The Pediatric Motor Activity Log-Revised: Assessing Real-World Arm Use in Children With Cerebral Palsy

Gitendra Uswatte, Edward Taub, Angi Griffin, Laura Vogtle, Jan Rowe, and Joydip Barman
University of Alabama at Birmingham

Objective: Widely accepted models of disability suggest that actual use of an impaired upper extremity in everyday life frequently deviates from its motor capacity, as measured by laboratory tests. Yet, direct measures of real-world use of an impaired upper extremity are rare in pediatric neurorehabilitation. This paper examines how well the Pediatric Motor Activity Log-Revised (PMAL-R) measures this parameter, when the PMAL-R is administered as a structured interview as originally designed. Design: Parents of 60 children between 2 and 8 years of age with upper-extremity hemiparesis due to cerebral palsy completed the PMAL-R twice. Additionally, the children were videotaped during play structured to elicit spontaneous arm use. More-affected arm use was scored by masked raters; it was thought to reflect everyday activity since no cues were given about which arm to employ. Testing sessions were separated by 3 weeks, during which 29 children received upper-extremity rehabilitation and 31 did not. Results: The PMAL-R had high internal consistency (Cronbach’s alpha = .93) and test–retest reliability (r = .89). Convergent validity was supported by a strong correlation between changes in PMAL-R scores and more-affected arm use during play, r(53) = .5, p < .001. Conclusions: The PMAL-R interview is a reliable and valid measure of upper-extremity pediatric neurorehabilitation outcome.

Keywords: arm, function, daily living activity, hemiparesis, cerebral palsy

Impact and Implications

- This paper introduces a structured interview measuring what children with upper-extremity hemiparesis due to cerebral palsy (CP) actually do with their more-affected arm in everyday life, as opposed to what they can do when tested in the laboratory.
- The results add to the literature by establishing the convergent validity of this interview against an objective proxy of everyday activity for the more-affected arm.
- Researchers should consider using the revised Pediatric Motor Activity Log (PMAL-R) to measure real-world use of the more-affected arm in young children with upper-extremity hemiparesis due to CP.

Introduction

Approximately one third of children with cerebral palsy (CP) exhibit motor deficits in their more-affected arm (Hagberg, Hagberg, Beckung, & Uvebrant, 2001). Although measures of upper-extremity function in the laboratory or clinic have been available (Gilmore, Szakowski, & Boyd, 2010), no instruments existed until recently that assess how much such children actually use their more-affected arm in daily life. To address this gap, Taub and coworkers modified a structured interview developed for measuring real-world more-affected arm activity in adults after stroke (Taub, Ramey, DeLuca, & Echols, 2004). In the adult version, termed the Motor Activity Log (MAL), patients rate how they use their more-affected arm for 30 upper-extremity activities in daily life over a specified period (e.g., the past week; Taub et al., 1993; Uswatte & Taub, 2005; Uswatte, Taub, Morris, Light, & Thompson, 2006; Uswatte, Taub, Morris, Vignolo, & McCulloch, 2005). In the version for young children, the Pediatric Motor Activity Log (PMAL), primary family caregivers report on their child with upper-extremity hemiparesis (Taub et al., 2004).

In adult neurorehabilitation, four developments have contributed to the generation of methods for assessing activity of an impaired upper extremity in daily life (Uswatte & Taub, 2005). First, randomized controlled trials (RCTs) supporting the efficacy of Constraint-Induced Movement therapy (CIMT; Taub et al., 1993; Taub et al., 2006; Wolf et al., 2006) for rehabilitating real-world arm use after stroke (reviewed in Langhorne, Coupar, & Pollack, 2009), along with evidence of substantial brain plasticity after this treatment (reviewed in Taub & Uswatte, 2009) have led to a greater emphasis on ameliorating more-affected arm use, as opposed to teaching compensatory strategies (Morris & Taub, 2001; Uswatte et al., 2005). Second, there has been growing acceptance...
of the learned nonuse formulation (Taub, 1976, 1980; Taub, Us- 
watte, Mark, & Morris, 2006), which predicts that after neurolog- 
ical injury, the degree of motor recovery measured by laboratory 
performance tests will frequently overestimate how much an im-
paired extremity is actually used in daily life. Third, the World 
Health Organization [WHO] model of disability has promoted 
measuring impairment and activity separately (WHO, 2001). 
Fourth, the rigor with which outcomes are measured has received 
increasing attention in the rehabilitation sciences (Callahan & 
Barisa, 2005).

Similar considerations in pediatric neurorehabilitation have in-
spired a parallel interest in measuring more-affected arm activity 
in children with upper-extremity hemiparesis. Wallen, Bundy, 
Pont, and Ziviani (2009) examined the psychometric charac-
teristics of the original version of the PMAL (Taub et al., 2004) for 
measuring this parameter in children with CP between 6 months 
and 8 years of age when the PMAL was administered as a paper-
and-pencil parent report. Wallen and coworkers conducted a Rasch 
analysis that found the rating scale structure of the PMAL was 
disordered. Hence, they collapsed the number of PMAL rating 
scale steps from the original six to three. After rescoring the 
PMAL using the three-step scale, Wallen and coworkers found 
that, as predicted, children with low levels of manual ability in 
daily life, as measured by the Manual Abilities Classification 
System (MACS; Eliasson et al., 2006), had lower PMAL scores 
than children with high levels.

This paper examines the test–retest reliability, convergent va-
licity, and other psychometric characteristics of a revised version 
of the PMAL in young children with upper-extremity hemiparesis 
due to CP. The method here departs from the Wallen et al. study 
by (a) administering the PMAL-R as a structured interview, as 
originally designed, rather than a correspondent report; (b) scoring 
the PMAL-R using both the original six-step and modified three-
step scales; and (c) validating the PMAL-R against an index of 
everyday activity specific to the more-affected arm rather than a 
categorization system whose scoring scheme does not take into 
account which arm is employed, that is, the MACS.

Method

Participants

Children (N = 60) were between 2 and 8 years of age with mild 
to severe paresis of an upper extremity subsequent to CP. As 
shown in Figure 1, their data were assembled from three prospec-
tive studies: (a) RCT comparing 6.5 hr of pediatric CIMT every 
weekday for 3 weeks (n = 10) to customary care (n = 10; Taub et 
al., 2011); (b) dose response trial testing two forms of pediatric 
CIMT administered for shorter periods (N = 19); and (c) test– 
retest reliability study of the PMAL-R with only customary care 
and no CIMT between the two testing sessions (N = 21). Families 
of the children in the two CIMT studies either contacted our 
laboratory because of their interest in CIMT or were referred by 
local rehabilitation professionals. Consecutive children who met 
the eligibility criteria (see Table 1) were enrolled. Families of the 
children in the test–retest reliability study were recruited from a
pediatric outpatient physical rehabilitation facility. They were paid $50 for participating. Consecutive children who met the eligibility criteria (see Table 1) were enrolled.

Data from the CIMT groups in the two clinical trials were combined into a CIMT supergroup (n = 29) for this paper because the treatment outcomes in all three groups were equivalent, F(2, 25) = 0, p = .97. In the RCT (Taub et al., 2011), children in the CIMT group received systematic training in use of their more-affected arm for 6 hr daily for 15 consecutive weekdays and wore a bivalved cast on their less-affected arm to limit its use. A package of behavioral techniques, termed the Transfer Package, was employed to transfer gains made during training to daily life (Taub et al., 2007; Taub et al., 2006; Taub, Uswatte, Mark, & Morris, 2006). This CIMT component, on average, took 30 min on each treatment day. In the dose response study of CIMT, children were randomized to 3.5 hr of treatment every weekday for either 2 (n = 10) or 3 weeks (n = 9). The efficacy of pediatric CIMT is supported by a recent review that evaluated data from 23 studies (Huang, Fetters, Hale, & McBride, 2009).

Data from the test–retest reliability study and the control group in the RCT were combined into a control supergroup (n = 31) for this paper, because all of these participants received customary care between PMAL-R administrations. Customary care in the study locale ranged from no physical rehabilitation to 2 hr per week of outpatient physical rehabilitation with a maximum 1 hr per week of occupational therapy for the upper extremities. Such treatment has been shown previously not to affect everyday use of the more-affected arm in children with CP (Taub, Ramey, DeLuca, & Echols, 2004). Furthermore, both of sets customary care children in this paper showed no changes on the PMAL-R from the first to second administration (see Results and Discussion).

Measures

PMAL-R. This structured interview elicits information regarding actual use of the more-affected arm for completing 22 activities commonly carried out in daily life (see Table 2). The protocol is the same as for the adult MAL (Uswatte et al., 2006; Uswatte et al., 2005) except for three departures: (a) the point of reference for evaluating function of the more-affected arm is the less-affected arm instead of the same arm before brain injury; (b) on items that can include a wide variety of tasks, such as Item 17 (push large object across floor), the tester notes the particular activity assessed when administering the test the first time and asks about that specific activity on subsequent administrations; (c) primary caregivers of the children being treated are questioned rather than the patients themselves.

During the interview, parents or guardians are asked to rate how well (HW scale; Table 3) and how often (HO scale) children use their more-affected arm to accomplish each activity over a specified period. Both scales, which range from 0–5, are anchored at 6 points by abbreviated phrases and longer definitions; interviewees may select scores halfway between the anchors. As noted in the Introduction, Wallen et al. (2009) recommend the following transformation for HW items scores (0, 1, or 2 = 0; 3 = 1; 4 or 5 = 2) and HO item scores (0 or 1 = 0; 2 or 3 = 1; 4 or 5 = 2) based on a Rasch analysis suggesting that the rating scale structure is disordered. In the original or Wallen scheme (M. Wallen, personal communication, March 27, 2011), the mean of the item scores can be used to derive the total scores for each scale.

In our CIMT efficacy studies, both scales are administered immediately before and after treatment and weekly for a month following treatment. Interviewees report about participants’ behavior over the previous week for all these occasions except posttreatment, for which the typical observation period is 3 days. During treatment, only the HW scale is administered because the less-affected arm of the participants is placed in a cast to induce use of the more-affected arm. In addition, only half of the items are completed at the beginning of each treatment day (e.g., Items 1–11 on Day 1 and Items 12–22 on Day 2) to permit at least 48 hr between reassessment of each activity.

Of note, half of the items, that is, Items 2–5, 7–9, 11, 17, 19, and 22, in the current version differ from the original (Taub et al., 2004). Items 1, 5, 15, 16, 17, and 20 from the original version are not included in the current one because those activities are accomplished using many different movement patterns even by children without upper-extremity impairment. Items 2, 3, 10, 18, 19, and 21 from the original version are not included because those activities are not performed commonly by children in the age range for which the current version is intended, that is, 2–8 years. In addition, probing of interviewee responses is added to the current protocol for testing occasions subsequent to the initial administration, so that the PMAL-R protocol parallels that for the adult MAL. In this procedure, the interviewer, using scripted phrases, probes responses that differ from those given on the previous...
testing administration to determine whether such scores reflect an actual change. The purpose of this procedure is to prevent errors in recall, misunderstanding of the scale levels, and enthusiasm about treatment gains (i.e., a halo effect) from resulting in final responses that are not veridical. A study of the adult MAL suggests that probing does not bias scores (Uswatte et al., 2005). The PMAL-R testing manual is available online (Taub, Griffin, & Uswatte, 2012b).

**Pediatric Arm Function Test (PAFT) limb preference scale.**

This component of the PAFT (Taub et al., 2007), which contains eight of the 26 items on the full test, involves behavioral observation of children at play structured to elicit use of one of their arms. No cues are given about which arm to employ. On Item 1, for example, the tester positions a ball 6 in. directly above children’s heads and says “Touch the ball.” Examples of other items are pouring a ball out of a cup and grasping a puzzle piece with a large knob. Trained, masked raters score whether or not a child attempts to use the more-affected arm to accomplish each item from video of the test. The test score is the percentage of completed items that are attempted with the more-affected arm. Although the test is conducted in the laboratory, it is designed to reflect how much children use their arm in everyday life, since the choice of which arm the child employs to accomplish the item is spontaneous and testing is conducted in a play setting. Data from

### Table 2

**Characteristics of the PMAL-R Items (N = 60)**

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Not applicable</th>
<th>Item–total correlation</th>
<th>Type of activity</th>
<th>Finger movement required</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Eat finger foods</td>
<td>0%</td>
<td>.67</td>
<td>Basic</td>
<td>Unimanual</td>
</tr>
<tr>
<td>2</td>
<td>Pick up a small item</td>
<td>0%</td>
<td>.73</td>
<td>Basic/instrumental</td>
<td>Unimanual</td>
</tr>
<tr>
<td>3</td>
<td>Self-feed with a fork/spoon</td>
<td>0%</td>
<td>.65</td>
<td>Basic</td>
<td>Unimanual</td>
</tr>
<tr>
<td>4</td>
<td>Brush teeth</td>
<td>0%</td>
<td>.63</td>
<td>Basic</td>
<td>Unimanual</td>
</tr>
<tr>
<td>5</td>
<td>Gestures</td>
<td>0%</td>
<td>.51</td>
<td>Instrumental</td>
<td>Unimanual</td>
</tr>
<tr>
<td>6</td>
<td>Push arm through sleeve of clothing</td>
<td>4.5%</td>
<td>.46</td>
<td>Basic</td>
<td>Unimanual</td>
</tr>
<tr>
<td>7</td>
<td>Turn a page in a book</td>
<td>0%</td>
<td>.44</td>
<td>Instrumental</td>
<td>Unimanual</td>
</tr>
<tr>
<td>8</td>
<td>Point to a picture</td>
<td>4.5%</td>
<td>.58</td>
<td>Instrumental</td>
<td>Unimanual</td>
</tr>
<tr>
<td>9</td>
<td>Reach for an object above head</td>
<td>4.5%</td>
<td>.55</td>
<td>Basic/instrumental</td>
<td>Unimanual</td>
</tr>
<tr>
<td>10</td>
<td>Push a button or key</td>
<td>4.5%</td>
<td>.62</td>
<td>Instrumental</td>
<td>Unimanual</td>
</tr>
<tr>
<td>11</td>
<td>Steady self</td>
<td>4.5%</td>
<td>.54</td>
<td>Basic</td>
<td>Uni/bimanual</td>
</tr>
<tr>
<td>12</td>
<td>Open a door or cabinet</td>
<td>4.5%</td>
<td>.72</td>
<td>Instrumental</td>
<td>Unimanual</td>
</tr>
<tr>
<td>13</td>
<td>Turn a knob</td>
<td>9%</td>
<td>.39</td>
<td>Basic/instrumental</td>
<td>Unimanual</td>
</tr>
<tr>
<td>14</td>
<td>Use arm to move across the floor</td>
<td>4.5%</td>
<td>.50</td>
<td>Basic</td>
<td>Bimanual</td>
</tr>
<tr>
<td>15</td>
<td>Take off shoes</td>
<td>4.5%</td>
<td>.69</td>
<td>Basic</td>
<td>Uni/bimanual</td>
</tr>
<tr>
<td>16</td>
<td>Take off socks</td>
<td>13.6%</td>
<td>.78</td>
<td>Basic</td>
<td>Uni/bimanual</td>
</tr>
<tr>
<td>17</td>
<td>Push a large object across the floor</td>
<td>4.5%</td>
<td>.62</td>
<td>Instrumental</td>
<td>Bimanual</td>
</tr>
<tr>
<td>18</td>
<td>Hold a ball</td>
<td>0%</td>
<td>.62</td>
<td>Instrumental</td>
<td>Unimanual</td>
</tr>
<tr>
<td>19</td>
<td>Throw a ball or other object</td>
<td>4.5%</td>
<td>.63</td>
<td>Instrumental</td>
<td>Unimanual</td>
</tr>
<tr>
<td>20</td>
<td>Use a cylindrical object</td>
<td>9%</td>
<td>.43</td>
<td>Instrumental</td>
<td>Unimanual</td>
</tr>
<tr>
<td>21</td>
<td>Hold a handle while riding, pulling, or pushing a toy</td>
<td>0%</td>
<td>.64</td>
<td>Instrumental</td>
<td>Bimanual</td>
</tr>
<tr>
<td>22</td>
<td>Placement of an object</td>
<td>9%</td>
<td>.53</td>
<td>Basic/instrumental</td>
<td>Unimanual</td>
</tr>
</tbody>
</table>

### Table 3

**PMAL-R Arm Use Scale**

<table>
<thead>
<tr>
<th>Rating</th>
<th>Anchor</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td><em>Not used</em> – Your child did not use the weaker arm for the activity.</td>
</tr>
<tr>
<td>1</td>
<td><em>Very poor</em> – Your child had very little functional use of the weaker arm for the activity. The arm may have moved during the activity but was of no real functional help.</td>
</tr>
<tr>
<td>2</td>
<td><em>Poor</em> – Your child had minor functional use of the weaker arm for the activity. The arm actively participated in the activity, but the stronger arm/caregiver did most of the work.</td>
</tr>
<tr>
<td>3</td>
<td><em>Fair or moderate</em> – The weaker arm was used to accomplish the activity, but the performance was very slow and/or involved great difficulty.</td>
</tr>
<tr>
<td>4</td>
<td><em>Almost normal</em> – The weaker arm was able to accomplish the activity independently but did so with some difficulty and/or inaccuracy.</td>
</tr>
<tr>
<td>5</td>
<td><em>Normal</em> – The weaker arm did the activity normally.</td>
</tr>
</tbody>
</table>

*This rubric is also known as the How Well scale.*
the first testing occasion in this study indicate that the internal consistency of the Limb Preference scale, as indexed by Cronbach’s alpha, is .85. Test–retest reliability, that is, the intraclass correlation (ICC; Shrout & Fleiss, 1979) between limb preference scores for control participants from the first and second testing occasions, is .66. Interrater reliability is described in the Procedure section. Indirect evidence of the validity of the limb preference score for indexing arm function comes from a recent study reporting that gains in limb preference scores after pediatric CIMT in children with cerebral palsy are strongly correlated with increases in amount of gray matter, r(6) = .68, p < .05 (Sterling et al., 2011). Instructions for administering, videotaping, and scoring the PAFT are in the testing manual, which is available online (Taub, Griffin, & Usвате, 2012a).

Manual Ability Classification System (MACS). This widely employed instrument orders children according to their ability to manipulate objects regardless of whether they use the more-affected arm, less-affected arm, both, or neither. (A child, e.g., might move a toy by picking it up with his teeth.) The categories range from Level I (“handle objects easily and successfully”) to Level V (“does not handle objects”). Evidence of reliability includes high interrater reliability between health professionals (ICC = .97, n = 168) and between health professionals and family caregivers (ICC = .96, n = 25; Eliasson et al., 2006). Although classification using the MACS is ideally done by health professionals in consultation with family members who best know the child’s daily behavior, adequate agreement between family caregiver and health professional MACS scores (ICC, range = .73 – .85) has been reported, even when a substantial portion of children have been classified by the health professionals solely on the basis of retrospective chart review (range = 40–61%; Morris, Kurinczuk, Fitzpatrick, & Rosenbaum, 2006).

Procedure

To permit assessment of the PMAL-R’s responsiveness to treatment, primary family caregivers of CIMT children completed PMAL-R interviews before treatment (Test 1) and afterward (Test 2). To allow calculation of test–retest reliability, stability, and accuracy, caregivers of control children completed the PMAL-R twice with no intervening CIMT. There were approximately three weeks between the first and second occasion (Tests 1 and 2, respectively). The testing, which typically took 30 min, was carried out by occupational therapists (OTs) or psychology graduates who received approximately 8 hr of training. Training involved reading the PMAL-R testing manual, shadowing an experienced tester, and conducting two to three supervised tests.

To permit assessment of convergent validity, children in both groups also completed the PAFT at Tests 1 and 2. The PAFT, which typically took 30 min, was administered by pairs of pediatric OTs who received approximately 7 hr of training. The training procedures for the PAFT involved the same three steps as for the PMAL-R. After testing, the PAFT was scored independently from video by either a pediatric physical therapist or OT. They each received approximately 25 hr of training, which involved: (a) reading the testing manual; (b) scoring several PAFT sessions jointly under the supervision of an experienced tester; and (c) scoring eight sessions independently followed by joint review of each session under the supervision of the experienced tester. Before scoring study data, the raters demonstrated high interrater reliability on an additional set of 10 tests (ICC = .98). When scoring the study data, the raters were masked to group assignment, testing occasion, and participants’ PMAL-R scores.

To describe the children’s level of everyday manual ability, an OT gave each child a MACS score based on a retrospective review of the children’s records. All testing was conducted in a treatment room in a pediatric outpatient physical rehabilitation facility at an urban medical center in the southeastern United States. Informed consent was received from all legal guardians; the study procedures were approved by our Institutional Review Board.

Data Analysis

An item analysis of the PMAL-R was conducted to determine which, if any, items to eliminate (see Table 2). When scoring the PMAL-R, an item was deemed not applicable, and dropped from the calculation of the test score for a child if it was impossible to do (Taub et al., 2012b). For example, items that were physically impossible for a child to do, never carried out in their family, or developmentally inappropriate were treated this way. To identify items that had to be dropped frequently, the percentage of children for whom each item was not applicable was calculated. To identify items that measured a parameter that was different from the bulk of the items, the Pearson correlation between each item score and the mean of the other item scores was calculated (Kline, 2000). A priori criteria for eliminating an item from the test were having >20% interviewees identify it as not applicable or having an item–total correlation value < .3 (Usвате et al., 2006).

A content analysis was conducted to determine whether an appropriate variety of upper-extremity tasks was represented on the test. To this end, the items were classified along three dimensions: (A) basic, instrumental, or either (depending on the context) according to definitions given by the National Center for Health Statistics (2004); (B) unimanual, bimanual, or either according to how the items are performed typically by individuals without impairment (Usвате et al., 2006); and (C) requiring finger movement if the item called for objects to be manipulated by the fingers to be successfully accomplished (Usвате et al., 2006). Classification of the items along each of these dimensions was done independently by two pairs of pediatric occupational therapists. Interrater agreement between the two pairs for each dimension was >91%. Classification of items for which there was disagreement was adjudicated by authors Usвате, Taub, and Griffin: Dimension A, Item 13; Dimension C; Items 5 & 12.

After completing the item and content analyses, several statistics were calculated to evaluate the psychometric characteristics of the test score. To index internal consistency, Cronbach’s alpha was calculated using Test 1 data from all participants. Test–retest reliability was evaluated by the ICC, Type II (Shrout & Fleiss, 1979) between Tests 1 and 2 scores in the control group. Stability was evaluated by paired t tests of Tests 1 and 2 scores in the same group. Accuracy was indexed by the standard error of measurement (SEM) and minimum detectable change (MDC) with a 95% confidence interval (CI) following the method in Fritz, Blanton, Usвате, Taub, & Wolf, 2009. Responsiveness to change was indexed by the standardized response mean (SRM), that is, by dividing the mean change in scores from Test 1 to Test 2 for the CIMT group by the standard deviation (SD) of the corresponding change scores in the control group (Cohen, 1988;
Guyatt, Walter, & Norman, 1987; Uswatte et al., 2005). Convergent validity was evaluated by the Pearson correlation between PMAL-R and PAFT limb preference scores from all participants at Test 1. In addition, the partial Pearson correlation between PMAL-R and PAFT scores from all participants at Test 2 was calculated to assess whether PMAL-R and PAFT data converge when evaluating change across time. For this purpose, Test 1 scores from both instruments were entered into the statistical model as covariates. Characterization of the size of the correlations was based on Cohen, 1988: large \( r \geq .5 \); moderate \( r \geq .3 \); small \( r \geq .1 \). If data were missing from a participant for a test (see Figure 1), that participant was eliminated from all analyses that required the missing data. To permit comparison of the original rating scale structure and the Wallen structure, all of the test score analyses were conducted with scoring done both ways.

**Results**

**Relationship Between the PMAL-R How Well and How Often Scales**

PMAL-R HW and HO scores were highly correlated at Test 1, \( r(58) = .92, p < .001 \), and at Test 2 after controlling for Test 1 scores, partial \( r(54) = .87, p < .001 \), suggesting that these two scales provide redundant information. Hence, results are presented for the PMAL-R HW scale only, hereafter referred to as the *Arm Use scale* (see Table 3).

**Participant Characteristics and Safety**

The children’s characteristics are summarized in Table 4. According to both the PMAL-R and PAFT Limb Preference scale, the control group had better use of their paretic arm in daily life than the CIMT group at the outset of the study, \( p \)'s < .05. Other characteristics were similar. Administration of the PMAL-R or PAFT did not result in any adverse events.

**Item and Content Analysis**

The item analysis did not suggest removing any of the activities from the PMAL-R (see Table 2). The content analysis indicated that basic and instrumental activities of daily living made up 36% and 45% of the items, respectively (see Table 2). Remaining items were classified as basic/instrumental, meaning that they could fall in either category depending on the context. Unimanual and bimanual activities made up 72% and 9% of the items, respectively. Remaining items were judged to be commonly performed either way. Finger movement was judged necessary for 73% of the items, while it was not for 27%.

**Reliability, Stability, Accuracy, and Responsiveness to Change of PMAL-R With Original Six-Step Scale Structure**

Internal consistency of the PMAL-R Arm Use scale was high (Cronbach’s alpha = .93), as was test–retest reliability (ICC = .89). The mean score for the control group did not change from Test 1 to Test 2: mean change = 0 points, \( SD = 0.5, t(27) = 0.1, p = .92, 95\% CI = [−0.2, 0.2] \). The SEM was 0.15. The MDC was 0.42. In contrast, the mean change in the CIMT group was 2.2, \( SD = 0.5, t(28) = 2.19, p < .001, 95\% CI [2.0,2.4] \), resulting in a SRM of 4.4. On the PAFT Limb Preference scale, the mean change in the control group was 0 percentage points, \( SD = 16, t(28) = −0.2, p = .95, 95\% CI = [−6, 6] \), while the mean change...
Convergent Validity of PMAL-R With Original Six-Step Scale Structure

For changes in scores from Test 1 to Test 2, there was a large correlation between the PMAL-R Arm Use and PAFT Limb Preference scales, partial \( r(53) = .5, p < .001 \). For Test 1 scores, the correlation between the PMAL-R and PAFT was moderate, \( r(58) = .36, p = .006 \).

Psychometric Characteristics of PMAL-R With Modified Three-Step Scale Structure

Table 5 lists the psychometric characteristics of the PMAL-R when employing the three-step scale structure recommended by (Wallen et al., 2009). The results are very similar to those for the original six-step scale structure. For the three-step scale, the mean change in the control group was 0 points, \( SD = 0.2, t(28) = -0.3, p = .77, 95\% CI = [-0.1, 0.1] \), while the mean change in the CIMT group was 0.9, \( SD = 0.4, t(28) = 11.7, p < .001, 95\% CI = [0.8, 1.1] \).

Discussion

The PMAL-R Arm Use scale with the original six-step structure exhibited high reliability, stability, accuracy, and responsiveness to change in children between 2 and 8 years of age, with a wide range of severity of upper extremity hemiparesis due to CP. Participants in the control group, who received little or no upper extremity rehabilitation between testing occasions, had Arm Use scores at Test 2 that were similar in rank and absolute value to those at Test 1. The Cronbach’s alpha value obtained was high, suggesting that the items all assess a single underlying construct. The signal-to-noise ratio, that is, change after an efficacious therapy relative to error in measurement, was remarkably large. The mean gain on the PMAL-R in the CIMT group was 5.2 times larger than the MDC, that is, the smallest change that can be confidently detected by the instrument.

Validity of the PMAL-R for measuring real-world changes in more-affected arm use was supported. There was a strong correlation between change on the PMAL-R and PAFT Limb Preference scale after CIMT. However, the validity of the PMAL-R for measuring absolute levels of real-world, more-affected arm use was not. On the first testing occasion, the correlation between the PMAL-R and PAFT was only moderate. A possible explanation of the last result is that parents reported changes in their child’s arm use accurately but that parents of children with similar levels of arm use at Test 1 did not select similar PMAL-R scores. In other words, parents had similar sensitivity to the relative amount of change in arm use due to treatment, but did not share a common frame of reference for rating arm use at baseline. To remedy this possible shortcoming, a video for parents depicting arm use that is typical of each rating scale anchor is available from the authors, as for the adult MAL (Uswatte et al., 2006; Uswatte et al., 2005). Of note, additional, albeit indirect, support for the validity of the PMAL-R for measuring changes in arm function was reported in a recent study of CIMT in children with upper-extremity hemiparesis due to CP. PMAL-R gains after treatment were found to be strongly correlated with increases in amount of gray matter in contralateral primary sensorimotor cortex, \( r(8) = .63, p < .05 \) (Sterling et al., 2011).

The findings did not change if Wallen’s transformation of item scores was used instead of the original six-step scale structure (see Table 5). Wallen et al. (2009) might have found disorder in the original rating scale structure because they asked parents to fill out the PMAL as a paper-and-pencil test. (Wallen and coworkers mailed the PMAL to parents for completion in many cases.) In this study, the PMAL was administered as a structured interview, as originally designed. In the interview, the tester explains the test, solicits questions about it, verifies parents’ responses by repeating them back, gives parents feedback on the initial testing occasion if a score departs markedly from what might be expected of their child, and asks parents on subsequent testing occasions if a score represents a genuine change if it departs from the score for that item on the previous testing occasion. With such guidance, it appears that parents are able to reliably discriminate between behaviors that warrant a 0, 1, or 2, for example. Among other applications, such discrimination is important when making treatment plans for patients, since a 0 represents no use of the more-affected arm, while a 2 represents poor use.

A recent paper by another group (Lin et al., 2012) on the validity of the original version of the PMAL when administered as an interview and using the six-step scale structure found that Arm Use

Table 5

| Psychometric Characteristics of PMAL-R With Six- and Three-Step Versions of Arm Use Scale |
|-----------------------------------------------|-----------------|-----------------|
| Clinimetric characteristic                     | Six-step version| Three-step version|
| Reliability, accuracy, and responsiveness to change | 0.93            | 0.9             |
| Internal consistency, Cronbach’s \( \alpha \)     | 0.89            | 0.9             |
| Test–retest reliability, ICC, type 2.1           | 3               | 3               |
| SEM, percent of scale range                      | 8               | 10              |
| MDC, percent of scale range                      | 4.4             | 3.9             |
| Responsiveness ratio\( ^{a} \)                   | 0.5             | 0.46            |
| Convergent validity between PMAL-R Arm Use scale and PAFT Limb Preference Scale | 0.36            | 0.41            |
| Change from tests 1 to 2, partial Pearson correlation | 0.36            | 0.41            |

Note. CIMT = Constraint-Induced Movement therapy; ICC = intraclass correlation; MDC = minimum detectable change; PAFT = Pediatric Arm Function Test; PMAL-R = Pediatric Motor Activity Log-Revised; SEM = standard error of measurement.
scores were correlated with the Peabody Developmental Motor Scales-Second Edition (PDMS-2; Folio & Fewell, 2000) Grasping ($r_{39} = .32, p < .05$) and Visual-Motor Integration ($r_{39} = .48, p < .01$) subscales and WeeFIM (McCabe & Granger, 1990) self-care subscale ($r_{39} = .43, p < .01$). These validity coefficients were smaller than the one reported here, which might be expected given that Lin et al.'s criterion measures were less direct measures of more-affected arm use in daily life than the PAFT. The PDMS-2 subscales in question assess capacity to manipulate objects in the laboratory setting regardless of which arm is used, while the WeeFIM self-care subscale assesses need for assistance when performing activities of daily living regardless of which arm is used (see discussion of pediatric neurorehabilitation measures below). Lin et al. also reported that the Arm Use score was highly responsive to change ($SRM = 0.99$), although their MDC value (0.8 points) was larger than reported here, which might be due to differences between the original and revised version of the PMAL.

A limitation of our study was that the control group, on average, had better use of their more-affected arm in everyday life at Test 1 than the CIMT group (see Participants). This raised the possibility that the findings on PMAL-R test–retest reliability, stability, and error in measurement might have been different if the control and CIMT groups had the same initial characteristics. However, the findings did not change when analyzing data only from participants in the control group who were similar to those in the CIMT group. For 10 control participants who were from the customary care group in a RCT of CIMT (see Figure 1), PMAL-R test–retest reliability ($ICC = .72$) and stability, mean change from Test 1 to Test 2 = 0.1, $SD = 0.3$, $n(9) = -0.7, p = .51, 95\% CI = [-0.1, 0.3]$, were supported, and measurement error was small ($SEM = .14$). Corresponding values for the entire control group ($n = 31$) are in the Results.

Another study limitation was the use of the PAFT Limb Preference scale as the convergent measure. The PAFT is not a direct measure of more-affected arm use in everyday life; observation of spontaneous behavior in the laboratory setting during structured play, which is thought to reflect arm use during everyday life, is employed. When ambulatory assessment methods for measuring arm use in young children in everyday life becomes available, it would be desirable to validate the PMAL-R against such instruments. It is worth noting that the adult MAL, which the PMAL-R parallels, is strongly correlated ($r > .5, p < .05$) with both laboratory-based behavioral observation of spontaneous arm use (i.e., Actual Amount of Use Test; Uswatte & Taub, 2005) and ambulatory assessment of arm use in daily life (Uswatte et al., 2006; Uswatte et al., 2005). Another drawback of the PAFT is that it is not an established measure. However, the PAFT was selected rather than the Assisting Hand Assessment (AHA; Krumlinde-Sundholm & Eliasson, 2003) because of the limitations with that instrument discussed below.

The learned nonuse formulation explains that after an injury nonuse of an upper-extremity in excess of the impairment present will prevail whenever (a) there is marked initial impairment that resolves gradually; (b) use of the impaired upper-extremity soon after injury results in punishment, for example, failure to accomplish a task; and (c) use of compensatory behaviors, for example, employing the other arm, results in reward (Taub, 1976; Taub et al., 2006). As predicted, studies with monkeys after unilateral forelimb deafferentation (Taub, 1976), adults after stroke (Andrews & Stewart, 1979; Dromerick et al., 2006; Sterr, Freivogel, & Schmalohr, 2002; Taub et al., 2006; Uswatte & Taub, 2005), and children with CP (Fedrizzi, Pagliano, Andreucci, & Oleari, 2003) show that tests of motor capacity overestimate spontaneous use of the more-affected upper extremity. In children with CP, learned nonuse is termed “developmental disregard” because they never develop use of their more-affected arm for many activities, unlike adults after stroke who stop using their more-affected arm to conduct activities that they employed that arm for prior to their CNS injury (Taub et al., 2007; Taub et al., 2004).

The established measures in upper-extremity pediatric neurorehabilitation assess function in the treatment setting, which is in the WHO model of disability is termed maximum capacity (WHO, 2001). For example, on the Quality of Upper Extremity Skills Test (DeMatteo et al., 1992) and Melbourne Assessment of Unilateral Upper Limb Function (Johnson et al., 1994), children are asked to conduct movements or tasks with their more-affected arm in the treatment setting, and the quality of movement of that extremity is scored by a therapist. Thus, these instruments capture what children can do with their more-affected arm when asked to use it.

There are established instruments that measure what children actually do spontaneously in their daily life, which is in the WHO model of disability is termed performance (WHO, 2001). However, these instruments do not account for whether the more-affected arm is employed. For example on two widely employed instruments, that is, the Pediatric Evaluation of Disability Inventory (Haley, Coster, Ludlow, Haltiwanger, & Andrello, 1992) and WeeFIM (McCabe & Granger, 1990), parents evaluate how much assistance their child needs to accomplish everyday activities, including those requiring use of the arms. The arm employed when children complete activities is not recorded. Therefore, changes in test scores could be due to changes in function of the more-affected arm, changes in compensatory strategy, or both. Other well known scales (e.g., ABILHAND-KIDS; Arnould, Penta, Renders, & Thonnard, 2004) focus on how effectively children accomplish activities in everyday life but do not regard which arm is employed, which results in the same drawback as for first two instruments in this class.

Some recent instruments have been designed to measure what children actually do with their more-affected arm in the real world. The AHA (Krumlinde-Sundholm & Eliasson, 2003) extrapolates how effectively children from 1.5 to 12 years of age use their more-affected arm in everyday life by observing its spontaneous use in the laboratory or clinic. Therapists rate how effectively children use the more-affected arm to assist the other when performing bimanual activities during a 10- to 15-min, semistructured play session. No cues are given about which arm to use. Although several studies report good reliability and validity, an important limitation is that assessment of more-affected arm use is restricted to bimanual activities. Attempts to carry out bimanual activities force a child to try to use the more-affected arm, while tasks that can be carried out effectively by just one hand do not impose this constraint on behavior, raising the possibility that the AHA overestimates unimanual use of the more-affected arm. Furthermore, the AHA has not been validated against other indices of real-world use of the more-affected arm. The Video Observations Aarts and Aarts (Aarts et al., 2007; Aarts, Jogerius, Geerdink, & Geurts, 2009), another behavioral observation system, has similar limitations. A third behavioral observation system, the Shriners Hospital
Upper Extremity Evaluation Spontaneous Functional Analysis (Davids et al., 2006), contains both unimanual and bimanual activities but has not been validated against real-world indices or shown to be sensitive to change. Nevertheless, it would be of interest in future studies to evaluate how scores on the PMAL-R relate to scores these instruments.

Conclusion
In contrast to most other pediatric upper-extremity neurorehabilitation measures, the PMAL-R interview specifically evaluates use of the more-affected arm. Furthermore, only the PMAL-R elicits information about children’s actual behavior in daily life and includes both unimanual and bimanual activities. The availability of a reliable and valid method, such as the PMAL-R, for measuring changes in real-world, more-affected arm use advances pediatric neurorehabilitation by permitting clinicians and researchers to distinguish between existing interventions that only improve motor capacity in the treatment setting and those that in addition improve motor function in the domain that matters most, that is, actual use of that capacity in children’s daily lives.

References