Goal attainment scaling (GAS) is an individualized criterion-referenced measure of change that has several unique features, compared with the behavioral objective, and five possible levels of attainment for each goal. The validity of GAS as a measure of motor change was investigated in 65 infants, 3 to 30 months of age, with motor delays. For each infant, two goals to be attained within a 6-month period of intervention were established. After the 6-month period, the mean GAS T-score of 55.4 was significantly higher than the expected mean of 50. This finding indicated that the infants' motor change exceeded the therapists' expectations. Neither the type nor the category of goal influenced the therapists' ability to select outcomes that the infants were capable of achieving. Because moderate and low correlations were found between GAS T-scores and Peabody gross motor and fine motor change scores, the two assessments appear to measure different aspects of motor development. Selected child, family, and program variables were examined as sources of bias in GAS T-scores. Motor delay was the only variable that biased GAS T-scores; infants who were less delayed had higher GAS T-scores. The results indicate that GAS was responsive to change in individualized motor goals and support the model of GAS as an idiosyncratic measure. The unique features of GAS offer advantages for measurement of motor change compared with the behavioral objective and norm-referenced developmental scales. [Palisano RJ, Haley SM, Brown DA. Goal attainment scaling as a measure of change in infants with motor delays. Phys Ther. 1992;72:432-437.]

Key Words: Early intervention, Evaluation, Goal attainment scaling, Measurement, Physical therapy.

Selection of clinically relevant outcome measures that are responsive to the changes made by children with developmental disabilities is essential for evaluating the effectiveness of physical therapy. Responsiveness, or the ability to measure minimal clinically important changes over time, is a type of validity that is necessary for a test that is used to evaluate the effects of treatment. A test based on normal development cannot automatically be used to evaluate response to treatment. In particular, the Bayley Motor Scale and Peabody Developmental Motor Scales (PDMS) are two norm-referenced scales administered...
by physical therapists that are designed to identify children with motor delays. Neither scale, however, has been validated as responsive to motor change in children with developmental disabilities.

An alternative to norm-referenced motor scales is individualized criterion-referenced measures of change. The behavioral objective is an individualized criterion-referenced measure that is familiar to physical therapists, especially those who work in educational settings. A behavioral objective is a statement that specifies a therapy goal in measurable terms. Goal attainment scaling (GAS)5 is another type of client-centered measure that was originally designed to evaluate mental health services, but recently has been applied to the evaluation of children with developmental disabilities,6,7 including infants receiving physical therapy as part of early intervention services.8 Like the behavioral objective,9 GAS requires selection of goals that are observable and repeatable, specification of the conditions under which performance is measured, statement of the criteria for success in measurable terms, and a time frame for achievement of the goals.

There are several unique features of GAS that are advantageous for measurement of qualitative change and small, but clinically important, improvement in motor development of children receiving physical therapy. The GAS format uses five possible outcomes, as opposed to only one outcome for the behavioral objective. For GAS, the expected level of attainment for each goal is stated, as are two possible levels of attainment that are less favorable and two possible levels that are more favorable. This allows the physical therapist to determine whether a child has made progress despite not achieving the desired goal and whether a child's progress has surpassed the expected outcome. Another unique and useful feature of GAS is the ability to compute a change score for multiple goals, providing an aggregate measure of change. The composite change score can be expressed as a T-score with a mean equal to 50 and a standard deviation of 10. The T-score provides a method for quantitating change over time and across individuals. The physical therapist can also weight the relative importance of each goal so that the composite change score reflects the emphasis placed on each goal during treatment.

Heavlin et al10 have developed a model that describes the psychometric properties of GAS. Central to the model is the construct of GAS as an idiosyncratic measure and bias. An idiosyncratic measure is defined as one in which each client's outcome at follow-up is independent of any predictive factors that are determinable at baseline. This type of measure is in contrast to norm-referenced motor development scales in which high correlations between age, baseline, and follow-up scores are desirable based on the construct that normal motor development has a predictable sequence and rate. Children with motor development delays and disorders are a diverse group who do not demonstrate a normal rate of motor development and frequently do not follow the normal developmental sequence. An idiosyncratic measure, therefore, by virtue of being independent of factors associated with normal motor development, is well suited for evaluation of change in children who receive physical therapy. Heavlin et al10 propose that the idiosyncratic nature of GAS should result in low to moderate concurrent validity with norm-referenced scales.

Bias refers to a potential source of measurement error that is unique to an idiosyncratic measure and poses a threat to validity.10 For GAS, the clinical judgment of the physical therapist(s) who determine the expected level of attainment for each goal is a potential source of bias. The clinician or researcher must make accurate judgments regarding a child's potential for change and the impact of the intervention. Bias occurs when the effect of baseline factors that influence the GAS score are not accurately anticipated; hence, baseline factors are not independent of GAS scores. As a result, GAS scores may be systematically underestimated or overestimated. For infants receiving physical therapy, age, diagnosis, degree of motor delay, and frequency of treatment are examples of factors that might bias the GAS T-score. Heavlin and associates10 recommend that researchers investigate for bias by linear multiple regression.

Research on the use of GAS by physical therapists is limited to one study. Stephens and Haley11 examined concurrent validity of GAS and the PDMS in infants with motor delays by selecting for 54 infants either one or two goals that were related to achievement of gross and fine motor milestones. A purpose of their study was to examine an innovative method of adjusting the difficulty of goals for each infant's developmental rate. The method of determining the GAS change score, therefore, differed from convention. For each goal, the levels of attainment were converted to an estimated age equivalent based on developmental scales and related literature. The GAS change score for each goal was expressed as the gain (in months) between the initial level of attainment and the level of attainment at the end of 6 months. Correlations of .35 and .14 were reported between GAS change scores for the two goals and change in scaled scores for the PDMS. The results indicate low concurrent validity, which supports the model of GAS as an idiosyncratic measure.10

The purpose of our study was to examine the ability of GAS to measure change (responsiveness) in motor development in infants with motor delays and the model of GAS as an idiosyncratic measure. Responsiveness was examined by determining whether the infants achieved the expected level of attainment for two goals, the equivalence of a T-score of 50. The model of GAS as an idiosyncratic measure was examined in two ways. First, concurrent validity with the PDMS was examined using the GAS T-score. To accentuate the unique features of GAS, we included
not only goals that measured development of a new posture or movement but also goals that measured improved performance of an existing posture or movement, improved function, and improved range of motion. Second, we investigated selected factors that might bias GAS T-scores and hence pose a threat to the validity of GAS as an idiosyncratic measure. The following research questions were addressed:

1. Will the mean GAS T-score differ significantly from the expected mean of 50?

2. Will either the type or category of goal influence the ability of therapists to select outcomes that the infants are capable of attaining?

3. What is the relationship between the GAS T-scores and change in age-equivalent scores on the PDMS?

4. Does infant age; motor delay; time spent by therapists, parents, and teachers on activities related to the goals; or length of previous therapy bias GAS T-scores?

Method

Subjects

Sixty-five infants, 3 to 30 months of age ($\bar{X}=19.4, SD=7.1$), who were enrolled in one of nine early intervention programs in the Boston, Mass, metropolitan area, participated in the study. The subjects were the 54 infants included in the study by Stephens and Haley$^8$ and 11 additional infants. Infants eligible for the study had a motor delay as measured by the PDMS and were receiving physical therapy, occupational therapy, or both, as part of their early intervention program. Informed consent of a parent or guardian was obtained for each infant. The sample consisted of 43 boys and 22 girls. Twenty-six infants were born preterm, and 1 infant was born postterm. Fifteen infants had Down’s syndrome, 8 had cerebral palsy, 4 had myelomeningocele, and 1 had autism. The remaining infants demonstrated delayed motor development and had one or more of the following conditions: prenatal infection, congenital anomaly, hydrocephaly, microcephaly, or failure to thrive. Infants were enrolled in their respective programs an average of 11.4 months ($SD=6.5$, range=1-25) prior to the start of the study.

Tests and Procedures

The therapist for each infant identified two motor goals for a 6-month period of intervention. Goals were based on the therapists’ assessment and treatment plan and the progress they expected the infants to make in the 6-month intervention period. A total of 83 gross motor and 47 fine motor goals were developed. Goals were also grouped by category. Sixty-seven goals were categorized post hoc as inter-milestone (development of a new posture or movement), 57 as intra-milestone (improved performance of an existing posture or movement), 3 as functional (mobility or self-care tasks not included on standardized motor development scales), and 3 as physical parameter (eg, range of motion). Goals were put into GAS format, which involved specifying five measurable levels of attainment represented by scores of −2 to +2. A score of −2 represented the infant’s level of attainment at the start of the study, −1 represented improvement that was less than the expected level of attainment, 0 represented the expected level of attainment, and scores of +1 and +2 represented levels of attainment that exceeded expectations for the 6-month period. The second author assisted the therapists with the GAS format but not with the substance of the goals. Examples of GAS for two of the infants in this study are presented in the Appendix.

A physical therapist and an occupational therapist who did not provide treatment to any of the infants performed all of the testing. The two therapists established interrater reliability for GAS by independently scoring the performance of five subjects on both goals. Agreement was calculated using the Kappa statistic. The Kappa coefficient was 1.00, indicating exact agreement. The therapists also administered the PDMS at the beginning and end of the 6-month intervention period. The two therapists established interrater reliability by independently scoring the performance of nine subjects. The intraclass correlation coefficients (ICC[2,1]) for age-equivalent scores were .98 for the gross motor scale and .97 for the fine motor scale.

The two motor goals for each infant reflected the focus of intervention. Activities that addressed the goals were implemented by teachers, parents, and therapists, and the time spent on the activities was recorded. Twenty-four of the infants received home-based services, and the remaining infants received a combination of center- and home-based services. A detailed account of the methods of service delivery has been reported.$^{11}$

Data Analysis

The GAS T-scores were computed using the formula developed by Kiresuk and Sherman$^5$:

$$T = 50 + \frac{(10\Sigma W_i X_i)}{\sqrt{(1-r)\Sigma W_i^2 + r(\Sigma W_i)^2}}$$

where $W_i$ represents the weighting for a particular goal and $X_i$ represents the outcome score for each goal (a value of −2 to +2). The $r$ value in the formula reflects the estimated average intercorrelation for the outcome scores of multiple goals. The formula for computing the T-score assumes a correlation of .30 between goals. The actual correlation between the two goals for each infant in this study was .33, indicating the data met the assumption that outcome scores for the two goals are relatively independent of each other. In this study, the two goals for each infant received equal weight.

The Kolmogorov-Smirnov Goodness-of-Fit Test was used to examine whether the mean GAS T-score for both goals differed significantly from
a normal distribution having a mean of 50 and a standard deviation of 10. The Kolmogorov-Smirnov Goodness-of-Fit Test is a one-sample nonparametric test that can be used to compare the cumulative distribution function of a variable with a specified test distribution. The question of whether either the type or the category of goal influenced the ability of the therapists to select attainable outcomes was investigated for infants who had one goal of each type or category. Paired t tests were used to examine differences between mean GAS T-scores for (1) gross motor and fine motor goals and (2) inter-milestone and intra-milestone goals.

To determine the relationship between GAS T-scores and PDMS age-equivalent change scores, GAS T-scores for gross motor goals were correlated with Peabody gross motor scale age-equivalent change scores, and GAS T-scores for fine motor goals were correlated with Peabody fine motor scale age-equivalent change scores using the Pearson product-moment correlation coefficient.

A stepwise multiple regression was used to investigate whether infant age, motor delay at the start of the study (average of gross motor and fine motor delay); time spent by therapists, parents, and teachers on activities related to the goals; or the length of time the infant had received therapy prior to the study biased GAS T-scores. The .05 probability level was used to test for statistical significance.

**Results**

The infants' mean level of attainment was +0.6 (SD=1.4) for the first goal and +0.3 (SD=1.5) for the second goal. A histogram of the GAS T-scores is presented in the Figure. The GAS T-scores ranged from the minimal possible score of 25 to the maximal possible score of 75. Twenty infants (31%) attained a T-score of less than 50, 10 infants (15%) attained a T-score of 50, and 35 infants (54%) attained a T-score of greater than 50. The mean GAS T-score was 554 (SD=14.6), which was significantly higher than the expected mean of 50 (Z=2.48, P<.001).

For the 41 infants who had one gross motor and one fine motor goal, the mean gross motor GAS T-score was 58.0 (SD=12.5) and the mean fine motor GAS T-score was 54.4 (SD=15.2), a difference that was not significant (t=1.68, df=40, P=.10). For the 25 infants who had one inter-milestone and one intra-milestone goal, the mean intra-milestone GAS T-score was 56.0 (SD=16.3) and the mean inter-milestone GAS T-score was 50.8 (SD=14.1), a difference that was not significant (t=1.76, df=24, P=.09).

The correlation between GAS T-scores for gross motor goals and Peabody gross motor scale age-equivalent change scores was .44 (P<.001). The correlation between GAS T-scores for fine motor goals and Peabody fine motor scale age-equivalent change scores was .18 (P=.12). Peabody gross motor age-equivalent change scores explained 19% of the variance associated with gross motor GAS T-scores, whereas Peabody fine motor scale age-equivalent change scores explained only 3% of the variance associated with fine motor GAS T-scores.

Motor delay was the only variable that entered into the regression equation (F=15.7; df=1,62; P<.001). Motor delay accounted for 20% (adjusted R²=.19) of the variance attributed to GAS T-scores. Infants who were less delayed in motor development at the start of the study had higher GAS T-scores.

**Discussion**

The responsiveness of GAS is dependent on the therapist selecting goals and levels of attainment for each goal that represent clinically important change that the infant is capable of making. In our study, only one infant did not demonstrate change on either goal and 31% of the infants had GAS T-scores less than 50. In contrast, 20% of the infants attained scores of +2 for both goals and 69% had GAS T-scores of 50 or higher. The mean GAS T-score of 55.4 was significantly higher than the expected mean of 50, indicating that the motor change made by the infants exceeded the therapists' expectations.

Although the ability of therapists to select exact levels of attainment is desirable, of equal importance in examining the responsiveness of GAS is the ability of therapists to select levels...
of attainment that measure the change made by clients who do not achieve the expected outcome. Inspection of individual scores indicates that the difference between the actual and expected mean T-scores is attributable to the 13 infants who attained the maximum GAS T-score of 75. Consequently, with the possible exception of the infant who attained the minimum GAS T-score and the 13 infants who attained the maximum GAS T-score, the goals were scaled by the therapists to reflect the infants' potential for change. Furthermore, significant differences were not found between the mean GAS T-scores for either gross motor and fine motor goals or inter-milestone and intra-milestone goals. This finding indicates that neither the type nor the category of goal influenced the therapists' ability to select levels of attainment that the infants were capable of achieving. The results are encouraging given the challenge involved in predicting the development of infants whose rate and, in some cases, sequence of motor development deviated from normal.

The correlation results suggest that GAS and the PDMS measure different aspects of motor development. Items on the PDMS represent the broad spectrum of postures and movements that comprise the motor development sequence, and the age-equivalent score is based on the abilities of infants demonstrating normal development. The Peabody age-equivalent score measures change in development over time, but gains may not reflect the goals of treatment. In contrast, GAS measures attainment of a small number of individualized goals. In our study, the Peabody age-equivalent scores provided a global measure of change in motor development, whereas GAS measured change in specific postures and movements that were directly related to the goals of intervention. Our findings support the conclusion of Stephens and Hale,

Conclusions

The results of our study support the validity of GAS as a responsive measure of motor change in infants with motor delays. Goal attainment scaling has several advantages compared with the behavioral objective and is recommended for use by clinicians. Advantages include the ability to document change in infants who do not achieve or exceed the expected level of attainment, to weight the relative importance of goals, to obtain an aggregate score for multiple goals, and to obtain a T-score that provides a method for quantitating change over time and across individuals. Goal attainment scaling is also recommended for use by interdisciplinary early intervention or rehabilitation teams. Goals for multiple domains of development can be prioritized and scaled, providing a comprehensive measure of change.

There are issues to consider before using GAS in research to evaluate the effects of physical therapy. A disadvantage of GAS is the therapist's reliance on clinical judgment in determining levels of attainment that represent important changes that the child is capable of achieving. The validity of GAS is dependent on the therapist's ability to accurately assess the child and predict the expected level of change. To minimize bias, Ottenbacher and Cusick suggest that the therapist who sets the goals should not be the same therapist who provides treatment. Goals should be established prior to random group assignment and scored by an independent examiner. The method of assigning weights to goals does not address the conflict that arises when the most important goal for a child is also the most difficult. To address this concern, Clark and Caudrey have proposed computing the T-score by assigning weights to each goal according to both importance and difficulty. Only six of the goals selected by the therapists in our study were in the functional and physical parameter categories. Functional and physical parameter domains are not included on norm-referenced motor development scales and are important areas.
Appendix. Examples of Goal Attainment Scaling

Gross Motor—Inter-Milestone
Child's age: 30 months  Diagnosis: cerebral palsy

In 6 months, the child will:
-2=When placed, stand independently holding onto her walker for 5 seconds
-1=When placed, stand independently holding onto her walker and take 2-4 steps while guarded by an adult
0=Independently pull to stand on furniture and, when given her walker, take 2 to 4 steps while guarded by an adult
+1=Independently pull to stand through half-kneeling on furniture and, when given her walker, independently take 2 to 4 steps
+2=Independently pull to stand through half-kneeling on the walker and then walk 18.3 m (6 ft)

Fine Motor—Intra-Milestone
Child's age: 12 months  Diagnosis: Down's syndrome

In 6 months, the child will grasp a raisin:
-2=Using a raking pattern
-1=Between a flexed thumb and side of curled index finger
0=Between thumb and ventral surface of index finger
+1=Between distal pads of opposed thumb and index finger
+2=Between tips of opposed thumb and index finger

to evaluate in children with moderate to severe motor disabilities. Consequently, the use of GAS to measure change in these two domains should receive greater emphasis in future research. Implementation of these recommendations is encouraged to further investigate the validity of GAS as a measure of motor change in children receiving physical therapy.

Acknowledgments

We are grateful to the parents and staff involved in the following programs: Bay Cove Mental Health Center Early Intervention Program, Concord Area Early Intervention Program, Dorchester Counseling Center Early Intervention Program, First Program, Professional Center for Handicapped Children Inc, Winchester Early Intervention Program, Waltham Early Intervention Program, Waltham-Weston Hospital, Step One, and New England Memorial Hospital Early Intervention Program. Special appreciation is given to Terry Stephens and Andrea Larsen for their efforts in data collection and to Louisa Seraydarian for her statistical consultation.

References