
Hsiang-Hui Wang, Hua-Fang Liao and Ching-Lin Hsieh

PHYS THER. 2006; 86:1351-1359.

The online version of this article, along with updated information and services, can be found online at: http://ptjournal.apta.org/content/86/10/1351

Collections
This article, along with others on similar topics, appears in the following collection(s):
- Cerebral Palsy
- Cerebral Palsy (Pediatrics)
- Tests and Measurements

E-Letters
To submit an e-Letter on this article, click here or click on "Submit a response" in the right-hand menu under "Responses" in the online version of this article.

E-mail alerts
Sign up here to receive free e-mail alerts

Background and Purpose. The psychometric properties of the Peabody Developmental Motor Scales–Second Edition (PDMS-2), a revised motor test to assess both gross motor and fine motor composites in children with cerebral palsy (CP), are largely unknown. The purpose of this study was to examine the test-retest reliability and the responsiveness of the PDMS-2 for children with CP.

Subjects. A sample of 32 children who had CP (age = 27–64 months) and who received intervention participated in this study.

Methods. The PDMS-2 was administered to each child 3 times (at the beginning of the study, at 1 week, and at 3 months later) by a physical therapist. The agreement between the first 2 measurements was used to examine the reliability. The change between the first and the third measurements was used to examine the responsiveness.

Results. The composite scores on the PDMS-2 had good test-retest reliability (intraclass correlation coefficient = 0.88–1.00). The sensitivity-to-change coefficients ranged from 1.6 to 2.1, and the responsiveness coefficients ranged from 1.7 to 2.3.

Discussion and Conclusion. Our results provide strong evidence that the 3 composites of the PDMS-2 had high test-retest reliability and acceptable responsiveness. The PDMS-2 can be used as an evaluative motor measure for children with CP and aged 2 to 5 years. [Wang HH, Liao HF, Hsieh CL. Reliability, sensitivity to change, and responsiveness of the Peabody Developmental Motor Scales–Second Edition for children with cerebral palsy. Phys Ther. 2006;86:1351–1359.]

Key Words: Cerebral palsy, Measurement: applied, Measurement: basic theory and science, Reliability of results.

Hsiang-Hui Wang, Hua-Fang Liao, Ching-Lin Hsieh
During the past 20 years, physical therapists have had considerable interest in the development and evaluation of health status outcome measures. Outcome measures are used by researchers and clinicians to assess changes in patients' abilities before and after health care to promote the accountability of health care services. Outcome measures must have the psychometric properties of reliability and responsiveness. The low intrasubject variation in stable subjects reflects the test-retest reliability of a measure. Only measures with high test-retest reliability can detect real change and reduce the bias caused by measurement error. The responsiveness of a measure is defined as the ability to assess clinically important change over time. Thus, evidence supporting the test-retest reliability and responsiveness of an outcome measure must be established before its use in research or clinical settings.

Cerebral palsy (CP) describes a group of disorders of the development of movement and posture, causing activity limitation, that are attributed to nonprogressive disturbances that occurred in the developing fetal or infant brain. To evaluate the effectiveness of treatment for the motor domain, clinicians need a motor evaluative tool. The Gross Motor Function Measure (GMFM) and the Peabody Developmental Motor Scales (PDMS) are the 2 most well-known motor instruments for children with CP. However, the GMFM measures the gross motor domain only. For measurement of the fine motor domain, the GMFM is inadequate as an evaluative tool.

A previous responsiveness study with the gross motor (GM) composite of the PDMS (PDMS-GM) for infants with CP showed that the PDMS-GM had limitations when used as an evaluative measure for infants with CP. The PDMS has been revised to the Peabody Developmental Motor Scales–Second Edition (PDMS-2), with new norms, revised testing materials, more precise scoring criteria, and more information on norm samples. Each item of the PDMS-2 was evaluated with both conventional item analyses and modern differential item functioning analyses to select the appropriate items. New normative data on the PDMS-2 were collected through 1997 and 1998 for a sample of 2,003 children residing in the United States and Canada; not only children without disabilities but also 10% of children with various types of disabilities were included in the sample. There are also more reliability and validity data for the PDMS-2 than for the PDMS. Therefore, the PDMS-2 is potentially appropriate for investigating the progress of the gross and fine motor domains for children with CP because it assesses both GM and fine motor (FM) composites and incorporates both quantitative and qualitative rating criteria.

The concurrent validity studies of the standard scores on the PDMS-2 showed high correlations with the PDMS or the Mullen Scales of Early Learning: AGS Edition in the GM or the FM composite (r = .80–.91) for children for whom detailed information on health conditions was not available. For children with developmental delays, although the developmental quotients (DQ) of the PDMS-2 were significantly correlated with the Bayley Scales of Infant Development–Second Edition, the classification agreement between these 2 tests was poor. The construct validity of the PDMS-2 was established by confirmatory factor analyses, and the results showed that the GM and the FM composites are 2 separate constructs within general movement. Another construct validity study of the PDMS-2 demonstrated high correlations...
between age and subtest raw scores. One recent study showed that the overall diagnostic accuracy of the PDMS-2 was high, with an area under the receiver operating characteristic curve of 0.98 for children with motor disabilities. These results indicate that clinicians could diagnose motor disabilities correctly 98% of the time with the test results of the PDMS-2. One of the purposes of the PDMS-2 is to evaluate a child’s progress after intervention. However, norm-referenced motor assessments should not be used as evaluative measures until they are validated to have acceptable responsiveness for children with motor dysfunction. Because the responsiveness of the PDMS-2 for children with CP is still unknown, one purpose of this study was to investigate the responsiveness of the PDMS-2 for children with CP.

The reliability of PDMS-2 scores was investigated by the test developers. In their study, 3 types of error variance—internal consistency, test-retest reliability, and inter-scorer reliability—were investigated. All of the reliability coefficients for 3 composites and 6 subtests of the PDMS-2 (Cronbach’s \( \alpha = 0.89–0.97 \), test-retest \( r = 0.82–0.93 \), and interscorer \( r = 0.96–0.99 \) ) showed that the PDMS-2 is a reliable tool for the assessment of motor development in children. However, only children without disabilities were recruited for that reliability study. Because the reliability levels may vary for different populations, the reliability of PDMS-2 scores for children with CP needs further investigation. The test-retest reliability coefficients are often thought of as stability coefficients; however, they do not reveal how much variability should be expected on the basis of measurement error. Thus, for estimating the confidence intervals of test scores, the standard error of measurement (SEM) of the PDMS-2 for children with CP needs to be calculated. Therefore, the other purpose of this study was to examine the test-retest reliability and SEM of the PDMS-2 for children with CP.

**Method**

**Participants**

Previous studies showed that the score change of motor tests was affected by age and CP severity in children with CP. There was an age and severity interaction on the amount of change in the GMFMA-88 scores. Children who were young and had mild CP demonstrated greater change in the GMFM-66 scores than did children who were older and had more severe CP over a period of time. To make the age (less than 48 mo or greater than or equal to 48 mo) and severity (mild or severe) levels evenly distributed in the present study sample, a quota sample of 32 children with CP was recruited.

Children were recruited from 2 developmental centers and 7 hospitals in the northern and eastern areas of Taiwan. To be eligible to participate in the study, children had to meet the following criteria: a confirmed medical diagnosis of CP from the attending pediatrician, age ranging from 24 to 65 months at the first evaluation, receiving physical therapy or occupational therapy at least twice per month during the study period, and written informed consent of the caregivers or guardians. The exclusion criteria were as follows: a medical problem that might prevent participation in therapy programs and progressive neurological disorders or medical conditions for which progress in motor development would not be expected over a 3-month period. In epidemiology studies, ages ranging from 2 to 10 years have been chosen as the age of ascertainment for CP diagnosis. Furthermore, the upper limit of the suitable age for testing children with the PDMS-2 is 71 months. Therefore, we set the minimum age for children at 24 months and the maximum age at 65 months.

The severity of CP in the children with CP was measured according to the Gross Motor Function Classification System (GMFCS) and was rated by the physical therapists treating those children and confirmed by one senior physical therapist before the PDMS-2 assessment. In this study, children at GMFCS levels I and II were classified as having mild CP, and those at GMFCS levels III to V were considered to have severe CP.

The mean age, body height, body weight, CP severity level, and sex of the children at the first evaluation are shown in Table 1. Their ages ranged from 27 to 64 months. The clinical types of CP in these children were spastic hemiparesis \( (n=5) \), spastic diplegia \( (n=14) \), spastic triplegica \( (n=4) \), spastic quadriplegia \( (n=6) \), and ataxia \( (n=3) \). The ages \((X \pm SD)\) of the fathers and mothers were \(37 \pm 5.8\) and \(34 \pm 5.6\) years, respectively. The education levels of the fathers and mothers, respectively, were graduate school \( (n=2\ and\ 0) \), university or college \( (n=11\ and\ 9) \), senior high school \( (n=11\ and\ 17) \), junior high school \( (n=4\ and\ 3) \), and primary school \( (n=2\ and\ 2) \). The occupations of the fathers and mothers, respectively, were professional or central administrators \( (n=6\ and\ 1) \), semiprofessional workers \( (n=10\ and\ 6) \), technical workers \( (n=10\ and\ 2) \), and semitechnical or nontechnical workers \( (n=4\ and\ 22) \). For 1 child,
information on the social or economic status of his family was not available.

Study Design
The single-group design method for examining both reliability and responsiveness was used in this study.26 One physical therapist assessed each child 3 times; the period between the first and second measurements was about 1 week, and the duration between the first and third measurements was 3 months. The agreement between the first 2 measurements was used to examine test-retest reliability. The change between the first and third measurements was used to examine responsiveness.

A previous study27 indicated that responsiveness studies involving children with CP need to be at least of 3 months’ duration. In line with this finding, we set the duration between the first and third measurements at 3 months.

Usually a measure must be sensitive to change before it can be responsive.6 Sensitivity to change is the capacity of a measure to assess change over time.6 Thus, 2 types of change indexes were used in this study; one was the sensitivity-to-change coefficient, and the other was the responsiveness coefficient. The responsiveness coefficient can be calculated from the differences of score change between groups of subjects who have and subjects who have not experienced “clinically important change” on the basis of retrospective judgment.28 To examine responsiveness in this study, a caregivers’ rating scale to detect a retrospective global rating of change was designed on the basis of a previous study.29

Assessment and Instruments
The PDMS-2 is a standardized, norm-referenced test.11 The GM composite of the PDMS-2 includes 151 items from 4 subtests: reflexes, stationary, locomotion, and object manipulation. The FM composite comprises 98 items from 2 subtests: grasping and visual-motor integration. The total motor (TM) composite includes 249 items from all subtests. Items of the PDMS-2 are scored with a 3-point score (0, 1, and 2); a score of 2 is assigned when the child performs the item according to the specified item criterion, a score of 1 indicates that the behavior is emerging but that the criterion for successful performance is not fully met, and a score of 0 indicates that the child cannot or will not attempt the item or that the attempt does not show that the skill is emerging. Therefore, the maximum raw scores of the subtests are different, ranging from 16 to 198.

From the results of the raw scores on each subtest of the PDMS-2, the standard scores and developmental age equivalents on each subtest can be obtained from the norms in the manual for the PDMS-2. The DQs for the GM, FM, and TM composites then are derived by summing the subtest standard scores and converting them to a quotient with a mean of 100 and a standard deviation of 15.11 Folio and Fewell11 suggested that to make important decisions about diagnosis and placement for children, the clinician should rely primarily on the results of composites rather than subtests. Therefore, this study focused on composite scores only. For the 3 composites of the PDMS-2, only the raw scores, percentile scores, and DQs can be obtained from the PDMS-2 manual. In clinics, the percentile scores and DQs for the 3 composites of the PDMS-2 can be used to share the test results with others and to identify the risk for or severity of the motor developmental delay. Clinicians should know the possible magnitudes of the measurement errors of these scores. Therefore, we analyzed the test-retest reliability coefficients and SEMs of the raw scores, percentile scores, and DQs for the 3 composites.

Change indexes for measures usually were calculated from raw scores, percentage scores, and scaled scores in previous responsiveness studies for children.27,30,31 Percentile scores and DQs are scores adjusted by age and are not suitable to be used as outcome indexes.27 The PDMS-2 does not provide information on scaled scores; therefore, only raw scores on the 3 composites of the PDMS-2 were used to calculate change indexes in this study.

The caregivers’ rating scale is composed of 3 items that are closed-ended questions that ask the main caregivers’ perception about overall change in GM, FM, or TM areas in the previous 3 months and are based on a 7-point Likert scale (much better, better, somewhat better, about the same, somewhat worse, worse, and much worse). The rating scale was self-administered by the

Table 1.
Demographic Data for Children With Cerebral Palsy (CP)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Younger</th>
<th>Older</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mild CP</td>
<td>Severe CP</td>
</tr>
<tr>
<td>Age at first test (mo)</td>
<td>34.0±5.9</td>
<td>39.8±2.8</td>
</tr>
<tr>
<td>Age at start of intervention (mo)</td>
<td>13.1±10.5</td>
<td>10.6±4.8</td>
</tr>
<tr>
<td>Body height (cm)</td>
<td>88.8±5.8</td>
<td>92±8.1</td>
</tr>
<tr>
<td>Body weight (kg)</td>
<td>13.1±2.4</td>
<td>13.9±2.9</td>
</tr>
<tr>
<td>No. of boys/girls</td>
<td>5/3</td>
<td>7/1</td>
</tr>
</tbody>
</table>
main caregiver of the child (usually the mother) at the time of the third measurement. The test-retest reliability of the caregivers’ rating scale for motor change within 1 week was analyzed by the quadratic weighted kappa coefficient test.35 The test-retest reliability (kappa coefficient) values of the caregivers’ rating scale were .63 (GM), .43 (FM), and .54 (TM), indicating moderate reliability.14 Because of the moderate test-retest reliability of the caregivers’ rating scale scores, we administered the caregivers’ rating scale 2 times with a 1-week interval to achieve more stable ratings. For calculating the responsiveness coefficient, only children who were rated “somewhat better,” “better,” or “much better” at both times were classified as having clinically important change.

**Procedure**

All caregivers of the children tested were informed of the procedure and purposes of the study and signed consent forms. The PDMS-2 assessments were administered by following the standard procedures outlined in the test manual.11 At the third assessment, the caregivers’ rating scale was administered. All of the testing was performed by a physical therapist (with 2 years of working experience with children with CP) who was familiar with the PDMS-2 and had good interrater reliability with a senior physical therapist (intraclass correlation coefficient [ICC] = .99–1.00 for raw scores or DQs for the 3 composites). Most assessments were performed in the places at which the children received treatments regularly. A few assessments were performed at a local child assessment laboratory because of a lack of appropriate space for assessments at the original treatment area. Each child received 3 assessments at the same time during the day.

**Data Analysis**

In order to attain even contributions of subtests for each composite, the raw scores on each subtest were transformed to the percentage score, as has been done for GMFM-88 scores.17 For example, the percentage score on the stationary subtest equaled the raw scores on the stationary subtest divided by the maximum raw score on the stationary subtest multiplied by 100. The percentage score on the GM composite was the average of the percentage scores on 4 subtests (reflexes, stationary, locomotion, and object manipulation). The percentage score on the FM composite was the average of the percentage scores on 2 subtests (grasping and visual-motor). The percentage score on the TM composite was the average of the percentage scores on all 6 subtests.

Statistical analyses were performed with SPSS (Statistical Package for the Social Sciences) version 10.0.* Test-retest reliability and change indexes were computed as follows.

**Test-retest reliability.** The ICC(2,1) was used to analyze the test-retest reliability of the raw scores, the percentage scores, the percentile scores, and the DQs for the 3 composites between the first and second assessments.33 In general, values of ICC of less than .5 can be interpreted as indicating poor reliability, those between .5 and .75 can be interpreted as indicating moderate reliability, and those above .75 can be interpreted as indicating good reliability.14 The SEMs for the different scales of the 3 composites of the PDMS-2 also were calculated.15

**Sensitivity to change.** Four statistical analyses were performed to calculate the sensitivity-to-change coefficient: the paired t test, the effect size (ES), the standardized response mean (SRM), and the Guyatt responsiveness index (GRI) for sensitivity to change (GRI-S). The paired t test is used to analyze data originating from a 1-group repeated-measures design and concludes whether a statistically significant change in the measures over time exists or not. The ES is a standardized measure of change obtained by dividing the average change between initial and follow-up measurements by the SD of the initial measurement.26 In this study, the ES was calculated by dividing the average change between the first and third tests by the pooled SDs of the first and third tests. The value of the ES is interpreted as trivial (ES of <0.2), small (ES of ≥0.2 <0.5), moderate (ES of ≥0.5<0.8), or large (ES of ≥0.8) according to the well-known thresholds of Cohen.34 The SRM equals the mean change in scores divided by the SDs of subjects’ difference scores.35 Therefore, in this study, it was calculated by dividing the average change between the first and third tests by the SDs of the score differences between the first and third tests. To interpret the value of the SRM for each composite, the ES thresholds (0.2, 0.5, and 0.8) proposed by Cohen34 were converted to SRMs according to the correlation coefficients between the scores on the first and third tests in this study and the formula proposed by Middel and van Sonderen36; then, the magnitude of the SRM was interpreted as trivial, small, moderate, or large according to the derived values. The GRI represents the ratio of observed change (or clinically important difference, if it is known) in a group of subjects expected to undergo a change to the variability in stable subjects.37 For sensitivity to change, in this study, the GRI-S was calculated by dividing the average change between the first and third tests by the standard

* SPSS Inc, 233 S Wacker Dr, Chicago, IL 60606.
deviation of the score differences between the first and third tests.26

Responsiveness. One of the GRI-Rs,37 which reflects the extent to which change in a measure relates to corresponding change in a reference measure of clinical or health status,35 is referred to as the GRI for responsiveness (GRI-R) in this study. The GRI-R is calculated by dividing the change in the group expected to undergo a change by the variability in a stable group.26 We calculated the GRI-R by dividing the mean change score between the first and third tests for subjects classified as having a clinically important change on the basis of the caregivers’ rating scales by the standard deviation of the score differences between the first and second tests for the entire group.26

Results

The DQs for the 3 composites of the PDMS-2 for the children with CP at the initial assessment are shown in Table 2. All children had DQs of less than 85 for the GM composite and the TM composite. The means of the percentage scores on the GM, FM, and TM composites of the PDMS-2 had good test-retest reliability.

Table 2. Test-Retest Reliability and Standard Error of Measurement (SEM) for Developmental Quotients, Percentile Scores, Raw Scores, and Percentage Scores for the Gross Motor (GM), Fine Motor (FM), and Total Motor (TM) Composites of the Peabody Developmental Scales—Second Edition for Children With Cerebral Palsy

<table>
<thead>
<tr>
<th>Parameter</th>
<th>First Test</th>
<th>Second Test</th>
<th>Intraclass Correlation Coefficient (95% Confidence Interval)</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developmental quotients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM</td>
<td>54.0±9.9</td>
<td>54.6±10.4</td>
<td>.988 (.976–.994)</td>
<td>1.1</td>
</tr>
<tr>
<td>FM</td>
<td>71.7±17.1</td>
<td>73.2±18.9</td>
<td>.979 (.958–.990)</td>
<td>2.5</td>
</tr>
<tr>
<td>TM</td>
<td>57.9±12.5</td>
<td>58.9±13.5</td>
<td>.984 (.968–.992)</td>
<td>1.6</td>
</tr>
<tr>
<td>Percentile scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM</td>
<td>1.0±2.2</td>
<td>1.0±1.7</td>
<td>.954 (.909–.978)</td>
<td>0.5</td>
</tr>
<tr>
<td>FM</td>
<td>10.7±15.4</td>
<td>13.5±21.0</td>
<td>.919 (.840–.960)</td>
<td>4.4</td>
</tr>
<tr>
<td>TM</td>
<td>1.8±3.3</td>
<td>2.5±4.9</td>
<td>.878 (.765–.939)</td>
<td>1.2</td>
</tr>
<tr>
<td>Raw scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM</td>
<td>122.4±45.3</td>
<td>124.6±47.2</td>
<td>.996 (.991–.998)</td>
<td>3.0</td>
</tr>
<tr>
<td>FM</td>
<td>129.2±35.5</td>
<td>130.9±37.1</td>
<td>.993 (.985–.996)</td>
<td>3.0</td>
</tr>
<tr>
<td>TM</td>
<td>251.6±73.0</td>
<td>255.5±76.1</td>
<td>.996 (.992–.999)</td>
<td>4.7</td>
</tr>
<tr>
<td>Percentage scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM</td>
<td>49.8±15.9</td>
<td>50.4±16.7</td>
<td>.993 (.986–.997)</td>
<td>1.3</td>
</tr>
<tr>
<td>FM</td>
<td>69.4±17.5</td>
<td>70.2±18.5</td>
<td>.995 (.989–.997)</td>
<td>1.3</td>
</tr>
<tr>
<td>TM</td>
<td>56.4±14.7</td>
<td>57.0±15.6</td>
<td>.995 (.990–.998)</td>
<td>1.1</td>
</tr>
</tbody>
</table>

*S For all values, *p<.0001, as determined with the ICC(2,1) model.

Sensitivity to Change

The mean percentage scores on the 3 composites of the PDMS-2 at the first and second tests are shown in Table 2, and those at the third test are shown in Table 3. The percentage scores were significantly different between the first and third tests, with *t*(*df=31*) values of 4.98 to 7.35 (*p<.001*). The ES value was 0.2 for all 3 composites; this value met the minimum standard proposed by Cohen for indicating a small change.34 The correlation coefficients of the percentage scores on the GM, FM, and TM composites between the first and third tests were .978, .976, and .986, respectively. Therefore, the values of the SRMs were interpreted as trivial (SRM of <1.0), small (SRM of ≥1.0<2.4), moderate (SRM of ≥2.4<3.8), or large (SRM of ≥3.8) for the GM composite; trivial (SRM of <0.9), small (SRM of ≥1.0<2.3), moderate (SRM of ≥2.3<3.7), or large (SRM of ≥3.7) for the FM composite; and trivial (SRM of <1.2), small (SRM of >1.2<3.0), moderate (SRM of ≥3.0<4.8), or large (SRM of ≥4.8) for the TM composite according to previously described methods.36 The SRM values of the percentage scores on the PDMS-2 ranged from 1.6 to 2.1 (Tab. 3).

Responsiveness

The GRI-R values for the 3 composites of the PDMS-2 ranged from 1.7 to 2.3 (Tab. 4).

Discussion

The results of this study are the first to confirm not only good test-retest reliability of various scales of the PDMS-2 but also acceptable responsiveness of the percentage scores on the 3 composites of the PDMS-2 for children with CP. These observations suggest that the PDMS-2 can be used as a set of evaluative tools for children with CP. Reliability is particularly important for developmental tests, either as a diagnostic test to evaluate the severity of developmental delay in clinics16 or as an evaluative test to detect the progress of a child after intervention.26
Table 3. Sensitivity-to-Change Coefficients for Percentage Scores on the Gross Motor (GM), Fine Motor (FM), and Total Motor (TM) Composites of the Peabody Developmental Motor Scales–Second Edition for Children With Cerebral Palsy Over a 3-Month Interval

<table>
<thead>
<tr>
<th>Composite</th>
<th>X±SD Percentage Score at Third Test</th>
<th>t Value (P)</th>
<th>Effect Size</th>
<th>Standardized Response Mean</th>
<th>GRI-Sa</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM</td>
<td>52.8±15.5</td>
<td>4.98 (&lt;.001)</td>
<td>0.2</td>
<td>0.9</td>
<td>1.6</td>
</tr>
<tr>
<td>FM</td>
<td>73.3±17.3</td>
<td>5.68 (&lt;.001)</td>
<td>0.2</td>
<td>1.0</td>
<td>2.1</td>
</tr>
<tr>
<td>TM</td>
<td>59.6±14.7</td>
<td>7.35 (&lt;.001)</td>
<td>0.2</td>
<td>1.3</td>
<td>2.1</td>
</tr>
</tbody>
</table>

aGRI-S=Guyatt responsiveness index for sensitivity to change.

Table 4. Responsiveness Coefficients for Percentage Scores on the Gross Motor (GM), Fine Motor (FM), and Total Motor (TM) Composites of the Peabody Developmental Motor Scales–Second Edition for Children With Cerebral Palsy Over a 3-Month Interval

<table>
<thead>
<tr>
<th>Composite</th>
<th>Mean Change Scorea</th>
<th>SD of Differencesb</th>
<th>GRI-Rc</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM</td>
<td>3.2</td>
<td>1.9</td>
<td>1.7</td>
</tr>
<tr>
<td>FM</td>
<td>4.4</td>
<td>1.9</td>
<td>2.3</td>
</tr>
<tr>
<td>TM</td>
<td>3.4</td>
<td>1.5</td>
<td>2.3</td>
</tr>
</tbody>
</table>

aMean change score between the first and third tests for children rated more than or equal to “somewhat better” (percentage score).
bSD of differences=standard deviation of score differences between the first and second tests for the entire group (percentage score).
cGRI-R=Guyatt responsiveness index for responsiveness.

Usually, DQs and percentile scores can be used to evaluate the severity of developmental delay, and raw scores and percentage scores can be used for quantifying the effect of intervention. In this study, the reliability of the DQs, percentile scores, raw scores, and percentage scores of the PDMS-2 was investigated, and high levels of test-retest reliability were demonstrated for children with CP. A previous study showed that the test-retest reliability coefficients of the DQs for the PDMS-2 were .73 to .89 for children developing typically and aged 2 to 11 months, and .93 to .96 for those aged 12 to 17 months. Because of differences in samples, the reliability coefficients are not directly comparable. Previous studies did not provide information on test-retest reliability for children with CP. As indicated by the results of this study, various scales of the PDMS-2 are reliable for use in clinics for motor skill acquisition or development for children with CP.

We found that the SEMs of the PDMS-2 obtained in this study were rather small, indicating that the error band of the observed scores was limited. Compared with the SEMs for the DQs of the norm samples of the PDMS-2 for children aged 24 to 72 months (3–4 for GM, 2–5 for FM, and 2–3 for TM), the SEMs for children with CP in this study were lower. Because the SEM is inversely related to the reliability coefficient, a relatively higher reliability coefficient may cause a lower SEM. The value of the SEM for a measure is useful for interpreting whether a change or difference in scores is beyond measurement error (ie, reaching real change or difference) in clinical settings. A higher criterion (SEM of 1.96) has been suggested for determining whether a change for a child with CP is real (ie, beyond measurement error). For example, the raw score on the TM composite for a child with CP should change more than 9.2 (ie, 1.96×4.7) for the change to be claimed as a real change with a 95% confidence level. On the other hand, if a child with CP has a change in the TM composite raw score of less than 9.2, it cannot be interpreted as a real improvement because such a change may be caused by measurement error.

Note that a change beyond measurement error does not necessarily indicate clinical relevance. Change beyond measurement error is the minimum level representing meaningful change. A clinically relevant change on a scale can be determined by combining both distribution-based methods (eg, SEM) and anchor-based methods (eg, parents’ or clinicians’ judgments). Although caregivers’ perceptions about overall change on the GM, FM, or TM composites were determined with an anchor-based method in this study, the lack of clinicians’ judgments and the modest sample size in this study limit the data for determining minimal clinically important change in the PDMS-2. Future studies to determine the benchmarks of minimal clinically important change in the PDMS-2 are warranted for clinicians to interpret their data.

This study also revealed that the percentage scores on the 3 composites of the PDMS-2 could be used for evaluating motor change in children with CP and receiving therapy. According to Liang, sensitivity to change is a necessary but insufficient condition for responsiveness. For a test to be relevant or meaningful to the decision maker, the responsiveness of the test should be provided. Our study revealed not only acceptable sensitivity to change but also acceptable responsiveness of the PDMS-2 for children with CP. The GRI-R values of the PDMS-2 for children with CP were 1.7 to 2.3 in our study. The magnitude of these statistics is comparable to that of values obtained for other outcome measures. For example, the GRI-R value of the motor component of the
Functional Independence Measure for stroke was 1.29,12 Few previous responsiveness studies for children with CP used many of the change indexes suggested by Husted et al35 to determine the validity of an evaluative tool.10,29 To select proper outcome instruments, clinicians should consider the child’s age and diagnosis, the purpose of testing, the reliability and responsiveness of the instruments, and the interpretability of the outcomes of the instruments.27 The previous study with the GM composite of the first edition of the PDMS (PDMS-GM) for infants with CP showed that the PDMS-GM had limitations when used as an evaluative measure for infants with CP.19 However, the change score on the PDMS-GM was not significantly different from that on the GMFM-88 for infants with CP over a 6-month period.27 Previous studies did not examine the change indexes of the PDMS-2. Our study provides sensitivity-to-change and responsiveness coefficients for the GM composite of the PDMS-2 as well as for the FM and TM composites. Our study also provides evidence for clinicians and researchers to confidently use the percentage scores of the PDMS-2 to detect a motor change for children with CP.

For sensitivity-to-change coefficients, the SRM may be preferred over the paired t value and the ES because the paired t value is influenced by sample size26,35 and the SRM, which uses the between-subject variability of individual change scores over time, provides more appropriate standardization than does the ES.37 Although a high between-subject variability of individual scores may have caused low ES values for children with CP in our study, the ES values for 3 composites of the PDMS-2 still met the minimum criterion (0.2) of Cohen.34 Guyatt et al37 suggested that the GRI was the most appropriate measure of responsiveness because it used the variability of change scores in stable subjects to standardize the clinically important difference; however, its assumption that the variance in stable subjects is approximately equal to the variance in an improved subject may induce biased estimation.41 At present, no single change index is superior. We provided 4 sensitivity-to-change coefficients and 1 responsiveness coefficient for the percentage scores of the PDMS-2 for children with CP in this study.

Advance knowledge of the responsiveness coefficient of an instrument would permit the accurate estimation of the sample size needed for adequate statistical power.37,43 If the GRI for the PDMS-2 is known, then the sample size needed for any experiment in which change over time in the PDMS-2 is the end point can be chosen immediately. According to the table in the report by Guyatt et al,37 for example, to detect a 3.2% mean change in the GM composite score (GRI-R=1.7), approximately 11 children per group would be required for a study with unpaired observations or 7 per group would be required for a study with paired observations.

To detect a 4.4% mean change in the FM composite score (GRI-R=2.3), the required sample size would be approximately 7 for unpaired observations or 5 for paired observations. To detect a 3.4% mean change in the TM composite score, the required sample size would be similar to that for the FM composite score.

The sample size in this study was modest, although it is reasonable for a clinical study. To improve the representation of the study sample, we tried to recruit children with different types of CP in this study. Furthermore, our evidence regarding test-retest reliability and responsiveness was acceptable. With the purposes that we proposed and the results that we obtained, our results might not be threatened by the modest sample size. In addition, we used a retrospective global rating scale with moderate reliability to calculate the responsiveness coefficient in this study. Although we used the scores from 2 repetitive rating scales to confirm the clinically important score change in this study, a retrospective computation of responsiveness has been criticized.28 The retrospective global rating scale was valued lower than the prognostic global rating scale by Stratford and Riddle.1 However, the prognostic global rating scale can be used only by clinicians and not by caregivers. The retrospective global rating scale used to assess the importance and magnitude of a measured change is critical if health status measures are to have an effect on patient care.31 Further studies are needed to develop a valid external criterion significant for both clients and clinicians. The further study of psychometric properties (eg, minimal clinically important change) is warranted to fully explore the utility of the PDMS-2 for children with different types of CP.

**Conclusion**

The results of this study showed that the PDMS-2 had good test-retest reliability over a 1-week period and acceptable sensitivity to change and responsiveness over a 3-month period for children with CP. The percentage scores on 3 composites of the PDMS-2 could be used over time to measure motor skill and motor development change over time for children with CP and aged from 2 to 5 years. The criterion of an SEM of 1.96 (GM=2.6%, FM=2.6%, and TM=2.2%) could be used to determine whether an individual achieves real change (ie, beyond measurement error).

**References**


36 Middel B, van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. Int J Integrated Care. 2002;2:1–22.


Hsiang-Hui Wang, Hua-Fang Liao and Ching-Lin Hsieh
PHYS THER. 2006; 86:1351-1359.

References
This article cites 30 articles, 8 of which you can access for free at:
http://ptjournal.apta.org/content/86/10/1351#BIBL

Cited by
This article has been cited by 4 HighWire-hosted articles:
http://ptjournal.apta.org/content/86/10/1351#otherarticles

Subscription Information
http://ptjournal.apta.org/subscriptions/

Permissions and Reprints
http://ptjournal.apta.org/site/misc/terms.xhtml

Information for Authors
http://ptjournal.apta.org/site/misc/ifora.xhtml