

Validity of the Peabody Developmental Gross Motor Scale as an Evaluative Measure of Infants Receiving Physical Therapy

Background and Purpose. The purpose of this study was to examine the validity of the Peabody Developmental Gross Motor Scale (PDMS-GM) as an evaluative measure of infants receiving physical therapy. **Subjects and Methods.** Infants who attended an early intervention program (N= 124) were grouped by diagnosis: cerebral palsy, Down syndrome, hydrocephalus, preterm with developmental delay, full term with developmental delay, and other. The PDMS-GM was administered to each infant three times over a 6-month period by a therapist who did not provide treatment. **Results.** Mean scaled scores and age-equivalent scores increased for each group. Individual change was examined using the reliable change index. The results indicated that the change in total raw score for 62% of the infants was greater than what could be attributed to measurement error. When minimal clinically important change was defined as 10 scaled score points, the index of responsiveness was equal to 0.5. This finding indicates that a sample size of 68 subjects per group would be needed when the PDMS-GM is used as an outcome measure in research. **Conclusion and Discussion.** The mean change scores for each group support the use of the PDMS-GM as an evaluative measure. For many infants, particularly infants with cerebral palsy, the PDMS-GM was not responsive to change over a 6-month period. The index of responsiveness suggests that the PDMS-GM should be used only as an outcome measure in large clinical trials. The PDMS-GM is not recommended for evaluating the direct effects of physical therapy but is recommended for providing a global measure of change in motor development as part of a multidimensional assessment. [Palisano RJ, Kolobe TH, Haley SM, et al. Validity of the Peabody Developmental Gross Motor Scale as an evaluative measure of infants receiving physical therapy. *Phys Ther.* 1995; 75:939-951.]

Key Words: Developmental disability, Gross motor development, Measurement, Responsiveness, Validity.

Robert J Palisano
Thubi H Kolobe
Stephen M Haley
Linda Pax Lowes
Sandra L Jones

Selection of a developmental assessment should be dependent on the ability of the physical therapist to identify the purpose of assessment and to critically review published information for evidence of a test's validity specific to the intended use. Standardized developmental assessments are administered for several different purposes, including screening, diagnosis, determination of eligi-

bility for public-funded early intervention programs, treatment planning, and measurement of change for evaluating the effectiveness of intervention.

Discriminative measures are intended to distinguish between individuals with or without a particular characteristic or function.¹ For example, norm-referenced assessments are based on the average performance of children

without impairments or disabilities and are thus designed to discriminate between infants with and without developmental delays. The use of a discriminative measure is appropriate when the purpose of testing is to screen, diagnose, or determine eligibility for early intervention services. Alternatively, evaluative measures are used to measure the magnitude of change in development over time or after inter-

vention.¹ Criterion-referenced assessments that are based on task analysis, sequential stages, or mastery of a content domain without regard to performance of other children of the same age are preferable when the purpose of assessment is treatment planning or evaluation of the effects of treatment.²

Although norm-referenced assessments of motor development are intended to discriminate between children with and without motor delay, they have frequently been used as evaluative measures in research that has examined the effectiveness of physical therapy for infants with cerebral palsy and those at risk for neuro-motor disabilities.³⁻⁵ This practice is controversial. Rosenbaum and colleagues⁶ stated that norm-referenced motor assessments should not be used as evaluative measures until they are validated as responsive to changes made by children with motor dysfunction.

Responsiveness (ie, the ability to measure clinically important changes over time) is a type of validity that is necessary for an evaluative measure.⁷ Rosenbaum et al⁶ and Boyce et al⁸

have described several psychometric properties that are unique to a measure that is intended to evaluate motor function outcomes of intervention for children with cerebral palsy. Items must reflect the goals of intervention and the changes that children with cerebral palsy are capable of making. There should be a sufficient number of items above a child's present level of ability and a rating scale that includes several response options for partial as well as complete accomplishment of each item. The method of scoring should be quantitative and allow detection of small changes in performance.

The Peabody Developmental Gross Motor Scale (PDMS-GM)⁹ is standardized and normed for ages from birth through 83 months and is intended for use by professionals from several disciplines, including physical therapists. Folio and Fewell constructed the PDMS-GM to accomplish several purposes, including (1) identification of children with delayed motor development, (2) identification of a child's unique strengths and needs, (3) assessment of motor development over time or in response to intervention, and (4) identification of motor objec-

tives and intervention strategies when used with accompanying activity cards. The PDMS-GM, therefore, was designed for use as both a discriminative and an evaluative measure.

The PDMS-GM contains 170 items equally divided among 17 age levels. Items are grouped into five skill categories (reflexes, balance, nonlocomotion, locomotion, and receipt and propulsion of objects) that in the test authors' opinion represent the clustering of items that place similar demands on the child. Items are scored on a three-point scale (0, 1, 2), with a score of 1 indicating that the behavior is emerging but that the criterion for successful performance is not fully met. The raw score for the gross motor scale can be converted into an age-equivalent, a percentile, or a standardized score. The raw score may also be converted into a scaled score. Scaled scores are normalized raw scores that are independent of age norms and, therefore, capable of measuring small changes in motor development. Scaled scores have a mean of 500, a standard deviation of 100, and a range of 200 to 800. A scaled score of 500 indicates that a child is at the midpoint in mastery of the items on the PDMS-GM but does not indicate what items were achieved. Although Folio and Fewell state that "a change of 25 scaled score points is the same no matter where it occurs on the scale,"^{9(p124)} evidence that items are ordered from least to most difficult and that the difficulty of each successive item increases in equal increments is not provided in the test manual.

Information on reliability and content, construct, and concurrent validity of the PDMS-GM is reported in the test manual.⁹ Evidence of the ability of the PDMS-GM to discriminate between children with and without delayed motor development (except for the 0- to 5-month age levels) supports construct validity.⁹ Studies have provided further evidence of concurrent validity of the PDMS-GM with the Bayley Motor Scale¹⁰ and of interrater reliability of PDMS-GM scores for 4- and 5-year-old children with and without motor delays.¹¹ An in-depth analysis of

RJ Palisano, ScD, PT, is Associate Professor, Department of Physical Therapy (Mail Stop 502), Medical College of Pennsylvania and Hahnemann University, Philadelphia, PA 19102-1192 (USA) (palisano@Hal.Hahnemann.edu), and Co-Investigator, Neurodevelopmental Clinical Research Unit, McMaster University, Hamilton, Ontario, Canada L8S 4K1. Address all correspondence to Dr Palisano at the first address.

TH Kolobe, PhD, PT, is Assistant Professor, Department of Physical Therapy, University of Illinois at Chicago, Chicago IL 60612.

SM Haley, PhD, PT, is Coordinator of Staff Development, Education, and Research, Rehabilitation Services, New England Rehabilitation Hospital, Woburn, MA 01801.

L Pax Lowes, PT, is a doctoral candidate, Program in Pediatric Physical Therapy, Hahnemann University.

SL Jones, PT, is Physical Therapist, Bridgeton Public Schools, Bridgeton, NJ 08302.

Support for this study was provided by a Career Scientist Award, Ontario Ministry of Health, awarded to Dr Palisano, a grant from the Foundation for Physical Therapy Inc awarded to Dr Haley, and a clinical research grant from the Section on Pediatrics, American Physical Therapy Association, awarded to Dr Kolobe. Ms Pax Lowes' participation was funded by a Leadership Preparation Grant, US Department of Education, Office of Special Education and Rehabilitation Services, awarded to Hahnemann University.

The results of this study were presented in platform format at the Second Joint Congress of the Canadian Physiotherapy Association and the American Physical Therapy Association, June 4-8, 1994; Toronto, Ontario, Canada.

The study was approved by the Hahnemann University Committee for Human Subjects.

This article was submitted October 21, 1994, and was accepted July 25, 1995.

the strengths and limitations of the PDMS-GM when used as a discriminative measure is provided by Hinderer et al.¹²

Although Folio and Fewell⁹ state that a purpose of the PDMS-GM is to measure change across time or after intervention for children with motor impairments or delays and the PDMS-GM has been used in physical therapy outcome research,^{13,14} the responsiveness of the PDMS-GM has not been investigated. The purpose of our study was to examine the validity of the PDMS-GM as an evaluative measure of infants receiving physical therapy. We were interested in examining change from three perspectives: mean change, individual change, and minimal clinically important change. The following research questions were examined: (1) Do infants grouped by diagnosis demonstrate an increase in PDMS-GM scaled scores and age-equivalent scores when tested three times over a 6-month period? (2) What percentage of infants demonstrate an increase in PDMS-GM raw score that exceeds the change that could be explained by measurement error? and (3) What is the sample size requirement when the PDMS-GM is used as an evaluative measure in treatment outcome research? The third question was addressed in two phases, which involved estimation of minimal clinically important change and determination of the index of responsiveness.⁷

Method

Subjects

The subjects were 124 infants who met the following eligibility criteria: (1) were enrolled in an early intervention program in which they received physical therapy, (2) had a motor delay as measured by a z score of -1.5 or below on the PDMS-GM, (3) did not have a medical condition that prevented participation in physical therapy, and (4) had no progressive neurological disorder or medical condition in which progress in motor development would not be expected over a 6-month period. Sixty-one infants attended one of nine early interven-

Table 1. Mean Age and Percentage of Gross Motor Delay at the Initial Test Session by Diagnosis

Diagnosis	N	Age (mo) ^a			Percentage of Motor Delay ^b		
		\bar{X}	SD	Range	\bar{X}	SD	Range
Cerebral palsy	36	16.6	6.7	28	65%	22%	74%
Developmental delay-preterm	21	12.8	6.3	23	41%	30%	84%
Developmental delay-full term	20	19.6	7.4	26	46%	19%	77%
Down syndrome	19	16.8	7.7	28	50%	14%	58%
Hydrocephalus	9	14.2	5.5	14	74%	17%	54%
Other	19	15.7	6.0	21	60%	23%	82%

^aAges of preterm infants adjusted for gestational age.

^b100% - (Peabody Developmental Gross Motor Scale age-equivalent score/Age).

tion programs in the Boston (Mass) metropolitan area, 42 infants attended one of six early intervention programs in the Chicago (Ill) metropolitan area, and 21 infants attended an early intervention program in Cumberland County, New Jersey. Informed consent of a parent or guardian was obtained for each infant.

The sample comprised 76 male infants and 48 female infants who ranged in age from 2 to 33 months (\bar{X} =16.2, SD=6.9) at the start of the study. The ages of the 54 infants born preterm were adjusted to account for gestational age at birth. One hundred infants were Caucasian, and 19 infants were African-American. The race of 5 infants was not reported.

Subjects were grouped by diagnosis based on review of intake forms and medical reports contained in each infant's early intervention program file (Tab. 1). Thirty-six infants had cerebral palsy, 19 infants had Down syndrome, and 9 infants had hydrocephalus. Twenty-one preterm infants and 20 full-term infants had a developmental delay in two or more domains, with no specific diagnosis. The criterion for developmental delay varied among the early intervention programs, but developmental delay was generally defined as an age-equivalent score that was a minimum of 25% below the

infant's chronological or adjusted age. The remaining 19 infants constituted a heterogeneous group that had the following diagnoses: spina bifida (n=5), genetic syndrome other than Down syndrome (n=4), microcephaly (n=4), congenital infection (n=2), congenital central nervous system malformation (n=1), macrocephaly (n=1), autism (n=1), and head trauma (n=1).

The mean age of each group ranged from 12.8 to 19.6 months (Tab. 1). The results of a one-way analysis of variance (ANOVA) indicated an age effect ($F=2.30$; $df=5,118$; $P=.05$). *Post hoc* analysis using Tukey's honestly significant difference multiple-comparison test indicated that the mean age of the full-term infants with developmental delays was greater than that of infants with other diagnoses ($P<.05$), whereas the mean age of the preterm infants with developmental delays was less than that of infants with other diagnoses, except for infants with hydrocephalus ($P<.05$). The mean percentage of gross motor delay at the start of the study ranged from 41% to 74% among the groups (Tab. 1).

This study did not investigate the effectiveness of physical therapy; therefore, no attempts were made to control the methods of service delivery, treatment frequency, or the treat-

ment approaches utilized among the physical therapists. In general, infants were seen by a physical therapist between twice a month and twice a week. The type of service delivery varied and included direct physical therapy in an isolated setting, group therapy, therapy that was integrated into educational and play activities, and consultation to parents and early intervention team members. The services received by infants who attended the early intervention programs in Massachusetts and New Jersey have been reported.^{15,16}

Procedure

The PDMS-GM was administered to the infants at the start of the study and at 3-month and 6-month intervals following the standardized procedures outlined in the test manual. All of the testing was performed by therapists who did not provide treatment to any of the infants. A physical therapist or an occupational therapist administered all of the assessments. Sixty-six percent of the assessments were performed at the center at which the infants received services, and 34% of the assessments were performed in the infants' homes.

The occupational therapist and the physical therapist who administered the PDMS-GM to the infants who attended the early intervention programs in Massachusetts established interrater reliability by independently scoring the performance of nine subjects prior to data collection. The intraclass correlation coefficient (ICC[2,1]) for age-equivalent scores was .98.

The first author (RJP) established interrater reliability with the author (THK) who administered the PDMS-GM to the infants in Illinois and the author (SLJ) who administered the PDMS-GM to the infants in New Jersey. Prior to testing the infants who lived in Illinois, six infants were independently scored (three from videotape and three from direct observation) and the ICC (2,1)

for age-equivalent scores was .95. To determine test-retest reliability, the author who administered the PDMS-GM to the infants in Illinois tested six infants twice within 1 week. The ICC (3,1) for age-equivalent scores was .99.

A similar procedure was followed prior to testing the infants who lived in New Jersey. Six infants were independently scored, three from videotape. The ICC (2,1) for raw scores of items administered was .96. One week later, the therapist readministered the PDMS-GM to three of the infants to determine test-retest reliability. The ICC (3,1) for raw scores of items administered was .98.

Data Analysis

The PDMS-GM raw scores were converted into scaled scores and age-equivalent scores using the tables contained in the test manual. Statistical analyses were performed using SPSS/PC+, version 3.0.*

The distributions of scaled scores and age-equivalent scores for each group at each of the three test sessions were examined for skewness and kurtosis using the *z* distribution and an alpha level of .001.¹⁷ The distributions of age-equivalent scores did not demonstrate skewness or kurtosis for any group. The distributions of scaled scores demonstrated skewness, kurtosis, or both, for three groups. For the preterm infants with developmental delays, the distribution of scaled scores for the 6-month test session demonstrated positive kurtosis and negative skewness. For the full-term infants with developmental delays and the infants with hydrocephalus, the distribution of scaled scores demonstrated positive kurtosis for two of the test sessions. Overall, the assumption of a normal distribution was not violated, and the data were analyzed using parametric statistics.

For each group, a one-way ANOVA for repeated measures was used to analyze differences in mean scaled scores and age-equivalent scores for the three test sessions. For all ANOVAs, *post hoc* analyses of effects were performed using the Tukey honestly significant difference multiple-comparison test. The .05 probability level was used to test for statistical significance.

The reliable change index (RCI) developed by Jacobson et al¹⁸ and modified by Christensen and Mendoza¹⁹ was used to examine whether each infant's change in raw score for the 6-month period exceeded what could be explained by measurement error. Application of the modified RCI to measurement of change in children receiving therapy services is discussed by Ottenbacher et al.²⁰ The formula for the modified RCI is

$$(1) RCI = \frac{\text{posttest score} - \text{pretest score}}{\sqrt{S_{\text{diff}}}}$$

where the S_{diff} is the standard error of the difference between pretest and posttest scores that would be expected if no actual change had occurred.

The S_{diff} for a particular test is computed from the standard error of measurement for a 95% confidence interval. The S_{diff} for the PDMS-GM was calculated using the formula:

$$(2) S_{\text{diff}} = \sqrt{(SE_1)^2 + (SE_2)^2}$$

where SE_1 is the standard error of measurement of the raw score for the age level of the PDMS-GM that corresponded to the infant's age at the initial assessment and SE_2 is the standard error of measurement for the raw score for the age level that corresponded to the infant's age at the 6-month assessment. Jacobson et al¹⁸ proposed that an RCI larger than ± 1.96 would be unlikely to occur without actual change ($P < .05$).

The index of responsiveness⁷ was calculated to determine sample size requirements when the PDMS-GM is used as an outcome measure in clini-

*SPSS Inc, 444 N Michigan Ave, Chicago, IL 60611.

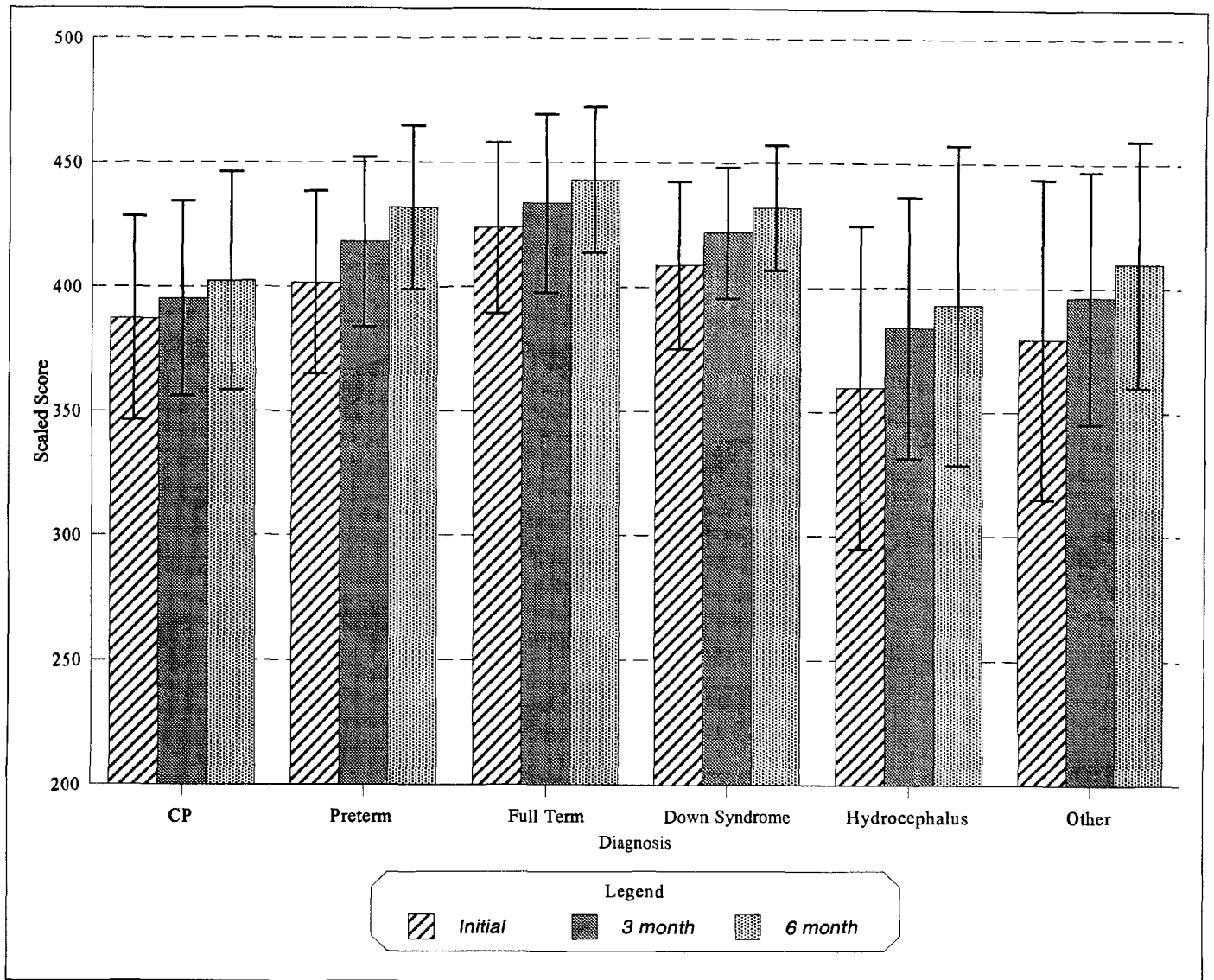


Figure 1. Mean Peabody Developmental Gross Motor Scale scaled scores for initial, 3-month, and 6-month assessments by diagnosis.

cal research. The equation for the index of responsiveness is

$$(3) \quad R = \frac{\Delta}{\sqrt{2 \times \text{MSE}}}$$

The numerator Delta (Δ) is the minimal clinically important change. The denominator is derived from the mean square error (MSE) of a univariate repeated-measures ANOVA and reflects within-subject variability. Delta has not been determined for the PDMS-GM scaled score. In this study, Delta was estimated as 10 points. This estimate is roughly equivalent to a 1-month gain in age-equivalent score

for the age levels of the infants in this study. The rationale for the estimate was also based on pilot work that involved having two physical therapists independently view videotaped excerpts of consecutive test sessions for five infants who made varying amounts of change during this study. The two infants judged as having made minimal clinically important change by both therapists had a 1-month gain in age-equivalent score.

Results

Mean scaled scores and age-equivalent scores for each group are presented in

Figures 1 and 2. For all groups, mean scaled scores and age-equivalent scores were higher for each successive test session. The mean change in scaled scores and age-equivalent scores for the 6 months for each group are presented in Table 2. The mean change in scaled scores ranged from 15.0 (infants with cerebral palsy) to 33.3 (infants with hydrocephalus). The mean change in age-equivalent scores ranged from 2.2 months (infants with cerebral palsy) to 4.0 months (preterm infants with developmental delays).

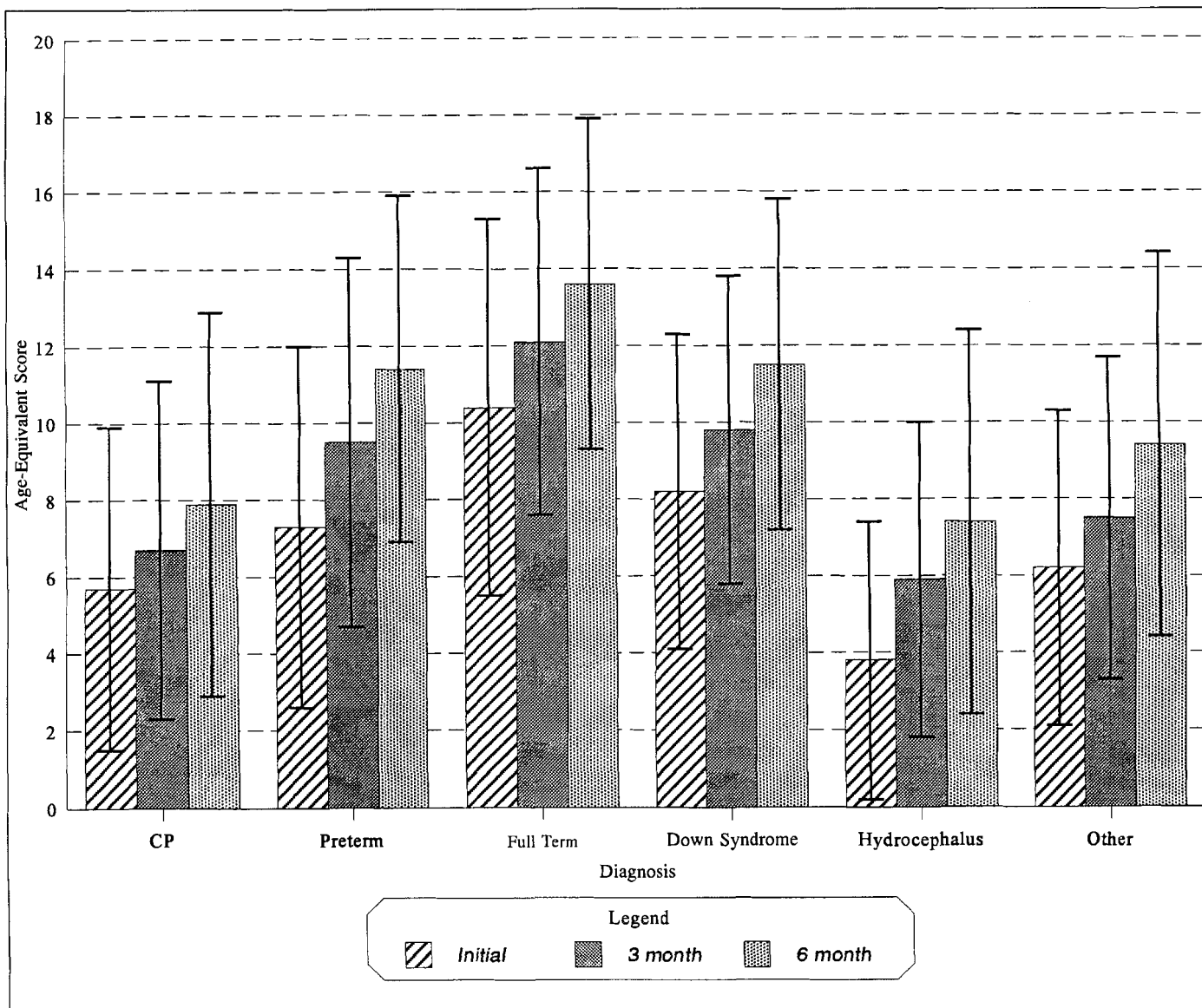


Figure 2. Mean Peabody Developmental Gross Motor Scale age-equivalent scores for initial, 3-month, and 6-month assessments by diagnosis.

Table 2. Mean Change in Peabody Developmental Gross Motor Scale Scaled Scores and Age-Equivalent Scores for the 6-Month Period by Diagnosis

Diagnosis	N	Scaled Score			Age-Equivalent Score		
		\bar{X}	SD	Range	\bar{X}	SD	Range
Cerebral palsy	36	15.0	14.7	65	2.2	1.7	6.0
Developmental delay-preterm	21	30.1	17.0	67	4.0	2.1	9.0
Developmental delay-full term	20	19.3	15.3	55	3.2	2.0	6.0
Down syndrome	19	23.1	15.5	64	3.3	1.7	5.5
Hydrocephalus	9	33.3	19.8	61	3.7	2.1	9.0
Other	19	30.2	27.8	122	3.2	2.1	8.0

The results of the one-way ANOVAs for repeated measures are summarized in Table 3. Both mean scaled scores and mean age-equivalent scores increased for each group. The results of the Tukey multiple-comparison tests indicated that mean scaled scores and mean age-equivalent scores were higher for each successive test session for all groups, except infants with hydrocephalus ($P < .05$). The mean scores of infants with hydrocephalus increased between the initial and 3-month test sessions and the initial and 6-month test sessions but not

Table 3. Results of One-Way Analyses of Variance for Repeated Measures Done to Examine Differences in Mean Peabody Developmental Gross Motor Scale Scaled Scores and Age-Equivalent Scores for the Three Test Sessions by Diagnosis

Diagnosis	df	Scaled Score		Age-Equivalent Score	
		F	P	F	P
Cerebral palsy	2,70	31.75	<.001	49.29	<.001
Developmental delay-preterm	2,40	55.91	<.001	59.22	<.001
Developmental delay-full term	2,38	20.92	<.001	38.92	<.001
Down syndrome	2,36	31.95	<.001	41.23	<.001
Hydrocephalus	2,16	16.02	<.01	8.90	<.001
Other	2,36	17.11	<.001	29.46	<.001

between the 3- and 6-month test sessions.

A graph of the RCI value for each infant's initial and 6-month raw scores rank ordered by group is presented in Figure 3. The RCI values ranged from -0.46 to 9.06. Seventy-seven infants (62%) had an RCI greater than 1.96, indicating that their change in raw score was greater than what could potentially be attributed to measurement error ($P < .05$). The RCI of the remaining 47 infants (38%) was less than 1.96, indicating that their change in raw score could potentially represent measurement error and not actual change. The percentage of infants in each group with an RCI of greater than 1.96 was 86% for preterm infants with developmental delays, 63% for

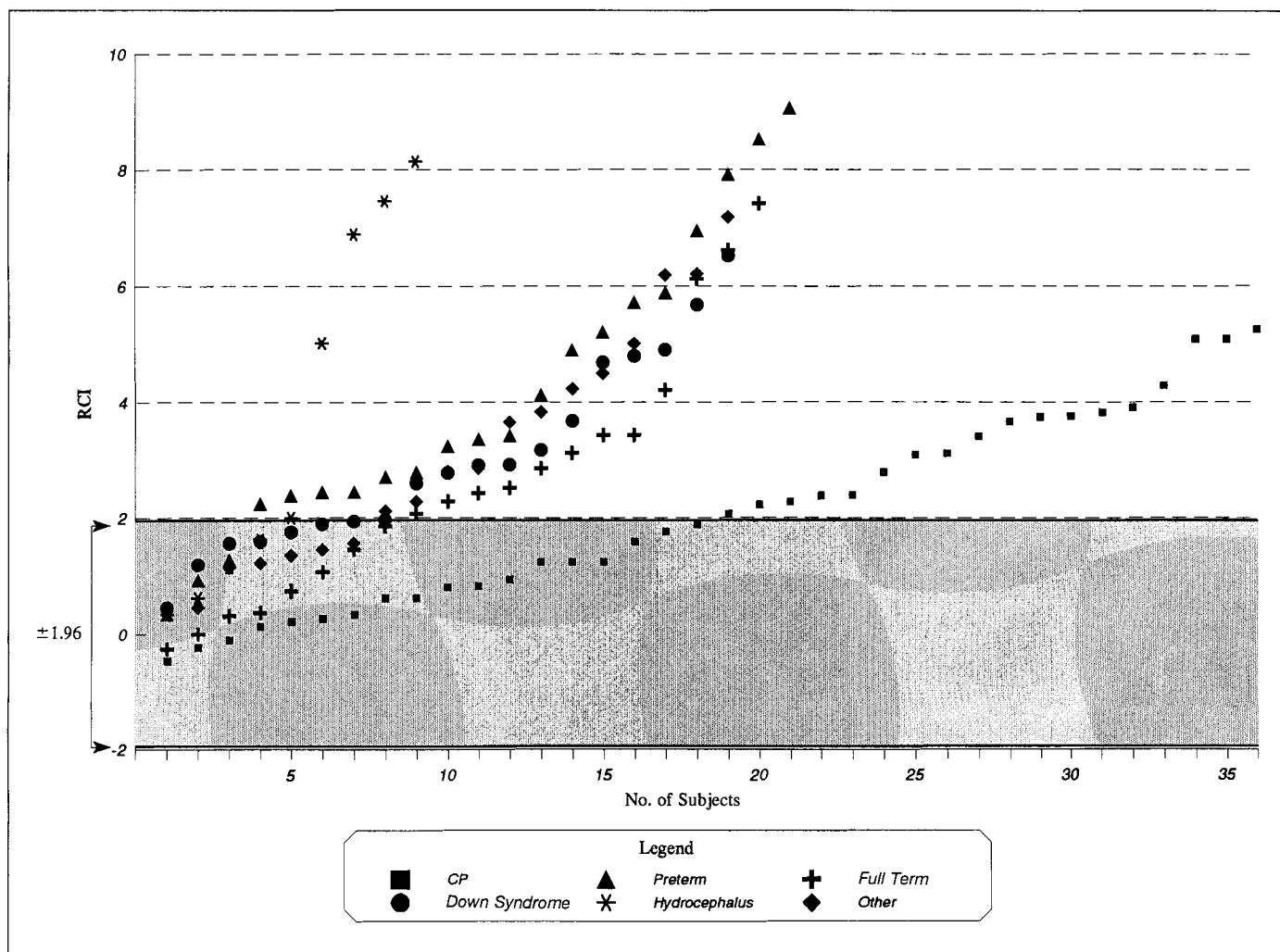


Figure 3. Reliable change index (RCI) values of individual infants rank ordered by diagnosis.

infants with Down syndrome, 63% for infants with other diagnoses, 60% for full-term infants with developmental delays, 56% for infants with hydrocephalus, and 50% for infants with cerebral palsy.

The equation for the index of responsiveness (R) was solved when the numerator Delta (minimal clinically important change) was equal to 10 scaled score points (equivalent to a 1-month gain in age-equivalent score). The MSE for the repeated-measures ANOVA for initial and 6-month scaled scores for the entire sample was 182.01. Entering this value into the responsiveness equation resulted in a denominator of 19.1 and an R of 0.5. Using a table of sample size requirements for various responsiveness levels, a sample size of 68 subjects per group would be required when $R=0.5$.⁷ This sample size requirement is based on an alpha of .05 (one tailed), a beta of .10, and the assumption that the samples are independent.⁷

Discussion

The purpose of this study was to examine the validity of the PDMS-GM when used as an evaluative measure in infants receiving physical therapy from three perspectives: mean change, individual change, and minimal clinically important change. The changes in mean scores for each group provide initial support for the validity of the PDMS-GM when used as an evaluative measure for infants receiving physical therapy. Although the scaled score is recommended when the purpose of testing is to measure change in gross motor development, the infants in this study also made gains in age-equivalent scores. Although an increase in raw score of 1 or 2 points increased the scaled score, a gain of between 5 and 12 points was necessary to increase the age-equivalent scores achieved by the infants in this study. The finding that the mean age-equivalent scores for each group increased between a mean of 1.0 and 2.1 months for the 3-month intervals and between a mean of 2.2 and 4.0 months for the 6-month interval sug-

gests that the changes were of clinical importance.

Each subject's change in PDMS-GM raw score was also examined because decisions in clinical practice are made on an individual basis. Furthermore, Jacobson et al¹⁸ contend that statistical analyses based on group means provide no information on the proportion of subjects who benefited from the treatment. The RCI was developed as one criterion for determining improvement in individual clients based on the premise that for individual change to be considered statistically significant, it must also be statistically reliable.¹⁸ The RCI values for this study indicate that 62% of the infants demonstrated an increase in raw score for the 6-month period that could not be attributed to either random variation or measurement error associated with the PDMS-GM. Although the majority of infants made true gains in raw scores, the changes made by 38% of the infants were not large enough to rule out measurement error. This finding suggests that for many infants, particularly those with neuromotor impairments such as cerebral palsy, the PDMS-GM is not responsive to change over a 6-month period. Therapists, therefore, should take into consideration the standard error of the difference between initial and follow-up assessment scores to determine whether the change in PDMS-GM raw score made by individual children exceeds what could potentially be attributed to measurement error.

The results suggest that the PDMS-GM has limitations when used as an evaluative measure in infants with cerebral palsy. The mean scaled scores and age-equivalent scores of infants with cerebral palsy increased over a 6-month period; however, the amount of change was less than the changes made by the other groups. Furthermore, only 50% of infants with cerebral palsy made changes in raw score that exceeded what could potentially be attributed to random variation or measurement associated with the PDMS-GM. The results are not surprising given the number and severity of motor impairments associated with

cerebral palsy and the fact that the construct of the PDMS-GM is based on motor development of children without motor delays or impairments. The Gross Motor Function Measure (GMFM)²¹ was recently designed and validated to measure change in gross motor function in children with cerebral palsy. Kolobe²² compared the GMFM and the PDMS-GM and reported that although infants made comparable changes on both measures, the construct, large number of items in the lying and rolling, sitting, and crawling and kneeling domains; a four-point rating scale, and separate items for the right and left sides suggest that the GMFM is the preferred measure for evaluating change in infants with cerebral palsy.

What constitutes minimal clinically important change for a 6-month period has not been determined for the PDMS-GM. Our estimate of minimal clinically important change was based on the impression that a gain in scaled score of 10 points would require the infant to demonstrate emerging ability or acquisition of several postures and movements that parents and health care professionals would perceive as important for motor function. When minimal clinically important change was defined as 10 scaled score points, the sample size requirement of 68 subjects per group exceeds the sample size of all but one intervention study that included a control or comparison group.^{5,23-25} This finding suggests that for research purposes, the PDMS-GM is appropriate for use as an outcome measure only when there is a large sample size or when evidence exists that the intervention will produce a moderate or large treatment effect.

Although minimal clinically important change was used in this study to determine sample size, the concept has broader implications that have not been addressed in physical therapy outcome research. Changes that are statistically significant are not necessarily of clinical importance, and vice versa. Jaeschke et al²⁶ define minimal clinically important difference as the smallest difference in score that patients perceive as beneficial and that

would mandate a change in the patients' management. Based on this definition, the question of what constitutes minimal clinically important change must reflect the perceptions of families and children who receive physical therapy. Ultimately, resolution of this issue may depend on the cost-to-benefit ratio of providing children and their families the services associated with achievement of minimal clinically important change.

The results of this study cannot be generalized to children with motor impairments who are above the age of 3 years. Subjects were selected based on the assumption that infants with motor impairments would be more likely to demonstrate changes on the PDMS-GM compared with children above 3 years of age. This assumption is based on the expectation that there are a sufficient number of items on the PDMS-GM below the 12- to 14-month age level to measure changes in postures and movements in supine, prone, quadruped, and sitting positions made over a 3- to 6-month period in infants with mild to moderate motor impairments, including infants with limited potential for short-term changes in standing and walking. For children with motor impairments who are 3 years of age and older, many items below the 12- to 14-month age level no longer reflect the focus of physical therapy, and there is a ceiling effect above this age level for children who are unable to stand and walk independently.

Although the mean change scores and to a lesser extent the individual change scores support the use of the PDMS-GM as an evaluative measure, the gains measured may not be directly related to the goals of physical therapy. The PDMS-GM is based on the motor development sequence and the change needed to pass successive items for a posture or movement may exceed the potential of infants with motor impairments and disabilities, especially when change is measured over a short period of time. The PDMS-GM was not constructed to measure changes in quality of movement, amount of caregiver assistance,

the need for assistive devices and orthoses, or changes in motor function within contexts that are important for daily routines. When these areas are the focus of physical therapy, alternative or additional evaluative measures are recommended. Previous research supports the use of individualized outcome measures such as goal attainment scaling to measure small, but clinically meaningful, changes that are directly related to the focus of physical therapy.^{16,27} We recommend using the PDMS-GM to provide a global measure of change in motor development but not as the primary method of evaluating change in infants receiving physical therapy, especially infants with cerebral palsy.

Conclusions

This study was an initial attempt to examine the validity of the PDMS-GM when used as an evaluative measure of infants receiving physical therapy. Infants grouped by diagnosis made gains in mean scaled scores and age-equivalent scores over a 6-month period, and the magnitude of the mean changes was judged as clinically important. These findings provide support for use of the PDMS-GM as an evaluative measure. The measurement error associated with change scores of individual subjects was examined using the RCI. The gains in raw score made by 38% of the infants were not of sufficient magnitude to conclude that an actual change in motor development had occurred. This finding suggests that for many infants who receive physical therapy, the PDMS-GM is not responsive to changes made over a 6-month period. In particular, the changes made by infants with cerebral palsy were less than the changes made by the other groups, and only 50% of infants with cerebral palsy made changes in raw score that exceeded what could potentially be attributed to random variation or measurement error associated with the PDMS-GM.

The concept of minimal clinically important change has important implications for treatment outcome research and warrants further investigation.

When minimal clinically important change for a 6-month period is defined as 10 scaled score points, the sample size requirement would be 68 subjects per group. This requirement suggests that the PDMS-GM should be used as an outcome measure only in large clinical trials or when evidence exists that the intervention will produce a moderate or large treatment effect. In accordance with the recommendations of Neisworth and Baginato,²⁸ we advocate that the PDMS-GM should be used as a global measure of change in motor development as part of an assessment that includes multiple dependent measures and derives data from multiple sources and contexts.

Acknowledgments

The assistance of Louisa Seraydarian, PhD, with data analysis and the advice of Kenneth J Ottenbacher, PhD, on use of the RCI are greatly appreciated.

References

- 1 Kirshner B, Guyatt GH. A methodological framework for assessing health indices. *J Chronic Dis.* 1985;38:27-36.
- 2 Montgomery PC, Connolly BH. Norm-referenced and criterion-referenced tests: use in pediatrics and application to task analysis of motor skill. *Phys Ther.* 1987;67:1873-1876.
- 3 Palmer FB, Shapiro BK, Wachtel RC, et al. The effects of physical therapy on cerebral palsy: a controlled trial in infants with spastic diplegia. *N Engl J Med.* 1988;318:803-808.
- 4 Carlsen PN. Comparison of two occupational therapy approaches for treating the young cerebral-palsied child. *Am J Occup Ther.* 1975;29:268-272.
- 5 Piper MC, Kunos VI, Willis DM, et al. Early physical therapy effects on the high risk infant: a randomized controlled trial. *Pediatrics.* 1986;78:216-224.
- 6 Rosenbaum PL, Russell DJ, Cadman DT, et al. Issues in measuring change in motor function in children with cerebral palsy: a special communication. *Phys Ther.* 1990;70:125-131.
- 7 Guyatt GH, Walter SD, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis.* 1987;40:171-180.
- 8 Boyce WF, Gowland C, Rosenbaum PL, et al. Measuring quality of movement in cerebral palsy: a review of instruments. *Phys Ther.* 1991;71:813-819.
- 9 Folio MR, Fewell RR. *Peabody Developmental Motor Scales and Activity Cards.* Dallas, Tex: DLM-Teaching Resources; 1983.
- 10 Palisano RJ. Concurrent and predictive validities of the Bayley Motor Scale and the Pea-

body Developmental Motor Scales. *Phys Ther*. 1986;66:1714-1719.

11 Schmidt LS, Westcott SL, Crowe TK. Interrater reliability of the gross motor scale of the Peabody Developmental Motor Scales with 4- and 5-year-old children. *Pediatric Physical Therapy*. 1993;5:169-175.

12 Hinderer KA, Richardson PK, Atwater SW. Clinical implications of the Peabody Developmental Motor Scales: a constructive review. *Physical & Occupational Therapy in Pediatrics*. 1989;9:81-106.

13 Harris SR. Effects of neurodevelopmental therapy on motor performance of infants with Down's syndrome. *Dev Med Child Neurol*. 1981;23:477-483.

14 Jenkins JR, Sells CJ, Brady D, et al. Effects of developmental therapy on motor impaired children. *Physical & Occupational Therapy in Pediatrics*. 1982;2:19-28.

15 Haley SM, Stephens TE, Larsen AM. Patterns of physical and occupational therapy implementation in early motor intervention. *Topics in Early Childhood Special Education*. 1988;7(4):46-63.

16 Palisano RJ. Validity of goal attainment scaling in infants with motor delays. *Phys Ther*. 1993;73:651-658.

17 Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 2nd ed. New York, NY: Harper & Row, Publishers Inc; 1989:72-74.

18 Jacobson NS, Follette WC, Revenstorf D. Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance. *Behavior Therapy*. 1984;15:336-352.

19 Christensen L, Mendoza JL. A method of assessing change in a single subject: an alteration of the RC index. *Behavior Therapy*. 1986;17:305-308.

20 Ottenbacher KJ, Johnson MB, Hojem M. The significance of clinical change and clinical change of significance: issues and methods. *Am J Occup Ther*. 1988;42:156-163.

21 Russell D, Rosenbaum P, Gowland C, et al. *Gross Motor Function Measure Manual*. 2nd ed. Owen Sound, Ontario, Canada: c/o Mary Lane, Pediatric Physiotherapy Services, RR #4; 1993.

22 Kolobe THA. *Responsiveness and Concurrent Validity of the Peabody Developmental Motor Scale, the Bayley Motor Scale, and the Gross Motor Function Measure in Infants With Cerebral Palsy and Infants With Motor Delays*. Philadelphia, Pa: Hahnemann University; 1993. Doctoral dissertation.

23 Palisano RJ. Research on the effectiveness of neurodevelopmental treatment. *Pediatric Physical Therapy*. 1991;3:143-148.

24 Piper MC. Efficacy of physical therapy: rate of motor development in children with cerebral palsy. *Pediatric Physical Therapy*. 1990;2:126-130.

25 Campbell SK. Efficacy of physical therapy in improving postural control in cerebral palsy. *Pediatric Physical Therapy*. 1990;2:135-140.

26 Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407-415.

27 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.

28 Neisworth JT, Bagnato SJ. Assessment in early childhood special education. In: Odom SL, Karnes MB, eds. *Early Intervention for Infants and Children With Handicaps*. Baltimore, Md: Paul H Brookes Publishing Co Inc; 1988:23-49.

Invited Commentary

The desire to create and validate an instrument that can be used with confidence to conclusively demonstrate the effectiveness of treatment in infants receiving physical therapy is a laudable goal. Palisano et al have shown the merit of a sequential, step-by-step process in establishing the measurement properties of the Peabody Developmental Gross Motor Scale (PDMS-GM) for this purpose. They describe a detailed study in which they answer key questions for evaluative measures: Does the measure show changes in group scores over time? Do most infants show changes in their individual scores? Is the measure responsive to clinically important differences in gross motor skills, and can data from the study be used for sample-size calculations in clinical trials? This article, however, should be making only tentative conclusions for a variety of methodological reasons.

In establishing validity of an evaluative measure, it is commonly agreed that a new measure must demonstrate its

ability to record changes in subjects who are *actually* changing. Without a "gold standard" to assess whether individuals in the sample are truly changing, one cannot conclude that the change scores that may be observed actually reflect true change.¹ One may merely be observing measurement error or random variation in gross motor skills unless the study also collects data from an external measure that is applied concurrently with the test measure. Alternatively, one could divide the sample into groups known to be "stable" and "changing" by some other criterion, such as parent or therapist report, and then compare the ability of the test to show differential change between the groups.^{2,3} Finally, one could apply the test measure to a sample that receives an intervention of known efficacy and observe whether the test measure correctly detects a change in motor skills.

The authors have not incorporated any of these comparative techniques in the study, and thus we are left

wondering whether the observed change scores represented true change in motor skills. They appear to rely on the fact that the sample was recruited from children receiving therapy, as if that guaranteed that the children would actually be changing their skills. Even this assumption, however, does not appear to hold for the reported 38% of children who did not record a criterion reliable change index value of 1.96 and thus whose change scores may have been due to measurement error. It is questionable whether the authors' conclusion that the PDMS-GM can detect true change in motor skills in the diagnostic groups can be fully supported.

A similar lack of comparative information is evident in the sample selection. The authors use the PDMS-GM itself to classify children with motor delay. They also report, however, that the measure has not been validated for this purpose in the 0- to 5-month-old population. If the PDMS-GM is not sensitive in discriminating delay in this