

# Comparison of the Melbourne Assessment of Unilateral Upper Limb Function and the Quality of Upper Extremity Skills Test in hemiplegic CP

**K Klingels\*** MSc, Department of Rehabilitation Sciences, Faculty of Kinesiology and Rehabilitation Sciences;  
**P De Cock** MD PhD, Department of Paediatrics, Faculty of Medicine;  
**K Desloovere** PhD, Department of Rehabilitation Sciences, Faculty of Kinesiology and Rehabilitation Sciences, Katholieke Universiteit Leuven;  
**C Huenaerts** MSc, Clinical Motion Analysis Laboratory CERM, University Hospital Leuven;  
**G Molenaers** MD PhD, Department of Orthopaedics, Faculty of Medicine;  
**I Van Nuland** MSc;  
**A Huysmans** MSc;  
**H Feys** PhD, Department of Rehabilitation Sciences, Faculty of Kinesiology and Rehabilitation Sciences, Katholieke Universiteit Leuven, Belgium.

*\*Correspondence to first author at Department of Rehabilitation Sciences, Katholieke Universiteit Leuven, Tervuursevest 101, 3001 Heverlee (Leuven), Belgium.  
E-mail: Katrijn.Klingels@faber.kuleuven.be*

DOI: 10.1111/j.1469-8749.2008.03123.x  
Published online September 22 2008

This study investigated interrater reliability and measurement error of the Melbourne Assessment of Unilateral Upper Limb Function (Melbourne Assessment) and the Quality of Upper Extremity Skills Test (QUEST), and assessed the relationship between both scales in 21 children (15 females, six males; mean age 6y 4mo [SD 1y 3mo], range 5–8y) with hemiplegic CP. Two raters scored the videotapes of the assessments independently in a randomized order. According to the House Classification, three participants were classified as level 1, one participant as level 3, eight as level 4, three as level 5, one participant as level 6, and five as level 7. The Melbourne Assessment and the QUEST showed high interrater reliability (intraclass correlation 0.97 for Melbourne Assessment; 0.96 for QUEST total score; 0.96 for QUEST hemiplegic side). The standard error of measurement and the smallest detectable difference was 3.2% and 8.9% for the Melbourne Assessment and 5.0% and 13.8% for the QUEST score on the hemiplegic side. Correlation analysis indicated that different dimensions of upper limb function are addressed in both scales.

Hemiplegia is the most common type of cerebral palsy (CP) among children born at term and the second most common after diplegia among preterm infants.<sup>1</sup> Children with hemiplegia experience difficulties in arm and hand functions, such as the ability to reach, manipulate, grasp, and release. To evaluate the effectiveness of treatment, it is crucial to use validated measurements comprising the most important components of upper limb function. The two scales most frequently used to measure unilateral upper limb function in children with CP are the Quality of Upper Extremity Skills Test (QUEST)<sup>2</sup> and the Melbourne Assessment of Unilateral Upper Limb Function (Melbourne Assessment).<sup>3</sup>

The QUEST is a criterion-referenced measurement tool, developed to evaluate upper extremity quality of movement in children with CP aged from 18 months to 8 years. So far, this test has most frequently been used in intervention studies.<sup>4–9</sup> DeMatteo et al.<sup>10</sup> demonstrated good interobserver and test–retest reliability. The QUEST correlates strongly with the Peabody Developmental Fine Motor Scales, supporting its concurrent validity.<sup>10,11</sup> A more recently developed scale is the Melbourne Assessment.<sup>3</sup> This scale is an evaluative tool that measures unilateral upper extremity quality of movement in children with neurological impairments aged from 5 to 15 years. It has also proved to be a reliable and valid measurement<sup>3,12,13</sup> and has been used in several intervention studies.<sup>9,14–17</sup> The Melbourne Assessment demonstrates a strong correlation with the Pediatric Evaluation of Disability Inventory.<sup>17</sup>

The age group covered in both scales is different, but there is an overlap between 5 and 8 years. To the best of our knowledge, the relationship between these scales has not been assessed previously. Also, the reliability of both scales in the specific group of children with hemiplegia has only been evaluated to a limited extent. Most of the participants included in previous reliability studies were children with spastic diplegia or quadriplegia, ataxia, and athetosis, but few children with hemiplegia.<sup>3,10,12,13</sup> In addition, the reliability of the QUEST items has never been defined. For the QUEST, the standard error of measurement (SEM) and the smallest detectable difference (SDD) have not yet been investigated. For the Melbourne Assessment, the SEM has only been assessed in a group of children with quadriplegia and hemiplegia between 5 years 5 months and 12 years.<sup>13</sup>

The aims of this study were: (1) to evaluate the interrater reliability of the Melbourne Assessment (sub-items and total score) and the QUEST (items, domains, and total score) in children with hemiplegia between 5 and 8 years; (2) to determine the SEM and SDD; and (3) to assess the relationship between the scales.

## Method

### PARTICIPANTS

Participants were included in the study if they were diagnosed with a predominantly spastic type of hemiplegic CP, aged between 5 and 8 years, and able to understand the test instructions.

Twelve participants were recruited from the Clinical Motion Analysis Laboratory of the University Hospital in Pellenberg, four participants from special education schools, and six participants from private practices. The study population consisted of 22 participants. Ethical approval for this

study was obtained from the Ethics Committee of the University Hospital of Leuven and consent was obtained from the parents.

#### ASSESSMENT

The QUEST,<sup>2</sup> the Melbourne Assessment,<sup>3</sup> and the House Classification<sup>18</sup> were performed for all participants.

The QUEST comprises 34 items related to quality of movement in four domains that are essential components of normal developmental patterns between birth and 18 months (dissociated movements, grasp, weight bearing, and protective extension). For each domain, a percentage score is calculated. The scores of the four domains can be summed to a total percentage score. According to the instruction manual, the total QUEST score and the scores of the domains are based on the summation of the hemiplegic and the non-hemiplegic side. To create the possibility to compare the scores on the QUEST with the Melbourne Assessment, which is a unilateral evaluation, the raw and percentage scores were calculated for the hemiplegic side separately.

The Melbourne Assessment is based on 16 items comprising tasks that are representative of the most important components of unilateral upper limb function (reach, grasp, release, and manipulate). Most items are further subdivided in two to four sub-items that represent an aspect of the required movement, such as range of movement, fluency, target accuracy, speed, and quality of movement. The total score can range from 0 to 122 points and can be converted to a percentage.

The House Classification describes the functional level of the hemiplegic arm in nine levels, ranging from non-use to spontaneous and completely independent use during activities. In a previous unpublished study in 30 participants with hemiplegia, we found a high interrater (intraclass correlation coefficient [ICC]=0.86) and very high test-retest reliability (ICC=0.96).

#### PROCEDURE

The tests were administered and videotaped by two therapists. The items of the Melbourne Assessment and the QUEST that are performed in the same position were combined. This made the protocol more practicable and reduced the test duration. In the item order, there was no advantage for either test. Before being rated, the tapes were edited to separate the test items on the Melbourne Assessment and the QUEST. To ensure blindedness to each participant's performance on the alternate measure, two physiotherapists, different from the two therapists who had administered and videotaped the assessments, scored the tapes independently in a randomized order. Participants were randomly allocated to two groups (A and B) using an envelope system. Rater 1 first scored the Melbourne Assessment in group A and the QUEST in group B. Rater 2 first scored the QUEST in group A and the Melbourne Assessment in group B. Both raters were trained in the scoring of both scales.

#### STATISTICAL ANALYSIS

##### *Interrater reliability of Melbourne Assessment and QUEST scores*

Interrater reliability was assessed by percentage of agreement, Cohen's kappa ( $\kappa$ ), weighted kappa ( $\kappa_w$ ), and ICC

(1.1), depending on the type of variable. The 95% confidence intervals (CIs) for  $\kappa$ ,  $\kappa_w$ , and ICC were calculated. According to the guidelines of Landis and Koch,<sup>19</sup> kappa values of more than 0.80 are considered as excellent, between 0.61 and 0.80 as substantial, between 0.40 and 0.60 as moderate, and less than 0.40 as poor to fair. For the ICC, values above 0.80 are considered as very high, between 0.60 and 0.79 as moderately high, between 0.40 and 0.59 as moderate, and less than 0.40 as low.<sup>20</sup>

##### *Standard error of measurement and smallest detectable difference of the Melbourne Assessment and the QUEST*

The SEM and the SDD were calculated for the Melbourne Assessment, the total QUEST score, and the QUEST score on the hemiplegic side by using following formulae:  $SEM = SD \times \sqrt{(1-ICC)}$  and  $SDD = SEM \times 1.96 \times \sqrt{2}$ .<sup>21</sup> The consistency of measurements was examined using the method of Bland and Altman.<sup>22</sup>

##### *Relationship between scores on the Melbourne Assessment and the QUEST*

Pearson's correlation coefficient was used to assess the correlation between the scales. The normality of the data distribution was verified by the Shapiro-Wilk statistic. Total scores of the Melbourne Assessment and the QUEST were normally distributed.

## Results

#### PARTICIPANT CHARACTERISTICS

One participant could not complete the test owing to lack of cooperation. The remaining 21 participants (15 females, six males) with a mean age of 6 years 4 months (SD 1y 3mo, range 5–8y) were included for further analysis. Nine participants had left and 12 right hemiplegia. Fourteen participants attended mainstream school and seven were in special education.

According to the House Classification, three participants were classified as level 1 (use the hemiplegic arm as stabilizing weight), one participant as level 3 (hold and stabilize object for contralateral hand function), eight as level 4 (actively grasp and hold an object weakly), three as level 5 (actively grasp an object and stabilize it well), one participant as level 6 (actively grasp an object and manipulate it), and five participants as level 7 (perform bimanual activities easily and occasionally uses the hand spontaneously).

#### INTERRATER RELIABILITY OF MELBOURNE ASSESSMENT AND QUEST SCORES

The ICC for the total score of the Melbourne Assessment was 0.97 (95% CI 0.93–0.99). Percentage of agreement of the 32 sub-items varied between 35% and 95% (Table I). Almost half of the sub-items showed a substantial (>0.60) to excellent (>0.80) agreement, calculated by  $\kappa_w$ . Thirteen sub-items had a moderate agreement (0.40–0.60) and five fluency sub-items a poor to fair agreement (<0.40).

For the total QUEST score and for the total score on the hemiplegic side, the ICC was 0.96 (95% CI total score 0.90–0.98; 95% CI hemiplegic side 0.92–0.99; Table II). The ICCs for the four domains varied between 0.89 and 0.95. Reliability coefficients of the QUEST items for the hemiplegic side are presented in Table III. Eleven (weighted) kappa values exceeded 0.80, 15 were between

0.61 and 0.80, and for three items  $\kappa_w$  ranged between 0.40 and 0.60. On the non-hemiplegic side, all participants scored almost perfect to perfect for all items.

#### SEM AND SDD OF THE MELBOURNE ASSESSMENT AND THE QUEST

The SEM was 2.6% for the Melbourne Assessment, 3.2% for the total QUEST, and 5.0% for the QUEST score on the hemiplegic side. The SDD was 8.9% for the Melbourne Assessment, and 7.11% and 13.8% for the total QUEST score

**Table I: Percentage of agreement and weighted kappa with 95% confidence intervals (CI) for interrater agreement of the sub-items of the Melbourne Assessment**

	Percentage of agreement	Weighted kappa	95% CI
<i>Reach forwards</i>			
ROM	95	0.92	0.76–1.00
Target accuracy	65	0.52	0.23–0.81
Fluency	35	0.23	0.01–0.45
<i>Reach forwards elevated</i>			
ROM	85	0.75	0.49–1.00
Target accuracy	85	0.78	0.54–1.00
Fluency	50	0.33	0.04–0.63
<i>Reach sideways elevated</i>			
ROM	86	0.82	0.63–1.00
Target accuracy	71	0.60	0.31–0.90
Fluency	62	0.47	0.20–0.74
<i>Grasp crayon</i>	76	0.81	0.64–0.98
<i>Drawing grasp</i>	86	0.85	0.70–1.00
<i>Release crayon</i>			
ROM	76	0.74	0.53–0.95
Quality of movement	57	0.56	0.29–0.83
Target accuracy	95	0.90	0.70–1.00
<i>Grasp pellet</i>	71	0.76	0.58–0.94
<i>Release pellet</i>			
ROM	76	0.77	0.58–0.96
Quality of movement	91	0.81	0.64–0.98
Target accuracy	86	0.80	0.59–1.00
<i>Manipulation</i>			
Finger dexterity	71	0.68	0.47–0.88
Fluency	57	0.17	0.00–0.44
<i>Pointing</i>			
Red	71	0.69	0.45–0.92
Green	57	0.56	0.31–0.80
Yellow	67	0.65	0.41–0.88
Blue	62	0.59	0.33–0.85
<i>Brush from forehead to neck</i>			
ROM	62	0.60	0.37–0.82
Fluency	52	0.31	0.05–0.58
<i>Palm to bottom</i>			
ROM	81	0.77	0.58–0.95
Fluency	52	0.19	0.00–0.41
<i>Hand to hand transfer</i>	57	0.51	0.24–0.78
<i>Reach to opposite shoulder</i>			
ROM	67	0.60	0.33–0.87
Target accuracy	76	0.69	0.43–0.94
Fluency	52		
<i>Hand to mouth and down</i>			
ROM	71	0.58	0.27–0.89
Target accuracy	67	0.54	0.21–0.88
Fluency	76	0.54	0.22–0.86
Speed	67	0.42	0.00–0.83

ROM, range of motion.

and the score on the hemiplegic side of the QUEST respectively.

The Bland–Altman plots for interrater agreement for the Melbourne Assessment and the QUEST on the hemiplegic

**Table II: Intraclass correlation coefficient (ICC) and 95% confidence intervals (CI) for interrater agreement of domains and total score of the QUEST**

Domains	Hemiplegic side		Total QUEST	
	ICC	95% CI	ICC	95% CI
Dissociated movements	0.95	0.88–0.98	0.91	0.79–0.96
Grasp	0.93	0.83–0.97	0.93	0.84–0.97
Weight bearing	0.92	0.81–0.97	0.89	0.76–0.95
Protective extension	0.91	0.80–0.96	0.92	0.82–0.97
Total	0.96	0.92–0.99	0.96	0.90–0.98

QUEST, Quality of Upper Extremity Skills Test.

**Table III: Percentage of agreement, kappa ( $\kappa$ ), or weighted kappa ( $\kappa_w$ ), and 95% confidence intervals (CI) for interrater agreement of items of the QUEST for the hemiplegic side**

	Percentage of agreement	Kappa/ weighted kappa	95% CI
<i>Dissociated movements</i>			
<i>Shoulder</i>			
Item 1	91	$\kappa_w=0.71$	0.44–0.97
Item 2	86	$\kappa_w=0.73$	0.44–1.00
Item 3	76	$\kappa_w=0.70$	0.46–0.94
Item 4	91	$\kappa_w=0.84$	0.61–1.00
<i>Elbow</i>			
Item 1	95	$\kappa_w=0.91$	0.74–1.00
Item 2	84	$\kappa_w=0.62$	0.25–1.00
Item 3	68	$\kappa_w=0.47$	0.16–0.78
Item 4	89	$\kappa_w=0.89$	0.76–1.00
<i>Wrist</i>			
Item 1	71	$\kappa_w=0.63$	0.36–0.90
Item 2	76	$\kappa_w=0.68$	0.43–0.93
Item 3	81	$\kappa_w=0.75$	0.52–0.98
Item 4	71	$\kappa_w=0.47$	0.13–0.82
Item 5	86	$\kappa_w=0.69$	0.33–1.00
<i>Fingers</i>			
Item 1	100	$\kappa=1.00$	1.00–1.00
Item 2	100	$\kappa=1.00$	1.00–1.00
Grasp cube	81	$\kappa_w=0.71$	0.45–0.98
Release cube	90	$\kappa_w=0.84$	0.62–1.00
<i>Grasp</i>			
Posture	66	$\kappa_w=0.45$	0.17–0.72
Cube	81	$\kappa_w=0.85$	0.70–0.99
Cereal	81	$\kappa_w=0.88$	0.77–0.99
Pencil	90	$\kappa_w=0.93$	0.85–1.00
<i>Weight bearing</i>			
Weight bearing	76	$\kappa_w=0.79$	0.60–0.98
With reach	95	$\kappa=0.90$	0.71–1.00
Forwards	90	$\kappa_w=0.76$	0.54–0.98
Sideways	71	$\kappa_w=0.66$	0.40–0.92
Backwards	76	$\kappa_w=0.75$	0.53–0.98
<i>Protective extension</i>			
Forwards	79	$\kappa_w=0.72$	0.45–0.99
Sideways	85	$\kappa_w=0.78$	0.52–1.00
Backwards	84	$\kappa_w=0.90$	0.78–1.00

QUEST, Quality of Upper Extremity Skills Test.

side are shown in Figure 1. The mean difference of the total Melbourne scores was  $-1.1\%$  (95% CI  $-8.5\%$  to  $10.6\%$ ). For the total score on the hemiplegic side of the QUEST, the mean difference was  $1.5\%$  (95% CI  $-11.9\%$  to  $15.0\%$ ). For the Melbourne Assessment, Figure 1a shows that rater 1 scored slightly lower than rater 2 for most participants with a total score in the mid range. For the three participants with the best performance on the Melbourne Assessment, rater 1 scored clearly higher than rater 2. For the total QUEST scores on the hemiplegic side, there was no systematic bias between both observers (Fig. 1b).

#### RELATIONSHIP BETWEEN SCORES ON THE MELBOURNE ASSESSMENT AND THE QUEST

The scores of the first observer were used to compare the two scales. The Melbourne Assessment showed a mean value of  $56.9\%$  (SD  $20.6$ , range  $30.3$ – $98.4\%$ ). The mean score of the total QUEST was  $73.1\%$  (SD  $13.2$ , range  $49.0$ – $97.2\%$ ), and for the hemiplegic side  $50.6\%$  (SD  $25.6$ , range  $3.1$ – $96.7\%$ ).

The Pearson's correlation coefficients between the Melbourne Assessment and the QUEST (total score and score on the hemiplegic side) were  $0.83$  and  $0.81$  respectively ( $p < 0.001$ ). Comparison of the Melbourne Assessment with the four QUEST domains on the hemiplegic side (dissociated movements, grasp, weight bearing, protective extension) showed coefficients ( $r$ ) of  $0.87$  ( $p < 0.001$ ),  $0.83$  ( $p < 0.001$ ),  $0.50$  ( $p = 0.021$ ), and  $0.36$  ( $p = 0.114$ ).

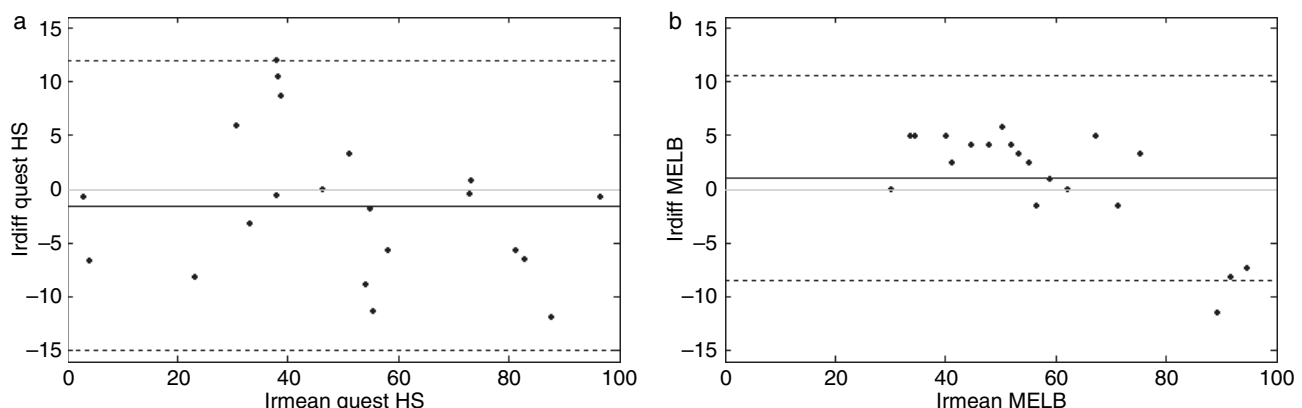
#### Discussion

This study assessed the interrater reliability of the Melbourne Assessment and the QUEST in a group of young children with hemiplegic CP. For the Melbourne Assessment, a very high ICC value was found which was comparable to the values reported in the studies of Randall et al.<sup>3,12</sup> and Cusick et al.<sup>13</sup> It should be noted that the age group and the type of CP were different in those studies. In the studies of Randall et al.,<sup>3,12</sup> participants were aged between 5 and 16 years and were diagnosed with spastic diplegia or quadriplegia, athetosis, or ataxia. The study of Cusick et al.<sup>13</sup> included participants aged between 5 and 12 years with spastic hemiplegia or quadriplegia. Our study confirms the high interrater reliability for the younger age group (between 5 and 8y) with

a diagnosis of hemiplegic CP. Randall et al.<sup>3</sup> reported that the reliability of the 32 sub-items varied considerably. In the present study, most sub-items had moderate to good agreement. There was a lower reliability of the fluency sub-items and the speed sub-item of 'Hand to mouth and down' which was also found in the study of Randall et al.<sup>3</sup> Clearly, these aspects are more difficult to judge on videotapes.

For the QUEST, interrater reliability of the total score and the domains was very high and in agreement with the study of DeMatteo et al.,<sup>10</sup> in which only the domain 'Protective extension' showed a clearly lower reliability. The reliability analysis of the items is a unique aspect of this study and, therefore, this could not be compared with the literature. Also, very high ICC coefficients were demonstrated for the total score and the domains of the QUEST for the hemiplegic side. To our knowledge, this has not been reported in previous studies.

For intervention studies and longitudinal follow-up, knowledge of the SEM and the SDD is important to ascertain which participants show a 'true' improvement rather than a difference in score due to measurement error. The SEM for the Melbourne Assessment was  $3.2\%$ , which is slightly higher than the SEM of the trained group ( $2.6\%$ ) in the study of Cusick et al.,<sup>13</sup> however, the study population differed in type of CP and age group. The SDD for the Melbourne Assessment was  $8.9\%$ , indicating that two assessments of the same child should differ by more than approximately  $9\%$  to reflect a true change in function rather than a 'chance' occurrence or an error in measurement. In the study of Randall et al.,<sup>12</sup> a 95% interrater repeatability range of  $14\%$  was estimated. To the best of our knowledge, no previous study has reported the SEM or SDD of the QUEST. For the total QUEST score, the SDD was  $7.1\%$ , which is lower than the SDD of the hemiplegic side ( $13.8\%$ ). This can be explained by the range of scores of the total QUEST, which was between  $49\%$  and  $97.2\%$  because of the maximal or nearly maximal scores on the non-hemiplegic side. Although the SDD of the QUEST on the hemiplegic side was slightly higher than the SDD of the Melbourne Assessment, these values correspond to reported measurement errors of other scales of upper limb function in stroke patients, such as the Action Research Arm Test<sup>23</sup> and Brunnström-Fugl-Meyer Test.<sup>24</sup> For the Bland-Altman limits



**Figure 1:** A Bland-Altman plot of the range of differences of total scores between raters ( $IR_{DIFF}$ ) against range of mean scores of the two raters. ( $IR_{MEAN}$ ) is shown for (a) the Quality of Upper Extremity Skills Test (QUEST HS) score on the hemiplegic side and for (b) the Melbourne Assessment of Unilateral Upper Limb Function (MELB). Solid line represents mean of the differences, and dashed lines limits of agreement (mean difference of  $2SD$ ).



of agreement of the Melbourne Assessment, the plot showed a systematic bias between the observers. Inspection of the data showed that for the three children with the highest Melbourne scores, rater 2 scored lower for all fluency items than rater 1. These results illustrate that Blant–Altman plots give important information on possible systematic differences between raters and contribute to optimizing the reliability. The findings of Cusick et al.<sup>13</sup> do not suggest that face-to-face training intervention is essential for reliable scoring of the Melbourne Assessment by novice users. Nevertheless, in our study, both observers were trained with 10 videotapes on previously assessed children. We recommend this training intervention to standardize the scoring system between raters and to aim for optimal reliability of the assessment.

The correlation between Melbourne Assessment and QUEST has not been assessed previously and was another aim of our study. Our results showed a high correlation between the two. This correlation supports the concurrent validity of the scales. However, inspection of the scales' content indicates that they measure different aspects of upper limb function. The QUEST emphasizes the body function level, including analytic movements in the domain 'Dissociated movements' and gross motor function in the domains 'Protective extension' and 'Weight bearing'. The Melbourne Assessment is a more activity-based measurement tool according to the International Classification of Functioning, Disability and Health, and is, therefore, more appropriate to determine the status of a child's functional and fine motor skills. Furthermore, the Melbourne Assessment evaluates other aspects of movement, such as speed, accuracy, and fluency, that are not considered in the QUEST. Still, correlation analysis shows a high association between the QUEST domain 'Dissociated movements' and the Melbourne Assessment. This finding stresses the importance of selective arm and hand movements for functional tasks. In contrast, weight bearing and protective extension, providing information on postural responses, were poorly correlated with arm and hand function measured in the Melbourne Assessment.

The QUEST and the Melbourne Assessment both measure the capacity of unilateral hand function. Apart from this aspect, it is also crucial to gain more insight into the performance of bimanual activities. This can be evaluated with a more recent scale, the Assisting Hand Assessment,<sup>25</sup> which measures the spontaneous use or performance of the affected hand in bimanual activities in a semi-structured play session. By comparing the results of the capacity of the hemiplegic side with the performance of the hemiplegic side in bimanual activities, instructive conclusions can be drawn about the developmental non-use of the hemiplegic side. This can be of interest for further research.

This study warrants some critical reflections. The combination of the tests resulted in a long test duration of approximately 1 hour. Still, all children except one (this participant had behavioural problems and failed to cooperate until the end) easily completed the test procedure. The study was limited to interrater reliability; intrarater and test–retest reliability were not evaluated. Further studies are required to evaluate other psychometric properties such as responsiveness, predictive validity, and Rasch anal-

ysis. This was the first study that compared the two most frequently used scales for unilateral upper limb function in a homogeneous study sample, including statistical properties.

## Conclusion

This study showed that the Melbourne Assessment and the QUEST are reliable measurement scales to evaluate unilateral upper limb function in young children with hemiplegic CP. The scales showed comparable measurement errors. Correlation analysis indicated that different dimensions of upper limb function are addressed in both scales.

*Accepted for publication 2nd May 2008.*

## References

1. Hagberg B, Hagberg G, Olow I, Van Wendt L. The changing panorama of cerebral palsy in Sweden. VII. Prevalence and origin in the birth year period 1987–1990. *Acta Paediatr* 1996; **85**: 954–60.
2. DeMatteo C, Law M, Russell D, Pollock N, Rosenbaum P, Walters S. Quality of Upper Extremity Skill Test. Ontario: Neurodevelopmental Clinical Research Unit, 1992.
3. Randall M, Johnson L, Reddihough D. The Melbourne Assessment of Unilateral Upper Limb Function: test administration manual. Melbourne: Royal Children's Hospital, 1999.
4. Fehlings D, Rang M, Glazier J, Steele C. An evaluation of botulinum-A toxin injections to improve upper extremity function in children with hemiplegic palsy. *J Pediatr* 2000; **137**: 331–37.
5. Law M, Russell D, Pollock N, Rosenbaum P, Walter S, King G. A comparison of intensive neurodevelopmental therapy plus casting and a regular occupational therapy program for children with cerebral palsy. *Dev Med Child Neurol* 1997; **39**: 664–70.
6. Naylor CE, Bower E. Modified constraint-induced movement therapy for young children with hemiplegic cerebral palsy: a pilot study. *Dev Med Child Neurol* 2005; **47**: 365–69.
7. Lowe K, Novak I, Cusick A. Low-dose/high-concentration localized botulinum toxin A improves upper limb movement and function in children with hemiplegic cerebral palsy. *Dev Med Child Neurol* 2006; **48**: 170–75.
8. Deluca SC, Echols K, Law CR, Ramey SL. Intensive pediatric constraint-induced therapy for children with cerebral palsy: randomized, controlled crossover trial. *J Child Neurol* 2006; **21**: 931–38.
9. Wallen M, O'Flaherty SJ, Waugh MA. Functional outcomes of intramuscular botulinum toxin type A and occupational therapy in the upper limbs of children with cerebral palsy: a randomized controlled trial. *Arch Phys Med Rehabil* 2007; **88**: 1–10.
10. DeMatteo C, Law M, Russell D, Pollock N, Rosenbaum P, Walter S. The reliability and validity of the quality of upper extremity skills test. *Phys Occup Ther Pediatr* 1993; **13**: 1–18.
11. Law M, Cadman D, Rosenbaum P, DeMatteo C, Walter S, Russell D. Neurodevelopmental therapy and upper-extremity inhibitive casting for children with cerebral palsy. *Dev Med Child Neurol* 1991; **33**: 379–87.
12. Randall M, Carlin JB, Chondros P, Reddihough D. Reliability of the Melbourne Assessment of Unilateral Upper Limb Function. *Dev Med Child Neurol* 2001; **43**: 761–67.
13. Cusick A, Vasquez M, Knowles L, Wallen M. Effect of rater training on reliability of Melbourne Assessment of Unilateral Upper Limb Function scores. *Dev Med Child Neurol* 2005; **47**: 39–45.
14. Wallen M, O'Flaherty SJ, Waugh MCA. Functional outcomes of intramuscular botulin toxin type A in the upper limbs of children with cerebral palsy: a phase II trial. *Arch Phys Med Rehabil* 2004; **85**: 192–200.
15. Speth LAWM, Leffers P, Janssen-Potten YJM, Vles JSH. Botulinum toxin A and upper limb functional skills in hemiparetic

- cerebral palsy: a randomized trial in children receiving intensive therapy. *Dev Med Child Neurol* 2005; **47**: 468–73.
16. Bourke-Taylor H. Melbourne Assessment of Unilateral Upper Limb Function: construct validity and correlation with the Pediatric Evaluation of Disability Inventory. *Dev Med Child Neurol* 2003; **45**: 92–96.
  17. Ozer K, Chesher SP, Scheker LR. Neuromuscular electrical stimulation and dynamic bracing for the management of upper-extremity spasticity in children with cerebral palsy. *Dev Med Child Neurol* 2006; **48**: 559–63.
  18. House J, Gwathmey F, Fidler M. A dynamic approach to the thumb-in-palm deformity in cerebral palsy. *J Bone Joint Surg* 1981; **63**: 216–25.
  19. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **86**: 420–28.
  20. Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparing measurement sensitivity of short and longer health status instruments. *Med Care* 1992; **30**: 917–25.
  21. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998; **26**: 217–38.
  22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **i**: 307–10.
  23. Van der Lee JH, De Groot V, Beckerman H, Wagenaar RC, Lankhorst GJ, Bouter LM. The intra- and interrater reliability of the Action Research Arm test: a practical test of upper extremity function in patients with stroke. *Arch Phys Med Rehabil* 2001; **82**: 14–19.
  24. Van der Lee JH, Beckerman H, Lankhorst GJ, Bouter LM. The responsiveness of the Action Research Arm test and the Fugl-Meyer Assessment scale in chronic stroke patients. *J Rehab Med* 2001; **33**: 110–13.
  25. Krumlinde-Sundholm L, Eliasson AC. Development of the Assisting Hand Assessment: a Rasch built measure intended for children with unilateral upper limb impairments. *Scand J Occup Ther* 2003; **10**: 16–26.