child Neurol* **51**: 46–52.

15. Dekker J, Dallmeyer AJ, Lankhorst GJ. (2005) Clinimetrics in rehabilitation medicine: current issues in developing and applying measurement instruments. *J Rehabil Med* **37**: 193–201.

16. Streiner DL, Norman GR. (2008) *Health Measurement Scales. A practical guide to their development and use*. Fourth Edition. Oxford University Press.

17. Cusick A, McIntyre S, Novak I, Lannin N, Lowe K. (2006) A comparison of goal attainment scaling and the Canadian Occupational Performance Measure for paediatric rehabilitation research. *Pediatr Rehabil* **9**: 149–157.

18. Surveillance of Cerebral Palsy in Europe. (2002) Prevalence and characteristics of children with cerebral palsy in Europe. *Dev Med Child Neurol* **42**: 816–824.

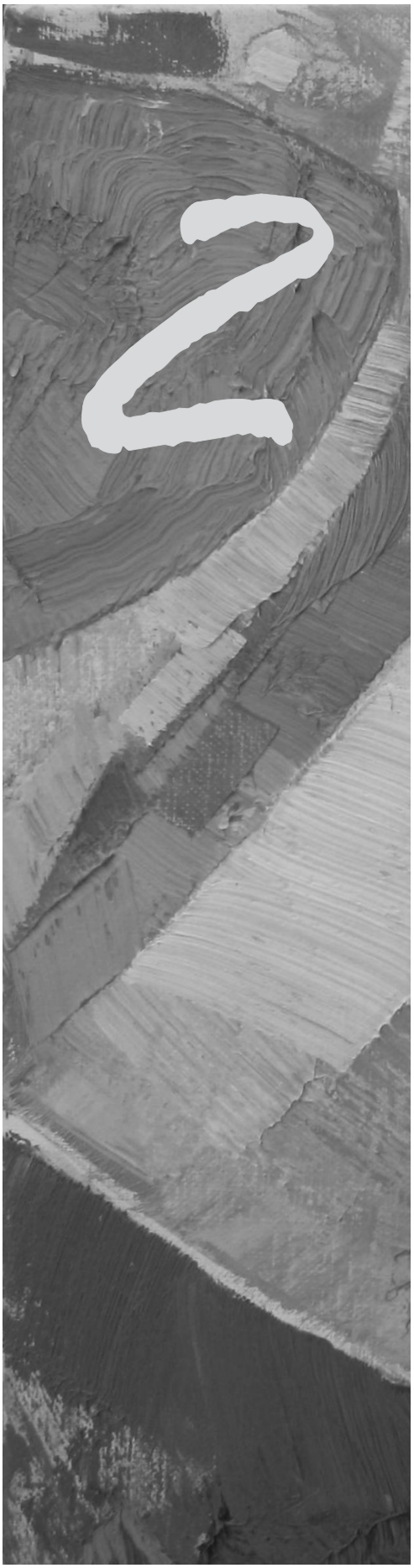
19. Wichers MJ, van der Schouw YT, Moons KG, Stam HJ, van Nieuwenhuizen O. (2001) Prevalence of cerebral palsy in The Netherlands (1977-1988). *Eur J Epidemiol* **17**: 527–532.

20. Rosenbaum P, Paneth N, Leviton A, Goldstein M, Bax M, Damiano D, Dan B, Jacobsson B. (2007) A report: the definition and classification of cerebral palsy April 2006. Dev Med Child Neurol Suppl 109: 8–14. Erratum in: *Dev Med Child Neurol* **49**: 480.

21. *CBO Richtlijn diagnostiek en behandeling van kinderen met spastische Cerebrale Parese*. (2006) Nederlandse Vereniging van Revalidatieartsen.

22. Gorter JW, Verschuren O, van Riel L, Ketelaar M. (2009) The relationship between spasticity in young children (18 months of age) with cerebral palsy and their gross motor function development. *BMC Musculoskelet Disord* **10**: 108.

23. Gorter JW. (2009) Rehabilitative therapies for the child with cerebral palsy: focus on family, function and fitness. *Minerva Pediatr* **61**: 425–440.



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

D. Steenbeek, A. Meester-Delver,
J.G. Becher, G.J. Lankhorst

Clin Rehabil 2005,
19: 274–282.

ABSTRACT

Objective: To measure the effect of botulinum toxin type A (BTX-A) treatment in children with Cerebral Palsy with regard to individual goals concerning functional abilities, using Goal Attainment Scaling.

Design: A single-blind randomized multiple baseline/treatment phase design across subjects.

Setting: The paediatric department of a rehabilitation centre.

Subjects: Eleven children with Cerebral Palsy participated.

Intervention: BTX-A treatment of the lower extremity.

Main measures: A six-point Goal Attainment Scaling of three individual treatment goals at the level of functional abilities. Standardized video-tapes of each goal were recorded weekly for a period of 14 weeks. Rating on the predetermined Goal Attainment Scaling was blinded.

Results: Nine of the 11 subjects showed significant improvement in 18 out of 33 goals. Seven subjects showed clinically relevant improvement (at least 2 points on the Goal Attainment Scaling) in 11 goals. Testing the difference between all medians of baseline measurements (after correction for improvement during baseline) and the medians of the treatment phase measurements for all Goal Attainment Scaling scores (n=33) resulted in significant improvement (p<0.001). Tested at subject level (medians of the three Goal Attainment Scaling scores per assessment, n=11), a significant improvement was also found (p=0.005). The change in Goal Attainment Scaling score was related to the moment of treatment with BTX-A.

Conclusion: Clinically relevant improvement in individual rehabilitation goals at ability level, achieved with the treatment of BTX-A in children with Cerebral Palsy, were demonstrated using the Goal Attainment Scaling method.

INTRODUCTION

In Cerebral Palsy, spastic paresis is the most common motor disorder. Spasticity is defined as a motor disorder characterized by a velocitydependent increase in tonic stretch reflexes with exaggerated tendon jerks, resulting from hyperexcitability of the stretch reflex, as one component of the upper motor neurone syndrome.[1] Botulinum toxin A, works by blocking the neuromuscular junction in the treated muscles. Since the 1990s, several studies have investigated the effects of botulinum toxin type A in the treatment of spasticity of the lower extremities in Cerebral Palsy.[2–11] In addition to spasticity, other disorders in muscle function, such as paresis, can seriously affect the abilities of a child. Thus, a reduction in spasticity does not always lead to an improvement in abilities. In most studies, tools that measure spasticity in a

clinical way, such as the Modified Ashworth Scale,[12] the Modified Tardieu Scale and the range of motion,[13] have been used to evaluate the effects of botulinum toxin A treatment. For the most part the results of these studies were extremely positive at the level of impairment. But these tools should be used with caution when evaluating changes in young children with Cerebral Palsy, because of variety in test/retest results.[13] At the level of abilities, the Gross Motor Function Measurement (GMFM) is often used, but with the GMFM[14] as primary outcome measure in botulinum toxin A studies, conflicting results have been reported.

Failure to detect improvement with a general quantitative measurement instrument could be caused by lack of sensitivity to a change at individual level. In paediatric rehabilitation, measurement of the treatment outcome at the level of abilities and participation is preferable above measurement of the effects at impairment level, because the main goal of rehabilitation is to achieve an improvement in abilities, despite existing impairments. According to the International Classification of Functions (ICF),[15] functional treatment goals need to be well defined and tailored to the individual needs of the patient and the parents.[16] This can be achieved with Goal Attainment Scaling (GAS). Goal Attainment Scaling is a potentially reliable method of outcome measurement in the domain of abilities, focusing on individual goals (see the Methods section).

The aim of the present study was to measure the effect of lower extremity treatment with botulinum toxin A on the functional abilities of children with Cerebral Palsy, focusing on goals at individual level.

METHODS

Design

A single-blind randomized multiple baseline/treatment phase design across subjects was applied. All subjects were measured weekly for a period of 14 weeks. To demonstrate the time-related relationship between improvement and injection with botulinum toxin A, the subjects were randomly assigned to one of two groups. In the first group the baseline phase (phase A) lasted eight weeks and the treatment phase (phase B) lasted six weeks. In the second group the baseline phase lasted six weeks and the treatment phase lasted eight weeks (**Figure 2.1**).

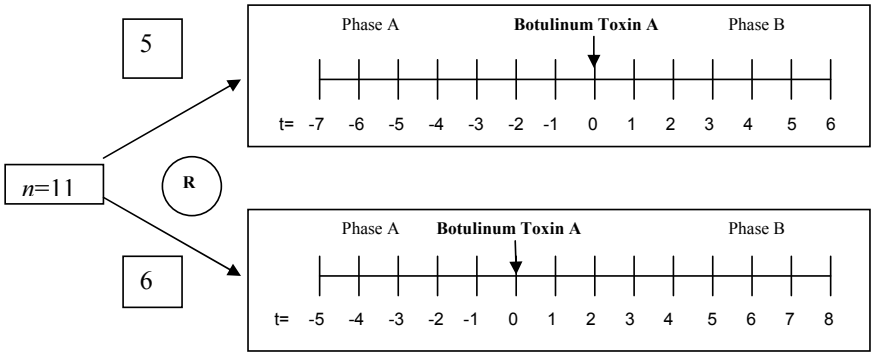


Figure 2.1 Design. R, Random allocation to one of the two groups.

Outcome measurement

Goal Attainment Scaling was used to measure the effect of the treatment. The original Goal Attainment Scaling is a 5-point scale on which the expected level of success is scored as 0. Better than expected is a score of +1, and the most favourable outcome is a score of +2. Less than expected, and the least favourable outcome are scores of –1 and –2, respectively. Most users score no change as –2. Goals should be specific, measurable, achievable, resource sensitive and timed (SMART). Kiresuk and Sherman[17] first introduced Goal Attainment Scaling in 1968 to evaluate mental health services. They transformed the goal set to a single *T* score, intended as a sensitive functional outcome measure, in which each goal is weighted and the (mostly) intuitive average correlation is corrected.

Since its introduction, Goal Attainment Scaling has been used in a variety of settings and for many purposes. Many writers attest to its value as a process for developing the quality of delivery of a service.[18] However, in children with Cerebral Palsy there is little experience with this method. Palisano[19] measured the validity of Goal Attainment Scaling as a measure of motor change in 65 infants (only eight with Cerebral Palsy). Satisfactory interrater reliability (r=0.51–0.91)[20] of Goal Attainment Scaling has been reported.[21–26] Due to the idiosyncratic nature of Goal Attainment Scaling, concurrent validity with norm-referenced measures is low to moderate. However, compared with standardized measures, Goal Attainment Scaling shows better responsivity and sensitivity to change and, in case of acceptable content validity, it can demonstrate more clearly the level of clinical relevance (an increase of at least two points). It has even been suggested that

Goal Attainment Scaling, itself, could have a therapeutic effect,[27] but that phenomenon can be a possible advantage for the patient as well as a possible bias for trials.

Problems with Goal Attainment Scaling are: (1) There are concerns about the selection of goals and the specification of expected outcomes. How will we know whether the selected goals are relevant and realistic, and whether they can be reproduced? There can be a high risk of therapist bias in the scoring. (2) Caution is essential in the interpretation of any results after weighting goals, because of the subjectivity. (3) A fundamental problem is that the *T* scores are usually treated as interval data, with the use of mean scores for groups. However at best this will be an ordinal scale of assessed value and only nonparametric statistics are acceptable.[18]

In this study, Goal Attainment Scaling was used in such a way as to avoid the above-mentioned problems. For each subject, three different abilities that needed to be improved were defined. The Goal Attainment Scaling was based on a 6-point scale (**Table 2.1**), instead of the original 5-point scale. As deterioration in ability can be provoked by botulinum toxin A treatment, the –3 score for deterioration was added to retain all original goal attainment levels. For each score, a clear improvement in ability was described (e.g.,

Table 2.1 Example of Goal Attainment Scaling: walking on an irregular surface for a four-year-old girl with Cerebral Palsy, spastic hemiplegia, GMFCS level I

Goal Attainment Scaling	Score	Goal description
Worse than start (deterioration)	–3	She cannot walk through the rungs.
Equal to start	–2	She does succeed with help, but she needs to place both feet in the space between all rungs.
Less than expected	–1	She walks with help for at least the half of the distance and with alternating feet between the rungs.
Expected goal	0	She walks the whole distance with alternating feet between the rungs with help.
Somewhat more than expected	+1	She walks at least half of the distance without help, but often needs to place a second foot next to the first in order to step over the next rung.
Much more than expected	+2	She walks the whole distance without help and places alternating feet over the next rung.

In order to create an irregular surface a ladder is placed horizontally at a height of 15 cm above the floor. The girl is asked to walk barefoot without an ankle-foot orthesis as quickly as possible through the rungs of the ladder over a distance of 8 m. Only if she begins to fall will a therapist help by holding one of her hands. The expected effect of botulinum toxin A treatment is improvement in the co-ordination of distal muscle function.

climbing four stairs without manual support). All subjects received a Goal Attainment Scaling for each of the three goals. Clinical relevance was defined as an improvement of at least 2 points in at least one of the three goals.

For each assessment, one assessor made a standardized video-recording during trials for each of the three functional ability goals. To mask the sequence of the 14 measurements the recording procedure was identical for all measurements. The scoring of the video on the predetermined Goal Attainment Scale was blinded by randomizing the sequence of clips. The second assessor, who rated the subject's performances from the video-recordings, was an independent consultant who was neither involved in the treatment of the subjects, nor in the recording procedure. All other confounding circumstances, such as the frequency of physical therapy, were controlled as much as possible during the trial. Physical and occupational therapy were applied regularly during both the baseline and the treatment phase. Caregivers were instructed to pay an equal amount of attention to the defined abilities in both phases.

The secondary outcome consisted of Goal Attainment Scaling scoring carried out by the treating paediatric physiotherapist, independent of the video-scoring. The aim of these measurements was to compare the non-blinded therapist's opinion with the visual information of the videobased scores. The Modified Ashworth Scale[12] was used to control the effect of treatment on spasticity and to detect possible nonresponders. The scale was scored by one nonblinded investigator.

Participants

The inclusion criteria were serious disabling spasticity of the lower extremity, causing functional problems, ability to carry out instructions, and adequate knowledge of the Dutch language.

The protocol was approved by the Medical Ethics Committee of the VU University Medical Centre in Amsterdam. Full written informed consent, according to the Declaration of Helsinki, was obtained from all the parents.

Intervention

By means of a physical examination, video gait analysis, and analysis of the functional problem, muscles were selected for injection with botulinum toxin A. If a serious spastic motor disorder was related to the functional problem, the injection sites were defined. The intervention consisted of treatment with botulinum toxin type A (Botox®), 4–6 U/kg per muscle group, with a

maximum of 50 U per injection site. Five multilevel treatments were performed under general anaesthesia (psoas, adductor, hamstrings and gastrocnemius muscles bilaterally), and six monolevel treatments with local anaesthesia (gastrocnemius muscles monoor bilaterally).

Statistical analysis

The design of this study makes statistical evaluation of group treatment effects possible, as well as evaluation of individual treatment effects (single-case experiments). Single-case methodology requires stepwise analysis of the ordinal data that are collected. Statistics were performed in SPSS-pc, release 10. In the present study it was decided not to use the popular *T* sum score in order to preserve the ordinal nature of the data, and also because the clinically relevant effect was defined as an improvement in at least one goal.

Individual effects were analysed by nonparametric testing of differences between the baseline conditions and the treatment phase with the Wilcoxon Mann-Whitney test.[28] Clinical relevance was defined as a significant positive difference of at least 2 points between the median of the baseline phase and the median of the treatment phase. Correction was made for improvement (e.g., due to the natural course or training) during baseline measurements, by subtraction of the difference between the start and the end of the baseline from the median of the treatment phase conditions.

Group effects were demonstrated by testing the difference between all medians of the treatment phase and the medians of the baseline phase for each Goal Attainment Scaling score, as well as for each subject (two-tailed Wilcoxon signed ranks test for two related samples). The median of the best score out of three goals per subject was tested, since clinical relevance was defined as an effect demonstrated in at least one of the three Goal Attainment Scaling scores. Because that assumption is arbitrary, the median of the best two scores was also tested.

Differences in the secondary outcome parameter, the physiotherapist's scoring of Goal Attainment Scaling, and the video-scoring of Goal Attainment Scaling, were also tested with a two-tailed Wilcoxon signed ranks test for two related samples. The correlation between the two was tested by applying a weighted kappa.

RESULTS

Eleven children, aged from 3 to 12 years (mean 5.0 years), with Cerebral Palsy (6 hemiplegic and 5 diplegic) participated in the study. The Gross Motor FunctionMeasurement (GMFM-88) scores in domain E (walking, running, jumping) had a mean of 45% (SD 32, range 0–89), Gross Motor Function Classification System (GMFCS)[29]: 4 × level I; 3 × level II; 3 × level III and 1 × level IV.

Of the 462 video fragments of 33 goals, 22 did not meet the requirements: 14 (all of one Goal Attainment Scale for one subject) because the scale was not properly recorded on video, and 8 (of four subjects) because the masking of the sequence of the recordings was unsuccessful (e.g., because of a change in haircut). For all missing data, a score of –2 was assigned in the analysis.

Effects at individual level

Nine of the 11 subjects showed statistically significant improvement in the video Goal Attainment Scaling scores from the first week after the botulinum toxin A injections. A statistically significant improvement in score was achieved for 18 out of 33 goals. Of these 18 significantly improved goals, 11 goals of 7 subjects were clinically relevant (i.e., the improvement was at least 2 points). Correction for an increase in the baseline score (1 or 2 points) was necessary for six baselines scores of four subjects.

The functional goals were classified as: climbing stairs, walking on an irregular surface, walking on a slope, standing balance (e.g., exercises with a ball) and walking balance (e.g., carrying a tray).

The frequencies and changes in Goal Attainment Scaling scores are shown in **Table 2.2**. None of the subjects demonstrated any deterioration after treatment.

Table 2.2 Classification of goals and improvement after treatment

Ability	Total number	Change between median of baseline and treatment phase in Goal Attainment Scaling		
		+0, 0.5	+1, 1.5	+2, 2.5, 3
Climbing stairs	5	2	1	2
Walking on an irregular surface	10	1	7	2
Walking on a slope	4	1	0	3
Standing balance	9	3	3	3
Walking balance	5	4	0	1
Total	33	11	11	11

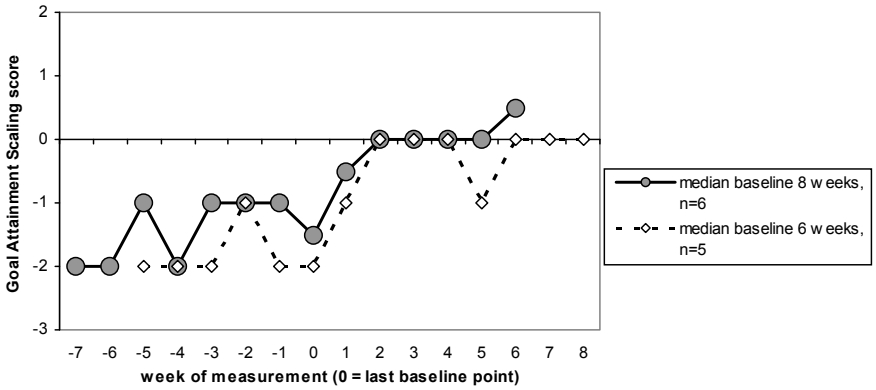


Figure 2.2 Median Goal Attainment Scaling scores of 11 subjects. Median of Goal Attainment Scaling scores for all three goals of all subjects, before (t=–7–0) and after (t=1–8) treatment with botulinum toxin A. The improvement in the score is related to the moment of treatment (week 0).

Group effects

Testing the differences between all medians of the baseline measurements and the medians of the treatment phase measurements (two-tailed Wilcoxon signed ranks test for two related samples) for the total of all Goal Attainment Scaling scores (n=33) showed a significant improvement (p<0.001). When testing at subject level (medians of the three Goal Attainment Scaling scores per assessment, n=11), a significant improvement was also found (p<0.005).

The median of all data per assessment, divided into the two groups with different periods of the baseline phase, is shown as a graph in **Figure 2.2**. The change in Goal Attainment Scaling score is related to the moment of treatment with botulinum toxin A. The median per measurement point of all Goal Attainment Scaling scores, the best two scores and the best score out of the three goals per subject is shown in **Figure 2.3a–c**. Comparing the data obtained from the video fragments (V) and the physiotherapist's scoring of the Goal Attainment Scaling (P), showed a good weighted kappa correlation (0.63). A two-tailed Wilcoxon signed ranks test, applied to compare the V and P data per Goal Attainment Scaling, resulted in only five goals out of 33 with significant differences between the V and the P data. These goals did not show any significant improvement in the data of the videoscoring of the Goal Attainment Scaling. The spasticity score on the Modified Ashworth Scale (MAS) did show an improvement in all 11 subjects. The baseline measurements (n=144) of the treated muscles per subject showed a median

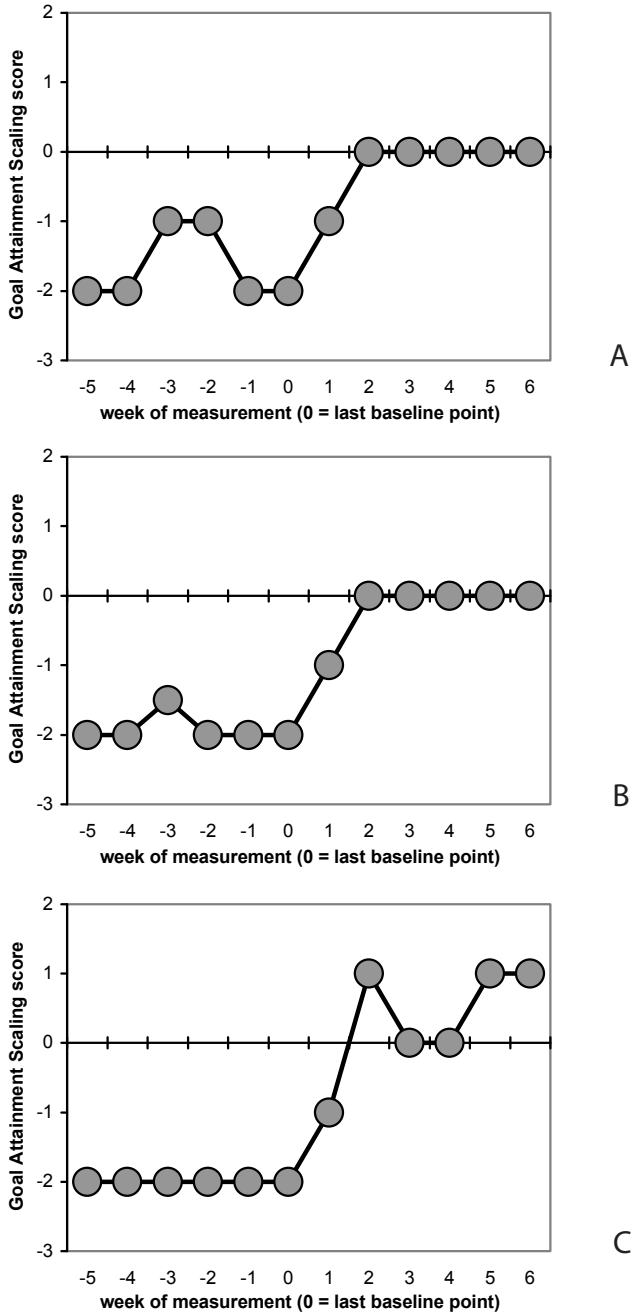


Figure 2.3 (a) Median of all Goal Attainment Scaling scores (33) per measurement point (t=–5 to t=6). (b) Median of the best two scores per subject (22) per measurement point. (c) Median of the best score per subject (11) per measurement point.

MAS of 2 (interquartile range 0.5–3). The treatment phase measurements (n=114) showed a median MAS of 1 (interquartile range 0–2).

DISCUSSION

In this study, 11 out of 33 (33%) functional ability goals were achieved by 7 out of 11 children with Cerebral Palsy after treatment with botulinum toxin A. The design of the study made statistical evaluation of individual effects possible, using subjects as their own control. Despite the small number of children, treatment effects could be demonstrated. With group level analysis of all 33 goals, a clinically relevant effect could also be demonstrated. Improvement in the two groups after different periods of the baseline phase indicated that the effect was not caused by spontaneous development or by physical or occupational therapy (**Figure 2.2**).

The findings agree with the results of a study carried out by Paolicelli,[30] in which Goal Attainment Scaling was used to evaluate the effect of botulinum toxin A on walking disorders in children with Cerebral Palsy: 85% of 54 children managed to achieve the expected functional goals. The difference could be explained by a lack of blinded scoring in that study.

With regard to the design, there is only one previous study, carried out by Garcia Ruiz et al.,[31] in which a blinded evaluation was made of gait on video-recordings, according to a 5-point scale. In that study, all patients showed a progressive improvement in their gait pattern over a period of 33 months. It is difficult to compare the outcome of that study with the results of the present study, because there is an important difference between goals in quality of gait and functional goals. Moreover, in the Garcia Ruiz study there was no control for spontaneous development.

In the present study, on the one hand the experience with Goal Attainment Scaling in testing the effect of botulinum toxin A in Cerebral Palsy was promising; on the other hand the study provided insight into concerns about the use of Goal Attainment Scaling. The study population (i.e., children with Cerebral Palsy) is a very heterogeneous group with regard to motor impairment and abilities. The method used in this study was found to be useful to evaluate the effect of an intervention at individual level, thus resolving the problem of the heterogeneous population. However, in research on Cerebral Palsy very little is known about the method of scale and goal selection, the specificity of the scale content, the appropriateness and relevance of the scales (content validity and content reliability), and

the interrater reliability of scoring on the predetermined scales. The quality of the goal-setting does influence the results. Goals should be realistic and achievable, but ambitious enough to justify the impact of the treatment with botulinum toxin A on a child. That is a subjective measure, and although the design of the present study made it possible to control for therapist-bias in *scoring* on the Goal Attainment Scaling, the *content* of the scales and the effect *size* were determined by the team of professionals.

Another limitation of Goal Attainment Scaling was the difficulty of implementation in the rehabilitation team. Defining six realistic, distinct levels of outcome for each child's own functional goals, with no gaps or overlap between levels, and with full agreement of the children and their parents, was a time-consuming procedure. Finally, one limitation of the video-recording method is that the Goal Attainment Scaling could only simulate the child's own functional setting, instead of reflecting actual daily practice. However, all the children and their parents were convinced that the outcome of the study was representative of their own setting. Nine of the 11 children were very satisfied with the treatment results.

Despite the questions that remain about the use of Goal Attainment Scaling, the results confirm that it can possibly be a reliable method with which to assess the outcome of botulinum toxin A treatment in children with Cerebral Palsy.

Future research on the psychometric properties of Goal Attainment Scaling in patients with Cerebral Palsy is necessary. A comparative study, based on individual treatment goals and generic outcome measures, such as the GMFM, should be carried out to compare sensitivity to change as part of the construct validity. The content validity, content reliability and inter-rater reliability must also be measured.

CLINICAL MESSAGES

- Clinically relevant effects of botulinum toxin type A treatment to improve the level of functional abilities were demonstrated in 7 out of 11 children with Cerebral Palsy, by individual functional goal-setting.
- Goal Attainment Scaling can be used to measure these effects at individual level.
- Future research on the psychometric properties of Goal Attainment Scaling in patients with Cerebral Palsy is necessary.

ACKNOWLEDGEMENTS

The authors wish to thank all participants from the rehabilitation centre 'De Trappenberg' in Huizen and, are grateful to Tjeerd van der Ploeg for his statistical support.

REFERENCES

1 Lance JW. (1980) The control of muscle tone, reflexes, and movement: Robert Wartenberg Lecture. *Neurology* **30**: 1303–1313.

2 Boyd RN, Dobson F, Parrott J, Love S, Oates J, Larson A, Burchall G, Chondros P, Carlin J, Nattrass G, Graham HK. (2001) The effect of botulinum toxin type A and a variable hip abduction orthosis on gross motor function: a randomized controlled trial. *Eur J Neurol* **8**: 109–119.

3 Love SC, Valentine JP, Blair E M, Price CJ, Cole JH, Chauvel PJ. (2001) The effect of botulinum toxin type A on the functional ability of the child with spastic hemiplegia: a randomized controlled trial. *Eur J Neurol* **8**: 50–58.

4 O'Donnell ME. (2001) Randomized double-blind placebo controlled trial of the effect of botulinum toxin on walking in cerebral palsy. *J Pediatr* **139**: 163.

5 Desloovere K, Molenaers G, Jonkers I, De Cat J, De Borre L, Nijs J, Eyssen M, Pauwels P, De Cock P. (2001) A randomized study of combined botulinum toxin type A and casting in the ambulant child with cerebral palsy using objective outcome measures. *Eur J Neurol* **8**: 75–87.

6 Ade-Hall RA, Moore AP. (2000) Botulinum toxin type A in the treatment of lower limb spasticity in cerebral palsy. *Cochrane Database Syst Rev* CD001408.

7 Ubhi T, Bhakta BB, Ives HL, Allgar V, Roussounis SH. (2000) Randomised double blind placebo controlled trial of the effect of botulinum toxin on walking in cerebral palsy. *Arch Dis Child* **83**: 481–187.

8 Koman LA, Mooney JF III, Smith BP, Walker F, Leon JM. (2000) Botulinum toxin type A neuromuscular blockade in the treatment of lower extremity spasticity in cerebral palsy: a randomized, doubleblind, placebo-controlled trial. BOTOX Study Group. *J Pediatr Orthop* **20**: 108–115.

9 Corry IS, Cosgrove AP, Duffy CM, Taylor TC, Graham HK. (1999) Botulinum toxin A in hamstring spasticity. *Gait Posture* **10**: 206–210.

10 Edgar TS. (2001) Clinical utility of botulinum toxin in the treatment of cerebral palsy: comprehensive review. *J Child Neurol* **16**: 37–46.

11 Chutorian A, Root L, BTA Study Group. (1995) A multicentered, randomised, double-blind placebocontrolled trial of botulinum toxin A in the treatment of lower limb spasticity in pediatric cerebral palsy. *Mov Disord* **10**: 364.

12 Pandyan AD, Johnson GR, Price CIM, Curless RH, Barnes MP, Rodgers H. (1999) A review of the properties and limitations of the modified Ashworth Scales as measures of spasticity. *Clin Rehabil* **13**: 373–383.

13 Fosang AL, Galea MP, McCoy AT, Reddihough DS, Story I. (2003) Measures of muscle and joint performance in the lower limb of children with cerebral palsy. *Dev Med Child Neurol* **45**: 664–670.

14 Palisano RJ, Hanna SE, Rosenbaum PL, Russell DJ, Walter SD, Wood EP, Raina PS, Galuppi BE. (2000) Validation of a model of gross motor function for children with cerebral palsy. *Phys Ther* **80**: 974–985.

15 WHO. (2001) *International classification of functioning, disability and health (ICF)*. Geneva Switzerland: World Health Organization. Accessed from http://www3.who.int/icf/icftemplate.cfm.

16 Beckung E, Hagberg G. (2002) Neuroimpairments, activity limitations, and participation restrictions in children with cerebral palsy. *Dev Med Child Neurol* **44**: 309–316.

17 Kiresuk T, Sherman R. (1968) Goal attainment scaling: a general method of evaluating comprehensive community mental health programs. *Community Ment Health J* **4**: 443–453.

18 MacKay G, Somerville W. (1996) Reflections on goal attainment scaling (GAS): cautionary notes and proposals for development. *Educ Res* **38**: 161–172.

19 Palisano RJ, Haley SM, Brown DA. (1992) Goal attainment scaling as a measure of change in infants with motor delays. *Phys Ther* **72**: 432–437.

20 Donnely C, Carswell A. (2002) Individualized outcome measures: a review of the literature. *Can J Occup Ther* **69**: 84–94.

21 Zaza C, Stolee P, Prkachin K. (1999) The application of goal attainment scaling in chronic pain settings. *J Pain Symptom Manage* **17**: 55–64.

22 Fisher K, Hardie RJ. (2002) Goal attainment scaling in evaluating a multidisciplinary pain management programme. *Clin Rehabil* **16**: 871–877.

23 Rushton PW, Miller WC. (2002) Goal attainment scaling in the rehabilitation of patients with lower-extremity amputations: a pilot study. *Arch Phys Med Rehabil* **83**: 771–775.

24 Rockwood K, Graham JE, Fay S; ACADIE Investigators. (2002) Goal setting and attainment in Alzheimer's disease patients treated with donepezil. *J Neurol Neurosurg Psychiatry* **73**: 500–507.

25 Gordon JE, Powell C, Rockwood K. (1999) Goal attainment scaling as a measure of clinically important change in nursing-home patients. *Age Ageing* **28**: 275–281.

26 Shefler G, Canetti L, Wiseman H. (2001) Psychometric properties of goal-attainment scaling in the assessment of Mann's Time-Limited Psychotherapy. *J Clin Psychol* **57**: 971–979.

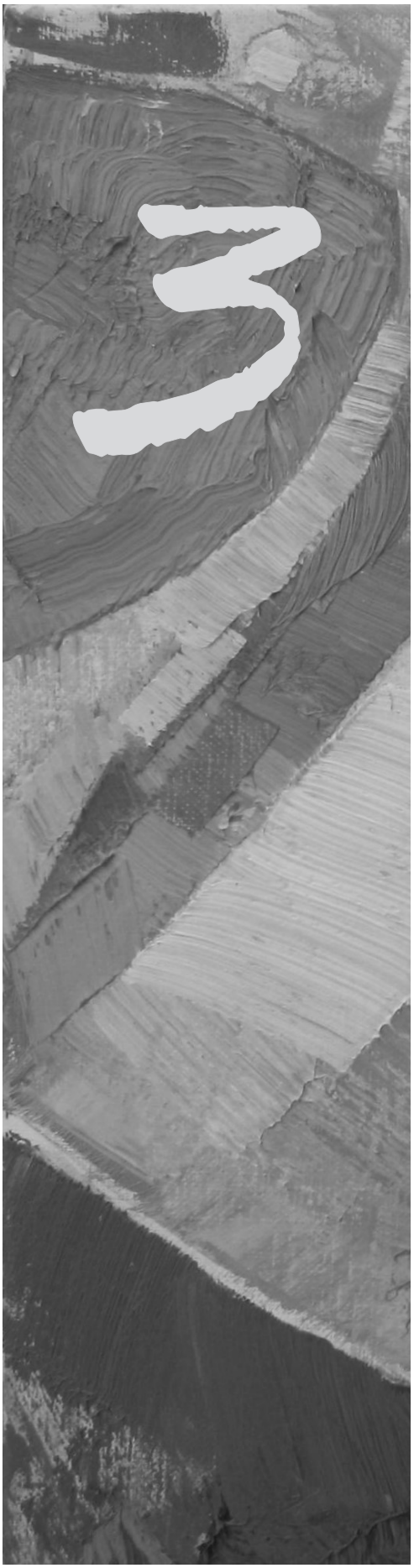
27 Smith A, Cardillo JE, Smith SC, Amezaga AM Jr. (1998) Improvement scaling (rehabilitation version). A new approach to measuring progress of patients in achieving their individual rehabilitation goals. *Med Care* **36**: 333–347.

28 Siegel S, Castallan NJ. (1988) *Nonparametric statistics for the behavioral sciences*. 2nd ed. New York: McGraw-Hill.

29 Wood E, Rosenbaum P. (2000) The gross motor function classification system for cerebral palsy: a study of reliability and stability over time. *Dev Med Child Neurol* **42**: 292–296.

30 Paolicelli PB. (2001) Use of botulinum toxin type A in walking disorders of children with cerebral palsy. *Eur Medicophys* **37**: 83–92.

31 Garcia Ruiz PJ, Pascual PI, Sanchez BV. (2000) Progressive response to botulinum A toxin in cerebral palsy. *Eur J Neurol* **7**: 191–193.



Goal Attainment Scaling in paediatric rehabilitation: a critical review of the literature

D. Steenbeek, M. Ketelaar,
K. Galama, J.W. Gorter

Dev Med Child Neurol 2007,
49: 550–556.

ABSTRACT

The aim of the study was to review the psychometric properties and use of Goal Attainment Scaling (GAS) in paediatric rehabilitation research. We performed a critical literature review searching: (1) all studies whose main focus was to assess the psychometric properties of GAS in paediatric rehabilitation; and (2) all effect studies in paediatric rehabilitation that used GAS as one of the outcome measures. Three articles in the first group and six in the second group met the inclusion criteria. None of the studies had investigated the content reliability of the scales. Interrater reliability had been investigated in one study and had been found to be good. Only one trial had assessed the content validity of the developed scales, which was found to be acceptable. Comparisons showed that GAS, because of its idiosyncratic nature, measures different constructs from those measured by some related instruments. Low concurrent validity was found. All included studies reported good sensitivity to change. We conclude that the literature supports promising qualities of GAS in paediatric rehabilitation. GAS is a responsive method for individual goal setting and for treatment evaluation. However, current knowledge about its reliability when used with children is insufficient. There is a need for further development of GAS and its application for children of different ages and disabilities, across therapists of different disciplines.

INTRODUCTION

In the paediatric rehabilitation setting, measuring children's progress towards individual goals is increasingly important. Such measurements must be individual, in view of the diversity of developmental disabilities, goals, and interventions. The heterogeneity of the population often induces researchers to use generic standardized measurement tools or health-related quality of life (QoL) measures; however, these include so many items that this limits their specificity and responsiveness to change. In contrast, in studies of homogeneous groups the sample size is often too small to detect convincing and clinically relevant differences between two treatment strategies.[1] A possible solution for measurement in heterogeneous groups is the use of individual measurement tools, one example of which is Goal Attainment Scaling (GAS).[2] GAS is a method of measuring individual progress towards individual goals. These goals should be specific, measurable, acceptable, relevant, and time-related (SMART). GAS was originally a 5-point scale, with 0 representing the expected level of success. If a patient achieves more than was expected, on the basis of pre-established criteria, a score of +1 or +2 is given, depending on the level of achievement; if a patient's progress is less than expected, also on the basis of pre-established criteria, a score of –1 or –2 is given (**Table 3.1**). Recently, authors have experimented with scales with 6 points (**Table 3.1**),[3] 7 points,[4] or 3 points.[5]

Table 3.1 Example of Goal Attainment Scaling for a 6-year-old male with spastic hemiplegia, Gross Motor Function Classification System Level II

	Score	Definition	Physical therapy	Occupational therapy	Speech therapy
Setting			A gymnasium with an obstacle course including wooden benches and climbing rack against the wall	Sitting at a table with a piece of paper that has a 22cm circle drawn on it and a pair of scissors. Keep verbal help to a minimum	Social communication in a group session; approx. 15 min
Task			Walk the obstacle course as fast as possible	Cut out the circle as accurately as possible	Ask fellow group members questions during a discussion
	−3	Worse than starting level (deterioration)	40s	Does not follow the line	Shows no interest
	−2	Equal to starting level	35s	Cuts along the line for over 5% of the circle	Shows interest but does not ask any questions
	−1	Less than expected	31s	Cuts along the line for over 50% of the circle with a lot of verbal stimulation	Asks a question when told to do so by the speech therapist
	0	Expected goal	27s	Cuts along the line for over 50% of the circle without verbal stimulation	Asks a question when the speech therapist suggests it
	1	Somewhat more than expected	25s	Cuts along the line for over 75% of the circle	Spontaneously asks a group member one question
	2	Much more than expected	23s	Cuts along the line for over 90% of the circle	Spontaneously asks his fellow group members more than one question

Kiresuk and Sherman[6] first introduced GAS to evaluate mental health services. They transformed several scales per subject into a single *T* score, in which each goal is weighted and a mostly intuitive average correlation is corrected. According to Sherman,[7] the mean of large series of *T* scores would be expected to converge to 50 (SD 10), so a score of 50 or more indicates attained goals.

Since its introduction, GAS has been used in a variety of settings and for many purposes.[8,9] The current debate about the use of GAS[8,10] includes issues such as: (1) sensitivity to subtle but important changes; (2) reliability of the scale development (content reliability) as well as of the scores (interrater reliability); (3) content validity, namely whether the content of the scales covers all aspects or elements of interest that are to be measured; and (4) construct validity, namely whether the scale measures the construct that it purports to measure. Definitions are given in **Table 3.2**. Evidence of the psychometric properties of GAS has come predominantly from the fields of mental health and geriatric medicine. Studies in these fields have reported that GAS has good content reliability,[10] an overall satisfactory interrater reliability (r=0.51–0.91), good content validity, good convergent validity compared with a commonly used QoL measure, and good responsiveness.[11,12]

In applying GAS, the selection of goals, the specification of outcomes, and the content of scales, and hence also the size of the effect, depend on the expectations of the team of professionals involved. The administration and scoring procedures in the paediatric rehabilitation setting cannot be compared with those in the field of mental health and geriatric medicine. How do we know if selected goals are relevant, realistic, and reproducible?

Two recent critical reviews of the use of GAS in rehabilitation show

Table 3.2 Definitions

Term	Definition
Interrater reliability	Correlation between the scores of two independent raters on the same GAS scale
Content reliability	Correlation in terms of textual content of GAS scales between two independent scale developers
Content validity	Value of the textual content of GAS scales in relation to the relevance to the subject
Concurrent validity	Correlation between the outcome in terms of GAS and other appropriate measures
Social validity	Subjective experience of users of the measure

GAS, Goal Attainment Scaling.

that little is yet known about its properties when used in this field. Schlosser[8] presented an overview of the issue in the field of communication disorders, and Donnelly and Carswell[9] compared the individualized outcome measures from the perspective of occupational therapy. Despite this lack of knowledge about its properties in rehabilitation medicine, GAS has been used in several effect studies as one of the outcome measures. Some examples are studies evaluating a comprehensive day-treatment programme,[13] botulinum toxin A (BTX-A) treatment,[14] lower-extremity amputations,[15] and pain management programmes.[16,17] These studies support the hypothesis that in the field of rehabilitation medicine, GAS provides a more responsive measure than other, more conventional outcome measures (such as the Barthel Index or the Locomotor Capabilities Index).

Thus, many authors have attested to the value of GAS as an outcome measure, confirming the assumption that GAS could be of special value in paediatric rehabilitation medicine, with its heterogeneity of therapy goals set by interdisciplinary teams working with children with various developmental disabilities. A critical review of the literature was, therefore, required, to weigh the positive expectations of GAS against the crucial questions about its use.

The aim of the present study was to review the available information about the content and interrater reliability of GAS as well as its validity, sensitivity to change, and use in paediatric rehabilitation research.

METHOD

A literature search and review was performed by the first author and critically checked by the second author. Two groups of studies were defined: group 1 included all studies whose main focus was to measure the psychometric properties of GAS in paediatric rehabilitation; and group 2 included all studies that examined the effect of a specific paediatric rehabilitation intervention using GAS as one of the outcome measures. Key words used in the search were goal attainment (scaling or scale) and children.

We searched the Medline and EMBASE electronic databases, without limitations, as well as reference lists from pertinent articles. Our review included all studies in the field of paediatric physical and occupational therapy, and communication disorders, excluding applications of GAS among children in school and psychiatric settings. This decision was based on the assumption that the psychometric properties of GAS in the interdisciplinary

therapeutic setting are specific and not comparable to those in school or psychiatric settings. Duplications of previously published studies or findings without new tests of its properties (e.g. to provide recommendations about its future use in a new field) were excluded.

The studies of the first group were evaluated in terms of the reliability, validity, and sensitivity to change of GAS. The studies of the second group were evaluated in terms of type of trial, population, and goals, parameters and effects measured with GAS, outcome of GAS, comments on scale development, and scoring techniques.

RESULTS

The Medline searches retrieved a total of 42 articles, eight of which were found to meet the inclusion criteria after a review of the full text.[3–5,18–22] EMBASE searches did not reveal any other relevant articles. Another relevant article was found from the reference lists.[23]

Psychometric properties of GAS in paediatric rehabilitation

Three studies met the inclusion criteria for this group[4,21,22] (**Table 3.3**). Palisano et al.[21] investigated the validity of GAS as a measure of motor change in 65 infants, aged 3 to 30 months, with motor delays. Their study used a 6-month period of intervention with standard physical and occupational therapy. Eighty-three gross motor and 47 fine motor rehabilitation goals were developed. The content validity, responsiveness, and concurrent validity were investigated in a supplemental study with 21 infants and two periods, with 42 scales each constructed by 10 physical therapists.[22] The third study in this group, by Cusick et al.,[4] reported on the use of GAS on data from an intervention study,[18] with a 3-month period starting with the 'proven treatment' with BTX-A and followed by occupational therapy, in a controlled design. The study included 162 specific upper limb and self-care goals developed in 41 children, aged 2 to 7 years, with spastic unilateral cerebral palsy (CP; hemiparesis). The relative value of GAS for paediatric rehabilitation research was compared with that of the Canadian Occupational Performance Measure (COPM).

The content reliability of GAS was not investigated in any of these studies. The interrater reliability of GAS was examined only by Palisano,[22] before and during data collection by scoring from videotape. He found a kappa coefficient of 0.89 between his own score and that of the examiner, indicating good interrater reliability. As regards validity, both Palisano[22] and

Table 3.3 Properties of Goal Attainment Scaling (GAS) in paediatric rehabilitation medicine

Study	Type	Purpose/ properties measured	Population	Design	Results	Conclusion of study	Comments on use of GAS
Cusick et al.[4]	Comparison of two individual outcome measures for paediatric rehabilitation research in an effect study with a 'proven intervention'[18]	Instrument sensitivity, convergent validity, goal/ problem profiles and administration were evaluated with GAS and COPM	41 children with spastic hemiplegic CP; evaluation of upper limb	Type B randomized study with occupational therapy only or in combination with BTX-A	GAS was sensitive to group change and detected significant changes between groups. Convergent validity assessment detected no correlation between any GAS or COPM measures. Scale development cost an average of 45min	Both GAS and COPM are sensitive to change after a proven intervention; the two evaluate different constructs	A 7-point self-constructed Likert scale variant of GAS was more sensitive than the traditional weighted GAS
Palisano[22]	Evaluation of a 6-month period of physical and occupational therapy	Content validity, responsiveness, concurrent validity	21 infants, 4–24 months of age, with motor delays, 42 goals for 2 periods of 3mo each	Prospective descriptive study	77–88% of ratings met the criterion for content validity. Responsiveness of GAS better than behavioural objective. No significant correlation between GAS and Peabody developmental gross motor scale	Initial support for content validity of GAS. Good responsiveness. GAS and Peabody measure different aspects	Classic procedure of scale development
Palisano et al.[21]	Evaluation of a 6-month period of physical and occupational therapy	Validity, responsiveness	65 young infants with motor delays, 130 goals	Prospective descriptive study	Moderate or low correlation between GAS and Peabody gross motor and fine motor change scores. 69% of the infants attained a *T* score of 50 or higher.	The unique features of GAS offer advantages for measurement of motor change compared with both other measures	Classic procedure of scale development by the treating therapist, independent rating

COPM, Canadian Occupational Performance Measure; CP, cerebral palsy; BTX-A, botulinum toxin A.

Cusick et al.[4] concluded that the instruments they compared were measuring different constructs, owing to the idiosyncratic nature of GAS.

As expected, a low concurrent validity was found. Palisano et al.[22] measured concurrent validity by comparing the GAS *T* scores with the Peabody gross and fine motor scale age-equivalent change score, and found low correlations (r=0.44 and 0.18, respectively). Cusick et al.[4] assessed concurrent validity by using a correlation matrix with COPM and GAS Likert scale data, and found no correlation between any of the GAS and COPM measures.

Palisano[22] found that between 77 and 88% of the therapists' ratings met the criterion for content validity in a random sample of 10 goals of 10 subjects. Ten paediatric physical therapists rated the 10 scales on three dimensions: the importance of the goals, the degree to which the expected level of goal attainment was achieved, and the extent of each level in the scales. Each goal was accompanied by a profile of the young children in the early intervention programme.

With regard to sensitivity to change, all three studies concluded that the responsiveness of GAS depends on whether therapists and parents select goals and levels of attainment for each goal that represent clinically important changes in future performance. Both Cusick et al.[4] and Palisano et al.[21] found a significantly higher mean attainment score than the expected mean of 50 ($p \leq 0.001$). Palisano et al.[21] found mean levels of attainment for two goals of +0.6 (SD 1.4) and +0.3 (SD 1.5) respectively, and a mean GAS *T* score of 55.4 (SD 14.6). Moreover, the supplemental study[22] showed that in 61% of the goals, change that could not be measured with the Behavioural Objective format was measured with the GAS format. Multiple regression was used to investigate variables as sources of bias, and it was found that infants who were less delayed in motor development had higher GAS *T* scores. Cusick et al.[4] described the ability of a weighted GAS *T* score and a 7-point selfconstructed Likert scale variant of GAS to detect change within and between groups, by calculating effect sizes and using linear regression. Between baseline and 3 months, they found large original GAS effect sizes in the BTX-A group compared with the control group. Ceiling or bottoming effects were avoided by means of the Likert coding option developed by Cusick et al.,[4] which resulted in larger effect sizes.

Goal Attainment Scaling used in effect studies in paediatric rehabilitation

As shown in **Table 3.4**, six other articles met the inclusion criteria for group 2 of our review, one of which was a sub-study of the study by Cusick et al.[4] discussed above.[18] In these studies, a total of 284 children were assessed with 859 GAS scales of varying quality, using a variety of scale development techniques in the different study designs. Five recent studies[3,18,19,20,23] evaluated BTX-A treatment. These studies found significant and clinically relevant effects of treatment with BTX-A as measured by GAS. One non-controlled cohort study[5] evaluated the effect of functional training in children with CP, and found that the children attained 77% of the goals. This study focused primarily on the effect of the introduction of goal-directed training rather than on outcome measurement with GAS.

Most of the goals set in the six studies were in the areas of activities and participation, but body function goals, such as gait pattern[19,23] and range of motion,[18] were also assessed. The results of these studies show that GAS is sensitive to change when used to measure an effect of BTX-A and an effect of functional training in children. Mall et al.[19] and Paolicelli[23] discussed the failure of the Gross Motor Function Measure (GMFM) to detect superiority of BTX-A treatment and the sensitivity of GAS to change at the International Classification of Functioning, Disability and Health (ICF) activity level. Another study, using a multiple baseline/treatment phase design and blinded scoring of scales by means of videotape recordings, found a high level of evidence of good responsiveness for GAS.[3] A change of more than 1 point in the median of GAS scores was related to the moment of treatment with BTX-A. Some authors also reported a very positive perception of parents working with GAS and discussed the possibility of improving the quality of rehabilitation services.[3,5]

Although GAS data are considered at best to be ordinal, seven studies used statistical methods treating the data as interval data. Only two studies analyzed the data as ordinal data.[3,23] Four studies provided insight into concerns about the use of GAS in the discussion section on the basis of their own impressions of issues relating to the method of scale and goal selection, the specificity of the scale content, the appropriateness and relevance of the scales and the possibilities of influencing the reliability of scoring GAS.[3,5,18,23] GAS was described as a time-consuming procedure, with scale development requiring about 45 minutes per child.[3,4]

Table 3.4 Goal Attainment Scaling (GAS) used in paediatric rehabilitation research

Study	Type	Purpose	Population	Variables and measures	Outcome on GAS	Conclusion of study	Comments on use of GAS/ statistics
Mall et al.[19]	RCT	Evaluation of the effects of BTX-A on adductor spasticity	61 children with CP	Knee-knee distance, Ashworth, GMFM, GAS	Significant superiority of the BTX-A group (p=0.037)	Four weeks after treatment, significant effect was measured in all variables except for GMFM.	At least 61 goals at the level of activity and participation, including pain and walking patterns, were defined by the doctor in attendance. Scale development and scoring method were not described. GAS data treated as interval data
Lowe et al.[18]	RCT	Evaluation of the effects of specific BTX-A on upper limb movement quality and function	42 children with hemiplegic CP	Quest, COPM, Parent GAS and therapist GAS, PEDI, Ashworth at 1, 3, 6 months	Significant superiority of the BTX-A group (GAS [parent] p=0.001, GAS [therapist] p<0.001)	All assessments found a significant effect, except for Quest at 6 months and PEDI caregiver assistance	162 specific upper limb and self-care goals. GAS data were considered ordinal but treated as interval data (parametric statistics)
Steenbeek et al.[3]	RCC	To determine the possibility of using GAS to measure effect of BTX-A treatment	11 children with CP	GAS in an evaluator-blinded randomized multiple baseline/ treatment phase design with standardized videotapes	9/11 participants showed significant improvement in 18 of 33 goals. Testing at group and individual levels resulted in significant improvement (p<0.001;0.005)	Clinically relevant improvement in individual rehabilitation goals (ability level) were demonstrated using GAS	33 functional goals in the area of abilities. A score of –3 for deterioration was added (6-point scale). No change was scored as –2. Blinded scoring of scales by means of videotape recordings. GAS data were treated as ordinal data

Ekström-Ahl et al.[5]	NCC	Evaluation of functional training in CP	14 young children with CP (1-6y of age)	GAS, GMFM, PEDI, MPOC, self-constructed questionnaire	77% of goals were fully attained, most of them relating to activities	Children with CP benefited from a functional goal-directed training approach in their development of gross motor function and everyday activities. All measures showed significant effects	98 specific and measurable goals were set. A self-constructed scale was used in which goal attainment was graded as complete (100%), partial (>50%) or no improvement (<50%). GAS data were treated as interval data. GAS was used as a quality improvement tool ('the cement that can assist in avoiding fragmentation of services') rather than as an outcome measure
Wallen et al.[20]	NCC	Phase II trial evaluating upper limb function after BTX-A treatment	16 children with CP (2-12y of age)	COPM, GAS, Melbourne, CHQ, MAS, Tardieu	Mean *T* scores 42.3 and 47.4 at 3 vs 6mo	BTX-A therapy was effective	74 upper limb goals were set. Scale development and method of scoring were not described. GAS data were treated as interval data
Paolicelli et al.[23]	NCC	Effect study of BTX-A for the management of walking disorders	54 ambulatory children with CP	GAS, GMFM, video recording, Ashworth, parental interview	Only 15% of the children did not achieve the expected results	BTX-A was effective	217 GAS scales were set for the standing and walking parameters. Scale development was not described. The scoring was not blinded. Data were treated as interval data

RCT, randomized controlled trial; RCC, randomized concurrent cohort; NCC, non-controlled cohort; CP, cerebral palsy; GMFM, Gross Motor Function Measure; QUEST, Quality of Upper Extremity Skills Test; COPM, Canadian Occupational Performance Measure; PEDI, Pediatric Evaluation of Disability Inventory; MPOC, Measure of Processes of Care; CHQ, Child Health Questionnaire; MAS, Modified Ashworth Scale; BTX-A, botulinum toxin A.

DISCUSSION

The aim of the present literature review was to study the properties of GAS in paediatric rehabilitation. The findings provide evidence for several favourable qualities of GAS as a method of assessing progress in children. The method provides clear goals and priorities for intervention and reflects a clientcentred perspective to service delivery.[2]

However, some properties have not yet been studied in detail. One of the most remarkable outcomes of our literature search was that the reliability of GAS is largely unknown. Twenty-four years after the introduction of GAS in rehabilitation by Clark and Caudrey,[24] very little is known about the reliability of the scale content and scoring of predetermined scales. Despite this serious lack of knowledge, GAS has been used extensively in recent outcome measurements among children with CP to derive assumptions about the functional benefits of certain interventions. This is especially true for measurements of the effect of BTX-A treatment.

One problematic aspect of reliability is that the clinical judgement of the therapists, who determine the expected level of attainment for each goal themselves, not only is a potential source of therapist bias (subjectivity) in scoring,[21] but also depends on many unpredictable factors. Extra caution is required when goals at both the ICF activity/participation level and body function level are used and weighted, as was done in the studies by Lowe et al.[18] and Mall et al.[19] For example,' knee–knee distance'[19] is not an outcome measure provided by the child itself, but it is more reliable than a GAS scale on climbing stairs.

In view of the psychometric properties, it is generally assumed that those who have to use GAS have an absolute need to be trained in its use. Nevertheless, none of the articles that met the inclusion criteria for the present study described any GAS training for the investigating team, although 7 years after the initial publications of Palisano et al.,[21] the Canadian group (King et al.[2]) reported on their many years of experience in conducting formal programme evaluations with GAS. The authors describe their strict instructions about training (7h of specific training), goal selection, scale development, and rating. Subjective elements include the relationship between body structure, activities and participation, the prediction of learning potential, intercultural differences, and the interpretation of clinical relevance in relation to the intensity of therapy. In addition, goals depend on the children's cognitive functions and developmental delay. This could explain the poorer outcome in terms of GAS data among children with more severe disabilities.[21] These

are all factors that necessitate training in GAS. Although a double-blind randomized design (such as that of Mall et al.[19]) can prevent therapist bias, even in the best trials the reliability of functional rating on GAS scales for children depends on the children's psychomotor control, fatigue, motivation, interpersonal interaction with the therapist, and behaviour. A multiple baseline/treatment phase design[3] is able to compensate for the changes in the children's psychomotor control, but puts a great burden on the participants.

The validity of GAS is ambiguous. The studies by Palisano et al.[21,22] and the studies using GAS as one of several measures have revealed the different constructs measured by GAS. This represents a potential source of measurement error that is unique to the idiosyncratic nature of GAS. Although it poses a threat to validity, it also offers an opportunity to measure exactly what one intends to measure. The method can ensure the ongoing relevance of the child's goals by basing them on the request for help by the child and the parents. As expected, however, low concurrent validity was found in comparisons with other measures. Nevertheless, this could also be an advantage, because GAS was more sensitive than the norm-referenced measures.

The sensitivity to change of GAS can be assumed to be better than that of the common standardized functional measures at the level of abilities and participation, despite the questions that remain about its reliability. For example, GAS can capture subtle but important changes in activity skills of the upper extremity that are not items of the GMFM nor the Pediatric Evaluation of Disability Inventory (PEDI). A study found that 14% of individual physical therapy treatment goals of children with CP remained undetected by PEDI and GMFM.[25] GAS is likely to be more sensitive in this respect than GMFM and PEDI, although both have shown responsiveness to change in activities and participation of children with CP (Gross Motor Function Classification System Level I or II), particularly children under 4 years of age.[1]

Three issues relating to the sensitivity to change need to be discussed in more detail: the statistical compilation of data, the weighting of goals, and bottoming effects.

(1) There are concerns about the statistical calculation of group effects with the traditional single numeric *T* score formula of Kiresuk and Sherman.[6] A fundamental problem is that the data are usually treated as interval data, using mean scores for groups, whereas at best they represent an ordinal scale of value judgements. Although recognizing traditional arguments favouring non-parametric treatments for this type of data,[26] most analyses, even in

recent studies, have used parametric statistics and means in the group data of *T* scores. This may be the ultimate problem with GAS: it does not fulfil its claim of being a parametric expression of non-parametric information. In fact, it gives a false impression of precision by yielding a standardized score with several decimal places and a false sensitivity to change. Solutions have been discussed[3] that involve treating the data in medians instead of means, with non-parametric statistics.

(2) Because of the subjectivity of the weighting process itself, caution must also be exercised in the interpretation of any results after weighting goals. The internal consistency depends on the average correlation between the different GAS scales per subject, and is not known. The interdependence of goals per subject could represent another risk of false sensitivity to change. It would probably be best to abandon the *T* score, 38 years after its introduction. Nonparametric techniques are being developed that provide statistical alternatives.

(3) A third potential bias resulting in false sensitivity to change could be that of bottoming effects. The original GAS by Kiresuk and Sherman[6] was scored before and after the intervention, and deterioration could only be scored if the start was defined as –1. This can introduce a bias for the calculation of group effects, because only three values then remain to score progress, the first of which is the expected goal (0). Most users score 'no change' by definition as –2, but this means that deterioration cannot be scored, which is also a potential bias. Experiments with an additional score of –3 for deterioration (6-point scale)[3] and a 7-point Likert scale[4] have indicated possibilities to retain all original goal attainment levels while avoiding the bottoming and ceiling effects.

All of the articles included in the present review reported teams and parents as being very satisfied about working with GAS, indicating an acceptable 'social validity'.[8] GAS was enthusiastically accepted by clinicians and parents; this was also confirmed by two case descriptions not included in our study.[27,28] Besides its use in physical, occupational, and speech therapy, there have also been reports on the use of GAS by other disciplines of the paediatric rehabilitation team, such as social workers.[29] GAS reflects the typical clinician's and parent's thinking as they judge actual outcome against what they would have expected at the time that treatment was started.[30] Although GAS has mostly been used as an evaluative outcome tool, the study by Ahl et al.[5] used it as a treatment tool to improve quality. It has also been suggested that GAS could have a therapeutic effect in itself,[31,32] probably because it focuses attention on a goal-directed interdisciplinary team approach.

CONCLUSIONS AND RECOMMENDATIONS

Although few publications have convincingly assessed the psychometric properties of GAS, this method does have promising properties in terms of measuring changes in activities and participation in children with disabilities. The method can be used to measure progress in heterogeneous populations with a variety of treatment goals. The reliability of GAS in paediatric rehabilitation remains largely unknown. It is generally assumed that training in the use of GAS is absolutely necessary to ensure its reliability, and research reports should include descriptions of such training to allow accurate reproduction of studies. The validity of GAS is ambiguous: whereas its concurrent validity is low, it offers good opportunities to measure exactly what needs to be measured. The sensitivity to change of GAS can be assumed to be better than that of common standardized functional measures at the level of abilities and participation. Teams of therapists and parents are very satisfied about working with GAS. Further development of GAS and its application in children of various ages and disabilities across therapists of different disciplines is necessary and of great importance for the field of childhood disability.

ACKNOWLEDGEMENTS

The authors wish to thank J. Klerkx for his textual improvements.

REFERENCES

1. Vos-Vromans DC, Ketelaar M, Gorter JW. (2005) Responsiveness of evaluative measures for children with cerebral palsy: the Gross Motor Function Measure and the Pediatric Evaluation of Disability Inventory. *Disabil Rehabil* **27**: 1245–1252.
2. King GA, McDougall J, Palisano RJ, Gritzan J, Tucker MA. (1999) Goal attainment scaling: its use in evaluating pediatric therapy programs. *Phys Occup Ther Pediatr* **15**: 31–52.
3. Steenbeek D, Meester-Delver A, Becher JG, Lankhorst GJ. (2005) 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* **19**: 274–282.
4. Cusick A, McIntyre S, Novak I, Lannin N, Lowe K. (2006) A comparison of goal attainment scaling and the Canadian Occupational Performance Measure for paediatric rehabilitation research. *Pediatr Rehabil* **9**: 149–157.
5. Ekström-Ahl E, Johansson E, Granat T, Carlberg EB. (2005) Functional therapy for children with cerebral palsy: an ecological approach. *Dev Med Child Neurol* **47**: 613–619.

6. Kiresuk T, Sherman R. (1968) Goal attainment scaling: a general method of evaluating comprehensive community mental health programs. *Community Ment Health J* **4**: 443–453.

7. Sherman RE. (1994) Keeping follow-up guides on target. In: Kiresuk T, Smith A, Cardillo J, editors. *Goal Attainment Scaling: Applications, Theory, and Measurement*. London: Erlbaum. p 279–282.

8. Schlosser RW. (2004) Goal attainment scaling as a clinical measurement technique in communication disorders: a critical review. *J Commun Disord* **37**: 217–239.

9. Donnelly C, Carswell A. (2002) Individualized outcome measures: a review of the literature. *Can J Occup Ther* **69**: 84–94.

10. Shefler G, Canetti L, Wiseman H. (2001) Psychometric properties of goal-attainment scaling in the assessment of Mann's time-limited psychotherapy. *J Clin Psychol* **57**: 971–979.

11. Rockwood K, Howlett S, Stadnyk K, Carver D, Powell C, Stolee P. (2003) Responsiveness of goal attainment scaling in a randomized controlled trial of comprehensive geriatric assessment. *J Clin Epidemiol* **56**: 736–743.

12. Rockwood K, Stadnyk K, Carver D, MacPherson KM, Beanlands HE, Powell C, Stolee P, Thomas VS, Tonks RS. (2000) A clinimetric evaluation of specialized geriatric care for rural-dwelling, frail older people. *J Am Geriatr Soc* **48**: 1080–1085.

13. Malec JF. (2001) Impact of comprehensive day treatment on societal participation for persons with acquired brain injury. *Arch Phys Med Rehabil* **82**: 885–895.

14. Ashford S, Turner-Stokes L. (2006) Goal attainment for spasticity management using botulinum toxin. *Physiother Res Int* **11**: 24–34.

15. Rushton PW, Miller WC. (2002) Goal attainment scaling in the rehabilitation of patients with lower-extremity amputations: a pilot study. *Arch Phys Med Rehabil* **83**: 771–775.

16. Fisher K, Hardie RJ. (2002) Goal attainment scaling in evaluating a multidisciplinary pain management programme. *Clin Rehabil* **16**: 871–877.

17. Zaza C, Stolee P, Prkachin K. (1999) The application of goal attainment scaling in chronic pain settings. *J Pain Symptom Manage* **17**: 55–64.

18. Lowe K, Novak I, Cusick A. (2006) Low-dose/high-concentration localized botulinum toxin A improves upper limb movement and function in children with hemiplegic cerebral palsy. *Dev Med Child Neurol* **48**: 170–175.

19. Mall V, Heinen F, Siebel A, Bertram C, Hafkemeyer U, Wissel J, Berweck S, Haverkamp F, Nass G, Döderlein L, et al. (2006) Treatment of adductor spasticity with BTX-A in children with CP: a randomized, double-blind, placebo-controlled study. *Dev Med Child Neurol* **48**: 10–13.

20. Wallen MA, O'Flaherty SJ, Waugh MC. (2004) Functional outcomes of intramuscular botulinum toxin type A in the upper limbs of children with cerebral palsy: a phase II trial. *Arch Phys Med Rehabil* **85**: 192–200.

21. Palisano RJ, Haley SM, Brown DA. (1992) Goal attainment scaling as a measure of change in infants with motor delays. *Phys Ther* **72**: 432–437.

22. Palisano RJ. (1993) Validity of goal attainment scaling in infants with motor delays. *Phys Ther* **73**: 651–660.

23. Paolicelli PB. (2001) Use of botulinum toxin type A in walking disorders of children with cerebral palsy. *Eur Medicophys* **37**: 83–92.

24. Clark MS, Caudrey DJ. (1983) Evaluation of rehabilitation services: the use of goal attainment scaling. *Int Rehabil Med* **5**: 41–45.

25. Engelen V, Ketelaar M, Gorter JW. (2007) Selecting the appropriate outcome in pediatric physical therapy; how individual treatment goals of children with cerebral palsy are reflected in GMFM-88 and PEDI. *J Rehabil Med* **39**: 225–231.

26. MacKay G, Somerville W. (1996) Reflections on goal attainment scaling (GAS): cautionary notes and proposals for development. *Educ Res* **38**: 161–172.

27. Mitchel T, Cusick A. (1998) Evaluation of a client-centered paediatric rehabilitation programme using goal attainment scaling. *Aust Occup Ther J* **45**: 7–17.

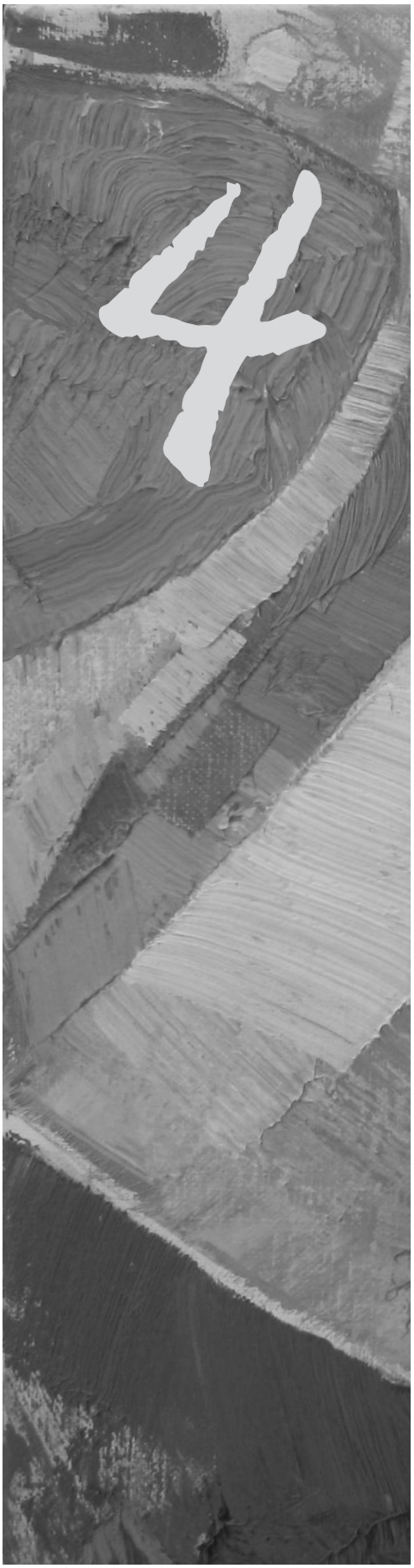
28. Mulligan H, Wilmshurst E. (2006) Physiotherapy assessment and treatment for an ambulant child with cerebral palsy after botox A to the lower limbs: a case report. *Pediatr Phys Ther* **18**: 39–48.

29. Rock BD. (1987) Goal and outcome in social work practice. *Social Work* **32**: 393–398.

30. Lewis AB, Spencer JH Jr, Haas GL, DiVittis A. (1987) Goal Attainment Scaling. Relevance and replicability in follow-up of inpatients. *J Nerv Ment Dis* **175**: 408–418.

31. Williams RC, Stieg RL. (1987) Validity and therapeutic efficacy of individual patient goal attainment procedures in a chronic pain treatment centre. *Clin J Pain* **2**: 219–228.

32. Smith A, Cardillo JE, Smith SC, Amezaga AM Jr. (1998) Improvement scaling (rehabilitation version). A new approach to measuring progress of patients in achieving their individual rehabilitation goals. *Med Care* **36**: 333–347.



Goal Attainment Scaling in paediatric rehabilitation: a report on the clinical training of an interdisciplinary team

D. Steenbeek, M. Ketelaar,
K. Galama, J.W. Gorter

Child Care Health Dev 2008,
34: 521–529.

ABSTRACT

Background: Goal Attainment Scaling (GAS) is a responsive method for individual goal setting and treatment evaluation. However, current knowledge about its reliability when used in paediatric rehabilitation treatment is insufficient and depends highly on standardization of the GAS method. A training programme was developed to introduce GAS to a team of 27 professionals from five disciplines. The purpose of the paper is to share the experiences of professionals and parents during this training.

Methods: The training consisted of three 2-h general discussion sessions and intensive individual feedback from the study leader (i.e. the first author). Feedback was given until the GAS scales met predetermined criteria of ordinality, described specific, measurable, acceptable, realistic abilities and activities in a single dimension, used the 'can-do' principle and could be scored within 10 min. Therapists and parents were asked to give their opinion by completing a questionnaire.

Results: One hundred and fifteen GAS scales were developed and scored by professionals. The development of a GAS scale remained a time-consuming procedure, despite the training: 45 (SD=27) minutes per scale. The content criteria of GAS were found to be useful by all participants. Common issues requiring revision of the initial scales were equal scale intervals, specificity, measurability and selection of a single variable. After the training, 70% of the therapists and 60% of the parents regarded GAS as a suitable tool to improve the quality of rehabilitation treatment. Examples of GAS scales developed by the various disciplines are presented and discussed.

Conclusions: The experiences reported in this paper support the further development of training procedures for GAS before it can be used as an outcome measure in effect studies. The findings may be helpful in introducing GAS in the field of childhood disability.

INTRODUCTION

In paediatric rehabilitation treatment, evaluation of patients' and families' progress towards individual goals at the level of activities and participation is increasingly important, in view of the diversity of developmental disabilities, goals and interventions. The literature shows that Goal Attainment Scaling (GAS) offers promising qualities as a method for individual goal setting and evaluation of treatment in paediatric rehabilitation.[1,2]

The use of GAS in children's treatment reflects the current debate in adult rehabilitation.[3] GAS appears to be a sensitive instrument that measures exactly what needs to be measured. The use of GAS in children's rehabilitation treatment has been recommended to improve transparency and co-ordination with parents during a child's rehabilitation process.[4] GAS

is a method that provides clear goals and priorities for intervention, ensuring the ongoing relevance of the child's goals, and reflecting a client-centred perspective to service delivery.[5] The original GAS developed by Kiresuk and Sherman[6] is a 5-point scale, with '0' representing the expected level of success. If a patient achieves more than expected, a score of +1 or +2 is given, depending on the level of achievement. If the patient's progress is less than expected, scores of –1 or –2 are given. Authors have recently experimented with scales using 6,[2,7] 7[8] or 3[9] points.

Surprisingly little is known about the test–retest reliability and concurrent validity of the content of the scales (goal setting) or about the reliability of the attainment.[2,10,11] It can be assumed that therapists who develop GAS scales and score the scales need to be trained in its use, to improve the reliability,[2,5] and future studies should consider the importance of training. Although some training in GAS was recommended for its use in mental health and recently in paediatric rehabilitation,[8] there are no specified guidelines for GAS development and scoring.

As long as there is no consensus about a standardized form of training for GAS, research reports should include descriptions of the training used, to allow accurate reproduction of studies. Nevertheless, none of six recently reviewed effect studies in paediatric rehabilitation using GAS as one of the outcome measures described how the investigation team was trained to use GAS.[2] The only specific description of a training in GAS for children receiving paediatric therapy was by King and colleagues.[5] This Canadian group reported on their experiences during several intervention studies.

It is generally assumed that the reliability and validity of the content of scales and the reliability of the scores on scales largely depend on the agreements made about scale development, scoring procedures and the experience gained with these procedures. What is now needed is further development of GAS and consensus about its application for children of various ages and disabilities across therapists of different disciplines, as well as further development of training procedures for the use of GAS. Consensus about training procedures can improve the consistency of studies measuring psychometric properties.

We developed a practical clinical training programme to introduce GAS into one interdisciplinary paediatric rehabilitation team at a medium-sized Dutch rehabilitation centre. For practical reasons, we started with the largest diagnosis group, namely children with Cerebral Palsy (CP). The purpose of the present paper is to share our experiences of this training.

Step by step introduction and training in GAS

The group of professionals was trained clinically during an 8-month period as part of their normal work in the children's department of the rehabilitation centre. The training was based on a step-by-step process of implementation.[12] Some *initial orientation* was provided by promoting awareness of the additional value of GAS in the process of goal setting. *Interest* and *involvement* were stimulated by examples. *Insight* was promoted by improving the participants' understanding of the necessary criteria for scale development. *Acceptance* of the use of GAS was promoted by stimulating a positive attitude towards adopting GAS.

In our training programme we first introduced the GAS method in a 2-h general discussion session. Based in part on the article by King and colleagues,[5] the participants agreed to adhere to the following criteria for *scale development*. All goals should belong to the areas of functional ability and activity level, and be based on the request for help. The goals should be meaningful and relevant to the children and their families. Each level should be described as clearly as possible. All levels should be specific, measurable, acceptable, realistic and time-specific (SMART). The scale should be ordinal, with intervals between the levels as equal as possible. Only a single dimension of change should be reflected in a GAS scale. In addition, two criteria for scoring GAS that had not yet been described were added: (1) in order to improve their feasibility, all scales should specify a behaviour that is observable within 10 min; (2) the scales should be developed as 'can-do' instead of 'do-do' scales, in order to reflect the professional's subjective judgement and to prevent unreliability because of children's whims (fatigue, motivation, interaction with the therapist and behaviour).

After the introduction, the second step in the training programme involved practicing scale development, with intensive individual feedback and two more general discussion sessions. The goals were set for a standard treatment period of 6 months. The 8-month programme included scale development and evaluation of the outcome. Before the start of the training the interdisciplinary team had already been trained in SMART goal setting.

In the GAS scales, the *setting* and *task* were defined separately at the start of the scale development procedure. A 6-point scale was used with an additional score of –3 for deterioration,[7] while –2 was defined as *equal to start*. The study leader gave individual feedback on all of the scales that were developed, until the scales met the criteria. As the feedback was given in the form of questions about the criteria, the final result was the product of the therapists themselves, experimenting with GAS scales in their daily practice. Although several

authors have stated that GAS rating should not be done by the person who developed the scale, each child's therapist in this study performed the goal-rating procedure 6 months after having created the scale, to allow them to fully benefit from the training programme. Experiences with the scoring system were consistently exchanged among the team members, allowing them to learn from each other and to offer regular positive feedback.[12]

Team and patients

In The Netherlands, a paediatric rehabilitation treatment team in a rehabilitation centre consists of a permanent group of cooperating professionals, who treat children eligible for interdisciplinary treatment, under the overall responsibility of a physician in rehabilitation medicine.

Twenty-seven therapists of the paediatric rehabilitation team (10 physical therapists, nine occupational therapists, four speech therapists, two child psychologists and two social workers) and one of the team physicians participated in the training programme, which was based on the GAS scales produced by 21 therapists. Six therapists did not have children with CP in therapy or were not employed for the entire training period. The professionals involved in the training programme had an average of 13.6-year (SD=9.1) experience in multidisciplinary rehabilitation care. The study leader (first author) was the physician in rehabilitation medicine of this team, who had gained experience with GAS in previous work with another team.[7]

The team members developed GAS scales for all children with CP who were expected to be in interdisciplinary therapy for at least the next 6 months. Results of the evaluation of the training are based on 115 GAS scales which were developed and scored for 58 children with CP, aged 2–18 years (mean=9.8 years, SD=1.4), being treated by the team. All Gross Motor Function Classification System levels[13] were represented: level I (n=10); level II (n=13); level III (n=13); level IV (n=9) and level V (n=13). The GrossMotor Function Classification System classifies the motor involvement of children with CP based on their functional abilities and their need for assistive technology and wheeled mobility. Level I represents the best functional abilities and level V the most limited.

Methods of evaluation

The time needed for individual feedback and common issues for the revision of the scales was recorded by the study leader. The training was evaluated using a questionnaire, constructed specifically for this purpose, which was

completed by therapists and parents of the patients involved. The parents had been informed about GAS, the training programme and participated in scale development as the scales should be based on their request for help. The therapists were asked about their experience with GAS, the efficiency of GAS and the quality and effect of the training. The questionnaire was completed twice: (1) the first time, immediately after the team members had gained their first experience with scale development (t=0); and (2) the second time, after the training had been completed (t=1). Therapists and parents were also asked after the end of the training period to indicate their satisfaction about working with GAS. Respondents were asked to be critical in their answers, and the questionnaire was anonymous.

The questionnaire included open questions as well as statements that respondents could agree or disagree with. The questionnaire included the following items:

1. A question to assess whether participants had any experience with GAS before the training and the number of scales made after the training.
2. The amount of time the therapists required to construct their first GAS scale and the time they required after the training period.
3. A statement: 'the training has resulted in my needing less time to develop and score a GAS'.
4. A statement: 'the GAS method is suitable for the for the purpose of team meetings with parents'.
5. Opinions about the quality of the training (open question).
6. A statement: 'the quality of the scales I construct has improved as a result of the training'.
7. A statement: 'GAS is a good way to assess treatment results'.
8. A statement: 'the use of GAS improves the quality of rehabilitation treatment'.
9. After the training, both therapists and parents were asked their opinion about the participation of the parents in scale development.

The statements were scored on a 5-point Likert scale: strongly agree, agree, neither agree nor disagree, disagree and strongly disagree.

The questionnaire measured: the level of experience with GAS (item

1); the influence of the training programme on the efficiency (items 2, 3, 4), opinions about the quality of the clinical training (item 5); the influence of the training on the quality of the scales (item 6); satisfaction about working with the GAS method (items 7, 8) and parent participation (item 9). Finally, respondents were invited to give their opinion on any aspect they wanted.

Descriptive statistics were used where applicable.

RESULTS

All team members agreed to continue their participation in the training after the introductory session. The findings reported below are based on 115 GAS scales. The individual feedback sessions with the study leader took a total of 30 h. The time needed per session varied from up to 20 min for the first scales to no feedback at all being required for the scales constructed by therapists who had constructed 10 or more during the training period. Common issues for individual revision of the scales were the SMART formulation of all levels, the use of a single variable and the use of equal scale intervals. Examples of the development of GAS scales by all participating disciplines before and after individual feedback are presented in **Tables 4.1–4.5**. The distribution of the GAS scores after 6 months (in percentages) is shown in **Table 4.6**. The median of the GAS scores was '0', representing acceptable goal attainment.[7]

The therapists completed all questionnaires, while the response rate of the parents was 29/58 before and 20/58 after the training. The distribution of answers to questionnaire items consisting of a statement and 5 options at t=1 is presented in **Tables 4.7 and 4.8**. Answers at t=0 (bracketed in the tables) were generally the same as those at t=1, indicating that the expectations of most therapists about working with GAS were fulfilled.

Before the training, none of the therapists had any experience with GAS (item 1). The time required to construct a scale was 49 (SD=22) minutes for the first scale and 45 (SD=27) minutes after the training (item 2). The time used can be divided into a mean of 25 min to create the scale and 9 min to verify the current functional level and adjust the scale.When constructing their first scale, the therapists used an additional 15 min, on average, to confer with an expert, compared with 11-min per scale after the training. Most respondents reported that the training had not made a clear difference in the amount of time that they had to spend to develop a GAS scale (item 3). Seventy-nine percent of the therapists reported that the GAS scales were suitable for team meetings with parents (item 4).

Table 4.1 Example of Goal Attainment Scaling by a physical therapist for a 5-year-old boy with spastic Cerebral Palsy, unilateral (hemiplegia), Gross Motor Function Classification System level II. Revision relating to the use of one single variable

	Score	Definition	First scale development	After revision	Explanation
Setting			Standard climbing rack against the wall.	A gymnasium with an obstacle course including a climbing rack against the wall.	The score would be wrong if P learns to place his arm correctly before he learns to place his leg correctly or if he learns to climb quickly and doesn't place his arm and leg correctly.
Task			Climb up the rack as good as you can, and climb down again.	Walk the obstacle course as fast as possible.	
	–3	Worse than start (deterioration)	P climbs up the rack but is afraid to climb back down.	More than 2 min.	
	–2	Equal to start	P climbs up the rack and back down but climbs very slowly because he has trouble placing his spastic arm and leg.	Between 2 min and 1 min and 40 s.	
	–1	Less than expected	P climbs up and down the rack and places his spastic leg correctly but has trouble placing his spastic arm.	Between 1 min and 40 s and 1 min and 20 s.	
	0	Expected goal	P places his spastic arm and leg correctly.	Between 1 min and 20 s and 1 min.	
	1	Somewhat more than expected	P climbs up and down in 1 min.	Between 1 min and 50 s.	
	2	Much more than expected	P climbs up and down in 50 s.	Less than 50 s.	

Table 4.2 Example of Goal Attainment Scaling by an occupational therapist for a 9-year-old girl with spastic bilateral Cerebral Palsy, Gross Motor Function Classification System level V. An intensive upper extremity training session in the computerized wheelchair had been planned after a botulinum toxin treatment. Revision relating to specificity and the aspect of equal scale intervals

	Score	Definition	First scale development	After revision	Explanation
Setting			Adremo wheelchair with computer with joystick control. Computer program involving children's drawing and typing.	Adremo wheelchair with computer with joystick control. Computer program involving children's drawing and typing.	Typing is a complex cognitive task that measures many other aspects as well as arm–hand function. It is therefore impossible to judge if the intervals between the scores are equal. Further specification makes measurement more reliable.
Task			Control a computer with a joystick.	Colour as many spaces as you can in the next 10 min.	
	–3	Worse than start (deterioration)	No control over the computer.	None.	
	–2	Equal to start	She has some control with help in controlling and selecting but is too slow to be functional.	1–4 with verbal help.	
	–1	Less than expected	She can choose and colour pre-existing spaces.	1–4 without help.	
	0	Expected goal	She can draw and colour figures.	5–10.	
	1	Somewhat more than expected	She can use an on-line computer keyboard to type but not enough for written communication.	11–20.	
	2	Much more than expected	She can type a short note.	More than 20.	

Table 4.3 Example of Goal Attainment Scaling by a speech therapist for a 5-year-old girl with spastic unilateral Cerebral Palsy and serious dysarthria. Revision relating to increasing the ordinality of the scale and reproducibility of the score

	Score	Definition	First scale development	After revision	Explanation
Setting			Social communication with classmates during the day.	Social communication with classmates during the day. Paying attention to clear simple sentences.	A uniform rating of the scale is made possible by further specification of the task.Measurement by the speech therapist during a therapy session is not possible. The first scale could be wrong if complex sentences were clear enough to be understood during therapy before simple sentences were clear in other situations (insufficiently ordinal).
Task			Talk as well as you can.	The opinion of the class teacher is asked with regard to spontaneous speech.	
	–3	Worse than start (deterioration)	Speech so garbled that he can not be understood.	Speech so garbled that he can not be understood.	
	–2	Equal to start	Simple sentences are often poorly formed. Says a few single words clearly enough to be understood.	Simple sentences are so unclear that they aren't understood. Single words are clear enough to be understood.	
	–1	Less than expected	Simple sentences are clear enough to be understood with a picture, for example a girl eating an apple.	Simple sentences are clear enough to be understood with a picture.	
	0	Expected goal	Simple sentences are clear enough to be understood during structured play with classmates.	Simple sentences are clear enough to be understood during structured play with classmates.	
	1	Somewhat more than expected	During spontaneous speech, simple sentences are often clear enough to be understood.	Simple sentences are clear enough to be understood during a conversation.	
	2	Much more than expected	More complex sentences are clear enough to be understood.	Simple sentences are always clear enough to be understood.	

Table 4.4 EExample of Goal Attainment Scaling by a paediatric social worker for a 3-year-old boy with spastic bilateral Cerebral Palsy, Gross Motor Function Classification System level III, mental retardation and serious behavioural disturbance. Revision relating to a larger dimension of measurability

	Score	Definition	First scale development	After revision	Explanation
Setting			Session with parents without child.	Session with parents without child.	The scale remains difficult to score for an independent social worker even after revision. The social worker's impression is necessary as well as the parents' trust in the social worker. The use of behavioural questionnaires is probably the only way for this scale to be objectively scored.
Task			Ask the parents how the child is behaving.	Impression from the social worker.	
	−3	Worse than start (deterioration)	Serious persistent behavioural disturbance and the parents don't know why.	The parents ask too much from the child and are not motivated to discuss the problem.	
	−2	Equal to start	The parents begin to understand the reason for their child's behaviour. For example too much is being asked from the child.	The parents are willing to discuss the problem of asking too much from the child.	
	−1	Less than expected	The parents can cope with the child's behaviour part of the time.	The parents discuss the behaviour disturbance that they see at home in relationship to the problem of asking too much from the child.	
	0	Expected goal	The parents can reasonably cope with the behaviour of the child.	The parents are motivated to make a plan to address the problem of asking too much from the child.	
	1	Somewhat more than expected	The parents can cope with the child's behaviour.	The parents report the results of the plan and make adjustments for the following period.	
	2	Much more than expected	The child's behaviour is normal for his age.	The parents believe that with help the child will not be asked too much and behavioural disturbance will be reduced.	

Table 4.5 Example of Goal Attainment Scaling by a paediatric psychologist for a 13-year old girl with spastic unilateral Cerebral Palsy and serious age-related problems in accepting her disability and coping. Revision relating to the aspect of equal scale intervals

	Score	Definition	First scale development	After revision	Explanation
Setting			Adolescent therapy session.	Adolescent therapy session.	A Visual Analogue Scale is also subjective but in this case the subjective opinion of the patient measures exactly what needs to be measured.
Task			P is asked about her mood in relationship to her hemiparesis.	P is asked about her mood in the last month. VAS scale from 0% = very bad to 100% = very good.	
	–3	Worse than start (deterioration)	She had two very depressed months without any positive time.	0–15%	
	–2	Equal to start	She had a bad time but had some good days. She still has a strong wish to be a different person without the hemiparesis.	16–30% (21%)	
	–1	Less than expected	She doesn't know if the days were bad or good.	31–45%	
	0	Expected goal	There is generally reasonable acceptance of the hemiparesis.	46–60%	
	1	Somewhat more than expected	Most normal days were good.	61–75%	
	2	Much more than expected	Life was excellent.	86–100%	

Table 4.6 Frequencies of outcomes for all Goal Attainment Scaling scales

Score	Percentage
–3	1%
–2	6%
–1	23%
0	42%
1	19%
2	9%

4

Table 4.7 Distribution of outcomes on questions with 5 options for therapists at *t*=1 (after the training). The data at *t*=0 are in brackets

Strongly agree	Agree	Neither agree/ nor disagree	Disagree	Strongly disagree
3. 'Thanks to the training I need less time to develop and score a GAS'				
0%	30%	45%	25%	0%
4. 'The GAS method can be directly used for the purpose of the team meetings with parents'				
22% (37)	57% (57)	9% (0)	12% (6)	0% (0)
6. 'The quality of the scales improved as a result of the training'				
5%	55%	30%	10%	0%
7. 'GAS is a good way to measure treatment results'				
14% (14)	64% (64)	14% (14)	8% (8)	0% (0)
8. 'The use of GAS leads to an improvement in the quality of rehabilitation treatment'				
13% (0)	57% (64)	22% (36)	4% (0)	4% (0)
9. 'The parents closely participated in GAS development'				
0%	44%	26%	26%	4%

Table 4.8 Distribution of outcomes on questions with 5 options for parents

Strongly agree	Agree	Neither agree/ nor disagree	Disagree	Strongly disagree
7. 'GAS is a good way to measure treatment results'				
5%	55%	30%	10%	0%
8. 'The use of GAS leads to an improvement in the quality of rehabilitation treatment'				
20%	40%	40%	0%	0%
9. 'The parents closely participated in GAS development'				
0%	5%	20%	30%	45%

The overall quality of the training was assessed as good (item 5). All therapists reported that they were able to meet the GAS criteria. An aspect that was criticized was that the amount of experience the therapists gained with GAS during the training period depended on the number of children with CP who were being treated, and was therefore unevenly divided among the team members. Therapists who treated few or no children with CP complained that they had had too little training. Sixty percent of the therapists reported that the quality of the scales they constructed had improved as a result of the training (item 6). Seventy-eight percent of the therapists and 59% of the parents regarded GAS as a good method to assess treatment results (item 7). Seventy percent of the therapists and 60% of the parents stated after the training that the use of GAS improved the quality of rehabilitation treatment (item 8). After the training period, two therapists were more negative about the value of GAS for the quality of treatment. Answers to the open questions revealed the concerns of these two therapists about the subjectivity of outcome measurement using GAS.

Forty-four percent of the therapists and only 5% of the parents reported that the parents had participated closely in the GAS procedure (item 9). The amount of time involved and the difficulty of scale development were the most important obstacles perceived as regards working with GAS (item 10).

DISCUSSION

The 8-month training period to introduce GAS to this paediatric treatment team was favourably evaluated; however, there are some aspects of GAS that need improvement or special attention. Although the team of professionals had previously been well trained in setting SMART goals as a basic skill in a treatment setting, the children's rehabilitation unit had had no experience with GAS before the training. As a result of the training, GAS has now become an important issue for the unit.

The quality of evaluation of the training was restricted. There is no standard tool to measure the effect of training in a new method for rehabilitation treatment professionals, as has also been concluded by Russell and colleagues.[14] The decision to use a purpose-made questionnaire was based on the view that the subjective opinions of the professionals would be the best measure of the training. The conclusions must therefore be interpreted while keeping in mind that no validated tools could be used. Moreover, there was a rather low response rate from the parents. Another

limitation of this study is that any change in the quality of the GAS scales because of the training was an ongoing and immeasurable process. The process was based on enthusiasm about working with a new method and changes were produced by the therapists themselves. Although there is no tool for measuring the quality of a GAS scale, most participants and the study leader felt that the quality of the scales gradually improved.

Defining six realistic, distinct levels of outcome for each child's own functional goals, in an ordinal structure with no gaps or overlap between levels, and with full agreement from the children and their parents, is a time-consuming procedure and requires much practice. Various subjective elements have to be considered, including the relationship between body structure, activities and participation, the prediction of learning potential, intercultural differences, interpretation of fatigue, motivation and the interpretation of clinical relevance in relation to the intensity of therapy. These are all factors which necessitate training in GAS and are probably the reason why the therapists did not report time gains as a result of the training. Nevertheless, there was a trend in the correlation between experience in scale development and efficacy. Time was also gained by using the scales for the purpose of the team meeting reports. Our finding that it took about 45 min to develop a single GAS scale corresponds to those of King et al.[5] and one of our own previous studies.[7] The scales of the psycho-social disciplines in particular had to be discussed before they could be improved. A substantial proportion of the time was used to test a draft scale with the child to check if –2 corresponded with the child's present level of performance. Almost all therapists reported that the first draft scale undervalued the child's special achievement, even though the therapist and the child had often known each other for years. This continues to provide food for thought.

A remarkable and unexpected outcome of the evaluation was the difference of opinion between therapists and parents about the parents' participation in the GAS procedure. Whereas parents reported having scarcely participated, therapists reported that the parents were involved often and intensively. Although there was a low response rate on the part of the parents, this outcome can be regarded as reliable, because the most involved parents were more likely to have answered the questionnaires. In our opinion, the use of GAS should always be based on the request for help and expectations with regard to the outcome, which makes it impossible to apply the procedure without involving the parents. The parents probably received too little information about their position in the procedure. Apart from their own role, the parents reported that working with GAS was useful, facilitating

insight into the treatment results and enhancing the quality of treatment. These findings support previous recommendations to use GAS to improve transparency and co-ordination with parents.[4] The level of satisfaction among the teams and parents about working with GAS corresponded to the findings reported in all of the articles previously reviewed.[2]

For practical reasons, only children with CP participated in this study: they represent the largest group in a paediatric rehabilitation treatment setting, and that only they were included helped to increase the homogeneity of the scales. The consequence of this decision was, however, that those therapists who treated more children with CP gained more experience with GAS than others who treated fewer children with CP, decreasing the homogeneity of the subjects in the evaluation (the group of therapists). Applying the training procedure to more than one diagnosis might have prevented this problem.

The training did not guarantee that the professionals would continue to use GAS as outlined in the model for inducing change in professional behaviour.[12] Integrating GAS into existing routines and embedding the method in the organization of a rehabilitation centre could be difficult, especially in view of the time investment. However, the process of developing a GAS scale is mainly a cognitive process. If the phases described by Grol and Wensing[12] are all implemented and GAS becomes structurally embedded in the treatment procedures, the cognitive process should become part of regular therapy, and then a serious time gain can be expected. The training changed the professionals' awareness from being unaware to being aware, although still struggling with their competence to use GAS. The clinical training programme did not provide the next steps from'can-do' to 'do-do' and 'going on doing it' or being aware of and competent to apply the new method.

Goal Attainment Scaling is fundamentally ready for all daily rehabilitation treatment practice, but further development of its application in intervention studies among children is necessary before it can be reliably used as an outcome measure in research. The opportunities it offers for children of various ages and disabilities across therapists of different disciplines make this development of great importance. The trained team that took part in this study is ready for a follow-up measurement of psychometric properties.

CONCLUSIONS

The results of this study strongly indicate the need to train professionals in the use of GAS before it can be used as an outcome measure in effect studies with children in paediatric rehabilitation, confirming the necessity of further discussion about training procedures in GAS around the world. The GAS training procedure described in this article, involving specific agreements, general training sessions and individual feedback, can be regarded as a successful attempt to introduce GAS. The criteria for the use of GAS, namely the use of SMART goals at the functional ability and activity levels, ordinal scales with a single dimension in each scale, using the 'can-do' principle and allowing scoring to be completed within 10 min, were proved to be realistic and useful for all participants. Special attention is needed for the participation of parents in the development of the scales.We recommend adding a programme to ensure that the procedure becomes embedded in the treatment process, in order to reduce the time cost of scale development. This study may be valuable in that it describes a training procedure in ongoing rehabilitation research in which GAS is one of the outcome measures, and may prove helpful in further development of GAS in this field.

KEY MESSAGES

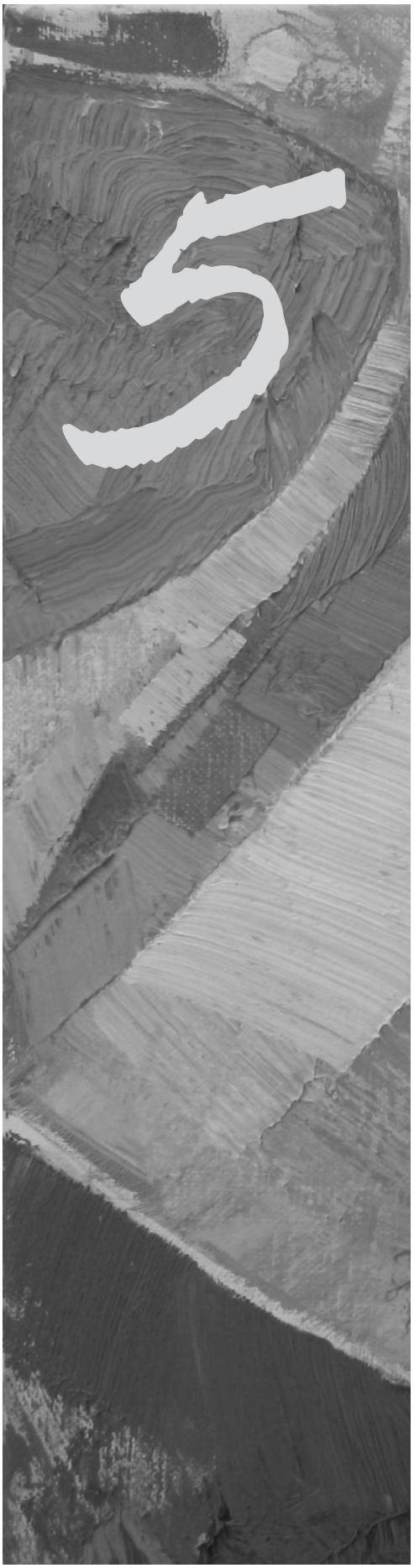
- Goal Attainment Scaling is regarded as a valuable but timeconsuming procedure.
- The use of GAS requires training first, before it can be implemented. The paper reports on a clinical training.
- Agreements about the criteria to be applied for GAS development are necessary to enhance its reliability.
- Further development of GAS and its application in the treatment of children is necessary and of great importance for the field of childhood disability.

ACKNOWLEDGEMENTS

We would like to thank all participating children, parents and therapists of the children's department of the rehabilitation centre at Breda, The Netherlands.

REFERENCES

1. Mailloux Z, May-Benson TA, Summers CA, Miller LJ, Brett-Green B, Burke JP, Cohn ES, Koomar JA, Parham LD, Roley SS, Schaaf RC, Schoen SA. (2007) Goal Attainment Scaling as a measure of meaningful outcomes for children with sensory integration disorders. *Am J Occup Ther* **61**: 254–259.
2. Steenbeek D, Ketelaar M, Galama K, Gorter JW. (2007) Goal Attainment Scaling in paediatric rehabilitation: a critical review of the literature. *Dev Med Child Neurol* **49**: 550–556.
3. Hurn J, Kneebone I, Cropley M. (2006) Goal setting as an outcome measure: a systematic review. *Clin Rehabil* **20**: 756–772.
4. Siebes RC, Ketelaar M, Gorter JW, Wijnroks L, De Blécourt AC, Reinders-Messelink HA, Van Schie PE, Vermeer A. (2007) Transparency and tuning of rehabilitation care for children with Cerebral Palsy: a multiple case study in five children with complex needs. *Dev Neurorehabil* **10**: 193–204.
5. King GA, McDougall J, Palisano RJ, Gritzan J, Tucker MA. (1999) Goal Attainment Scaling: its use in evaluating pediatric therapy programs. *Phys Occup Ther Pediatr* **15**: 31–52.
6. Kiresuk T, Sherman R. (1968) Goal Attainment Scaling: a general method of evaluating comprehensive community mental health programs. *Community Ment Health J* **4**: 443–453.
7. Steenbeek D, Meester-Delver A, Becher JG, Lankhorst GJ. (2005) 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* **19**: 274–282.
8. Cusick A, McIntyre S, Novak I, Lannin N, Lowe K. (2006) A comparison of Goal Attainment Scaling and the Canadian occupational performance measure for paediatric rehabilitation research. *Pediatr Rehabil* **9**: 149–157.
9. Ekström-Ahl L, Johansson E, Granat T, Carlberg EB. (2005) Functional therapy for children with Cerebral Palsy: an ecological approach. *Dev Med Child Neurol* **47**: 613–619.
10. Donnelly C, Carswell A. (2002) Individualized outcome measures: a review of the literature. *Can J Occup Ther* **69**: 84–94.
11. Schlosser RW. (2004) Goal Attainment Scaling as a clinical measurement technique in communication disorders: a critical review. *J Commun Disord* **37**: 217–239.
12. Grol R, Wensing M. (2004) What drives change? Barriers to and incentives for achieving evidence-based practice. *Med J Aus* **180**: S57–S60.
13. Palisano RJ, Cameron D, Rosenbaum PL, Walter SD, Russell D. (2006) Stability of the Gross Motor Function Classification System. *Dev Med Child Neurol* **48**: 424–428.
14. Russell DJ, Leung KM, Rosenbaum PL. (2003) Accessibility and perceived clinical utility of the GMFM-66: evaluating therapists' judgements of a computer-based scoring program. *Phys Occup Ther Pediatr* **23**: 45–58.



Inter-rater reliability of Goal Attainment Scaling in rehabilitation of children with Cerebral Palsy

D. Steenbeek, M. Ketelaar,
E. Lindeman, K. Galama, J.W. Gorter

Arch Phys Med Rehabil
2010, **91**: 429–435.

ABSTRACT

Objectives: To determine the inter-rater reliability of Goal Attainment Scaling (GAS) in the routine practice of interdisciplinary rehabilitation of children with Cerebral Palsy, and to examine the difference in the inter-rater reliability of the scores between GAS scales constructed by the children's own therapists and the scales constructed by independent therapists.

Design: Individually tailored GAS scales, based on predetermined criteria, were constructed at the start of a 6-month rehabilitation period. The outcome was rated independently by 2 therapists at the end of the treatment period. Two different data sets were acquired, one consisting of scores on GAS scales constructed by the children's own therapists, the other of scores on GAS scales constructed by matched independent raters of the same profession.

Setting: A children's unit of a medium-sized rehabilitation center in The Netherlands.

Participants: Physical therapists (n=8), occupational therapists (n=8), and speech therapists (n=4) participated in pairs. They constructed 2 sets of 64 GAS scales each, for 23 children with Cerebral Palsy.

Interventions: A 6-month interdisciplinary pediatric rehabilitation program.

Main outcome measure: Inter-rater reliability was assessed using linear-weighted Cohen's kappa.

Results: The scales constructed by the children's therapists had an inter-rater reliability of .82 (95% confidence interval [CI], .73–.91). The inter-rater reliability for scales constructed by the independent raters was .64 (95% CI, .49–.79). The main reason for disagreement between raters was discrepancies in the professionals' interpretation of the children's capacities versus their actual performance during assessment.

Conclusions: The inter-rater reliability of GAS used under optimal conditions was good, particularly for scales constructed by the children's own therapists.

INTRODUCTION

An important aspect of pediatric rehabilitation is to evaluate patients' and families' progress towards activity and participation goals. Given the diversity of developmental disabilities and therapy goals, this requires the use of individual measurement tools. GAS is an increasingly popular individual instrument for progress assessment in rehabilitation. Kiresuk and Sherman[1] introduced GAS to evaluate adult mental health services. In its original form, it is a 5-point scale, constructed before an intervention period, with "0" representing the expected level of functioning after a predefined period. If a patient achieves more than is expected, a score of +1 or +2 is given, depending on the level of

achievement. If the patient's progress is less than expected, a score of –1 or –2 is given. Several recent studies have experimented with scales using 6,[2] 7,[3] or 3[4] points, and in most research with GAS, "no change" is scored as –2. In a previous article, we presented arguments in favor of using a score of –3 for deterioration.[5] In this version of GAS, the additional score of –3 allows all original levels to be attained and avoids bottoming effects. This removes one of the causes of false sensitivity to change in calculating group effects.[5]

The sensitivity of GAS to change is assumed to be better than that of standardized functional measures,[5,6] while the reliability of the method of scale development (content reliability) and the reliability of the scores (inter-rater reliability) in predetermined scales are still subject to debate. Despite this, GAS has been used extensively in recent intervention studies, for example, to assess botulinum toxin A treatment in the pediatric population.[7-9] Research into adult mental health and geriatric medicine has provided evidence about the psychometric properties of GAS, reporting the content reliability of GAS to be good and its inter-rater reliability satisfactory.[10–12]

Knowledge about the reliability of GAS in rehabilitation care is limited, and further validation and reliability studies to justify the use of GAS are warranted.[5,13] Palisano[14] examined the inter-rater reliability of GAS in pediatric physical therapy in the context of a validation study. He reported good inter-rater reliability, with kappa coefficients of .89 for 10 scales scored before the start of a validation study, and .75 for 16 scales scored during the validation study. In a previous study by Palisano et al.,[15] a physical therapist and an occupational therapist independently scored the performance of 9 subjects on a GAS scale, and their scores all matched. These authors constructed GAS scales using information from the therapists, and then scored them simultaneously with an independent rater from a video. The sample sizes in both studies were, however, small. Despite Palisano's recommendation to replicate inter-rater reliability studies across therapists and clients of various ages and disabilities, no further studies have been performed in this field in the last 15 years.

As shown in a previous study,[16] the reliability of the scores largely depends on the agreements made about scale development and scoring procedures, and on experience with GAS. Further studies showing that GAS has acceptable inter-rater reliability in practical settings are essential if this method is to be used in rehabilitation practice.

The reliability of GAS may be affected by the risk of so-called therapist bias. One potential source of therapist bias is that of therapists' expectations about their patients' level of attainment for each GAS scale. Another source

of bias arises when therapists score their own scales, given the probable dependence of the therapists and their interest in a good outcome.[15] To decrease the risk of therapist bias, stringent adherence to the original GAS protocol dictates that goal setters and raters be independent of the treatment process.[17] However, the influence of goal setters and/or raters being involved in the treatment process has never been examined, and independent construction and scoring procedures for GAS are, in fact, impractical and inefficient in routine rehabilitation. The functional relevance of independent GAS construction is debatable, as goals should be set together with the children and their families and should take children's changeability into account. The therapists are the people who have optimal knowledge of their patients' history and treatment results, and this knowledge is essential for establishing the actual goals and different GAS scale levels.

The aims of this study were (1) to determine the inter-rater reliability of GAS in the routine practice of children's therapists working in a team of trained professionals in interdisciplinary pediatric rehabilitation and explore the reasons for discrepancies between scores; and (2) to examine the difference in the inter-rater reliability of the scores between GAS scales constructed by the children's own therapists and that of scores on scales constructed by independent therapists.

METHODS

Design

A convenience sample of 20 members of a trained team of professionals, consisting of physical, occupational, and speech therapists, participated in the reliability study with a pretest-posttest design. Two team members of each of the 3 disciplines were selected as independent raters. Fixed pairs were formed consisting of each primary child's therapist and one of the independent raters. If one of the selected independent raters worked as the child's therapist, that rater filled the role of the child's therapist, and the second independent rater of that discipline was paired with him/her. Demographic baseline data on the raters (age and years of professional experience) were recorded.

The cases included were children with Cerebral Palsy (CP) of various degrees of severity. Children were eligible for this study if they were between the ages of 2 and 14 years and their physician expected them to be in interdisciplinary therapy for at least 6 months. The distribution of severity

of CP was evaluated by the GMFCS[18] and the Manual Ability Classification System.[19]

The study leader constructed a case summary for each child, based on the request for help and the expectations of each child and the child's parents. Each case summary included separate domain descriptions for the different disciplines. The child's therapist and the independent rater constructed GAS scales individually, both having been asked to base their scale development on the study leader's case summary and interpretation of the child's functioning. The child's therapists constructed GAS during regular therapy, whereas the independent raters constructed GAS during a separate 1-hour observation of the child, who was accompanied by his/her parents. Goals were set for a 6-month period. After all GAS scales for a child had been constructed, the scales were shared with the primary therapists and the child's parents. After the 6 months, the child's therapist and the independent rater both scored the GAS scale constructed by the child's therapist, as well as the GAS scale constructed by the independent rater (**Table 5.1**). The child's therapists scored GAS during regular therapy, whereas the independent raters scored the scales in a separate 30-minute session after the therapy period.

Table 5.1 Design of the reliability study and actual number of GAS scales per discipline

GAS scales	Scored by the child's therapist			Scored by the independent rater		
	PT	OT	ST	PT	OT	ST
Constructed by the child's therapist						
PT	23	ND	ND	23	ND	ND
OT	ND	23	ND	ND	23	ND
ST	ND	ND	18	ND	ND	18
Totals A		64			64	
Constructed by the independent rater						
PT	23	ND	ND	23	ND	ND
OT	ND	23	ND	ND	23	ND
ST	ND	ND	18	ND	ND	18
Totals B		64			64	

The first aim of this study was to determine the inter-rater reliability within the GAS scales constructed by the child's therapist (i.e., the totals of A). The second aim was to compare the reliability data for the GAS scales constructed by the child's therapists (i.e., the totals of A) with those for the GAS scales constructed by the independent raters (i.e., the totals of B). ND, no data; OT, occupational therapy; PT, physical therapy; ST, speech therapy.

To guarantee independence between the members of each pair, the child's therapist and independent rater were asked to avoid exchanging any details about the child's progress during therapy. During a meeting with the parents, the study leader explicitly asked the parents not to discuss the construction of the scales or the scoring, to help ensure the independence between the child's therapist and independent rater.

Goal Attainment Scaling method used

Before the start of the present study, an 8-month practical training period was used to introduce GAS in the children's rehabilitation unit of a medium-sized rehabilitation center in The Netherlands. In the training program, which we described and evaluated in detail in a previous article,[16] professionals practiced the construction and scoring of GAS scales. Six-point GAS scales were constructed, using –3 for deterioration and –2 for no change relative to the level determined at the start of treatment, as described in Steenbeek et al.[16] The participants agreed to adhere to the following criteria for scale development. All goals should be based on the request for help in the area of a child's capacity for daily activities. The goals must be important for and relevant to the children and their families, and describe the main aim of therapy for each discipline involved. Each level should be defined as clearly as possible. All levels of the scales should be specific, measurable, acceptable, realistic, and time-specific. The therapists were asked to construct ordinal scales with incremental steps of equal intervals. Each GAS scale had to reflect a single dimension of change. In addition, it had to be possible to score a scale within 10 minutes to ensure that it was practicable. The setting and task were described explicitly for each GAS scale.

Scale rating

All raters were asked to base the scores on their insight as a professional of goal achievement rather than on the child's actual performance at the time of the assessment. The reflection of the professionals' knowledge and experience in rating goal attainment was introduced to minimize the possible influence of children's whims (fatigue, motivation, interaction with the therapist, and behavioral issues). The raters' professional judgment of goal attainment was supported by observing the child or by conferring with parents or teachers, and the method was stipulated in each GAS scale description. When conferring was stipulated, the raters informally asked the parents or teachers their opinion about the child's functioning.

Evaluation of Goal Attainment Scaling scores

To meet the first aim of the study, we compared the scores given by the pairs on the GAS scales constructed by the children's own therapists, to evaluate the inter-rater reliability. For the second aim, the scores given by the pairs on GAS scales constructed by the independent raters were used as a second data set. The inter-rater reliability found for the 2 data sets were then compared to determine the influence of independent scale construction on the inter-rater reliability of the scores. The design was such that the 2 raters rating the same GAS scale had different perspectives, for example, in terms of their knowledge of the child's history. An additional evaluation of the scores was performed when the raters of a pair disagreed, for a better understanding of the consequences of independent scale construction.

After scoring had been completed, therapists were informed about the discrepancies in their scoring, thus ending the independence within the pairs. The pairs were then systematically interviewed to explore reasons for the scoring discrepancies.

Further aspects of GAS construction and scoring that might have influenced the inter-rater reliability were explored. We calculated the inter-rater reliability per discipline to gain insight in the separate CIs, as scores from different disciplines for the same child could be dependent. The influence of the type of GAS scale on the inter-rater reliability was also determined. After scoring was complete, the study leader classified all GAS scales into 2 categories, one with scales rating physical function and the other rating higher cognitive function. Scales rating physical function, such as running distance, are more concrete and observable, whereas scales rating higher cognitive function, such as recognizing clothes when self-dressing or self-efficacy in maneuvering a wheelchair, depend more on interpretation. This subdivision was checked by one of the authors (KG). Inter-rater reliability was then calculated for each type of GAS scale.

Data analysis

The inter-rater reliability was evaluated by calculating linear-weighted Cohen's kappa values with 95% CIs. The kappa values were interpreted as no (<0), very low (0.0–.20), low (.21–.40), moderate (.41–.60), good (.61–.80) or excellent agreement (.81–1.00).[20] We used the Wilcoxon signed ranks test for each data set to test whether the scores were more likely to be higher or lower when rated by the child's own therapists than when rated by the independent raters. Statistics were performed in SPSS/PC, release 14.0 and http://faculty.vassar.edu/lowry/kappa.html.

RESULTS

Eight physical therapists, 8 occupational therapists, and 4 speech therapists participated. The 20 professionals involved (age range, 28–59y; mean age ± SD, 40.3 ± 10.9y) had 4 to 30 years (mean ± SD, 14.8 ± 9.8y) of professional experience. Before the training, none of the therapists or independent raters had had any experience with GAS.[16]

GAS scales were constructed for 23 children in the reliability study. Patients ranged in age from 4 to 13 years, and severity of CP ranged from GMFCS level I to GMFCS level V (**Table 5.2**). Therapists treated 1 to 3

Table 5.2 Summary of characteristics of 23 children with CP

Patient	Age (y)	Sex	GMFCS level	MACS level	No. of disciplines*
1	12	M	IV	IV	3
2	13	F	V	V	2
3	7	F	II	I	3
4	13	F	V	V	3
5	5	M	I	III	3
6	7	M	III	II	3
7	9	F	IV	II	3
8	11	F	III	II	3
9	4	M	I	III	3
10	7	M	II	I	3
11	6	M	V	V	3
12	3	M	II	III	3
13	5	M	III	II	2
14	8	M	IV	IV	3
15	9	F	II	I	2
16	10	M	III	II	2
17	6	M	II	II	3
18	6	M	I	III	3
19	10	F	V	V	3
20	10	M	IV	II	3
21	2	M	II	III	3
22	11	F	V	V	3
23	5	M	II	I	2

GMFCS,[18] classifies the motor function of children with Cerebral Palsy (CP) based on their self-initiated movement with particular emphasis on sitting, walking, and wheeled mobility. Level I represents walking without limitations at 6–12 years. Level V represents no means of independent mobility at that age (http://www.canchild.ca). MACS[19] classifies how children with CP use their hands when handling objects as part of daily activities. Level I represents the best functional abilities and level V the most limited ones (http://www.macs.nu). M, male; F, female; GMFCS, Gross Motor Function Classification System; MACS, Manual Ability Classification System. * Number of disciplines involved: 3, physical therapy, occupational therapy, and speech therapy; 2, physical therapy and occupational therapy only.

children each. In 18 cases, one of the therapists appointed as independent rater was the child's primary therapist.

The total number of GAS scales constructed was 18 × 6 (3 therapists + 3 independent raters) plus 5 × 4 (2 therapists + 2 independent raters) for a total of 128 scales (see **Table 5.1**). **Table 5.3** shows an example of a case summary, with the domain description for the physical therapists, and the GAS scales constructed by a physical therapist and his matched independent rater. The different approaches and descriptions by the raters forming the pair illustrate their independence.

Table 5.3 Example of a GAS scale*

Case summary: L is a 5-year-old boy with a right-sided unilateral spastic CP, GMFCS level II. The main request for help is to fall less often. L falls a few times a day while standing on his right leg during normal walking. Treatment: a multilevel botulinum toxin intervention and physical therapy 4 times a week for 4 months.

	Score	Definition	Child's physical therapist	Physical therapist independent rater
Setting			A strictly specified gymnasium with an obstacle course including jumping and quick changes of walking direction. Safety is guaranteed by guidance. The therapist encourages L to complete the course within 3 min.	L is wearing his ankle-foot orthosis and shoes.
Scoring method			Observation.	Observation.
Task			Walk the obstacle course fast and don't fall.	Stand on your right leg only as long as possible.
	–3	Worse than start (deterioration)	L falls 5 times or more.	< 2 s.
	–2	Equal to start	L falls 4 times.	2 s.
	–1	Less than expected	L falls 3 times.	3–5 s.
	0	Expected goal	L falls 2 times.	6–8 s.
	1	Somewhat more than expected	L falls 1 time.	9–11 s.
	2	Much more than expected	L does not fall.	> 11 s.

* The GAS scale was constructed at the start of an intervention period by a child's own physical therapist and the physical therapist functioning as independent rater, based on the case summary provided by the study leader (No. 23 in Table 5.2). After 6 months of rehabilitation, the first resulted in agreement on the score, the second in disagreement

Table 5.4 is the cross table of the scores given by therapists and independent raters on the GAS scales constructed by the children's own therapists after 6 months of rehabilitation (n=64), showing 77% agreement between the raters forming the pairs and a linear-weighted Cohen's kappa of .82 (95% CI, .73–.91), indicating excellent agreement.

Table 5.5 is the cross table of the scores given by therapists and independent raters on GAS scales constructed by the independent raters after 6 months of rehabilitation (n=64), showing 64% agreement between raters forming the pairs and a linear-weighted Cohen's kappa of .64 (95% CI, .49–.79), indicating good agreement.

On the GAS scales constructed by the children's own therapists, the scores given by the two groups of raters matched for 49 scales, while the children's therapists gave higher scores on 6 scales, and the independent raters gave higher scores on 9 scales (see **Table 5.4**). Comparison of these scores revealed no statistically significant difference in scoring patterns (Wilcoxon signed ranks test, z=–.62; p=.54). Among the scales constructed by the independent raters, there were 11 on which the children's therapists gave a higher score, and 12 where the independent raters gave a higher score (see **Table 5.5**). There was no statistically significant difference in these scoring patterns either (z=–.16; p=.88). There were 15 discrepancies between raters scoring the scales constructed by the children's own therapists (2 × 2 points; 13 × 1 point [see **Table 5.4**]) and 23 between raters scoring the scales constructed by the independent raters (9 × >1 point; 14 × 1 point [see **Table 5.5**]).

Systematic interviews to evaluate discrepancies between the pairs in both data sets revealed that disagreement was largely due to different interpretations of the professionals' perception of the child's capacity, as

Table 5.4 Scores on the GAS scales constructed by the children's own therapists to calculate the inter-rater reliability*

		GAS score by the child's therapist						
	Score	–3	–2	–1	0	1	2	Totals
GAS score by the independent rater	–3	0	0	0	0	0	0	0
	–2	0	5	0	0	0	0	5
	–1	0	4	12	3	0	1	20
	0	0	0	0	8	0	0	8
	1	0	0	1	1	10	2	14
	2	0	0	0	0	3	14	17
	Totals	0	9	13	12	13	17	64

* Linear-weighted Cohen's kappa was .82 (95% CI, .73–.91).

opposed to what a child actually does when tested. In 12 scales, one rater unintentionally scored the actually observed performance, although he realized the child did not perform to the best of his ability. Differences in 10 scales were caused by a difference of opinion between the raters, when children were capable of more than they actually did during the assessment. Differences in 4 other scales were due to one rater misunderstanding the textual content of the scale (eg, a standard obstacle course in a gymnasium was not specific enough for the partner, who set up a more complex course). Causes of 12 discrepancies remained unclear.

The kappa values for each discipline are shown in **Table 5.6**. The mean kappa value corresponded with that of the 2 entire data sets. The kappa values for each type of scale are shown in **Table 5.7**. The study leader (DS) and another author (KG) agreed on the subdivision of the types of scales for 95% of the scales, and after discussion agreed on 100% of them. Thirty percent of the physical scales resulted in disagreement between the pairs of raters versus 29% of the higher cognitive function scales.

Table 5.5 Scores on the GAS scales constructed by the independent raters to compare the inter-rater reliability with that of Table 5.4*

		GAS score by the child's therapist						
	Score	–3	–2	–1	0	1	2	Totals
GAS score by the independent rater	–3	2	0	0	0	0	0	2
	–2	0	5	1	0	0	1	7
	–1	0	1	12	3	0	2	18
	0	0	1	2	7	2	1	13
	1	0	1	1	0	5	1	8
	2	0	0	1	1	4	10	16
	Totals	2	8	17	11	11	15	64

* Linear-weighted Cohen's kappa was .64 (95% CI, .49–.79).

Table 5.6 Kappa values (95% CIs) per discipline

Constructed by	Scales (*n*)	Scored by pairs of PTs	Scales (*n*)	Scored by pairs of OTs	Scales (*n*)	Scored by pairs of STs
Child's therapists	23	.73 (.55–.90)	23	.84 (.68–1.00)	18	.92 (.82–1.00)
Independent raters	23	.66 (.43–.87)	23	.61 (.35–.86)	18	.65 (.34–.97)

PTs, physical therapists; OTs, occupational therapists; STs, speech therapists.

Table 5.7 Kappa values (95% CIs) per type of GAS scales

Constructed by	Scales (*n*)	Physical	Scales (*n*)	Higher cognitive function
Child's therapists	21	.76 (.57–.95)	43	.85 (.75–.94)
Independent raters	18	.65 (.38–.92)	46	.63 (.46–.81)
All	39	.71 (.54–.87)	89	.74 (.63–.84)

DISCUSSION

This study found good-to-excellent inter-rater reliability of GAS when used by a group of trained therapists for children with CP. Similar to the studies on the reliability of GAS by Palisano,[14,15] this study provides further evidence for the inter-rater reliability of GAS in a regular rehabilitation setting. This is reassuring because GAS has been used extensively in recent rehabilitation research.

The present study confirms the necessity of further discussion of GAS training procedures, as previously recommended.[16] The reliability of the scores depends on the agreements made regarding scale development and scoring procedures, as well as the experience gained with these procedures. Because there were still misunderstandings between the members of the pairs of raters, even after training, the reliability of the GAS would most likely have been lower without training. The training period in this study was intended to improve the homogeneity of the method performed by all raters, as the inter-rater reliability was assessed between groups of raters with divergent levels of professional experience.

Critics of GAS have doubted the value of the patient's own therapists selecting goals and specifying outcomes. Cytrynbaum et al.[17] recommended that goal setters and raters be independent of the treatment process. Therefore, the second aim of this study was to examine the difference in inter-rater reliability of the scores on GAS scales constructed by the children's own therapists and those on scales constructed by therapists functioning as independent raters. The differences found between the groups (kappa, .82 vs .64) were not statistically significant because of an overlap in the 95% CIs. The difference can be explained by either coincidence or the difference in the content of the scales constructed by the 2 groups. Familiarity with a child's history, character, and preferences, as well as treatment results, is probably needed to construct GAS scales that can be rated reliably. These data are reassuring as regards the presumed therapist bias in constructing GAS scales, and contradict Cytrynbaum's advice.

No significant differences in kappa values were found between the various disciplines or between the types of scales. Contrary to expectation, the scales rating physical function did not show a higher percentage of agreement. The good inter-rater reliability for scales rating higher cognitive function could be because we measured the professionals' insight rather than the child's performance during the assessment. The main reason for disagreement between raters was the interpretation of the children's capacity when they did not perform to the best of their ability during the assessment. Using the professional's judgment of the therapists and independent raters for scoring was assumed to be beneficial to the inter-rater reliability, although misinterpretation of this criterion can also be detrimental to the reliability. This emphasizes the importance of clear instructions.

Study limitations

In terms of the first aim of our study, the results must be interpreted with caution, because the raters forming the pairs differed in their knowledge of the child's history and treatment results. Because GAS intends to measure a professional's perception of a child's level of functioning, the children's own therapists had more information they could use in scoring than did the independent raters. This difference may have biased the inter-rater reliability. However, if this were true, the inter-rater reliability would have been underestimated rather than overestimated. No systematic differences in scoring patterns were found between therapists and independent raters.

Another limitation in regard to the first aim is that a possible dependence of the data could present a problem with the CIs of the kappa value, because the calculation was based on 64 GAS scales constructed for 23 children. However, 3 therapists, of different disciplines, each constructed a GAS scale in a different, predefined domain that was distinctive for their discipline. Three different independent pairs of therapists only scored the GAS scale constructed by their own discipline, to determine the agreement within the pairs. Although the 6 *scores* per child are dependent, this procedure makes it unlikely that there was appreciable dependence between the 3 values of agreement per child, confirming the 95% CIs of the kappa values given in the Results section. The kappa values could be biased by floor and ceiling effects. If serious deterioration (−3) or excellent performance (+2) in all domains of functioning caused dependence, the correlations could be overestimated. We had no floor effects. Ceiling effects were improbable because 19 of all 24 scales scored as +2 by both raters were for different children. Only the influence on the corresponding CIs has to be taken into consideration, because the mean

kappa value of the separate data sets corresponds with the kappa of the entire data set. Multiple GAS scales could be constructed for a child, as usually more than one goal is set per child. GAS was originally characterized by the use of the *T* sum formula, a mathematical technique quantifying the achievement in several weighted goals per therapist. Although this formula theoretically could correct for dependence, the correction is too subjective to provide a real solution for the dependence issue. The formula also introduces false sensitivity to change, as GAS data are at best ordinal.[21] Therefore we prefer to evaluate individual GAS scores and analyze multiple raw change scores with nonparametric statistics.[2]

A limitation of the study regarding the second aim is that the study may be difficult to reproduce, because of the case summary provided by the study leader and the instructions influencing communication between parents and therapists. Although the information in the case summary was kept to a minimum, it was subjective and may have influenced the independence during GAS scale construction between the therapists and independent raters. This subjectivity was, however, unavoidable, as the GAS scales of both raters had to be based on the child's and parents' request for help and their expectations. The influence of the children themselves on the independence between the raters is unknown.

A second limitation regarding the second aim is that the therapy theoretically focused more on the goals set by the child's therapists than on those set by the independent raters, enabling more accurate rating of the goals set by the child's therapists. However, after construction of both scales was completed, the independent rater's goal was shared with the child's therapist, and no instruction was given as to whether therapy should focus on GAS. Moreover, the outcomes for both groups of GAS scales were equal, with a median of 0.

Finally, the procedure used to measure goal attainment could have influenced the inter-rater reliability. That parents or teachers could be consulted as part of the judgment of goal attainment may have meant that the professionals' knowledge and experience had less influence on the score. The independence of the raters could decrease as a result of them both conferring with the same responder. We divided all GAS scales into 2 categories: scales where the score was supported by observing the child, and scales where the score was supported by conferring with parents or teachers, as was stipulated in each GAS scale. Thirty-two percent of the scales scored after observation resulted in disagreement, compared with 18% of the scales scored after consulting with parents or teachers.

At the end of the study, all the parents were interviewed and asked about their experience, in order to gain insight into the impact of participating in a study like this for parents and children. The parents had an active role in keeping the pairs of raters independent during the study, and the GAS scales were shared with the parents right after their construction. The parents of all 23 children indicated that they were satisfied with the use of GAS in the treatment of their child and with their participation in this study. Future reliability studies on GAS in other rehabilitation settings or with other diagnostic groups should be considered, as the burden on participating families is minimal.

The measurement properties of the content of GAS scales and the sensitivity to change are subjects that remain to be explored.[5] A possibility for future research is to have a team of experts rate the overall value of therapists' goals. The present study showed that the GAS scales constructed by the child's own therapists in particular should be used for further validity research. Even if an independent assessor were essential to reduce bias in evaluating outcome, the involvement of the child's own therapist (and the child or child's parents) is still necessary when designing goals for therapy.

Despite the need for future research, GAS is a very promising tool that is allowing children and their families to set transparent and important functional goals together with their health care professionals.

CONCLUSIONS

In this study, GAS was used for children with CP, under optimal conditions (trained team using predetermined criteria) in the routine practice of a children's unit in a medium-sized rehabilitation center. The inter-rater reliability of GAS scores proved good to excellent. The results also suggest that scale construction by the child's own therapist as opposed to an independent rater has a positive influence on the inter-rater reliability of the scales. The inter-rater reliability can be further improved by standardizing the procedure. We found that discrepancies between the professionals' interpretation of the child's capacities and the child's actual performance during assessment were the main cause of disagreement between raters. Reliability studies in other rehabilitation settings and diagnostic groups are recommended. Further investigation of the content reliability and content validity in the construction of GAS scales and their sensitivity to change is also necessary and of great importance for rehabilitation.

ACKNOWLEDGEMENTS

We thank all participating children, parents, and therapists at the children's department of the Rehabilitation Center Breda, The Netherlands; Riekie H.C.W. de Vet, PhD, for her support as an expert on clinimetrics; and Tjeerd van der Ploeg for his statistical support.

REFERENCES

1. Kiresuk T, Sherman R. (1968) Goal attainment scaling: a general method of evaluating comprehensive community mental health programs. *Community Ment Health J* **4**: 443–453.
2. Steenbeek D, Meester-Delver A, Becher JG, Lankhorst GJ. (2005) 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* **19**: 274–282.
3. Cusick A, McIntyre S, Novak I, Lannin N, Lowe K. (2006) A comparison of goal attainment scaling and the Canadian Occupational Performance Measure for paediatric rehabilitation research. *Pediatr Rehabil* **9**: 149–157.
4. Ekström-Ahl L, Johansson E, Granat T, Carlberg EB. (2005) Functional therapy for children with cerebral palsy: an ecological approach. *Dev Med Child Neurol* **47**: 613–619.
5. Steenbeek D, Ketelaar M, Galama K, Gorter JW. (2007) Goal Attainment Scaling in Paediatric Rehabilitation: a critical review of the literature. *Dev Med Child Neurol* **49**: 550–556.
6. Mailloux Z, May-Benson TA, Summers CA, Miller LJ, Brett-Green B, Burke JP, Cohn ES, Koomar JA, Parham LD, Smith Roley S, Schaaf RC, Schoen SA. (2007) Goal attainment scaling as a measure of meaningful outcomes for children with sensory integration disorders. *Am J Occup Ther* **61**: 254–259.
7. Lowe K, Novak I, Cusick A. (2006) Low-dose/high-concentration localized botulinum toxin A improves upper limb movement and function in children with hemiplegic cerebral palsy. *Dev Med Child Neurol* **48**: 170–175.
8. Paolicelli PB, Ferrari A, Lodesani M, Muzzini S, Sassi S, Maoret A, Bianchini E, Cioni G. (2001) Use of Botulinum toxin type A in walking disorders of children with cerebral palsy. *Eur Med Phys* **37**: 83–92.
9. Mall V, Heinen F, Siebel A, Bertram C, Hafkemeyer U, Wissel J, Berweck S, Haverkamp F, Nass G, Döderlein L, Breitbach-Faller N, Schulte-Mattler W, Korinthenberg R. (2006) Treatment of adductor spasticity with BTX-A in children with CP: a randomized, double-blind, placebo-controlled study. *Dev Med Child Neurol* **48**: 10–13.
10. Sheffler G, Canetti L, Wiseman H. (2001) Psychometric properties of goal-attainment scaling in the assessment of Mann's time-limited psychotherapy. *J Clin Psychol* **57**: 971–979.

11. Rockwood K, Stadnyk K, Carver D, MacPherson KM, Beanlands HE, Powell C, Stolee P, Thomas VS, Tonks RS. (2000) A clinimetric evaluation of specialized geriatric care for rural dwelling, frail older people. *J Am Geriatr Soc* **48**: 1080–1085.

12. Rockwood K, Howlett S, Stadnyk K, Carver D, Powell C, Stolee P. (2003) Responsiveness of goal attainment scaling in a randomized controlled trial of comprehensive geriatric assessment. *J Clin Epidemiol* **56**: 736–743.

13. Schlosser RW. (2004) Goal attainment scaling as a clinical measurement technique in communication disorders: a critical review. *J Commun Disord* **37**: 217–239.

14. Palisano RJ. (1993) Validity of goal attainment scaling in infants with motor delays. *Phys Ther* **73**: 651–660.

15. Palisano RJ, Haley SM, Brown DA. (1992) Goal attainment scaling as a measure of change in infants with motor delays. *Phys Ther* **72**: 432–437.

16. Steenbeek D, Ketelaar M, Galama K, Gorter JW. (2008) Goal Attainment Scaling in paediatric rehabilitation: a report on the clinical training of an interdisciplinary team. *Child Care Health Dev* **34**: 521–529.

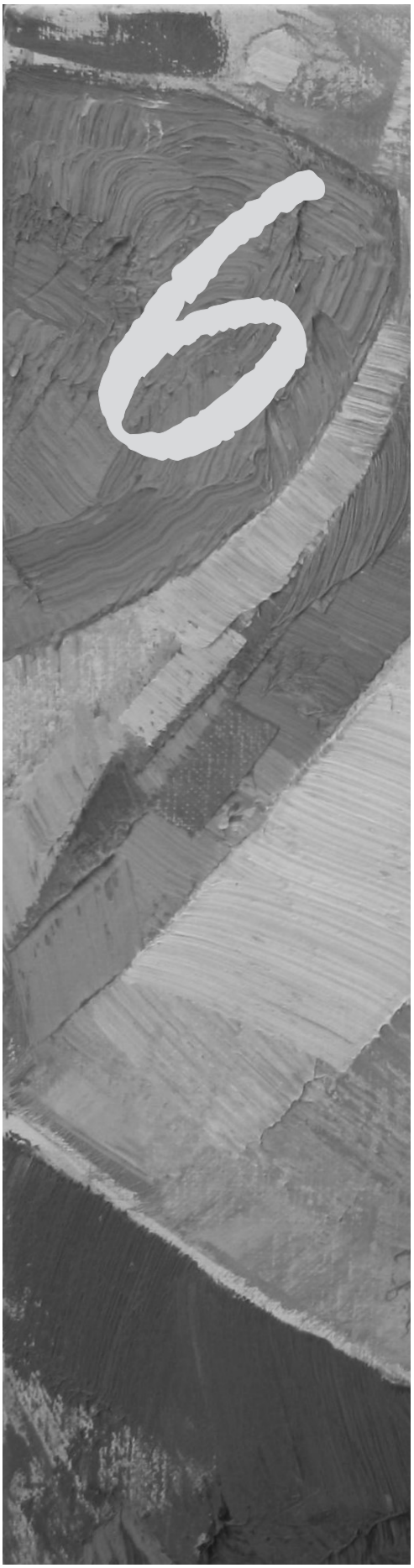
17. Cytrynbaum S, Ginath Y, Birdwell J, Brandt L. (1979) Goal attainment scaling: a critical review. *Eval Q* **3**: 5–40.

18. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. (1997) Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol* **39**: 214–223.

19. Eliasson AC, Krumlinde-Sundholm L, Rösblad B, Beckung E, Arner M, Öhrvall AM, Rosenbaum P. (2006) The Manual Ability Classification System (MACS) for children with cerebral palsy: scale development and evidence of validity and reliability. *Dev Med Child Neurol* **48**: 549–554.

20. Landis JR, Koch G. (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**: 159–174.

21. Tennant A. (2007) Goal attainment scaling: Current methodological challenges. *Disabil Rehabil* **15**: 1583–1588.



The added value of Goal Attainment Scaling for evaluation of outcome in rehabilitation of children with Cerebral Palsy

D. Steenbeek, J.W. Gorter, M. Ketelaar,
K. Galama, E. Lindeman

Submitted for publication.

ABSTRACT

Objectives: To explore the added value of Goal Attainment Scaling (GAS) to the Pediatric Evaluation of Disability Inventory (PEDI) and the 66-item Gross Motor Function Measure (GMFM-66) in inter-disciplinary rehabilitation practice.

Design: Observational study. Pre-test post-test design.

Subjects/Patients: Twenty-three children with Cerebral Palsy (age 2–13 years).

Methods: Physical, occupational and speech therapists constructed and scored 6-point GAS scales which met predetermined criteria, describing the most important functional goal per discipline. GAS, PEDI and GMFM-66 assessments were performed before and after a 6-month treatment period. The content of these measures were compared using the ICF-CY. Spearman's rho correlations between GAS change scores per discipline and the change scores of the PEDI Functional Skills scales and GMFM-66 were calculated. Complete goal attainment was compared to significant change on the standardized measures.

Results: Twenty percent of goals were not covered by the PEDI, or GMFM-66. Significant correlations were found between GAS and PEDI change scores. Of 39/64 GAS scales, scored as complete goal attainment, 16 individual PEDI scores did not show change on the related scale.

Conclusion: The GAS, PEDI and GMFM-66 were complementary in their construct and individual responsiveness. By using the standardized instruments alone, many goals achieved could have been missed.

INTRODUCTION

Measurement of change over time at the activities and participation level of the International Classification of Functioning, Disability and Health (ICF)[1] is very relevant in contemporary interdisciplinary rehabilitation care. Many standardized outcome measures evaluating patients with disabilities have demonstrated reliability and validity for specific populations. However, it has become clear in the last decade that proving clinically meaningful change using these measures is challenging. This is particularly true for measurement of individual patients' progress at the activities and participation level in rehabilitation practice.[2] Lack of responsiveness, defined in the present study as *an instrument's ability to detect clinically meaningful change over time*, may be caused by one of two issues. First, items in standardized measures may not match the individual rehabilitation goals.[3,4] They often include many items that are not relevant for that specific individual. Second, if items do

match therapy goals, the outcome may not represent goal attainment. For the purpose of rehabilitation research the large number of items of standardized measures intend to cover all aspects of functioning and should benefit the measure's sensitivity to change at a group level. In rehabilitation practice however, change may easily be missed if only a few items measure change against the noise of a larger number of unchanged items. Most populations in rehabilitation are heterogeneous and goals are often widely spread over the ICF domains, resulting in difficulty in interpreting the outcome when using generalized measures. Moreover, when measuring individual patients, the requirements for the quality of the instruments used are higher than in the research setting.[2,5] Failure to demonstrate improvement in an individual patient may be due to a lack of improvement, but it may also be the result of using an instrument with poor responsiveness.

Goal Attainment Scaling (GAS) is an increasingly popular individual instrument for assessment of progress in rehabilitation. GAS measures the extent to which individual goals are attained and is applicable for children,[6-8] adults[3,9-11] and the elderly.[12] Its responsiveness at the activity and participation level can be assumed to be better than that of common standardized functional measures, in particular in heterogeneous populations.[3,7,13] When GAS is used in combination with standardized outcome measures, it may overcome the limitations of the use of standardized measures alone. However, the application of GAS in a new field requires a lot of effort to apply the method properly[7] and to clarify its added value to standardized measures.

The present study focuses on the use of GAS in the paediatric rehabilitation setting, in particular on children with Cerebral Palsy (CP): the most prevalent condition. Capturing the most important individual goals is challenging in outcome measurement in rehabilitation of children with CP. However, little is known about the comparison of GAS and standardized measures in this field.

Both the Pediatric Evaluation of Disability Inventory (PEDI) and the Gross Motor Function Measure (GMFM) are examples of commonly used standardized and generic outcome tools in paediatric rehabilitation for children with CP. In an intervention study with 55 children with CP (age 2–7 years, GMFCS level I–III), individual treatment goals of the children receiving physical therapy were analyzed to find out how they were reflected in the PEDI and the GMFM.[4] Sixty percent of the treatment goals were covered by both measures, but 14% of the treatment goals were not covered at all. Treatment goals in the field of occupational and speech therapy have not as yet been researched.

The purpose of the present study was to explore the added value of GAS to the PEDI and the 66-item version of the GMFM (GMFM-66) in daily inter-disciplinary rehabilitation practice for children with CP. The added value was examined regarding (a) construct differences and (b) differences in responsiveness at an individual level.

METHODS

Participants and design

Children that met the following criteria were included: (a) A confirmed diagnosis of CP, (b) were between the ages of 2 and 14 years and (c) their physician expected them to be in interdisciplinary therapy for at least 6 months. The distribution of severity of CP was evaluated by the Gross Motor Function Classification System (GMFCS),[14] which classifies the motor function of children with CP based on their self-initiated movement (http://www.canchild.ca), and also with the Manual Ability Classification System (MACS),[15] which classifies how children with CP use their hands when handling objects in daily activity (http://www.macs.nu). In both classification systems, level I represents the highest functional ability and level V the most limited. After a training period in GAS,[8] eight physical, eight occupational and four speech therapists participated in this study with a pre-test post-test design. Changes were measured for each child over the 6-months therapy period using GAS, the PEDI and the GMFM-66. The child's therapists constructed GAS scales at baseline (t=0), scored the scales at 3 months (t=1) and after 6 months (t=2). The child's occupational therapists administered the PEDI and the child's physical therapists the GMFM-66 at t=0 and at t=2.

The GAS method used

Originally, GAS was developed by Kiresuk and Sherman in 1969[16] as a 5-point scale, with a score 0 representing the expected level of functioning after a predefined period, +1 or +2 representing achievement of more or much more than is expected, and –1 or –2 representing less and much less progress than expected. GAS was originally characterized by the use of the T sum formula: a mathematical technique quantifying achievement in several weighted goals per patient. Although different recent practical guides[10,17] use the original method, our research group adapted it. We introduced analysis of raw GAS scores, instead of using the T sum formula in order to prevent false sensitivity

to progress as GAS data are at best ordinal.[18] We introduced a –3 value for deterioration and determined that –2 was equal to start in order to prevent bottoming effects.[8]

For the present study six-point GAS scales were constructed. The participants agreed to adhere to the following criteria for scale development. (a) Goals should be based on the request for help, must be important for and relevant to the children and their families, and should describe the main aim of therapy for each discipline on the ICF activity- and participation level. (b) Levels of the scales should be specific, measurable, achievable, realistic/relevant and time related (SMART). (c) Scales should be constructed ordinally with incremental steps of equal intervals. To ensure ordinality each GAS scale should reflect a single dimension of change. (d) In addition, it had to be possible to score a scale within 10 minutes, in order to ensure that it was practicable.

Therapists scoring their own scales, proved to be more reliable than independent scoring.[6] The intermediate 3-month score was used to follow the course of goal attainment. The scores were based on the professionals' insight of goal achievement rather than on the child's actual performance at the time of the assessment[8] in order to minimise the influence of children's whims (fatigue, motivation, interaction, behaviour). The raters' insight (perception of goal attainment) was based on observing the child or interviewing parents or teacher.

The PEDI and GMFM-66

The PEDI[19,20] measures both capability and performance of functional activities in daily situations in three domains: self-care, mobility, and social function. Capability is measured by the identification of functional skills the child has mastered in these three domains. For the functional skills scale, the parents indicate whether the child is capable of performing each of 197 tasks in the three domains. Performance of daily functional activities is measured by the level of caregiver assistance the child needs to accomplish functional activities. For the present study only the three domains of the functional skills scale (FSS) of the Dutch version of the PEDI (PEDI-NL)[21,22] were used. The Dutch version is a cross-cultural adaptation of the USA version. This equivalent of the original USA version helped strengthen the construct of the PEDI for its use in The Netherlands,[22] improving the universalisation. For example bicycling was added to the mobility domain as it is the main means of transportation for going to school and visiting friends in The Netherlands.

The GMFM-66[23,24] is a 66-item version of the original 88-item standardized observational instrument designed and validated to measure change in gross motor function over time in children with CP. For this study we used a Dutch translation of the GMFM-66. The GMFM and PEDI are complementary to one another when used to attain a complete picture of the child and evaluate change over time.[25] Good reliability, validity and responsiveness was reported for both the PEDI and GMFM-66.[19,23] The psychometric properties of the Dutch translations are comparable to the original instruments.[21,22,26]

Evaluation and analysis

Three methods were used to explore the added value of GAS: two regarding construct differences (A, B) and one regarding individual responsiveness (C).

Method A. Both the GAS items and the PEDI and GMFM-66 items were classified by the study-leader (first author) and coded according to the ICF, Child's and Youth version (ICF-CY), using the 8 ICF linking rules of Cieza et al.[27] The combined use of the ICF and GAS was described in a recent publication by McDougall.[28] GAS items were compared to the PEDI and the GMFM-66 items and were defined as a match or no match.

Method B. Spearman's rho correlations were calculated separately for the GAS change scores of each discipline and the change scores of each domain of the PEDI and the GMFM-66 change scores, to explore the correlation between the outcomes per discipline. In order to understand the influence of construct differences on these correlations better, spearman's rho values were also calculated after excluding pairs of change scores where the GAS item did not match the items in the PEDI and GMFM-66.

Method C. Responsiveness for the three measures was approached as follows: minimal clinically meaningful individual improvement on GAS scales was defined as a score equal to or more than zero. Minimal clinically meaningful individual improvement on the PEDI and GMFM-66 was determined as being equal to the Smallest Detectable Change (SDC). According to the manuals of the methods,[20,24] the SDC was defined as a change in which the lower bound of the 95% confidence interval of the t=2 score did not overlap the upper bound of the 95% confidence interval of the t=0 score. The manuals of the measures were used to determine the 95% confidence interval for each score. The amount of individual change on the GAS scales and correlated scales of the PEDI and the GMFM-66 were compared in crosstabs.

In order to assess the possible consequences of the individual changes found, changes at group level were also considered, as research usually addresses group changes. Acceptable responsiveness for GAS at group level was defined as a median score of equal to or more than zero at t=2, if the median score at t=1 was between the median score at t=0 (by definition –2) and t=2 and if changes between t=0, t=1 and t=2 were statistically significant. Acceptable responsiveness for the PEDI and GMFM-66 was defined as statistically significant change between t=0 and t=2, as the minimal clinically meaningful change at group level is unknown. Changes at group level of all three measures were tested by using the Wilcoxon signed ranks test (2-tailed).

Statistics were performed in SPSS/PC, release 14.0.

RESULTS

Twenty-three children between the ages of 2 and 13 years participated (mean 7.6, SD 3.1; their GMFCS and MACS levels ranged from I to V (**Table 6.1**). Eighteen children were treated by 3 disciplines, i.e. physical therapy, occupational therapy and speech therapy, and 5 children by 2 disciplines, i.e. physical therapy and occupational therapy. Sixty-four GAS scales (18 × 3 + 5 × 2) were constructed and scored. For each child the GAS scores, the scale scores on the PEDI and GMFM-66 scores at baseline (t=0) and after 6 months (t=2) are shown (**Table 6.1**). No deterioration was found.

Classification with the ICF-CY

Table 6.2 displays the ICF-CY items used in the GAS scales and shows which of them matches the PEDI and GMFM-66. Thirteen out of 64 (20%) goals described in the GAS scales were not covered by the PEDI nor the GMFM-66.

Correlations

The GAS change scores from physical therapists and PEDI-FSS Mobility, and from the speech therapists and PEDI-FSS Social function were significantly correlated (**Table 6.3**). Another significant correlation coefficient was found between the GAS change scores from the occupational therapists and the PEDI-FSS Self-care change scores after exclusion of the non-matching pairs. Low correlations were found for all other combinations, including the GAS

Table 6.1 Summary of characteristics of the subjects and outcome

Child no.	Age (y)	Gender	GMFCS level	MACS level	GAS score PT *t*=2	GAS score OT** *t*=2	GAS score ST** *t*=2	PEDI FSS*** Self-care		PEDI FSS*** Mobility		PEDI FSS*** Social function		GMFM-66***	
								t=0	*t*=2	*t*=0	*t*=2	*t*=0	*t*=2	*t*=0	*t*=2
1	12	M	IV	IV	−1	1	1	50	56*	40	40	73	76	27	27
2	13	F	V	V	−1	1	ND	50	51	41	44	74	86*	23	26
3	7	F	II	I	−1	2	1	72	78*	65	66	74	79*	67	69
4	13	F	V	V	−1	0	0	32	32	36	38	64	66	28	30
5	5	M	I	III	2	1	0	58	65*	62	68*	69	72	69	74
6	7	M	III	II	−2	0	−1	51	58*	51	46	67	68	53	53
7	9	F	IV	II	−1	1	2	58	79*	67	70	86	86	54	57
8	11	F	III	II	−1	−2	−1	64	65	56	58	67	68	57	58
9	4	M	I	III	−1	2	−1	50	63*	51	68*	62	69*	64	70*
10	7	M	II	I	1	2	0	60	65*	79	83	69	72	76	78
11	6	M	V	V	2	−1	0	22	22	26	33*	55	55	35	35
12	3	M	II	III	−1	−1	1	50	53	61	60	67	69	64	64
13	5	M	III	II	2	1	ND	63	67*	59	66*	79	83	63	66
14	8	M	IV	IV	2	2	2	56	53	28	35*	63	68*	34	36
15	9	F	II	I	2	2	ND	71	71	85	90	74	89*	88	88
16	10	M	III	II	−1	0	ND	66	66	64	64	74	74	59	63
17	6	M	II	II	2	1	−1	57	61*	79	94*	65	64	75	80
18	6	M	I	III	−1	2	−2	72	77*	100	100	78	78	88	88
19	10	F	V	V	1	2	1	48	53*	40	38	68	76*	36	36
20	10	M	IV	II	−1	−1	−2	54	63*	48	42	73	73	29	31
21	2	M	II	III	1	0	2	45	48*	65	66	45	54*	53	55
22	11	F	V	V	−2	−1	1	41	44	39	39	76	83*	25	25
23	5	M	II	I	−1	2	ND	66	75*	66	70*	76	81*	83	83

Legend: *t*=0, baseline; *t*=2, 6 months after baseline; M, male; F, female; GMFCS, Gross Motor Function Classification System; MACS, Manual Ability Classification System. Level I represents the best functional abilities and level V the most limited ones. *Indication of significant positive individual change (no overlap of 95% CIs in scores *t*=0–*t*=2). **Indication of relevant group change (median GAS score ≥0). ***Indication of significant group change ($p \leq 0.05$). ND, No Data; speech therapy not involved in this child.

Table 6.2 List of ICF-CY items used in the 64 GAS scales and their coverage in the PEDI-NL and GMFM-66

Description of the activity	ICF-CY classification	Frequency of its use in the GAS scales	Goal covered by PEDI	Goal covered by GMFM-66
Using general skills and strategies of the writing process	d1700	2	–	–
Speaking	d330	6	+	–
Producing body language	d3350	1	+	–
Conversation	d350	4	+	–
Discussion with one person	d3550	1	–	–
Using writing machines	d3601	1	–	–
Using communication devices and techniques	d360	4	+	–
Changing basic body position	d410	3	+	+
Maintaining a sitting position	d4153	4	+	+
Maintaining a standing position	d4154	3	–	+
Transferring oneself	d420	1	+	+
Lifting and carrying objects	d430	1	+	–
Manipulating	d4402	6	–	–
Fine hand use	d440	2	–	–
Walking short distances	d4500	3	+	+
Walking on different surfaces	d4502	1	+	+
Climbing	d4551	3	+	+
Running	d4552	1	–	+
Moving around using equipment	d465	1	+	–
Driving human-powered transportation	d4750	3	+	–
Putting on clothes	d5400	1	+	–
Taking off clothes	d5401	2	+	–
Dressing	d540	4	+	–
Carrying out eating Appropriately	d5501	2	+	–
Basic interpersonal interactions	d710	2	+	–
Play	d9200	1	+	–
Sports	d9201	1	–	–
		Total 64		

Table 6.3 Spearman's rho correlations between the change scores of GAS and the most corresponding PEDI functional scales (FSS)

Δ GAS score	Measure	Spearman's rho	*p*-value	Recalculation after exclusion of the non-corresponding items	*p*-value
Physical therapists	Δ PEDI FSS Mobility	0.64 (*n*=23)	<0.01	0.57 (*n*=18)	0.01
Occupational therapists	Δ PEDI FSS Self-care	0.28 (*n*=23)	0.20	0.71 (*n*=12)	0.01
Speech therapists	Δ PEDI FSS Social function	0.55 (*n*=18)	0.02	0.73 (*n*=17)	<0.01

change scores from the physical therapists and the GMFM-66 change scores, even after exclusion of the non-matching pairs. Seven goals from the physical therapists, 11 goals from the occupational therapists and 1 goal from a speech therapist were not covered by the PEDI. The recalculated correlation coefficients – after exclusion of these pairs – can be found in **Table 6.3**.

Responsiveness

Looking at individual change, we found 9/23, 18/23 and 12/18 GAS scores equal to or more than zero for the physical, occupational and speech therapists respectively; 21 children (91%) scored zero or more on at least one of the GAS scales and 7 (30%) did on all GAS scales (**Table 6.1**). Twenty-one children (91%) improved on at least one of the PEDI functional scales and only 2 (9%) on all three scales. On the GMFM-66 individual change without overlap in the given 95% confidence intervals was found for only 1 child (**Table 6.1**).

The **Crosstabs 6.4a–c** show the comparison of individual change on correlated measures. Of the 39 GAS scales scored as 0 or higher, 16 individual PEDI scores did not show change on the related scale. On the other hand 3 out of 26 significant changes on the PEDI scales were not detected by GAS.

At group level the GAS scores after 3 months (t=1) had a median of –1 and after 6 months (t=2) a median of 0 with statistically significant differences from t=0 to t=1 and t=1 to t=2, indicating acceptable responsiveness of GAS at group level; **Table 6.5** shows the frequencies of scores and z values.

The changes at *group* level between the baseline scores and scores after 6 months (t=2) on the PEDI-FSS Self-care, the PEDI-FSS Mobility, the PEDI-FSS Social function, and GMFM-66 all were statistically significant (z=–3.62, z=–2.62, z=–3.68, and z=–3.33 respectively with all $p \leq 0.01$).

Table 6.4a–c Crosstabs comparing individual change on correlated measures

6.4a

Δ PEDI CI pre ≠ CI post GAS score ≥ 0		FSS Mobility +	FSS Mobility –	Totals
Physical therapy	+	5	4	9
	–	1	13	14
Totals		6	17	23

6.4b

Δ PEDI CI pre ≠ CI post GAS score ≥ 0		FSS Self-care +	FSS Self-care –	Totals
Occupational therapy	+	13	5	18
	–	1	4	5
Totals		14	9	23

6.4c

Δ PEDI CI pre ≠ CI post GAS score ≥ 0		FSS Social function +	FSS Social function –	Totals
Speech therapy	+	5	7	12
	–	1	5	6
Totals		6	12	18

6

Table 6.5 The frequency of GAS scores after 3 (t=1) and 6 months (t=2) and the statistically significance of differences at group level between t=0, t=1 and t=2 tested using the Wilcoxon signed ranks test

GAS scores value	Frequency		
	t=0	t=1	t=2
–3	0	1	0
–2	64	9	5
–1	0	24	20
0	0	16	8
1	0	12	14
2	0	2	17
Totals	64	64	64
Median	–2	–1	0
Wilcoxon t=0–t=1	z=–6.4; $p < 0.01$		
Wilcoxon t=1–t=2	z=–2.7; $p < 0.01$		
Wilcoxon t=0–t=2	z=–6.7; $p < 0.01$		

DISCUSSION

The present study explored the measurement construct and the responsiveness of GAS in comparison to the PEDI and GMFM-66. When used by a group of trained therapists, added value was found for both clinimetric properties. GAS detected important changes in activity skills that were not found using the generalized measures. This study is unique in that it measured the properties of GAS during the unrestrained daily work of a rehabilitation team.

Three out of 23 physical therapy goals and 13/23 occupational therapy goals were not covered by at least one of the measures. Examples are 2-handed activities like using a pair of scissors (child no. 10 in **Table 6.1**), the use of a joystick by a child with severe CP who uses an electric wheelchair (no. 11), and writing (no. 15 shown in **Table 6.6**). Almost all (17/18) speech therapy goals were covered by the PEDI Social Function domain. This can be explained by the criteria of GAS scale construction. Speech therapy goals at the activity and participation level usually have social communication as a main focus. The use of GAS by speech therapists has been shown to be valuable.[8,29] As far as we know, presently, there is no literature about the value of the PEDI in evaluating speech therapy.

The poor correlation between the GAS scores from the occupational therapists and the PEDI FSS Self-care may have been caused by the non-matching items, as the correlation improved after excluding these items. Discrepancies between GAS and the PEDI outcome can only in part be explained by the differences in textual content, as the correlations were not complete regarding the matching pairs. This finding supports the proposition that individualized and generalized measures measure different constructs. Contrary to the expectation, low and non-significant correlation coefficients (Spearman's rho 0.25, p=0.26) were found between physical therapists' GAS change scores and GMFM-66 change scores. This is probably due to the narrow distribution of the GAS data, as a median of –1 was found for the physical therapy scores. Another explanation may be a difference in the construct: the GMFM-66 measures a child's capacity in basic gross motor abilities and, in contrast, the GAS and PEDI measure the performance of activities.[30] An obvious limitation of the study, regarding the confidence intervals of Pearson's correlations, is the rather small number of subjects.

At an individual level, the responsiveness found for all three measures was ambiguous. This is directly related to our challenging definitions of clinically meaningful change for both GAS and the generalized measures. How much individual change is enough? There is still little consensus about

Table 6.6 Two examples of Goal Attainment Scaling

Patient no.		5	15
Discipline		Physical therapy	Occupational therapy
Setting		Training to ride a standard 16-inch 2-wheeled bicycle in the gymnasium.	Extra training in writing, geared to the methods of regular school. The table, chair, pencil are described.
Task		J, get on your bike and ride it.	H, write as accurately as possible in cursive.
Scoring method		Observation.	Observation.
–3	Worse than starting level (deterioration)	J needs help to get on and off his bike and loses his balance.	H does not succeed writing cursively at all. She can not reproduce the characters.
–2	Equal to starting level	With a lot of trial and error and variety of strategies, J gets on and off his bike without help, but looses his balance during the start of riding.	H writes 13 characters in the special workbook correctly.
–1	Less than expected	J has developed a method to get on and off, but looses his balance during the start of riding.	H writes all characters in the special workbook correctly.
0	Expected goal	J gets on and off easily, but still looses his balance during the first meters of riding.	H writes all characters on lined paper correctly.
1	Somewhat more than expected	J gets on and off easily and rides straight for 10 meters.	H succeeds writing short words.
2	Much more than expected	J gets on and off easily and rides without help in the gymnasium. The next goal is riding outside and in traffic.	H writes short sentences. The next goal is improving the writing speed.

Patient no. 5 (Table 6.1) Example of a GAS scale, constructed for a 5-y old boy with unilateral CP and spasticity in a flexion pattern of the upper extremity causing trouble steering his bicycle. The goal is covered by the PEDI-NL but not by the GMFM. GAS added the relevance of the change measured. *Patient no. 15 (Table 6.1)* Example of a GAS scale, constructed for a 9-y old girl with unilateral CP, illustrating GAS measuring relevant change in an activity such as writing that would not be detected by the PEDI-NL nor the GMFM-66.

this issue.[5,31] We evaluated GAS by setting the cut off point as a score of <0 versus a score of ≥0, because in general practice professionals are only satisfied when goals are attained completely. However, by doing this, another

construct was introduced (conform Palisano et al.[32]). Treating a range of ordinal data dichotomously threatens the sensitivity to different changes in goal attainment. Regarding the PEDI and in particular the GMFM-66, the poor individual changes found are also due to the strict demands for change, as dictated by each manual. On the other hand, due to the strict demands, we ignored the distinction between minimally detectable change and minimally important change.[5] Another explanation for poor individual change measured by the PEDI and GMFM-66 could have been the ages of the participants. Eleven out of the 23 enrolled children were older than 7 years. Both the PEDI and the GMFM are most responsive for children under the age of 7.[26] Despite the limitations due to the definitions, our results clearly show that GAS, the PEDI and GMFM were complementary. By using the PEDI alone, achieved goals could have been missed and by measuring complete goal attainment alone, significant change could have been missed. The crosstabs 6.4a–c suggest GAS to be more responsive than the PEDI measuring individual change.

Although change at group level was not the main focus of this study, the possibility of measuring group effects in heterogeneous groups with non-parametric statistics is one of the most important advantages of GAS. At group level the responsiveness of both the GAS method and the PEDI and GMFM-66 were good, supporting the conclusions of previous studies. Good *responsiveness* was found for the GAS scales from the occupational and speech therapists only. The reason for a disappointing responsiveness of GAS constructed by physical therapists is unknown.

Our results concur with those of other authors. In the field of paediatric physical therapy, moderate and low correlations were found between GAS and Peabody gross motor and fine motor change scores.[32] In the field of paediatric occupational therapy GAS was recently compared in several studies with another individualized measure: the Canadian Occupational Performance Measure (COPM).[13,33] When combined with GAS, a dynamic and interactive process may emerge, from individual needs to setting and implementing goals.[34]

Several studies in the field of adult rehabilitation care in different settings and diagnostic groups report GAS to be useful in functional outcome measurement.[3,9,11,12,17] Studies in neurorehabilitation comparing GAS and the Functional Independent/Assessment Measure (UK FIM+FAM)[3,11] or depression and self efficacy scales[35] reveal that up to one-third of goals set by GAS were not covered by these standardized measures.

What one measures when using GAS: the professional's expectation of

the patients' future functioning or the effect of a treatment period, remains a challenging question. GAS depends on the patient's ability to achieve the goals and the professional's skills to predict outcome, which requires knowledge and experience.[6,10] The outcome on GAS is dependent on the quality of scale construction. This represents a potential source of measurement error that is unique to the idiosyncratic nature of GAS. While potential therapist bias poses a threat to validity, it also offers an opportunity to use the professionals' insight in measuring what one intends to measure. Although previous studies have shown that this professional insight can be used reliably,[3,6,32] therapist bias has to be considered in each new trial setting. Many recent studies on GAS have been performed with the intention of increasing its reproducibility, validity and to standardize the method. Novel approaches to writing SMART goals to encourage uniformity of the application are examples.[3,17]

Finally, in contrast to the standardized measures, a property of GAS at group level as well as individual level, is that the score by definition describes the relevance of the change. When the outcome on standardized measures and GAS match, GAS has added value as it also indicates the relevance of the change measured. This is probably the most important additional value of GAS (first example given in **Table 6.6**).

In conclusion, this study shows that GAS as an individual measure and the PEDI and GMFM-66 as standardized generic measures were complementary as outcome tools for children with CP. The results show that there is added value regarding the construct and individual responsiveness of GAS. By using the generic measures alone, a substantial number of achieved goals could have been missed, even if the ICF item matched the items of the standardized measures. Moreover, by measuring complete goal attainment alone, smaller but significant change could have been missed. The added value of GAS, the possibility of measuring group effects in heterogeneous groups, the feasibility and satisfaction in working with GAS among patients, families and the teams of professionals, motivate the ongoing increase of its use in the rehabilitation care.

ACKNOWLEDGEMENTS

We would like to thank all participating children, parents and therapists of the children's department of the rehabilitation centre at Breda, The Netherlands.

REFERENCES

1. World Health Organization. (2001) *International classification of functioning, disability and health (ICF)*. Geneva: World Health Organization.
2. Dekker J, Dallmeyer AJ, Lankhorst GJ. (2005) Clinimetrics in rehabilitation medicine: current issues in developing and applying measurement instruments. *J Rehabil Med* **37**: 193–201.
3. Turner-Stokes L, Williams H, Johnson J. (2009) 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* **41**: 528–535.
4. Engelen V, Ketelaar M, Gorter JW. (2007) Selecting the appropriate outcome in paediatric physical therapy: how individual treatment goals for children with cerebral palsy are reflected in GMFM-88 and PEDI. *J Rehabil Med* **39**: 225–231.
5. de Vet HCW, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. (2006) Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes* **22**: 54. Review.
6. Steenbeek D, Ketelaar M, Lindeman E, Galama K, Gorter JW. (2010) The inter-rater reliability of Goal Attainment Scaling in Pediatric Rehabilitation Practice. *Arch Phys Med Rehabil* in press.
7. Steenbeek D, Ketelaar M, Galama K, Gorter JW. (2007) Goal Attainment Scaling in Paediatric Rehabilitation: a critical review of the literature. *Dev Med Child Neurol* **49**: 550–556.
8. Steenbeek D, Ketelaar M, Galama K, Gorter JW. (2008) Goal Attainment Scaling in paediatric rehabilitation: a report on the clinical training of an interdisciplinary team. *Child Care Health Dev* **34**: 521–529.
9. Hurn J, Kneebone I, Cropley M. (2006) Goal setting as an outcome measure: A systematic review. *Clin Rehabil* **20**: 756–772.
10. Turner-Stokes L. (2009) Goal attainment scaling (GAS) in rehabilitation: a practical guide. *Clin Rehabil* **23**: 362–370.
11. Khan F, Pallant JF, Turner-Stokes L. (2008) Use of goal attainment scaling in inpatient rehabilitation for persons with multiple sclerosis. *Arch Phys Med Rehabil* **89**: 652–659.
12. Bouwens SF, van Heugten CM, Verhey FR. (2008) Review of goal attainment scaling as a useful outcome measure in psychogeriatric patients with cognitive disorders. *Dement Geriatr Cogn Disord* **26**: 528–540.
13. Sakzewski L, Boyd R, Ziviani J. (2007) Clinimetric properties of participation measures for 5- to 13-year-old children with cerebral palsy: a systematic review. *Dev Med Child Neurol* **49**: 232–240.
14. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. (1997) Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol* **39**: 214–223.
15. Eliasson AC, Krumlinde-Sundholm L, Rösblad B, Beckung E, Arner M, Öhrvall AM, Rosenbaum P. (2006) The Manual Ability Classification System (MACS) for children with cerebral palsy: scale development and evidence of validity and reliability. *Dev Med Child Neurol* **48**: 549–554.

16. Kiresuk T, Sherman R. (1968) Goal attainment scaling: a general method of evaluating comprehensive community mental health programs. *Community Ment Health J* **4**: 443–453.

17. Bovend'eerdt TJ, Botell RE, Wade DT. (2009) Writing SMART rehabilitation goals and achieving goal attainment scaling: a practical guide. *Clin Rehabil* **23**: 352–361.

18. Tennant A. (2007) Goal attainment scaling: Current methodological challenges. *Disabil Rehabil* **29**: 1583–1588.

19. Haley SM, Coster WJ, Faas RM. (1991) A content validity study of the Pediatric Evaluation of Disability Inventory. *Pediatr Phys Ther* **3**: 177–184.

20. Haley SM, Coster WJ, Ludlow LH, Haltiwanger JT, Andrellos PJ. (1992) *Pediatric Evaluation of Disability Inventory: Development, standardization, and administration manual*. Boston: New England Medical Centre Inc and PEDI Research Group ed.

21. Wassenberg-Severijnen JE, Custers JW, Hox JJ, Vermeer A, Helders PJ. (2003) Reliability of the Dutch Pediatric Evaluation of Disability Inventory (PEDI). *Clin Rehabil* **17**: 457–462.

22. Custers JW, van der Net J, Hoijtink H, Wassenberg-Severijnen JE, Vermeer A, Helders PJ. (2002) Discriminative validity of the Dutch Pediatric Evaluation of Disability Inventory. *Arch Phys Med Rehabil* **83**: 1437–1441.

23. Russell DJ, Avery LM, Rosenbaum PL, Raina PS, Walter SD, Palisano RJ. (2000) Improved scaling of the gross motor function measure for children with cerebral palsy: evidence of reliability and validity. *Phys Ther* **80**: 873–885.

24. Russell DJ, Rosenbaum PL, Avery LM, Lane M. (2002) *Gross Motor Function Measure (GMFM-66 & GMFM-88). User's manual*. Mac Keith Press London.

25. Ketelaar M, Vermeer A, Helders PJ. (1998) Functional motor abilities of children with cerebral palsy: a systematic literature review of assessment measures. *Clin Rehabil* **12**: 369–380.

26. Vos-Vromans DCWM, Ketelaar M, Gorter JW. (2005) Responsiveness of evaluative measures for children with cerebral palsy: The Gross Motor Function Measure and the Pediatric Evaluation of Disability Inventory. *Disabil Rehabil* **27**: 1245–1252.

27. Cieza A, Geyh S, Chatterji S, Kostanjsek N, Ustun B, Stucki G. (2005) ICF linking rules: an update based on lessons learned. *J Rehabil Med* **37**: 212–218.

28. McDougall J, Wright V. (2009) The ICF-CY and Goal Attainment Scaling: Benefits of their combined use for pediatric practice. *Disabil Rehabil* **26**: 1–11.

29. Holsbeeke L, Ketelaar M, Schoemaker MM, Gorter JW. (2009) Capacity, capability, and performance: different constructs or three of a kind? *Arch Phys Med Rehabil* **90**: 849–855.

30. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. (2001) Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res* **10**: 571–578.

31. Palisano RJ. (1993) Validity of goal attainment scaling in infants with motor delays. *Phys Ther* **73**: 651–660.

32. Cusick A, McIntyre S, Novak I, Lannin N, Lowe K. (2006) A comparison of goal attainment scaling and the Canadian Occupational Performance Measure for paediatric rehabilitation research. *Pediatr Rehabil* **9**: 149–157.

33. Ostensjø S, Oien I, Fallang B. (2008) 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* **11**: 252–259.

34. Schlosser RW. (2004) Goal attainment scaling as a clinical measurement technique in communication disorders: a critical review. *J Commun Disord* **37**: 217–239.

35. Brock K, Black S, Cotton S, Kennedy G, Wilson S, Sutton E. (2008) Goal achievement in the six months after inpatient rehabilitation for stroke. *Disabil Rehabil* **26**: 1–7.

7

General discussion

INTRODUCTION

The main aim of this thesis was to contribute to the knowledge of the application and clinimetric issues of Goal Attainment Scaling (GAS) when used in interdisciplinary rehabilitation practice for children with Cerebral Palsy (CP). My interest in individualized outcome measurement at the child's activity and participation level evolved from my first experience in rehabilitation medicine as a physician (1996). I strongly felt that it was necessary to evaluate different team interventions individually to improve quality of rehabilitation services, using an approach that was uniform. This led to the development of the first study of this thesis in 1999, measuring the effect of treatment with Botulinum Toxin A (BTX-A) (Chapter 2). For this study, the research group assumed that relevant change in children's activities related to supplementary interventions to treat spasticity might be missed by only using common standardized measures such as the Pediatric Evaluation of Disability Inventory (PEDI) and the Gross Motor Function Measure (GMFM). We concluded that an individualized tool as primary outcome measure should be used and found GAS the most suitable method as introduced by Palisano and colleagues[1, 2] in paediatric rehabilitation. The application of GAS however was pragmatic, because research evidence supporting its reliability and validity when used for children was limited. During the intervention study on the effect of BTX-A treatment, we really discovered that GAS helped demonstrate clinically relevant change over time in individuals. We also concluded that the use and clinimetric properties of GAS for patients with CP needed further investigation. This was the foundation for the projects that followed and a series of specific research questions that needed to be addressed.

A critical review of the literature confirmed the promising qualities of GAS in paediatric rehabilitation, and that clinimetric properties should be examined in more detail (Chapter 3). The reliability, the added value and responsiveness of GAS should be specifically explored for paediatric rehabilitation. We decided to introduce GAS at Rehabilitation Centre Breda, The Netherlands, a middle-sized rehabilitation centre and to perform the required studies with the multidisciplinary paediatric rehabilitation team. We took advantage of the enthusiasm of the team members involved and the centre's policy of stimulating and supporting scientific endeavours. The projects were performed in cooperation with the Centre of Excellence for Rehabilitation Medicine Utrecht, a collaboration of Rehabilitation Centre De Hoogstraat and University Medical Centre Utrecht (UMCU), The Netherlands. The projects in Breda took place from 2004 to 2010 and

consisted of clinical training of a team of professionals (Chapter 4), a study on the inter-rater reliability of GAS in rehabilitation practice (Chapter 5) and a study exploring the added value of GAS for evaluation of outcome in rehabilitation of children with CP (Chapter 6).

We introduced content criteria for GAS and discovered during the training program that these criteria improved the reproducibility of GAS and were useful for rehabilitation professionals (Chapter 4). Most therapists and parents regarded GAS as a suitable tool for improving the quality of rehabilitation treatment.

The inter-rater reliability of GAS in the trained team was shown to be good (Chapter 5). Scale construction by the child's own therapist as opposed to an independent rater had no negative impact on the inter-rater reliability of scales. In contrast, when independent raters use GAS as was more or less prescribed by the original GAS method, extra care might be necessary. Finally, we showed that GAS has an important value when used in addition to the PEDI and GMFM (Chapter 6).

This thesis supports the use of GAS to detect changes relevant for the child and family in paediatric rehabilitation. However, several issues should be considered carefully before GAS can be used properly. I hope this thesis arouses the curiosity of professionals regarding GAS and that it provides a dynamic basis for practical use and future research. I will begin by summarizing the positive attributes and discussing the issues of GAS point for point based on the lessons learned during the projects that lead to this dissertation.

POSITIVE ATTRIBUTES OF THE USE OF GAS

The primary strength of GAS is the ability to evaluate individual longitudinal change (Chapter 1). In summary we have seen the following advantages of using GAS, in regard to the quality of care (1, 2), reproducibility (3), study design (4) and clinimetric properties (5, 6).

1. Grading Goal Attainment

In rehabilitation practice, the process of setting specific, measurable, acceptable, relevant and time-related (SMART) goals is an intricate part of what clinicians do. In order to determine whether a patient improved after treatment, professionals often rely on approaches like using the percentage of objectives attained, while patients usually discuss the extent to which a

goal was attained. GAS classifies levels of goal attainment and reflects the patients' experience properly.[3] In contrast to the standardized measures, a property of GAS at group level as well as individual level, is that the score by definition describes the relevance of the change. This contributes to parents' satisfaction when using GAS with their children. However, we also showed in Chapter 4 that special attention is needed for the participation of parents in the development of GAS scales (see *Implications for practice*).

2. Facilitating the quality of rehabilitation care

Several authors hypothesized that the GAS method itself may have a positive impact on goal attainment. Our evaluation of the training program (Chapter 4) supports previous recommendations from members of the Network for Childhood Disability Research in The Netherlands (*NetChild*). In her recent doctoral study, Siebes[4] recommended the use of GAS to improve transparency and coordination with parents. Kiresuk[5] has already postulated that clearly specified goals can mobilize a team to pursue relevant and feasible outcomes more coherently. GAS can easily be used with the Rehabilitation Activities Profile for Children (Children's RAP, Chapter 1) as it is useful in the reporting process and contributes to transparent team communication. Despite a decennium of experience with the Children's RAP in The Netherlands, turning needs into therapy goals and evaluation of individual goal attainment is still challenging in daily practice.[6, 7] In my opinion, the use of GAS offers tremendous opportunities to facilitate the quality of rehabilitation care.

3. Adaptability to the Activity and Participation level of the ICF

GAS can be adapted to virtually any of the domains and levels specified by the ICF. Measuring change over time at the activities and participation level, is an important advantage and highly relevant in contemporary interdisciplinary rehabilitation care for children with CP.[8] The study by Nijhuis and colleagues[6] reported that the content of needs, problems and rehabilitation goals of children with CP can be accurately described using the ICF-CY. According to McDougall et al.,[9] the combined use of GAS and the ICF-CY presents many benefits in all spheres for the special challenges of the child and his family, as well as for institutional systems or communities. We have also shown that goals in the activity and participation domain can be classified properly when using the ICF, improving the reproducibility of outcome (Chapter 6).

4. Comparability of goal attainment across goals and individuals

GAS allows legitimate comparisons of different goals for an individual and between different individuals. Rehabilitation research typically struggles with the dilemma of the heterogeneity of the populations in general practice. Even if changes in a homogeneous population can be proven, many results are not transferable to practice as the team has to deal with various requests for help. GAS offers a generic opportunity to measure change in heterogeneous populations under a variety of conditions. This thesis shows that the outcome between the disciplines of a team and between multidisciplinary teams can be compared using GAS (Chapter 5). Moreover, a promising focus for the future would be the comparison of the outcome of different departments of a rehabilitation centre or comparing rehabilitation settings. The same is true regarding the comparison of different goals attained per patient. Goals spread over several constructs will vary, necessitating many specific instruments per patient when using standardized measures. GAS in contrast permits generic measuring of a diversity of interests and allows comparability by using only one instrument.

5. Inter-rater reliability

In our setting, the inter-rater reliability of GAS was good (Chapter 5). We discovered that if the construct of measurement and the scoring procedure are determined and instructed properly in a team of professionals, the child's own therapists can score GAS scales reliably. We learned to trust the insight of professionals. Although this was suggested before by Palisano et al.,[1, 2] it took 17 years to verify their conclusions. Because our findings concur with the findings of many other authors in different fields of rehabilitation, in my opinion, good inter-rater reliability may be generalized and added to the positive attributes of GAS in paediatric rehabilitation.

6. Measuring progress that would otherwise not have been measured

Defining the major therapy goal individually offers the unique opportunity to measure exactly what one intends to measure. In general, GAS enables the measurement of the outcome of any rehabilitation intervention, including medical, therapeutic, social or educational, or any combination of these. However, GAS may generally improve the responsiveness of outcome measurement when measuring specific medical applications, for example supplementary management of spasticity as outlined in the first study of this thesis (Chapter 2). GAS was used by many other researchers to investigate

spasticity management strategies (Chapter 3). How can one measure the effect of multi-level surgery followed by an intensive rehabilitation program to prevent progression of serious flexion contractures, aiming to retain a child's ability to make transfers without help from wheelchair to toilet et cetera? Attaining this therapy goal obviously justifies the intensive and expensive treatment, as it improves the child's ability to travel and function in different environments. When a patient lift is needed it makes family transportation much more complex. In my opinion, GAS would be the first choice in enabling the measurement of the treatment effects in this example. The use of GAS can also be recommended, to measure the effect of Intrathecal Baclofen (ITB) treatment in children with severe spasticity (Chapter 1). Hoving recommended in her dissertation on ITB, that outcome measures include highly individualized measures because therapy goals for ITB vary greatly.[10] Problems with measurement during previous trials on spasticity management were solved too often by measuring outcome at impairment level, for example spasticity or range of motion,[11] while spasticity treatment in particular should be evaluated at the ICF activities and participation level. GAS provides a gentle solution for this measurement problem. Further examples of the added value have been discussed in Chapter 6.

LESSONS LEARNED ABOUT GAS IN PAEDIATRIC REHABILITATION

We discovered that GAS in paediatric rehabilitation practice has advantages and disadvantages. This thesis contributes to answers on some clinimetric questions about GAS, and provides a basis for future research and introduction in other paediatric rehabilitation teams. However, it also highlights the issues on its use that should be followed critically and evaluated as part of an ongoing learning process. In general, depending on the research question and the purpose of the studies combining GAS with standardized measures is recommended. When using GAS in trials and practice, I recommend that users consider the following matters on validity (1, 2), reliability (3, 4, 5), data analysis (6, 7), and other remaining questions (8, 9, 10).

1. Content validity of GAS

Content validity addresses the question of whether the content of the GAS scales covers the most important aspects or elements of the attribute of interest being

measured.[12] Inherent to the use of GAS is the crucial question if it measures outcome or only describes the therapist's prediction of future functioning. The content of the GAS scales describes the professional insight (Chapter 5). Critics of GAS believe that the scales mainly describe the future prediction, often based on the perception that the development of content of the scales is too subjective. In contrast, this thesis supports the evaluative properties of GAS, based on confidence in the objectivity of the professional insight. GAS measures the quantitative distance to the professional's expectation of outcome. Many influences determine the content of the scales, including cultural issues, intensity and length of the therapy period, patient's motivation, knowledge of prognosis, et cetera. Although we strive for uniform application of the GAS method, the differences around the world make results difficult to compare. Moreover, although our data on the issue of therapist bias were reassuring, the phenomenon of therapist bias has to be considered and reconsidered in each new setting when using GAS. Our trial setting in daily practice had no other interest than children's and family's well-being and measuring the properties of GAS. Clinicians are increasingly inclined to introduce measurement in clinical practice in order to evaluate individual treatment results.

However, service providers want to document effectiveness for reasons other than the family's well-being alone. Healthcare insurance companies for example increasingly require proof of the effects of interventions before allowing them. Although I certainly support the accountability of physicians for medical interventions, in my opinion and despite our reassuring findings on therapist bias, GAS is not suitable for monitoring achievements for external interests. If this were the case professionals may be tempted to set their goals at a less challenging level in order to show a better outcome.

2. Criterion and construct validity of GAS

If scales of the same or similar attributes were available, in developing GAS, it would be apparent to administer GAS and the other instruments to the same sample. This approach is described by several terms, including *convergent* validation, *criterion* validation, and *concurrent* validation.[12] In our study, *criterion* validity would address the correlation of the conclusions drawn from GAS with some other measures evaluating activity and participation in children with CP, ideally a 'gold standard'. However, a gold standard for functioning of children with CP does not exist. In Chapter 6, we suggested that there is low criterion validity when comparing GAS with the PEDI and GMFM, supporting the value of combining the measures.

When no other measure exists to compare the new measure with, how

can one acquire data to show that GAS is indeed measuring what is intended? The solution may be found with a broad set of approaches called *construct* validity.[12] This begins by linking the attribute that must be measured to some other attribute by a hypothesis or construct. This hypothetical construct should then be tested by applying GAS to the appropriate samples. If the expected relationship is found, the hypothesis and measure are sound, supporting construct validity. No data are available on the construct validity of GAS and in my opinion research on the construct validity of individualized measures would not yield applicable information.

The construct measured might have influenced the majority of GAS scores in my thesis, as is illustrated by diverse explanations of function within the activity dimension. In Breda the scales were scored as *can-do* instead of *does-do* scales, in order to reflect the professional's subjective judgement and to prevent unreliability due to children's whims. As a consequence the GAS score on the activity of controlling the computer of an Adremo wheelchair (Chapter 3, **Table 3.1b**), could vary for the same child on the same day between –3 when measuring performance and +2 when measuring capacity, depending on environmental variables such as the accuracy of software or personal variables such as actual physical function or motivation as this will fluctuate during the day in children with CP at GMFCS level V. This is in line with findings by Holsbeeke et al.[13] who showed that motor performance levels are only partly reflected by the motor capacity and motor capability levels in young children with CP. In retrospect the terminology in this thesis might be confusing. For future application of the GAS in research and in clinical practice I recommend following the definitions of Holsbeeke et al. Capacity, capability and performance are three constructs which can be measured depending on the purpose and should be described in detail when publishing the results. In this definition a person's capacity is what a person can do in a standardized, controlled environment, a person's capability is what a person can do in his/her daily environment, and a person's performance is what a person actually does do in his/her daily environment.

3. Content reliability of GAS

When relying on professional insight during GAS scale construction, the question of whether different professionals would develop similar scales rises. Critics of GAS have rightfully noted that the reliability data primarily addresses the accuracy of deriving GAS scores, but not the reliability of the process of constructing the scales. When considering this thesis, it is important to realize that we did not explore the content reliability, implicating the need for further

research. Our reliability study (Chapter 5) was restricted to the inter-rater reliability of pre-determined scales, addressing only the scores.

However, with our dataset, as an indirect measure of content reliability, all scores on scales constructed by the child's therapists and constructed by independent raters (Chapter 5, **Tables 5.4 and 5.5**) were compared. This resulted in a Wilcoxon signed ranks *p*-value of 0.035 (*Z* -2.11), suggesting that, despite the different textual contents of scales for the same child, the scores are related properly. However, in my opinion, this finding is too abstract to generalize.

4. The criteria of GAS scales

We determined criteria for GAS scales to improve the reproducibility of the studies in Breda. Although the criteria were useful for all participants (Chapter 4), some need to be discussed further.

We recommend users in our field follow 3 general criteria. First, the scales should be based on the main request for help, describing meaningful and relevant goals for the children and their families. Second, all goals should be at the ICF activity and participation level. Third, all levels should be Specific, Measurable, Acceptable, Realistic and Time-specific (SMART), according to customary goal setting in interdisciplinary rehabilitation care. SMART goal setting might be a time-consuming component of constructing GAS scales, but this competence should also be required without the use of GAS.

Other criteria need some discussion. We agreed that scales should be ordinal, with intervals between the levels as equal as possible. Therefore only a single dimension of change should be reflected in a GAS scale. This criterion required the most feedback during the training period (Chapter 4), as we discovered that defining six realistic, distinct levels of outcome with only one variable, and no gaps or overlap between levels, was challenging. This criterion suggests that we approximated interval data. Depending on the aim of the GAS use, it would be possible to leave this to the therapist's insight. Another criterion was that each level should be described as clearly as possible and be understandable for a colleague with the same profession. This was particularly important for our inter-rater reliability study, as the therapists scored each other's scales. A third debatable criterion is that in our studies all scales should specify a behaviour that is observable within 10 minutes. This improved the feasibility, but should be reconsidered with each new application. Finally, as described, we scored the scales as *can-do* instead of *does-do* and recommend carefully considering and describing this aspect, as different constructs can be measured by GAS for different purposes.

5. Rater selection

In Chapter 5, the results suggested that scale construction by the child's own therapist as opposed to an independent rater has a positive influence on the inter-rater reliability of the scales. In concurrence with the issues described above (content validity and content reliability), the generalisability of this finding is assumed to be dependent on the interest in the outcome of a study. It is interesting that in the other fields there has been an almost unanimous agreement among promoters and critics of GAS to only use independent raters to follow-up.[5, 14] However, in our field, even if an independent assessor were essential to reduce bias in evaluating outcome, the involvement of the child's own therapist (and the child or its parents) is still necessary when designing goals for therapy. I strongly believe that one of the positive attributes and added value of GAS is the result of the interaction between child and therapist. The child's therapist has more knowledge about the needs of the child and family than an independent therapist. Moreover, collaborative goal setting and construction of GAS scales by the parents, child and therapist promotes family-centered services. On the other hand, a broad introduction of GAS as outcome measure in a rehabilitation centre will be complicated by the implications of training and by how time consuming it is to start using GAS (therapists required 45 minutes per GAS procedure, Chapter 4). It is probably more practicable to train a few GAS experts to carrying out GAS construction and scoring for patients treated by other therapists. However, we still recommend cooperating with the child's therapists during construction and scoring of GAS scales. This introduces a dilemma of investing in training more team members versus investing in the time cooperation costs.

6. Computation and data analysis

As outlined in detail in Chapter 3, we strongly recommend that the *T* sum score, 42 years after its introduction be abandoned. Originally GAS was characterized by this sum score formula, typically transforming goal attainment into a standardized measure or *T* score per patient with a mean of 50 and standard deviation of 10 (Chapter 1). In my opinion, in small studies the formula deflects the positive attributes of GAS rather than contributing to the method. It does not fulfil its claim of being a parametric expression of non-parametric information. Little by little, due to the mutual peer-review process, the formula is criticized,[15, 16] but still commonly used.[17] The study of Turner-Stokes and colleagues[16] concerning neurorehabilitation of adults, was unique in its size (n=164) and its presumed normal distribution of outcomes,

providing the legality of parametric statistics. As far as I know, none of the studies that used GAS in paediatric rehabilitation were large enough to show normal distribution of GAS scores and legalize parametric statistics. Moreover, as introduced in Chapter 1, the *T* sum score with several decimal places would give false impression of precision and a false sensitivity to change, because this score compute data as interval data, whereas at best they represent an ordinal scale of values. Calculating with simulated cases, Tennant[15] supported the overestimation of outcome when using the formula: mathematical operations for many patients showed differences between the raw ordinal GAS data and linear-based *T* sum scores greater than the minimum clinically important difference (MCID). The second reason for abandoning the formula is that caution must be exercised in the interpretation of results after weighting goals, although optional in the formula. The weighting process itself is too subjective and threatens the reproducibility of GAS outcome and may be unnecessary according to the study of Turner Stokes et al.[16] as weighting did not have any influence on the outcome in their study. We even found studies weighting goals set in different ICF levels (Chapter 3).

We introduced analysis of raw GAS scores, instead of using the *T* sum formula in order to prevent false sensitivity to progress as GAS data are at best ordinal. As we have seen in Chapter 2, 5 and 6, even representing small amounts and non-normal distribution, different GAS data sets can be compared by using non-parametric statistics. We used the Wilcoxon signed ranks test, to compare ranks of dependent raw GAS data. It is important to realize that non parametric statistics are different from parametric equivalents. The differences are particularly relevant when small changes in small populations occur.

7. Scoring deterioration

Just before this thesis was finished, a study by Turner-Stokes and Williams[17] was published, comparing the method we introduced for scoring deterioration and an alternative 6-point GAS rating system. They were able to manipulate the existing dataset of the original 5-point GAS scales from the large study on neurorehabiliation[16] in order to attain two new datasets of 6-point scales. Version 1 set all baseline scores at '–2' and added '–3' to denote 'worsening', in reference to our work. Version 2 added a '–0.5' score to denote 'partial achievement' for goals starting at '–1'. While median achieved *T* scores were 50.0 for all three methods, version 1 underestimated and version 2 marginally overestimated goal attainment, in comparison to standard goal rating. They recommended the use of their '–0.5' version as it provided the closest match

to the standard rating. The study of Turner-Stokes demonstrates that this is an interesting time for GAS and that sharing our findings with the world helps to develop GAS. However, I don't share Turner-Stokes' opinion regarding this scoring issue. First, the conclusion is based on the similarity with outcome calculated with the *T* sum score that I just abandoned to prevent overestimating outcome. Second, the criticism of Turner on our –3 score version was that it probably underestimates outcome, but in contrast we introduced the -3 score in order to prevent bottoming effects and to gain a lower outcome. I think this is a matter of reality rather that underestimation.

8. Missing side effects

In general practice, GAS scales focus on the most prominent goals in order to represent outcome. Users should recognize that a GAS scale may be scored as +2 even while the patient regrets the treatment as a result of side effects. When side effects are expected, due to spasticity management for example[10] one should consider combining GAS with relevant standardized measures, or questionnaires targeting the potential side effects. Another example is measuring the outcome of wearing ankle-foot braces. Walking can improve while standing or sitting function can decrease, providing a clear advantage of the use of the GMFM-88 in combination with GAS.[18] Professional expertise and patients' opinions should weight individual therapy effects, rather than any outcome measure.

9. Training

Although test developers may demonstrate that a measure is reliable and valid in the context of a study, it does not necessarily mean that the measurement properties will be the same for all users under any condition. Regardless of who does the construction and follow-up, the degree to which the raters are trained influences the reliability and validity of GAS. As for each (new) instrument the critical questions include how much training and practice is necessary to ensure the user's competence in the administration and scoring of the measure. The question is how much training is needed and how do we know when someone is a reliable user? Although Chapter 4 of this thesis discussed our training method on GAS, these questions have not yet been answered.

10. Changing the intervention by introducing an outcome measure

Finally, although facilitation of the quality of rehabilitation care is considered to be a positive attribute of GAS, for research purposes, changing the quality of an intervention by the introduction of an outcome measure could also introduce a bias in outcome. We recommend designing controlled trials using GAS both in the intervention group and the control group, enabling a similar change in both groups.

LIMITATIONS AND POWERFUL ASPECTS OF THIS THESIS

Some powerful aspects are also the limitations of this thesis. Although the study in Chapter 2 was performed in a trial setting, the projects in Chapter 4, 5 and 6 were unique in their description of the use of GAS in the daily work of a paediatric rehabilitation team. We were driven by important and practically-oriented questions about GAS that our own team of professionals had. Almost all team members from the paediatric rehabilitation unit in Breda were involved, including physical therapists, occupational therapists, speech and language therapists, paediatric psychologists and social workers. SMART goal setting has become a routine activity of the interdisciplinary paediatric rehabilitation services, and the use of GAS in team conferences contributed to transparency and quality (Chapter 3). In the setting in Breda, the project succeeded in stimulating the scientific climate, as aimed for from the start.

However, performing research during daily teamwork also resulted in methodological limitations. Exploring the professionals' insight in the team of professionals of which the study leader is a member, may harm its reproducibility. We had rather small number of subjects as a consequence of describing rehabilitation team work in a medium sized centre. We presented large 95% confidence intervals as discussed in regard to the statistical analysis. That is particularly true for change measured in Chapter 2, the reliability in Chapter 5 when subdivided per discipline and the correlations between the instruments in Chapter 6. On the other hand, our findings concur with those of previous authors and our conclusions at group level are conservative.

Another limitation as well as powerful aspect is that it is spread out over time. We started collecting data in De Trappenberg in 2000. In accordance with other publications on our issues during the last decade, we had to change some terminology and insights. An example is that we changed 'sensitivity to change' (Chapter 3) to 'responsiveness' (Chapter 5) and functional ability and activity (Chapter 4) to activity and participation (Chapter 5, 6). On the

other hand, in my opinion, the extended time frame also contributed to the relevance of this thesis. It reflects the increased knowledge resulting from several discussions with the participating teams of therapists, the research group and invitational conferences on goal setting. The peer-review process and participation in and presentations at international conferences provided further inspiration. This thesis on therapy goals required the collaboration of the author with a team of rehabilitation professionals and needed the experience gathered over time.

FUTURE DIRECTIONS FOR RESEARCH: WHERE DO WE GO?

During the past decade, several projects have contributed to the knowledge of GAS. There is increasing interest in the technique. With some reservation due to issues on the validity of GAS and the small number of children examined, we feel there may be a potential to generalize our findings for other diagnostic groups, other rehabilitation teams, other age groups and other countries. The conclusions of previous authors in other fields support this. Future research is however of great importance. Two intervention studies supported by ZonMw in paediatric rehabilitation, started in The Netherlands in 2010, and will use our version of GAS as one of the outcome measures. One is called "Spacebop", a multi-centre randomized study on (cost) effectiveness of combined treatment with multilevel BTX-A and intensive functional physiotherapy with children with CP. The other is called "Learn to Move", a study on the efficacy of an activity stimulation program on performance of mobility in children with CP. This study will be performed with different age groups. The study with 2 and 3 year old children plans to use GAS.

Future research in rehabilitation practice should include the examination of the content reliability of GAS scales, for example having a team of experts rate the overall value of therapists' goals. The present thesis showed that the GAS scales constructed by the patient's own therapists in particular should be used for further validity research. In my perception, rehabilitation goals differ from therapist to therapist, from rehabilitation centre to rehabilitation centre and from country to country, as do GAS scales. Research on content reliability and validity should be a dynamic learning process and I recommend as much exchange of lessons learned as possible.

Another future direction is further exploration of the dynamic interaction between GAS and the Canadian Occupational Performance

Measure (COPM), as previously recommended by Cusick et al.[19] who reported that GAS and the COPM appeared to measure different constructs. With the COPM parent attention is drawn to a range of specific activity areas (productivity, leisure, and self-care) in addition to the child's performance and parent satisfaction. When using GAS, there is no pre-determined set of activities. When used in practice, the COPM might help to clarify the request for help as basis for constructing GAS scales. Further research is required to answer the question of whether one or both individualized measures should be selected, determined by the aim and logistic factors of a study. Although I feel some reservation in recommending research on the construct validity of GAS, further comparison of the constructs of GAS and the COPM following Cusick's work[19] would be an interesting future focus.

Comparing the outcome of departments of a rehabilitation centre or comparing different rehabilitation settings would be another promising future focus as GAS is unique in its ability to measure outcome in different populations uniformly. For this approach, although controversial, I also recommend using therapists' GAS scales rather than scales constructed and rated by independent raters. One can imagine several variables for research, including different working styles of teams and different aspects of programs.

IMPLICATIONS FOR PRACTICE

The use of GAS in the rehabilitation care is motivated by the added value of GAS, the possibility to measure group effects in heterogeneous groups and the satisfaction in working with GAS among patients, families and the teams of professionals. GAS might be ready for introduction in other rehabilitation teams and for other diagnostic groups to evaluate change and compare aspects of interventions. However, this thesis is only one example of the application of GAS. Our application can be used to discuss the many important issues on uniformity and stimulate the users' curiosity further, but I am not presenting this application as a truism.

Recently a local invitational conference stimulated the further development of GAS in different rehabilitation settings for children and adults. This should include working toward a national and preferably international consensus on the scaling, criteria and timing of GAS scale construction in rehabilitation programs. Next, a manual on GAS in rehabilitation medicine and training programs should be developed. Moreover, we recommend

the development of an implementation program, knowledge brokering for example, to ensure that the GAS method is embedded in paediatric rehabilitation services. In my opinion, with the current knowledge, further training procedures can be promoted in paediatric rehabilitation practice. I recommend broad exchange of any new information on how much training is necessary. Special attention is necessary for the participation of parents and, if possible children and youth themselves, in the development of scales for children. Parents' and children's satisfaction should stimulate their involvement and their enthusiasm should be used to improve participation. Families and children with disabilities know best what their needs are. Their involvement in goal setting and individual outcome evaluation is crucial in paediatric rehabilitation.

GAS is useful for families and professionals and measures outcome reliably when the proper conditions are provided.

REFERENCES

1. Palisano RJ, Haley SM, Brown DA. (1992) Goal attainment scaling as a measure of change in infants with motor delays. *Phys Ther* **72**: 432–437.
2. Palisano RJ. (1993) Validity of goal attainment scaling in infants with motor delays. *Phys Ther* **73**: 651–660.
3. Schlosser RW. (2004) Goal attainment scaling as a clinical measurement technique in communication disorders: a critical review. *J Commun Disord* **37**: 217–239.
4. Siebes, RC, Ketelaar, M, Gorter, JW, Wijnroks, L, De Blécourt, AC, Reinders-Messelink, HA, Van Schie, PE & Vermeer, A. (2007) Transparency and tuning of rehabilitation care for children with cerebral palsy: A multiple case study in five children with complex needs. *Dev Neurorehabil* **10**: 193–204.
5. Kiresuk T, Smith A, Cardillo JE (eds). (1994) *Goal Attainment Scaling: Applications, Theory and Measurement*. London, Erlbaum.
6. Nijhuis BJ, Reinders-Messelink HA, de Blécourt AC, Ties JG, Boonstra AM, Groothoff JW, Nakken H, Postema K. (2008) Needs, problems and rehabilitation goals of young children with cerebral palsy as formulated in the rehabilitation activities profile for children. *J Rehabil Med* **40**: 347–354.
7. Nijhuis BJ, Reinders-Messelink HA, de Blécourt AC, Boonstra AM, Calamé EH, Groothoff JW, Nakken H, Postema K. (2008) Goal setting in Dutch paediatric rehabilitation. Are the needs and principal problems of children with cerebral palsy integrated into their rehabilitation goals? *Clin Rehabil* **22**: 348–363.
8. Rosenbaum P, Stewart D. (2004) The World Health Organization International Classification of Functioning, Disability and Health: A model to guide clinical thinking, practice and research in the field of Cerebral Palsy. *Semin Pediatr Neurol* **11**: 5–10.

9. McDougall J, Wright V. (2009) The ICF-CY and Goal Attainment Scaling: Benefits of their combined use for pediatric practice. *Disabil Rehabil* **26**: 1–11.

10. Hoving MA, van Raak EP, Spincemaille GH, van Kranen-Mastenbroek VH, van Kleef M, Gorter JW, Vles JS. (2009) Safety and one-year efficacy of intrathecal baclofen therapy in children with intractable spastic cerebral palsy. *Eur J Paediatr Neurol* **13**: 247–256.

11. Baird MW, Vargus-Adams J. (2009) Outcome Measures Used in Studies of Botulinum Toxin in Childhood Cerebral Palsy: A Systematic Review. *J Child Neurol* **6** [Epub ahead of print].

12. Streiner DL, Norman GR. (2008) *Health Measurement Scales. A practical guide to their development and use*. Fourth Edition. Oxford University Press.

13. Holsbeeke L, Ketelaar M, Schoemaker MM, Gorter JW (2009). Capacity, capability, and performance: different constructs or three of a kind? *Arch Phys Med Rehabil* **90**: 849–855.

14. Cytrynbaum S, Ginath Y, Birdwell J, Brandt L. (1979) Goal attainment scaling: a critical Review. *Eval Q* **3**: 5–40.

15. Tennant A. (2007) Goal attainment scaling: Current methodological challenges. *Disability and Rehabil* **29**: 1583–1588.

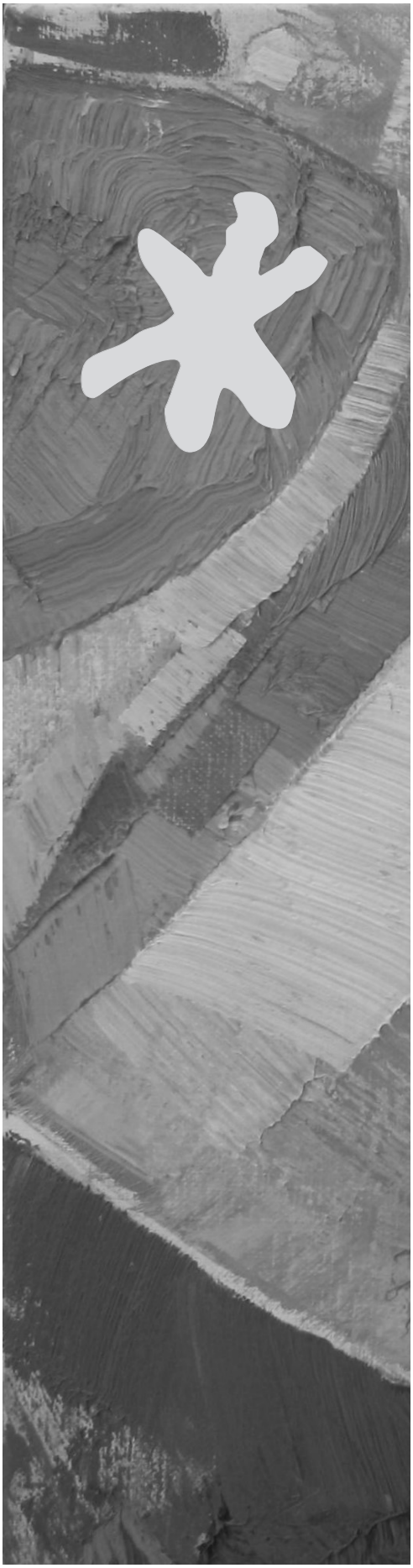
16. Turner-Stokes L, Williams H, Johnson J. (2009) Goal attainment scaling does it provide added value as a person-centred measure for evaluation of outcome in neurorehabilitation following acquired brain injury? *J Rehabil Med* **41**: 528–535.

17. Turner-Stokes L, Williams H. (2010) Goal attainment scaling: a direct comparison of alternative rating methods. *Clin Rehabil* **24**: 66–73.

18. Russell DJ, Gorter JW. (2005) Assessing functional differences in gross motor skills in children with cerebral palsy who use an ambulatory aid or orthoses: can the GMFM-88 help? *Dev Med Child Neurol* **47**: 462–467.

19. Cusick A, McIntyre S, Novak I, Lannin N, Lowe K. (2006) A comparison of goal attainment scaling and the Canadian Occupational Performance Measure for paediatric rehabilitation research. *Pediatr Rehabil* **9**: 149–157.

Summary



The main aim of this thesis was to investigate the use of a 6-point Goal Attainment Scaling (GAS) system to measure change over time in interdisciplinary rehabilitation practice for children with Cerebral Palsy (CP). This thesis begins (**Chapter 1**) with a brief introduction of research in childhood disability and more specifically, research in Dutch paediatric rehabilitation practice. The objectives of measures and measurement issues in general are introduced. GAS is a generic individualized evaluative criterion-referenced instrument. It can be used for measurement of changes in individual patients and in groups of patients, generating ordinal data. The definition and classification methods of CP, and common treatment strategies are presented in the General Introduction. The preeminent outcome in rehabilitation of children with CP concerns a child's activities and participation. The Children and Youth version of the International Classification of Functioning, Disability and Health (ICF-CY) is highlighted and discussed. It provides a useful framework for terminology, rehabilitation care policy and individual treatment goals.

Chapter 2 presents a study to measure the effect of Botulinum Toxin A treatment, performed using GAS. A single-blind randomized multiple baseline/treatment phase study across subjects was used for 11 children with CP. Each goal was recorded weekly with a standard video for a period of 14 weeks. Rating of pre-determined GAS scales was performed blinded. Nine of the 11 subjects showed improvement in 18 out of 33 goals. Seven subjects showed clinically relevant improvement in 11 goals. Testing the difference between baseline and treatment phase measurements for all GAS scores (n=33) resulted in improvement. An improvement was also found at subject level. The change in GAS score was related to the moment of treatment with BTX-A. We concluded that GAS could demonstrate clinically relevant improvement in individual rehabilitation goals, confirming the positive expectations of GAS. One of the key messages was the necessity for future research on the individual psychometric properties of Goal Attainment Scaling in patients with Cerebral Palsy. This study provided the foundation for the projects that followed.

A critical review of the literature about the content reliability, inter-rater reliability, validity, sensitivity to change of GAS and its use in paediatric rehabilitation research is presented in **Chapter 3**. Nine relevant articles reported promising qualities of GAS in paediatric rehabilitation, supporting the need for its further development. We concluded however, that current knowledge about its reliability, when used with children, was insufficient at that time. Moreover, its added value and responsiveness should be explored more specifically.

These conclusions resulted in the next project which was performed on the clinimetric properties of GAS. We started by developing a training program in the children's unit of the rehabilitation centre in Breda (**Chapter 4**). The training consisted of three 2-hour general discussion sessions. In addition I provided the participating physical, occupational, speech-language therapists, paediatric psychologists and social workers with feedback until the GAS scales met the predetermined criteria of ordinality, described specific, measurable, acceptable, realistic abilities and activities in a single dimension, used the 'can-do' principle and could be scored within 10 minutes. This resulted in 115 GAS scales constructed and scored by the professionals. Therapists and parents were asked to complete a questionnaire to express their opinions. The development of a GAS scale remained a time-consuming procedure with an average of 45 minutes per scale. The content-criteria of GAS were found to be useful by all participants. Most therapists and parents regarded GAS as a suitable tool to improve the quality of rehabilitation treatment.

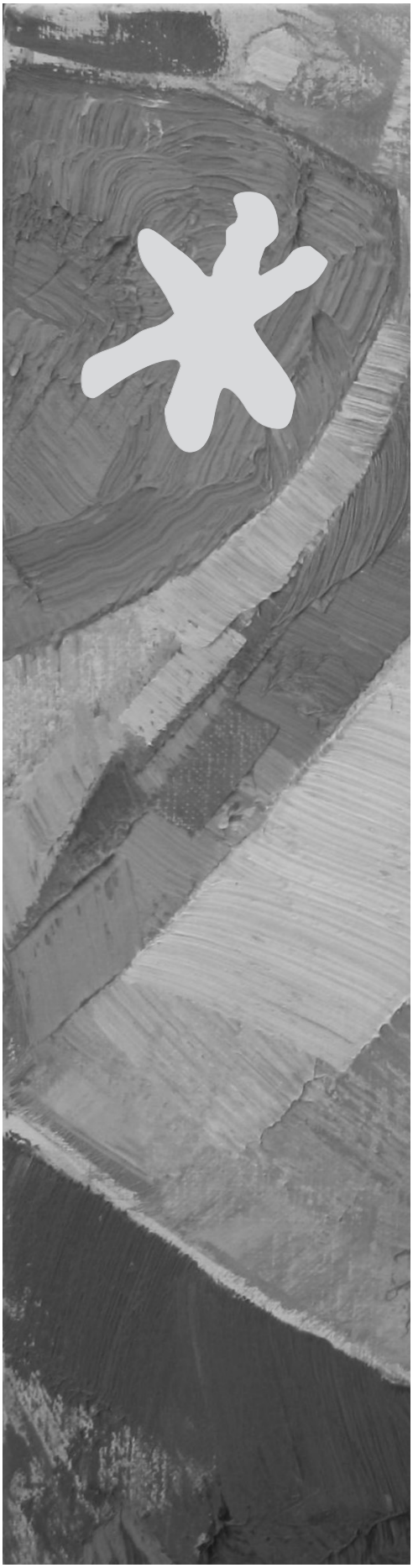
The first issue we investigated was the inter-rater reliability. **Chapter 5** describes a reliability study in which 23 children with CP and 20 therapists of 3 disciplines were enrolled. The child's own therapists, and for each therapist an independent therapist from the same discipline, constructed 64 GAS scales each. They scored both their own scale and their partner's scale resulting in 128 scores. The scales constructed by the children's therapists had an inter-rater reliability of 0.82 (Cohen's linear weighted kappa). The inter-rater reliability for scales constructed by the independent raters was 0.64. The major question that remained unanswered in literature was answered by this study: the inter-rater reliability of GAS was good, if professionals were trained. A key message was that these results also suggested that scale construction by the child's own therapist as opposed to an independent rater has a positive influence on the inter-rater reliability of the scales. This finding is reassuring regarding the phenomenon of "therapist bias".

The content of GAS scales was compared to the PEDI and GMFM-66 using the ICF-CY (**Chapter 6**). Twenty percent of goals were not covered by the PEDI or GMFM-66. Complete goal attainment was compared to changes on the standardized measures. Thirty nine out of 64 GAS scales were scored as complete goal attainment and of these 16 individual PEDI scores did not show change on the related scale. A key conclusion was that the instruments were complementary in construct and individual responsiveness. A substantial number of achieved goals could have been missed if only generic measures were used, even if the ICF-item matched

the items of the standardized measures. Moreover, by measuring complete goal attainment alone, smaller changes could have been missed. This final study of the thesis contributes to the knowledge of the added value of GAS and contemplates the questions regarding the responsiveness of GAS, as first discussed in **Chapter 2**.

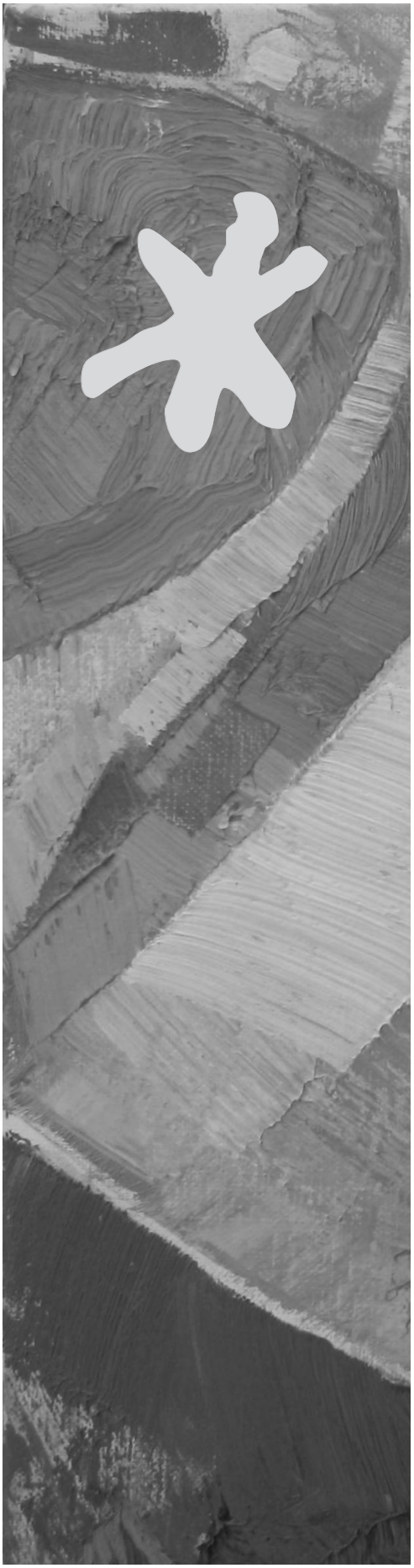
Chapter 7 highlights the lessons learned from this thesis. Positive attributes of GAS include quantifying the relevance of the change measured, facilitating the quality of rehabilitation care, the adaptability to the ICF-CY, the comparability of goal attainment across goals and individuals, the inter-rater reliability and the ability to measure progress that would otherwise not have been measured. Issues regarding GAS include reservations about the content validity related to the professional insight, the low construct and criterion validity and questions about the content reliability of GAS. Computation and data analysis should be performed preferably without the use of Kiresuk's *T* sum formula. Future directions for research and implications for practice are discussed.

List of abbreviations



Botox®	Botulinum toxin type A
BTX-A	Botulinum toxin type A
CanChild	Centre for childhood disability research in Canada
Children's RAP	Rehabilitation Activities Profile for Children
COPM	Canadian Occupational Performance Measure
CP	Cerebral palsy
CI	Confidence interval
FSS	Functional skills scale (of the PEDI)
GAS	Goal Attainment Scaling
GMFCS	Gross Motor Function Classification System
GMFM-66	66-item version of the Gross Motor Function Measure
GMFM-88	88-item version of the Gross Motor Function Measure
ICF	International Classification of Functioning, Disability and Health
ICF-CY	International Classification of Functioning, Disability and Health, Children and Youth version
MACS	Manual Ability Classification System
MAS	Modified Ashworth Scale
MCID	minimum clinically important difference
NetChild	Network for Childhood Disability Research in The Netherlands
OT	Occupational therapy/therapist
PEDI	Pediatric Evaluation of Disability Inventory
PEDI-NL	Dutch version of the Pediatric Evaluation of Disability Inventory
PERRIN	Pediatric Rehabilitation Research in The Netherlands
PT	Physical therapy/therapist
QoL	Quality of life
SD	Standard deviation
SDC	Smallest detectable change
SMART	Specific, measurable, achievable, resource sensitive and timed
SPSS	Statistical package for the social sciences
ST	Speech therapy/therapist
WHO	World Health Organization

About the author



Curriculum vitae

Duco Steenbeek was born in Den Dolder, The Netherlands, on June 30, 1967. After highschool, he studied physical therapy in Utrecht, The Netherlands, graduating in 1989. In 1987, he commenced his study of medicine at the Utrecht University, graduating as a physician in 1995.

After his graduation the author worked as a physician at the Wilhelmina Children's Hospital in Utrecht (Dept. of Paediatrics), the Jan van Breemen Institute in Amsterdam (Dept. of Rehabilitation Medicine), and the Academic Medical Center Amsterdam (Depts. of Rehabilitation Medicine and Neurology), The Netherlands.

Between 1998 and 2002 the author fulfilled his Specialty training in Physical Medicine and Rehabilitation at the Medical Centre Alkmaar, Rehabilitation Centre De Trappenberg, Huizen, and Rehabilitation Centre Amsterdam, The Netherlands. During his residency, he conducted studies on the effects of Botulinum Toxin treatment on functional abilities in children with Cerebral Palsy.

Since 2002, the author is affiliated with the Rehabilitation Centre Breda, The Netherlands as physician in Paediatric Rehabilitation Medicine. Between 2002 and 2010 he worked on his dissertation, studying Goal Attainment Scaling in paediatric rehabilitation practice. This thesis summarizes the results of this study.

Co-authors

Eline Lindeman, MD, PhD
Professor in Rehabilitation Medicine, Rehabilitation Centre De Hoogstraat, Utrecht, The Netherlands and University Medical Centre, Department of Rehabilitation, Nursing science and Sports, Rudolf Magnus Institute of Neuroscience, Utrecht, The Netherlands.

Jan Willem Gorter, MD, PhD
Associate Professor, *CanChild* Centre for Childhood Disability Research, McMaster University, Hamilton, Canada; Partner of *NetChild*, Network for Childhood Disability Research in The Netherlands.

Marjolijn Ketelaar, PhD
Associate Professor, Rehabilitation Centre De Hoogstraat, Utrecht, The Netherlands and University Medical Centre, Department of Rehabilitation, Nursing science and Sports, Rudolf Magnus Institute of Neuroscience, Utrecht, The Netherlands; Partner of *NetChild*, Network for Childhood Disability Research in The Netherlands.

Krys Galama, OT
Occupational therapist, Rehabilitation Centre Breda, The Netherlands.

Anke Meester-Delver, MD, PhD
Physician in Paediatric Rehabilitation Medicine, Academic Medical Center University Hospital, Amsterdam, The Netherlands.

Jules G. Becher, MD, PhD
Professor in Paediatric Rehabilitation Medicine, VU University Medical Centre, Amsterdam, The Netherlands.

Gustaaf J. Lankhorst, MD, PhD, FRCP
Professor emeritus in Rehabilitation Medicine, VU University Medical Centre, Amsterdam, The Netherlands.

Presentations regarding this thesis

Rameckers E, Steenbeek D, Dekkers K, Galama K, Ketelaar M. (2010) From needs to goals to steps for task specific therapy for children with physical disabilities. Instructional course. European Academy of Childhood Disability (EACD) 22nd annual multidisciplinary scientific meeting, Brussels, Belgium, May 26-29, 2010.

Steenbeek D, Keen I, van den Heuvel D, Braaksma E. (2010) Goal Attainment Scaling in de kinderrevalidatie: een waardevolle uitkomstmaat [Goal Attainment Scaling in paediatric rehabilitation, a useful outcome measure]. Free paper and workshop. Keypoint neurorehabilitation, Utrecht, The Netherlands, April 10, 2010.

Steenbeek D, Ketelaar M, Lindeman E, Galama K, Gorter JW. (2009) The value of Goal Attainment Scaling compared to a standardized and generic tool, measuring change over time of children with Cerebral Palsy. Poster presentation. American Academy for Cerebral Palsy and Developmental Medicine (AACPDM) 63rd Annual Meeting Scottsdale, Arizona, USA, Sept 23–26, 2009. Abstract: *Dev Med Child Neurol* **51** (suppl 5): 61.

Steenbeek D, Gorter JW, Ketelaar M, Galama K, Lindeman E. (2009) Goal Attainment Scaling as a reliable and useful outcome measure for children with Cerebral Palsy. Free Paper. American Academy for Cerebral Palsy and Developmental Medicine (AACPDM) 63rd Annual Meeting Scottsdale, Arizona, USA, Sept 23–26, 2009. Abstract: *Dev Med Child Neurol* **51** (suppl 5): 22.

Steenbeek D, Ketelaar M, Galama K, Gorter JW, Lindeman E. (2009) Goal Attainment Scaling in de kinderrevalidatie [Goal Attainment Scaling in paediatric rehabilitation]. VRA Required Courses, Clinimetrics. Utrecht, The Netherlands, May 15, 2009.

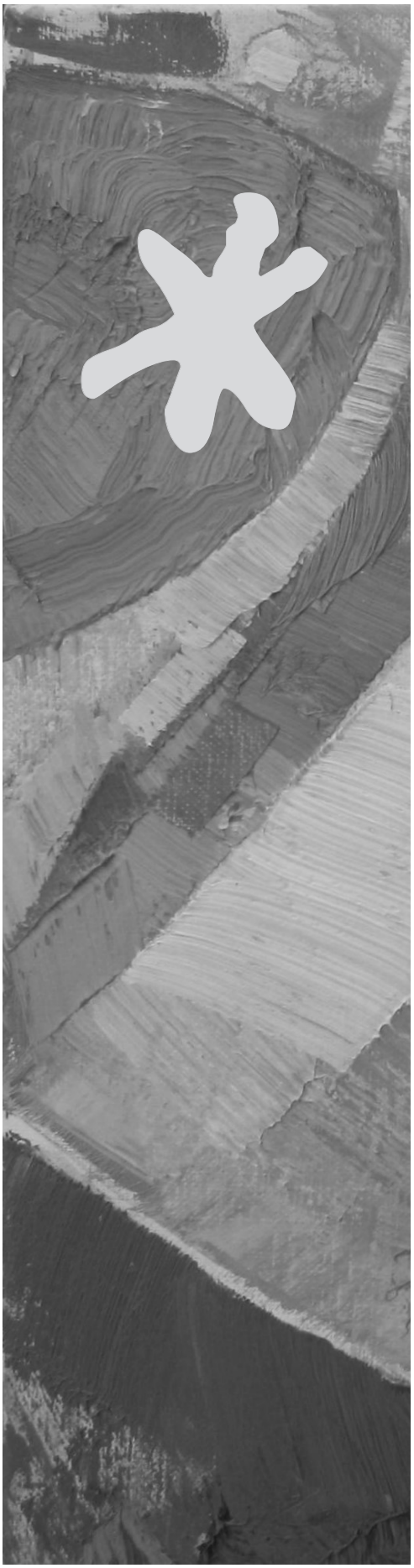
Steenbeek D, Ketelaar M, Galama K, Gorter JW. (2008) Goal Attainment Scaling in paediatric rehabilitation: the inter-tester reliability. Free Paper. VRA voorjaarssymposium Utrecht, The Netherlands, April 11, 2008. Abstract: *Clin Rehabil* **22**: 1135.

Steenbeek D, Galama K, Ketelaar M, Gorter JW. (2006) Goal Attainment Scaling in de kinderrevalidatie [Goal Attainment Scaling in paediatric rehabilitation]. Conference on Intervention and Communication, Huizen, The Netherlands, Nov 1, 2006.

Steenbeek D, Galama K, Ketelaar M, Gorter JW. (2005) Goal Attainment Scaling bij kinderen met Cerebrale Parese [Goal Attainment Scaling, the use for children with Cerebral Palsy]. Poster presentation VRA Conference, Arnhem, The Netherlands, Oct 28–29, 2005.

Steenbeek D, Meester-Delver A, Becher JG, Lankhorst GJ. (2002) Het effect op vaardigheidsniveau van behandeling met botulinetoxine-A bij kinderen met Cerebrale Parese [The effect of Botulinum Toxin type A treatment on the level of functional abilities in children with Cerebral Palsy: evaluation with Goal Attainment Scaling]. VRA Conference, Groningen, The Netherlands, Apr 2002. Abstract: *Ned Tijdschr Geneesk* **46**: 2170–2171.

Samenvatting
(Summary in Dutch)



Dit proefschrift heeft als doel het onderzoeken van het gebruik van Goal Attainment Scaling (GAS) in de dagelijkse interdisciplinaire kinderrevalidatie-praktijk. Bij GAS worden voorafgaand aan de behandeling schalen gemaakt met een beschrijving in vrije tekst van de mate waarin behandeldoelen worden behaald. Voor dit proefschrift is gebruik gemaakt van 6-punts GAS-schalen. Achteruitgang wordt gedefinieerd als een score '–3', het uitgangsniveau als '–2', minder dan het doel als '–1', het doel wordt als '0' gescoord en iets meer dan en veel meer dan het doel respectievelijk als '+1' en '+2'. Dit wijkt enigszins af van de originele methode die in 1969 geïntroduceerd werd, waarin nog geen sprake was van een score '–3'. In dit proefschrift wordt GAS gebruikt om doelen op vaardigheids-, activiteiten- en participatieniveau te formuleren bij kinderen met Cerebrale Parese (CP) en veranderingen over tijd te evalueren. Het gebruik van GAS in de kinderrevalidatie is relatief nieuw.

Hoofdstuk 1

Het proefschrift begint met een korte introductie in het onderzoek in de kinderrevalidatie en een beschrijving van de lopende projecten in Nederland. De verschillende doelen van meetinstrumenten en eigenschappen van meetinstrumenten in het algemeen komen aan de orde. Zo worden de termen generiek versus specifiek meten, gestandaardiseerd versus individueel meten en discriminatief versus evaluatief meten geïntroduceerd. GAS is een voorbeeld van een generiek individueel evaluatief meetinstrument waarmee zowel het niveau van functioneren van individuen alsook groepen gemeten kan worden. De schalen leveren ordinale data op: de orde van de scores is bepaald, maar rekenkundig gaat het niet om een getalsmatige uitkomst.

In hoofdstuk 1 worden ook de definitie en verschillende vormen van classificatie van CP besproken, alsmede de meest voorkomende behandelstrategieën. CP is een groep aandoeningen die gekenmerkt wordt door houdings- en bewegingsstoornissen, die het gevolg zijn van functiestoornissen van de hersenen die zijn ontstaan vóór, rondom of kort na de geboorte. De meest voorkomende vorm is de spastische CP, waarbij spasticiteit en een onnatuurlijke dysbalans in spieractiviteit het niveau van functioneren van kinderen beperkt. Het gaat om een zeer heterogene groep kinderen. Sommigen stellen doelen op het voetbalveld, anderen zijn volledig afhankelijk van een elektrische rolstoel of kunnen alleen met hun hoofd een computer besturen. Het niveau van functioneren is ook afhankelijk van geassocieerde problemen waaronder taal- en spraakstoornissen of cognitieve retardatie.

De WHO International Classification of Functioning, Disability and

Health, Children and Youth version (ICF-CY) wordt in hoofdstuk 1 toegelicht. Dit model biedt een waardevol kader voor terminologie, behandelfocus en doelstellingen. Bij kinderen met CP gaat het bij voorkeur om doelen waar zij daadwerkelijk zelf iets aan hebben. De revalidatiegeneeskunde en onze GAS-schalen kenmerken zich door het stellen van doelen op het activiteiten- en participatieniveau van de ICF.

In 2006 is een CBO richtlijn Cerebrale Parese gepubliceerd, die in Nederland de classificatie, diagnostiek, behandelstrategieën en evaluatie uniformeert. Deze richtlijn vat ook de evaluatiemethoden bij CP samen. In de dynamiek van de ontwikkeling van meetinstrumenten kan GAS een nieuwe aanvulling hierop zijn.

Hoofdstuk 2

Hoofdstuk 2 beschrijft een studie waarin het effect van een multi-levelbehandeling met Botulinetoxine A (BTX-A) gemeten werd door het gebruik van GAS. De toediening van BTX-A onder narcose in spiergroepen die vanwege spasticiteit het niveau van functioneren van kinderen bedreigen, is een belangrijke aanvullende spasticiteitbehandeling die de kinderrevalidatiearts ter beschikking staat. In deze studie werd een design gebruikt waarbij 6- of 8-wekelijkse metingen in een baselinefase werden vergeleken met 8- of 6-wekelijkse metingen na behandeling met BTX-A. Gedurende deze 14 weken werd wekelijks een videofragment opgenomen. De GAS-scores werden bepaald door het beoordelen van alle fragmenten in willekeurige volgorde, waarbij de beoordelaar geblindeerd was voor de volgorde van opnamen.

Elf kinderen namen deel en voor ieder kind werden voor drie doelen GAS-schalen geconstrueerd. Dat leidde tot 462 video's. De verandering tussen alle baseline- en behandelfasemetingen (33 doelen) was significant. Negen van de 11 kinderen toonden verbetering op 18 van de 33 GAS-schalen. Twee kinderen toonden geen verandering op de GAS-schalen. Van de negen kinderen die verbeterden, toonden er zeven klinisch relevante verbetering (meer dan 2 punten verschil) in 11 doelen. Randomisatie voor de lengte van de baseline maakte inzichtelijk dat het moment van de verandering gerelateerd was aan het moment van behandeling. We kwamen aldus tot de conclusie dat GAS klinisch relevante verbetering van het niveau van functioneren kon aantonen. De positieve verwachtingen van GAS werden in deze eerste studie bevestigd. Eén van onze boodschappen was dat verder onderzoek naar de klinimetrische eigenschappen van GAS noodzakelijk was, hetgeen de basis heeft gelegd voor dit proefschrift.

Hoofdstuk 3

Een kritische review van de beschikbare literatuur over de betrouwbaarheid van de inhoud van GAS-schalen in de kinderrevalidatie, de betrouwbaarheid van het scoren van GAS-schalen, de validiteit en de gevoeligheid voor verandering van GAS is beschreven in hoofdstuk 3. Ook hebben we gekeken naar de studies in de kinderrevalidatie die GAS hadden gebruikt als uitkomstmaat. Negen relevante publicaties rapporteerden de veelbelovende eigenschappen van GAS in de kinderrevalidatie en de auteurs adviseerden verdere ontwikkeling van de methode. Wij concludeerden echter wel dat de huidige kennis over de betrouwbaarheid van GAS in het gebruik voor kinderen te kort schoot. Ook de toegevoegde waarde ten opzichte van bestaande instrumenten en de responsiviteit verdienden nader onderzoek.

Deze conclusies resulteerden in het vervolgproject ter nadere beschouwing van de klinimetrische eigenschappen van GAS, dat werd uitgevoerd samen met de kinderunit van het Revalidatiecentrum Breda.

Hoofdstuk 4

Wij zijn in de kinderunit gestart met het ontwikkelen van een training. De training bestond uit drie centrale discussiebijeenkomsten van 2 uur. Daarnaast heb ik de betrokken kinderfysiotherapeuten, kinderergotherapeuten, kinderlogopedisten, maatschappelijk werkers en orthopedagogen van de kinderunit feedback gegeven op geconstrueerde schalen totdat deze schalen aan een aantal criteria voldeden die we samen ontwikkeld hadden. Deze criteria hadden betrekking op de mate waarin de schalen daadwerkelijk ordinaal waren met het gebruik van één enkele variabele, SMART doelen beschreven in vaardigheden, activiteiten en sociaal-maatschappelijke participatie, en om praktische redenen te scoren waren binnen 10 minuten. Afgesproken werd om de schalen volgens het 'can do' principe te scoren. Hiermee beoogden wij de weergave van het professioneel inzicht in het niveau van functioneren van de kinderen en toename van de betrouwbaarheid van de scores omdat de score verlost kon worden van de dagelijkse grillen van kinderen (motivatie, vermoeidheid, interactie).

Dit resulteerde in 115 GAS-schalen die geconstrueerd en na 6 maanden gescoord werden door de paramedici. De therapeuten en ouders werd door middel van enquêtes hun mening gevraagd over het werken met GAS en de training. Een belangrijke bevinding was dat het uitdenken en verwerken van een goede GAS-schaal veel tijd kost, bij ons gemiddeld 45 minuten per schaal. De afgesproken criteria ten aanzien van de inhoud van de schalen

werd waardevol bevonden door de ouders en paramedici. De meeste van hen beschouwden GAS als een bruikbare methode om de kwaliteit van de revalidatiebehandeling te verbeteren.

Hoofdstuk 5

De eerste klinimetrische eigenschap van GAS die wij nader beschouwd hebben, was de *inter-raterbetrouwbaarheid*: de mate van overeenkomst in de score wanneer twee onafhankelijke professionals één GAS-schaal scoren. Hoofdstuk 5 beschrijft een betrouwbaarheidsstudie waaraan 23 kinderen met CP en 20 therapeuten van drie disciplines (kinderfysiotherapie, kinderergotherapie en kinderlogopedie) deelnamen. Achttien kinderen hadden fysiotherapie, ergotherapie en logopedie, vijf kinderen hadden alleen fysiotherapie en ergotherapie. De behandelend therapeuten van de kinderen en onafhankelijke collega's van dezelfde discipline, die als onderzoeksmedewerker participeerden, maakten ieder één GAS-schaal per kind: dat werden 64 GAS-schalen per groep (18x3+5x2). Beiden scoorden zij een half jaar na het maken van de schaal zowel hun eigen schaal alsook de schaal van hun partner (2x128 scores).

De schalen die gemaakt waren door de behandelend therapeuten hadden een inter-raterbetrouwbaarheid van 0,82 (Cohen's linear weighted kappa). De inter-raterbetrouwbaarheid voor schalen die gemaakt waren door de onafhankelijke onderzoeksmedewerkers was 0,64. Aldus konden wij de belangrijkste in de literatuur opengebleven vraag beantwoorden met deze studie: de inter-raterbetrouwbaarheid van GAS in de kinderrevalidatie is goed mits de professionals getraind zijn en er duidelijke afspraken gemaakt zijn over de wijze van scoren. Wat ook bleek uit de data was dat het maken van de schalen door de behandelend therapeuten zelf, in plaats van door onafhankelijke therapeuten, mogelijk een positieve invloed heeft op de betrouwbaarheid. Die bevinding is geruststellend met betrekking tot het gevreesde fenomeen van 'therapist-bias': een structurele fout als gevolg van een positievere inschatting door de behandelaar van verbetering dan zich in werkelijkheid afspeelt. Dit proefschrift ondersteunt de betrouwbaarheid van het professionele inzicht.

Hoofdstuk 6

In hoofdstuk 6 werd de inhoud van de GAS-schalen vergeleken met die van twee andere wereldwijd gebruikte gestandaardiseerde uitkomstmaten voor kinderen met CP: de *Pediatric Evaluation of Disability Inventory* (PEDI) en de 66-item versie van de *Gross Motor Function Measure* (GMFM-66). Hiervoor

werden de doelen in de GAS-schalen en de items van de PEDI en GMFM-66 geclassificeerd volgens de ICF-CY.

Twintig procent van de GAS-doelen, die per definitie de belangrijkste ingang voor de revalidatiebehandeling beschrijven, werden niet gedekt door de beide gestandaardiseerde instrumenten. Het volledig behalen van het behandeldoel (een GAS-score van 0 of hoger) werd vergeleken met statistisch significante veranderingen op de PEDI en GMFM-66. Negenendertig van de 64 GAS-schalen werden 0 of hoger gescoord en één van de bevindingen was dat van deze 39 schalen de 16 betrokken kinderen op de PEDI geen verandering scoorden op de meest geassocieerde subschaal. Wij concludeerden dat de gekozen gestandaardiseerde instrumenten en GAS complementair waren in hun construct en individuele responsiviteit. Een belangrijk deel van volledig behaalde revalidatiedoelen kan worden gemist als alleen gemeten wordt met gestandaardiseerde instrumenten, zelfs als de behandeldoelen wel items zijn van deze instrumenten. Daarnaast is het waarschijnlijk dat wanneer alleen gescoord wordt op het volledig behalen van een behandeldoel, kleinere maar significante veranderingen in het niveau van functioneren gemist kunnen worden. Deze laatste studie van het proefschrift draagt bij aan de kennis over de toegevoegde waarde van GAS en geeft inzicht in de vragen over de responsiviteit die wij vanaf de eerste studie opwierpen.

Hoofdstuk 7

Hoofdstuk 7 belicht wat we hebben geleerd van de verrichte onderzoeken. Positieve eigenschappen van GAS betreffen het kunnen kwantificeren van de relevantie van de gemeten verandering, het faciliteren van de kwaliteit van zorg en het gebruik van de ICF-CY items voor activiteiten en participatie in de uitkomstmaat. Bovendien is een groot voordeel dat behandelresultaten vergeleken kunnen worden in de heterogene populatie die de kinderrevalidatie kenmerkt. De kinderen hebben alle hun eigen hulpvraag die tot zeer uiteenlopende behandeldoelen leidt. En we kunnen met GAS op een betrouwbare manier belangrijke veranderingen meten die op geen enkele andere manier meetbaar zijn.

Er blijven echter ook aandachtsgebieden bij de toekomstige toepassing van GAS. Voorbeelden zijn de volgende. De inhoud van de schalen dient steeds opnieuw kritisch beschouwd te worden. Meten we wat we willen meten? Zijn er geen andere belangen dan objectieve therapie-evaluatie? De betrouwbaarheid van de inhoud van de schalen is in dit proefschrift niet beschreven. GAS kent per definitie een lage construct- en criteriumvaliditeit omdat het een individuele uitkomstmaat is. Verder adviseren wij data-analyse

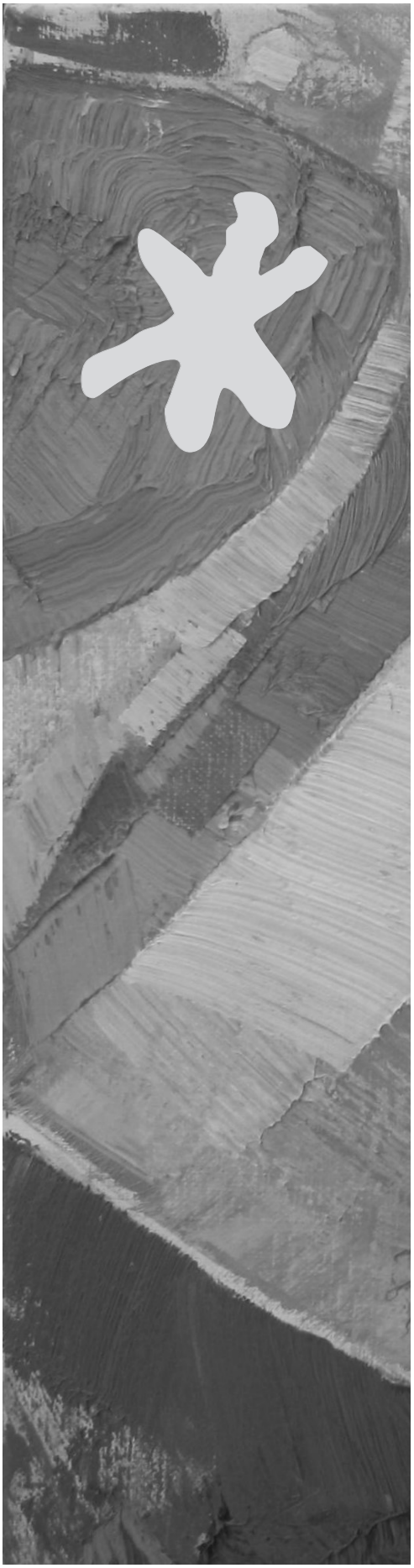
zonder het gebruik van de *T*-somscore die de originele GAS kenmerkt, hetgeen tot nieuwe discussies kan leiden over de dataverwerking. GAS-data voor kleinere populaties laten zich goed nonparametrisch verwerken, bij voorbeeld met de Wilcoxon signed ranks test.

Suggesties voor toekomstig onderzoek in GAS en het gebruik in de praktijk worden gegeven in hoofdstuk 7. Bij de beëindiging van dit proefschrift zijn er twee lopende onderzoeken in Nederland bij kinderen met CP, die onze uitwerking van GAS als uitkomstmaat gebruiken. De bevindingen over het gebruik zullen ons naar verwachting veel leren. Toekomstig onderzoek zal in eerste instantie gericht moeten zijn op de content-betrouwbaarheid: in hoeverre komen de vrije teksten van GAS-schalen gemaakt door verschillende professionals die over dezelfde informatie beschikken overeen? Verder is de dynamische interactie tussen de Canadian Occupational Performance Measure (COPM) en GAS een boeiend onderwerp voor verdere ontwikkeling. De COPM is een instrument dat behulpzaam kan zijn bij het exploreren van de hulpvraag als basis voor het maken van GAS-schalen. Ook kunnen verschillende revalidatieafdelingen en/of verschillende revalidatiesettingen met het gebruik van GAS vergeleken worden teneinde variabelen in de zorg te evalueren en uiteindelijk de kwaliteit van zorg te verbeteren.

De toegevoegde waarde van GAS, het meten van effectiviteit in groepen, en de tevredenheid van de gebruikers motiveert nu al tot het verdere gebruik in de praktijk. Dit proefschrift beoogt de nieuwsgierigheid te prikkelen in de toepassing van GAS. Wij hopen de discussie te voeden hoe we tot een uniforme werkwijze kunnen komen. Consensus over de criteria en timing in revalidatieprogramma's is een volgende stap, waarna een handleiding en implementatiemodel voor ons werkveld ontwikkeld kan worden.

GAS is een waardevolle uitkomstmaat voor de gezinnen van kinderen met CP wanneer de juiste voorwaarden worden gesteld.

Dankwoord
(Acknowledgements)



Het doet mij genoegen dit proefschrift te mogen besluiten met het bedanken van de grote groep waardevolle mensen om me heen die direct en indirect het project mogelijk hebben gemaakt. Het inzicht in het stellen van doelen bij kinderen met CP en het meten van uitkomst daarmee, komt voort uit talloze discussies met ouders, teamleden en mijn waardevolle onderzoeksbegeleiders. Daarnaast kost onderzoek doen tijd: ik ben velen dank verschuldigd die mij de tijd gegund hebben. De afronding is een overwinning voor alle betrokkenen.

Hoewel ik veel meer mensen dank verschuldigd ben dan ik kan noemen, zou ik graag bij de volgende groepen even stilstaan.

De deelnemende kinderen en hun ouders

Veel respect en waardering heb ik voor alle ouders van de kinderen met CP, die zo enthousiast medewerking hebben verleend. De meeste gezinnen participeren in meerdere onderzoeken. Het deelnemen aan onderzoek is een investering in de toekomst en dient meestal het belang van juist andere kinderen. De ouders van kinderen met CP zijn in de regel meer dan gemiddeld belast vanwege de extra zorgen om hun kind bij wie de normale activiteiten een uitdaging zijn. Juist deze ouders volgen vaak een geheel ander levenspad dan zij zich gewenst en voorgesteld hadden wat betreft werk en vrije tijd omwille van hun gezin. De ouders die dan extra investeren in hun medemens door deelname aan onderzoek, liggen mij na aan het hart. Allen zeer veel dank. Ook bedankt voor alle waardevolle feed-back op het werken met GAS. Daarnaast zijn wij natuurlijk veel dank verschuldigd aan de kinderen zelf die hebben meegedaan. Veel van hen vonden het gelukkig leuk. We hadden bij geen van de onderzoeken uitvallers, dus de kinderen hebben een tanig doorzettingsvermogen getoond.

De onderzoeksgroep van het promotietraject

Vervolgens wil ik graag mijn promotor Prof. Eline Lindeman, co-promotoren Marjolijn Ketelaar en Jan Willem Gorter en onderzoeksmaatje Krys Galama bedanken.

Eline, onder jouw leiding werd het onderzoek gebundeld tot deze dissertatie. Zeer bedankt voor het uitzetten van de lijn en het vormgeven van de samenwerking tussen de Hoogstraat en Revalidatiecentrum Breda. Zeer bedankt voor jouw altijd snelle en degelijke – op jarenlange ervaring gestoelde – supervisie van de data-analyses en de publicaties.

Marjolijn en Jan Willem, jullie waren bereid mijn onderzoeksplannen

samen te begeleiden en wat ben ik jullie daar onmetelijk dankbaar voor. Jullie zijn in mijn beleving een voorbeeld van een duo wetenschappers die er samen meer dan vier waard zijn door de wijze van samenwerken, elkaar inspireren, kritisch aanvullen, vertrouwen en vriendschap. Wat heb ik genoten van de vele gedachtenwisselingen, waarin jullie me geleidelijk helderheid boden in wat ik wilde zeggen. En wat hebben jullie veel energie in de publicaties gestopt als co-promotoren, gedreven door de wens antwoorden op vragen te vinden. Jan Willem, ik ben er trots op dat je mij ook na jouw verhuizing naar *CanChild* in Canada wilde blijven begeleiden. De diverse deadlines die we samen gehaald hebben omdat jij om 2.00 u 's nachts mijn aangeleverde tekst oppakte en verwerkte in de 6 uur tijdsverschil zodat ik 's ochtends weer verder kon, zal ik nooit vergeten. Marjolijn, jij bent onze centrale spil in Nederland als senioronderzoeker in de kinderrevalidatie. Ik ben er trots op dat jij ons project hebt begeleid te midden van jouw zo vele andere prominente projecten in de kinderrevalidatie. Jullie zijn wetenschappers van het allerhoogste niveau, duurzame inspirators en bovenmenselijk harde werkers met het hart op de juiste intermenselijke plaats. Ik heb veel van jullie geleerd.

Krys, vanaf de onderzoekstart in Breda heb jij het onderzoek vele ondersteunende diensten bewezen. Als unitmanager was jij in staat ten behoeve van de bepaling van de inter-rater-betrouwbaarheid van GAS heel veel werk te plannen bij twee paramedici, terwijl slechts één van hen gedeclareerd kon worden bij de verzekeraar. Desondanks, ik denk *dankzij* jouw aansturing, bloeide de kinderunit in Breda als nooit tevoren. Jij voedde de terechte trots van het team en bent als geen ander in staat om altijd snel de juiste krachten op de juiste plaats te ontplooien. Bij de publicaties speelde je een onmisbare rol als onze native speaker. Bedankt voor je ontelbare uren die jij in je vrije tijd aan de teksten hebt besteed. Ik zal niet vergeten hoe jij mij in Scottsdale, Arizona tussen de cactussen gecoacht hebt in het internationaal overbrengen van onze boodschap. Bedankt.

Onderzoeksgroep van het eerste project

De eerste publicatie is altijd de moeilijkste. Dank gaat uit naar mijn deelopleider Anke Meester-Delver en de wetenschappelijk begeleiders Prof. Guus Lankhorst en Prof. Jules Becher (VUMC, Amsterdam) voor het onderzoek dat ik in het kader van mijn opleiding tot revalidatiearts kon verrichten in de Trappenberg (hoofdstuk 2). Anke, ik kijk met veel voldoening op onze samenwerking terug en geniet nog steeds bij de gedachten aan de stapels en stapels VHS videobanden die jij destijds voor ons verzameld en gecodeerd hebt, zodat ik de banden geblindeerd kon beoordelen. Guus, bedankt voor de wetenschappelijke

begeleiding van mijn opleiding. Zonder jouw initiatie was dit proefschrift niet over GAS gegaan. Jules Becher, inmiddels onze Godfather van de kinderrevalidatiegeneeskunde in Nederland, bedankt voor je deskundige begeleiding en het samen uitwerken van de publicatie. Herhaaldelijk heb ik jouw motto als mantra mee laten spelen: "Gewoon doorgaan". Bedankt, dat doen we. Helaas ben jij verhinderd voor deelname aan de commissie op de voorgenomen promotiedatum.

Leden van de beoordelingscommissie

Prof. Dr. M.J. Jongmans, Prof. Dr. E.E.S. Nieuwenhuis, Prof. Dr. M.J. Schuurmans, Prof. Dr. K. Postema, Prof. Dr. M.W.G. Nijhuis-Van der Sanden, allen hartelijk dank voor het beoordelen van dit proefschrift en het voorbereiden van de oppositie.

De teamleden

Alle leden van de kinderunit van het Revalidatiecentrum Breda waren betrokken bij de trainingsfase in GAS en derhalve bij de vele inzichtgevende discussies. Ook de logopedisten verbonden aan de mytylschool De Schalm participeerden. De kinderlogopedisten, kindergotherapeuten en kinderfysiotherapeuten participeerden in de vervolgstudie naar de betrouwbaarheid en toegevoegde waarde van GAS.

In het bijzonder wil ik de onderzoeksmedewerkers ("independent raters") Inge Keen (fysiotherapeut) en Dieuwke van den Heuvel (ergotherapeut) bedanken. Libby Traas, Colette Buurlage, Els Braaksma, Marita Mechelinck waren de betrokken logopedisten die elkaar afwisselden als behandelaars en onderzoeksmedewerkers. Bedankt. De overige betrokken fysiotherapeuten waren Sandra Aarts, Karin Franken, Koen Dekkers, Ruud Touw, Corina Luyten, Jopie van Iersel en Frans van der Brugge. De overige betrokken ergotherapeuten waren Nanette Beunders, Pascalle Baljon, Fleur Paumen, Ankie Adema, Heleen Janssen, Wilma Bakker, Barbara van Riele, Jolanda Gommeren en Dianne van Rooij-Timmers. Aan de discussies tijdens de trainingsfase namen deel Mieke Coenen en Manon Merkx, maatschappelijk werkenden en Birgit Tennebroek, Ilze Vliegenberg en Cisca Cordier de Croust, othopedagogen. Het was niets geworden zonder Maartje Hoeve, onze vaste rots bij de planning. Na afronding van de dataverzameling is onze unit versterkt door een aantal nieuwe fysiotherapeuten en ergotherapeuten, die al even enthousiast het werken met GAS hebben opgepakt en tot heden uitdragen. Allen bedankt, ik geniet dagelijks van de leerzame samenwerking

en jullie enthousiasme. Jullie hebben de wereld bewezen dat professioneel inzicht betrouwbaar is.

Maarten Wojakowski, kinderfysiotherapeut in revalidatiecentrum De Trappenberg, na 10 jaar nog steeds en eigenlijk steeds opnieuw zeer hartelijk dank voor jouw prominente rol bij het organiseren, opnemen en verwerken van 462 video-opnamen. Alle door jouw hand gegaan, alle scherp en helder, alle met altijd vrolijke en gemotiveerde kinderen. We hadden geen enkele uitval en daarmee heb jij eigenlijk het onmogelijke gepresteerd. Ook ongelooflijk knap zoals je de volgorde van de video's wist te blinderen: altijd dezelfde kleding, geen kerstboom op de film.

Management

Graag bedank ik de Raad van Bestuur van wat voorheen de Stichting Revalidatiecentrum Breda (SRCB) was en nu – sinds de fusie met de revalidatie-instellingen in Zeeland – Revant geworden is.

Jan van Kampen (voormalig algemeen directeur), zeer hartelijk dank voor je persoonlijk enthousiasme om wetenschappelijk onderzoek te faciliteren. Jij was uniek in Nederland met de rekensom over de kosten van het aanvragen van subsidiegelden en je stelling dat je eerst moest starten met onderzoek en later over de financiering na moest denken. Niemand durft uit te rekenen wat het onderzoek gekost heeft, maar je hebt meer dan gelijk gekregen: de productie van de kinderunit is ruim gestegen tijdens de onderzoeksinspanningen. Christianne Lennards (huidig algemeen directeur Revant), zeer bedankt voor je inspiratie en support. Verwetenschappelijking van het klimaat is onder jouw leiding een van de beleidsmissies van Revant geworden. Bedankt voor het uitdragen van de meerwaarde van initiatie van onderzoek ook voor niet-academische revalidatiecentra.

Gijs Kuijpers (voormalig medisch manager SRCB) en Lou Corsius (voormalig sectormanager SRCB), enorm bedankt voor jullie consequent volgehouden vertrouwen en jullie initiatie van de samenwerking tussen het Revalidatiecentrum Breda en De Hoogstraat in Utrecht. Jullie hebben daarmee de randvoorwaarden voor dit onderzoek vormgegeven. Marlies Rommelse (huidig clustermanager Revant), bedankt voor het voorzetten van de beleidslijn in onderzoeksparticipatie.

Jan Willem Meyer (huidig medisch directeur Revant), zeer bedankt voor al je hulp en evaluaties van het promotietraject, veelal inhoud gegeven vanuit jouw eigen ervaring als promovendus destijds. Bedankt voor het mede vorm geven van de promotie in de oppositie en het vorm en inhoud geven van het beoogde symposium ter gelegenheid van de promotie.

Directie van de Trappenberg (Lily Heijnen), dank voor de facilitering van het onderzoek.

Directie van mytylschool De Schalm Breda (Ton Metselaar), dank voor de medewerking van de logopedisten verbonden aan school.

Het management van het Kenniscentrum Revalidatiegeneeskunde Utrecht, locatie revalidatiecentrum De Hoogstraat (Steven Berdenis van Berlekom), dank voor het altijd enthousiast faciliteren van de onderzoeksbegeleiding en de samenwerking tussen De Hoogstraat en Revalidatiecentrum Breda.

Hulp bij dataverwerking

Graag bedank ik Prof. Riekie de Vet (EMGO, Amsterdam) voor de hulp en inzichten in klinimetrie op het meest cruciale moment van het peer-reviewproces bij Archives van het vierde artikel. Graag bedank ik Tjeerd van der Ploeg (Hogeschool Alkmaar) voor het tot twee keer toe belangeloos helpen bij de statistische dataverwerking.

De medische staf Revalidatiecentrum Breda

Mijn directe collega's in Breda ben ik dank verschuldigd omdat ik af en toe onderzoeksbelangen voor moest laten gaan. Bedankt voor de wetenschapstijd, die wij als gehele staf dragen.

Agnes van Velzen (kinderrevalidatiearts, Revalidatiecentrum Breda), bedankt voor al het overgenomen werk, waarmee je het onderzoek een enorme dienst hebt bewezen. Bedankt voor je enorme kwaliteitsbevorderende invloed. Je steunde me door dik en dun en wist het team enthousiast te houden voor eigenlijk alles, dus ook voor ons onderzoek. We zijn samen een sterk duo en mogen trots zijn op een kinderunit in Breda waaraan vanaf de promotie een derde kinderrevalidatiearts zal worden toegevoegd.

Medewerking aan de vormgeving

Soms komt onverwacht hulp uit een hoek waar je niet meer op had durven hopen.

Renate Siebes, aan wiens werk als lid van *NetChild* in dit proefschrift gerefereerd wordt, is een bedrijf begonnen (Proefschrift.nu) in het vormgeven van proefschriften. Renate, bedankt voor je enorme hulp op tijdstippen waarop de meesten niet meer werken.

Minne de Groot (kunstschilder Amsterdam), bedankt voor het

beschikbaar stellen van De Zwaan. De Zwaan is mijn symbool voor betrouwbaarheid. Graag wijs ik de lezer nog een keer op jouw site: http://www.minnedegroot.ontheweb.nl/.

Medewerking aan het symposium ter gelegenheid van de promotie

Op de datum van de promotie is een symposium in Revalidatiecentrum Breda voorgenomen. Graag bedank ik de sprekers, behalve voor hun prominente bijdrage aan dit proefschrift, ook op voorhand voor hun bijdrage aan het symposium. Mijn promotor prof. Eline Lindeman, mijn co-promotoren Jan Willem Gorter en Marjolijn Ketelaar, mijn onderzoeksmaatje en collega Krys Galama, en de directeuren Jan Willem Meyer en Steven van Berlekom. Graag bedank ik ook de voorzitter Christianne Lennards en de organisatoren Marjan Leusink en Marlies Rommelse.

Paranimfen

Mijn zus, Romy Steenbeek, jij ging mij 11 jaar en 8 dagen geleden voor met je prachtige promotie op wilde Thomas's langurs die je jaren met passie hebt gevolgd in Indonesië. Je hebt me vooral bij de eerste publicatie enorm op weg geholpen en bent mij altijd blijven stimuleren. Je bent een wetenschapper van formaat en ik ben er trots op dat jij mijn paranimf wilt zijn. Krys Galama, ik val in herhaling, bedankt voor alles. De gehele onderzoeksgroep behoort op het podium dus ik ben blij dat ook jij mij terzijde wilt staan bij de verdediging.

Familie

Ook mijn andere broer en zus, Henk en Ingrid Steenbeek, bedankt voor jullie trots die ik toch zo'n 42 jaar als stimulans voel.

Mijn ouders, echt verder doordenken over een onderwerp leer je denk ik op jonge leeftijd. Dit en doorzettingsvermogen hebben jullie altijd gestimuleerd. Emmy, bedankt voor alle zorgen als oma, waardoor ik onder meer de congressen kon bezoeken.

En een vrolijke noot van nostalgie: ik denk graag aan mijn eigen oma die 14 jaar geleden overleden is. Bij haar op zolder aan de Oudegracht in Utrecht studeerde ik fysiotherapie en geneeskunde tegelijkertijd. Dat vond zij maar raar. Wat had ik gemoeten zonder haar onoverwinnelijk vermogen tot relativeren.

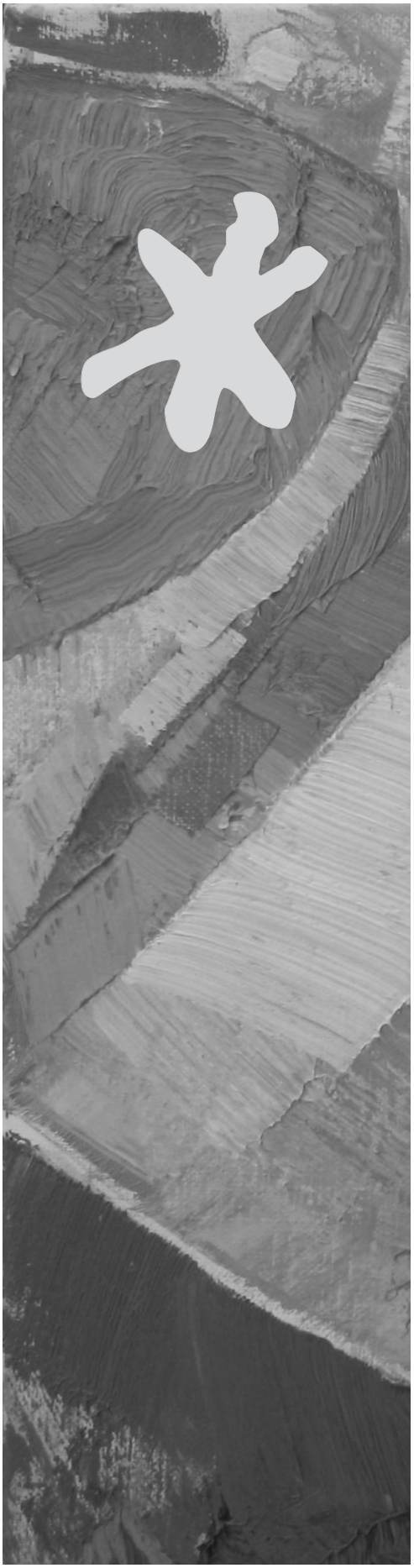
Mariëtte Bos, onze huisvriendin, bedankt voor al je zorgen voor ons, die mij zeker geholpen hebben om tijd te vinden voor dit onderzoek.

Het belangrijkste bewaar ik voor het laatst: mijn lieve gezin.

Mijn allerliefste Marije en twee stoere zoons Esli en Max: wat zijn jullie mij dierbaar! Jullie hebben als geen ander ondervonden dat het beoefenen van de wetenschap meestal niet tijdens werktijd lukt. Bedankt lieverds voor al jullie ondersteuning, aanmoediging, bijslijpen en de zo welkome tegenGAS als ik het te bont maak. Bedankt voor de energie die jullie mij geven om dit soort projecten mogelijk te maken. Marije, jij nam veel taken op je die van ons samen zijn en speelde mij vaak vrij voor dit onderzoek. Jullie verrijken mij ook met het benadrukken van de vele andere belangen dan wetenschap en houden mij in balans. We hebben dit verhaal samen geschreven. Ik ben trots op jullie.

Allen bedankt, Duco Steenbeek, april 2010.

De Zwaan



De Zwaan

"Ik benijd de zwanen om de onvermoeide ijver waarmee men ze gedichten wijdt." (Uit "Hoe-korter-hoe liever" van Theo van Baaren). Hun grote plaats in de poëzie komt voort uit de schoonheid, de kracht, de majestueuze verheven kalmte, die hen tot vorsten maken. De zwaan, de aan Apollo gewijde, geeft ons de kracht om in de toekomst te kijken. Zijn gratie en zijn ongereptheid maken hem tot een symbool voor de dichter die door zijn zangen de mensheid bekoort: Pindarus en Fenelon, Shakespeare en Vergilius, Leopardi, de zwarte zwaan van Racanetti, allen ontlenen aan hem een bijnaam die hun kunst karakteriseert in haar bekoring en vloeiende tederheid. Zo leeft de zwaan voort in de sage, in de letteren en het dagelijks leven. Franse dichters uit de 19e eeuw zien in hem het symbool van de onbegrepene. In de sprookjes van de Kelten zijn zwanen boodschappers van de goden. In de Hindoe mythologie uit India staan de zwanen met hun magische muziek die door de machtige wieken wordt gemaakt, voor het inademen en uitademen van de levensadem. Bij de Indianen brengen zwanen tijden van transformatie, veranderde staten van bewustzijn en ontwikkeling van de intuïtieve vermogens.

De zwaan als symbool voor betrouwbaarheid, zorgzaamheid, kracht, stijl en eenvoud mag dit proefschrift omarmen. Mits men hem doorgrondt, is hij als GAS en als ons professioneel inzicht.