Assessment of upper limb capacity in children with unilateral cerebral palsy: construct validity of a Rasch-reduced Modified House Classification

YVONNE GEERDINK1 | ROBERT LINDEBOOM2,3 | SANDER DE WOLF2 | BERT STEENBERGEN4 | ALEXANDER C H GEURTS5 | PAULINE AARTS1

1 Department of Pediatric Rehabilitation, Sint Maartenskliniek, Nijmegen; 2 Masterstudies Academisch Medisch Centrum-Universiteit van Amsterdam, Amsterdam; 3 Department of Clinical Methods and Public Health, Academic Medical Center, Amsterdam; 4 Behavioural Science Institute, Radboud University, Nijmegen; 5 Nijmegen Centre for Evidence Based Practice, Department of Rehabilitation, Radboud University Medical Centre, Nijmegen, the Netherlands.

Correspondence to Yvonne Geerdink, Department of Pediatric Rehabilitation, Sint Maartenskliniek, Postbus 9011, 6500 GM Nijmegen, the Netherlands. E-mail: Y.Geerdink@maartenskliniek.nl

This article is commented on by Charles on page 513 of this issue.

AIM The aim of this study was to test and improve the unidimensionality and item hierarchy of the Modified House Classification (MHC) for the assessment of upper limb capacity in children with unilateral cerebral palsy (CP) using Rasch analysis. The construct validity of the Rasch-reduced item set was evaluated.

METHOD Modified House Classification items were scored from 369 videotaped assessments of 159 children with unilateral CP (98 males, 61 females; median age 6y 6mo, range 2y 1mo–17y 5mo). Construct validity was tested in 40 other children with unilateral CP (21 males, 19 females; median age 8y 2mo, range 3y 3mo–17y 6mo) by comparing total scores with the Manual Ability Classification System (MACS) and the ABILHAND-Kids scale.

RESULTS Fifteen MHC items could be included in the Rasch analysis. The excluded items were either too easy or too difficult. Fourteen items fitted the unidimensional model ($\chi^2=41.3, df=39, p=0.37$). The hierarchy of these items was different from the original MHC. There was a significant correlation with the MACS ($r=–0.901, p<0.001$) and the ABILHAND-Kids scale ($r=0.558, p<0.001$).

INTERPRETATION The original item hierarchy of the MHC can be improved in order to use its sum score for the assessment of upper limb capacity in children with unilateral CP. The Rasch-reduced 14-item MHC with weighted sum score shows good construct validity to measure functional capacity of the affected hand in children with unilateral CP.

Children with unilateral cerebral palsy (CP), about 30% of the total CP population, typically present with a wide range of unilateral upper limb impairments. They have reduced capacity to handle objects with the affected hand and, consequently, are hindered in the performance of bimanual activities. To remediate these problems, various short and intensive training programmes for the upper limb in children with unilateral CP have been developed.1–4 Valid evaluation of treatment effects requires assessment tools that have adequate clinimetric properties and focus on improvement of functional skills rather than on reduction of impairments. In the case of children with unilateral CP, it is important to measure treatment effects on ‘performance’ (i.e. the actual performance of an activity in daily life)5 as well as to measure the ‘capacity’ (i.e. the execution of an activity in optimal conditions and standardized environment)5 to use the affected hand in bimanual tasks. Several instruments are available for the assessment of upper limb capacity in children with CP. However, as yet, there is no validated instrument that specifically evaluates the maximal capacity of the affected hand to participate in bimanual tasks.

In 1981, House et al.6 devised a nine-level functional classification system to describe the role of the assessed hand in children with CP as a passive or active assist in bimanual activities (Fig. 1). This classification provides categories for upper limb capacity. In 2008, Koman et al.7 introduced the Modified House Functional classification (MHC) with the intention of improving the score discrimination of the original classification, to make it better suited for monitoring patients and to evaluate treatment efficacy.7 Thirty-two items were added to the categories of the House classification (Fig. 1). Items were selected by means of a consensus process among experts. A sum score (i.e. number of passed items) represented the functional capacity of the assessed hand (range 0–32 points). In order to use the sum score of a scale, however, it is necessary that the scale is unidimensional (i.e. all items need to measure the same construct). Although the MHC was shown to be...
reproducible, its unidimensionality or hierarchical properties have never been tested.

The Rasch model for scale validation provides a means to evaluate both the unidimensionality and item hierarchy of a scale. One of the underlying assumptions for Rasch modelling is that items can vary from easy to difficult and, as such, can be ranked on one line, i.e. the logit unit scale (log-odds transformed probabilities). Second, individuals can be ranked from less to more able on the same scale. Thus, easy items can be performed by individuals of almost all ability levels, whereas more able persons are more likely to successfully execute the difficult items as well. Hence, in Rasch modelling, the ranked items and the item scores of the ranked subjects are compared and analysed. Items that do not add to the unidimensional construct or hierarchy can be identified and removed from the scale. The unidimensionality and item hierarchy of the remaining item set can then be tested using model fit statistics, thus asserting that the variation in sum scores can be attributed to differences in the evaluated construct only.

Furthermore, the recently developed Rasch models (e.g. the One Parameter Logistic Model [OPLM] we applied) can be used to weight the items according to their discriminative capacity, asserting a balanced and valid contribution to the sum score.

### METHOD

#### Participants

For the calibration cohort we used the videotaped assessments of all children with unilateral CP who had visited...
our paediatric rehabilitation department in the period September 2006 until April 2010. Children were included if (1) the assessment was videotaped and (2) it was performed and completed according to a standardized assessment protocol, and (3) written parental consent was available for use of the videotaped assessments in clinical research. This resulted in 373 videotaped assessments obtained from 159 children. All characteristics, shown in Table I, were obtained from the assessment reports and the medical records.

A second group of 40 children with unilateral CP was included to constitute the validation cohort. These children had attended our department for modified constraint-induced movement therapy combined with bimanual training between April 2011 and February 2013 (Table I). In this group MHC, MACS, and ABILHAND-Kids scores were assessed by an experienced occupational therapist as part of the standardized pre- and postintervention assessments.

### Assessment of the videotapes

Scoring criteria for the various MHC items have not been published by the developers. Hence, we consulted the authors of the study by Koman et al. to obtain the criteria they used for scoring each item as pass or fail. Six raters (five occupational therapy students and one registered occupational therapist) were instructed and trained to score the MHC items of all videotapes according to these criteria. The videotapes were blinded, so that no demographic or clinical information was disclosed to the raters.

### Table I: Demographic characteristics of the participants

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Calibration cohort (n=159)</th>
<th>Validation cohort (n=40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>98/61</td>
<td>21/19</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>6y 6mo</td>
<td>8y 2mo</td>
</tr>
<tr>
<td>Range</td>
<td>2y 1mo–17y 5mo</td>
<td>3y 2mo–17y 6mo</td>
</tr>
<tr>
<td>Affected side, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>71 (45)</td>
<td>20 (50)</td>
</tr>
<tr>
<td>Left</td>
<td>88 (55)</td>
<td>20 (50)</td>
</tr>
<tr>
<td>Manual Ability Classification System level, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>47 (29.6)</td>
<td>7 (17.5)</td>
</tr>
<tr>
<td>II</td>
<td>77 (48.4)</td>
<td>22 (55)</td>
</tr>
<tr>
<td>III</td>
<td>32 (20.1)</td>
<td>11 (27.5)</td>
</tr>
<tr>
<td>IV</td>
<td>3 (1.9)</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Modified House Classification sum score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>18.5 (5.5)</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>5–29</td>
<td></td>
</tr>
<tr>
<td>House category score (guided by Modified House Classification scores), n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4 (2.5)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3 (1.9)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11 (6.9)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>21 (13.2)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>43 (27)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>62 (39)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>9 (6.7)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6 (3.8)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0 (0)</td>
<td></td>
</tr>
</tbody>
</table>

Interrater reliability was established before the start of the study based on 10 videotaped assessments. The intraclass correlation (model 3, single measure) for the MHC sum score was 0.93 (95% confidence interval [CI] 0.84–0.98). A previous study investigating the use of standardized videotaped examinations to establish the House classification of 10 patients with unilateral CP found interrater reliability of 0.3 and intrarater reliability of 0.58. The good agreement of our raters supports the use of recorded assessments to score the MHC in the current study.

### Rasch analysis

Rasch analysis was used to test both the unidimensionality and item hierarchy of the MHC. More specifically, we used the OPLM, a hybrid between the one-parameter (Rasch) model and two-parameter item response theory (Binbaum) model. First, we excluded items that were passed by less than 5% or more than 95% of the calibration cohort. The difficulty of these items could not be reliably estimated because there was too little variance in the scores. Then, we examined the unidimensionality of the sum score and assigned item weights, according to an item’s discriminative ability, guided by the OPLM analysis. Item weights could range in discriminative ability from 1 (poor) to 5 (high) and were used to weight items before summation. Specific chi square-based goodness-of-fit statistics, implemented in the software, were used to examine the fit of the data to the measurement model. Data fitness tests are based on the OPLM’s expected versus observed proportions of patients with a positive score on the MHC items. Misfitting items, indicated by significant item-orientated tests (p<0.05), were deleted, starting with the item with the highest misfit (chi square-value) and continuing until the overall model fit indicated unidimensionality of the MHC item set. The overall model fit was tested with the chi square-based R1c statistic, which p-value should exceed 0.05 to indicate that the data fit the unidimensional OPLM. MHC item difficulties and person ability measures were determined using conditional maximum likelihood estimation and expressed on a logit unit scale, ranging from −3 (easy) to +3 (hard) in most practical applications. The scale is a log transformation of the odds of an MHC item being successfully carried out. Ultimately, we calculated Cronbach’s z of both the original MHC and the Rasch-reduced scale and compared the item hierarchy obtained by OPLM analysis with that of the current MHC.

### Construct validity

The weighted sum scores of the Rasch-reduced item set were first evaluated within the calibration cohort by relating them to the individual MACS scores and to age using Somer’s d and Spearman’s rho. It was hypothesized that the weighted sum scores would show a strong correlation with the MACS, a measure that classifies the ability to independently use the hands in daily activities. A low or non-existing relation with age was expected and considered...
important because this would confirm that the weighed sum scores were not biased by the age of a child. The correlations were then calculated for the validation cohort; the weighed sum scores with the MACS (Somer’s $d$) and with the ABILHAND-Kids (Spearman’s $r_{ho}$). We expected a substantial but somewhat lower correlation with the ABILHAND-Kids than with the MACS, because the ABILHAND-Kids score for manual ability is based not only on the capacities of the affected hand, but also on the use of compensatory strategies and environmental influences.13

RESULTS
The calibration cohort comprised 159 participants (98 males, 61 females; median age 6y 6mo, range 2y 1mo–17y 5mo; Table I). The MHC scale exhibited no floor or ceiling effects; there were no children with a maximum score and only 2.5% with a minimum score. Most children were classified at MACS levels 1 (29.6%) or 2 (48.4%) and, according to their MHC scores, were classed at House category 3, 4, or 5.

A flow chart of the analysis is shown in Figure 2. In a number of cases the items ‘pick up beans while holding one in ulnar side hand’ (item 25), ‘buttoning’ (item 27), ‘pick up, stabilize, and translate coins’ (item 28), and ‘unilaterally putting small pegs in pegboard’ (item 31) were not presented in the videos. As it is most likely that these items were not assessed because they were too difficult or frustrating for the child under evaluation, we excluded them from the Rasch analysis. Indeed, when available for scoring, these items were passed in 2.2%, 27.6%, 0%, and 3.8% of the cases (items 25, 27, 28, and 31 respectively). Another 13 items were excluded from further analysis because less than 5% or more than 95% of the children passed these items. In four of the 373 video-assessments, values for various MHC items were missing and thus these assessments were excluded, leaving 369 video assessments and 15 MHC items for the Rasch analysis.

Unidimensionality and hierarchy of the MHC
Table II summarizes the results of the Rasch analyses. The item fitness test identified one misfitting item (item 11: ‘gross grasp of small block’), which was subsequently removed. The remaining 14-item set showed goodness of fit to the unidimensional model ($R_{1c} \chi^2=41.3, df=39, p=0.37$), indicating that the 14 items worked well together to measure a single upper limb capacity construct. Item difficulties ranged from $-1.193$ to $1.309$ logits. The item pairs 16 and 18, 20 and 21, and 24 and 26 had similar item difficulties. The item difficulty hierarchy was different from the ordering suggested by Koman et al. This applies...
specifically to items 14, 19, and 23. Based on the Rasch analysis, items 14 and 19 both had a higher item difficulty than four originally higher ranked items, whereas item 23 was much easier than weighted in the original MHC item ordering. The discriminative capacity of the items, as indicated by the item weights, ranged from 2 to 5. Items 9, 18, and 22 had the lowest discrimination and item 20 (‘perform pad-to-pad pinch’) had the highest discrimination.

Internal consistency (Cronbach’s z) of the Rasch-reduced 14-item scale was 0.85, compared with 0.87 for the total MHC with 28 evaluated items.

### Construct validity

Within the calibration cohort the 14-item weighted sum score correlated well with the MACS ($r=0.688, p<0.001$). There was no correlation with age ($r=0.096, p=0.231$). Thirty children (8.1%) failed the easiest item of the reduced scale and 39 (10.6%) passed the two most difficult items. Only 2.4% failed all items, whereas 2.2% passed all items and subsequently reached the maximum score. In the validation cohort the weighted sum scores of the 14-item MHC scale were approximately normally distributed with a mean of 24.65 (SD 9.8). One child passed all items (maximum score). There was a strong correlation with the MACS ($r=0.901, p<0.001$) and a fair correlation with the ABILHAND-Kids ($r=0.558, p=0.001$). Assessment and scoring of the 14-item scale in this cohort took about 10 minutes per child.

### DISCUSSION

The purpose of this study was to evaluate whether the MHC could be validly used to measure functional capacity of the affected hand in children with unilateral CP. We examined the unidimensionality and item hierarchy of the MHC for this particular group and tested its construct validity.

The empirical evidence demonstrated that almost half of the MHC items were either too easy or too difficult for children with unilateral CP. The Rasch analyses showed that 14 of the original MHC items fitted a unidimensional model, implying that the variation in scores can be attributed to a single construct: the functional capacity of the affected hand. Reducing the scale to 14 items did not affect the internal consistency. We found that, for some items, the item hierarchy was different from in the original scale. Furthermore, the items did not discriminate with equal precision. Hence, if they are used in a scale they need to be weighted accordingly before summation to obtain a more reliable result. Taken together, for use in children with unilateral CP, the MHC could be reordered and reduced to a 14-item scale with a weighted sum score that expresses the functional capacity of the affected hand.

Instruments that are valid to assess upper limb capacity in children with CP are scarce and the frequently used ‘capacity’ instruments have several limitations. The Melbourne Assessment of Unilateral Upper Limb Function and the Quality of Upper Extremity Skills Test measure (overall) upper limb quality of movement, but have only a few items addressing grasping and releasing objects. In addition, the Quality of Upper Extremity Skills Test is not applicable to children of 8 years and older. The Jebson-Taylor Hand Function Test assesses unilateral efficiency in seven (timed) tasks focusing on unimanual capacity; however, it is a generic norm-referenced instrument and is often modified to meet the abilities of children with CP. The 14-item MHC scale specifically assesses grasping, holding, manipulating, and releasing objects. As such, this scale can fill a gap in the upper limb assessments in children with unilateral CP, because it specifically focuses on the capacities of the affected hand as an assist in handling objects.
In the calibration cohort the sum score showed a strong correlation with the MACS, confirming that children with a good MACS classification (e.g. level 1) are more likely to score high on the 14-item MHC scale. This strong convergent validity was confirmed in the validation cohort. We found that the relationship of the MHC scale with the ABILHAND-Kids was somewhat weaker, which can be explained by the characteristics of both instruments. The 14-item MHC scale assesses unilateral functional capacity, whereas manual ability assessed with the ABILHAND-Kids is a fusion of many aspects (including the capacity of the dominant hand and the use of compensatory strategies), of which the capacity of the affected hand is just one. There was no correlation with age, suggesting that the 14-item MHC scale can validly be applied in various age groups.

The item difficulties were evenly spread over the range of the scale. With 8.1% of the children failing the easiest item and 10.6% passing the most difficult item, this 14-item MHC scale has a floor and ceiling effect, although this is not reflected in the sum score. Less than 2.5% of the calibration cohort reached a minimum or maximum score. Our study also revealed three item pairs with almost identical item difficulties. Technically it would be better to remove the less discriminative item of each pair and recalculate the effects on the properties of the scale.

Limitations
The results of this study apply only to children with unilateral CP. Although our calibration cohort was typical of children with unilateral CP who are treated in rehabilitation centres, it did not represent the entire spectrum of children with CP for whom the MHC was originally developed. In particular, children with good upper limb capacity who are rarely treated in rehabilitation centres are not represented, which might account for the finding that only a few children in our study passed the items at the top of the MHC scale (i.e. items 25, 27–32), reflecting (near) normal hand function. As a consequence of the aim of study, children with bilateral CP who are more likely to have severe upper limb disabilities (e.g. MACS levels 4 and 5) were not included either, although the MHC was developed for these children as well. Future studies should, therefore, include other CP subtypes to identify the item difficulty and hierarchy of the lowest and highest MHC items.

Because this study was retrospective in nature, some data related to four items were missing from the calibration cohort and, therefore, could not be included in the Rasch analysis. We expect that the missing values were caused by these items being too difficult, as they concerned the highest ranked items of the MHC. Although not absolutely certain, this notion was supported by the percentage of passed items in the rest of the sample; for three of the four items being scored, the pass rate was less than 4%. Item 27 ‘able to fasten and unfasten a button bimanually’, for which the success rate among those in whom this item was presented was 27.6%, should be re-evaluated in future studies to establish its difficulty and contribution to the scale.

Finally, age, sex, and affected side may influence fine motor control of the upper limb in young children. Differential item functioning of the 14-item MHC scale should, therefore, be formally evaluated in various subgroups consisting of at least 200 patients each.

CONCLUSION
The results of this study suggest that the 14-item MHC scale might be useful in measuring functional capacity of the affected hand in children with unilateral CP. The reduced scale with reordered items has good internal consistency, but the item scores must be weighted for valid interpretation of the sum score.

Future studies are needed to further improve the applicability of the MHC scale in children with unilateral CP. In particular, its responsiveness should be established, differential item functioning should be tested in various subgroups, and construct validity should be strengthened by comparing the 14-item MHC scale with other well-established upper limb measures such as the Assisting Hand Assessment and the Melbourne Assessment.

ACKNOWLEDGEMENTS
We thank Simone Been, Mariëlle Brommet, Anja Harkink, Jeroen Lafieber, and Jorien Schuit for their excellent and time-consuming work scoring all video assessments used in this study. The authors have stated that they had no interests that might be perceived as posing a conflict or bias.

SUPPORTING INFORMATION
The following additional material may be found online:
Appendix SI: Conversion table of 14-item weighted sum score to obtain the functional capacity measure in logit units.

REFERENCES