Assessment of upper-limb capacity, performance, and developmental disregard in children with cerebral palsy: validity and reliability of the revised Video-Observation Aarts and Aarts module: Determine Developmental Disregard (VOAA-DDD-R)

ANNEMIEKE HOUWINK[1] | YVONNE A GEERDINK[2] | BERT STEENBERGEN[3] | ALEXANDER C H GEURTS[1] | PAULINE B M AARTS[2]

1 Radboud University Nijmegen Medical Centre, Nijmegen Centre for Evidence Based Practice, Department of Rehabilitation, Nijmegen; 2 Department of Pediatric Rehabilitation, Sint Maartenskliniek, Nijmegen; 3 Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, the Netherlands.

Correspondence to Dr Annemieke Houwink at Department of Rehabilitation (898), Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, the Netherlands.
E-mail: a.houwink@reval.umcn.nl

This article is commented on by Bialocerkowski. on pages 12–13 of this issue.

AIM To investigate the validity and reliability of the revised Video-Observation Aarts and Aarts module: Determine Developmental Disregard (VOAA-DDD-R).
METHOD Upper-limb capacity and performance were assessed in children with unilateral spastic cerebral palsy (CP) by measuring overall duration of affected upper-limb use and the frequency of specific behaviours during a task in which bimanual activity was demanded ('stringing beads') and stimulated ('decorating a muffin'). Developmental disregard was defined as the difference in duration of affected upper-limb use between both tasks. Raters were two occupational and one physical therapist who received 3 hours of training. Construct validity was determined by comparing children with CP with typically developing children. Intrarater, interrater, and test–retest reliability were determined using the intraclass correlation coefficient. Standard errors of measurement and smallest detectable differences were also calculated.
RESULTS Twenty-five children with CP (15 females, 10 males; mean age 4y 9mo [SD 1y 7mo], range 2y 9mo–8y; Manual Ability Classification System levels I–III) scored lower on capacity ($p=0.052$) and performance ($p<0.001$), and higher on developmental disregard ($p<0.001$) than 46 age- and sex-matched typically developing children (23 males; mean age 5y 3mo [SD 1y 5mo], range 2y 6mo–8y). The intraclass correlation coefficients (0.79–1.00) indicated good reliability. Absolute agreement was high, standard errors of measurement ranged from 4.5 to 6.8%, and smallest detectable differences ranged from 12.5 to 19.0%.
INTERPRETATION The VOAA-DDD-R can be reliably and validly used by occupational and physical therapists to assess upper-limb capacity, performance, and developmental disregard in children (2y 6mo–8y) with CP.

Children with unilateral spastic cerebral palsy (CP) have motor impairments such as muscle weakness and spasticity on predominantly one side of the body.[1,2] These motor impairments are important causes of activity limitations.[3,4] According to the International Classification of Functioning, Disability and Health, the 'activity' level can be subdivided into 'capacity' (i.e. the execution of an activity in a standardized environment) and 'performance' (i.e. the actual performance of an activity in daily life).[5] Children with CP not only experience limitations in their capacity, but they also tend to underuse their affected upper limb in daily life (i.e. limited performance) given their individual capacity. This lack of spontaneous use of the affected limb in developing children is also referred to as 'developmental disregard'.[6]

To design an individually tailored rehabilitation programme, detailed assessment of upper-limb disability is essential.[7] Therefore, it is important to assess bimanual activities because many children who have developmental disregard prefer to use their less-affected upper limb in unimanual tasks. They will only use their affected limb during bimanual tasks. However, tests of upper-limb use during bimanual activities are scarce,[8,9] and many functional measures focus on unilateral tasks.[10,11] Only the Assisting Hand Assessment[12] consists of semi-structured bimanual tasks to assess the effectiveness of use of the assisting upper limb. Although the Assisting Hand Assessment provides a summed frequency score of the effectiveness of upper-limb use, it does not assess the duration of spontaneous use. Because the overall duration of upper-limb

use takes into account all motor behaviours, including (unsuccessful) attempts to involve the affected arm and hand, it seems to be a more valid indicator of developmental disregard than merely counting the frequency of successful behaviours.

To assess both the overall duration and frequency of affected upper-limb use, the 'Video Observations Aarts and Aarts module: Determine Developmental Disregard' (VOAA-DDD) was developed.[13] It consists of two standardized tasks, 'stringing beads' and 'decorating a muffin', to assess upper-limb use. The beads task was designed to demand the use of both hands to accomplish the task, whereas the muffin task was designed merely to stimulate bimanual activity (the task is most efficiently performed with both hands). By using structured video observations and a custom-designed software program,[14] the tasks can be scored offline for the occurrence of specific motor behaviours (i.e. frequency) and the total duration of affected upper-limb use. When used by trained occupational and physical therapists, the VOAA-DDD was shown to be reliable and valid in children between 2 years 6 months and 8 years of age with unilateral spastic CP.[13] However, the scoring system of the VOAA-DDD was very elaborate and the numbers of subtasks and repetitions were not consistent in the two tasks.

Recently, the VOAA-DDD was revised (VOAA-DDD-R) to improve feasibility and interpretation. First, the distinction between the beads task (demanding bimanual hand use) and the muffin task (stimulating bimanual hand use) was made more pronounced. Second, the beads and muffin task now have the same number of subtasks, which is also the same for all ages. Third, the motor behaviours that need to be scored were reduced from 10 to the three most important behaviours (i.e. grasp, hold, release). These behaviours were shown to be essential to performing each subtask and did not differ in frequency between the dominant and non-dominant hand in typically developing children.[15] Finally, only three scores are used to reflect different aspects of upper-limb use: a capacity score (i.e. the frequency during the beads task), a performance score (i.e. the frequency during the muffin task), and a duration score (i.e. the difference in the duration of upper-limb use between the beads and the muffin task).[15] The last score was used as an operationalization of developmental disregard. These revisions required a new investigation of the psychometric properties of the VOAA-DDD-R. The goal of the present study was to investigate the construct validity and the intrarater, interrater, and test–retest reliability of the VOAA-DDD-R in children with unilateral spastic CP.

METHOD
Participants
Twenty-five children with CP were recruited from two rehabilitation centres in the Netherlands (Sint Maartenskliniek, Nijmegen, and Rijndam Rehabilitation Center, Rotterdam). This sample size was based on the results of our previous study. Inclusion criteria were (1) CP with unilateral spastic movement impairment, (2) age between 2 years 6 months and 8 years, and (3) Manual Ability Classification System[16] levels I, II, or III. Children were excluded when they could not understand or execute simple tasks because of intellectual

disability (i.e. developmental age below 2y). In addition, we recruited 46 age- and sex-matched typically developing children from two regular primary schools in Elst and Almere, and one pre-school playgroup in Nijmegen. Hand dominance of the comparison participants was determined based on parental information and on the hand the children used when they were asked to write their name or draw a picture. Legal caregivers provided written informed consent for all participants. All procedures in this study were approved by the regional medical ethics committee.

Raters
Two occupational therapists and one physical therapist experienced in the treatment of children with movement disorders performed the offline scoring of the videos, for which they received training for 3 hours.

Tasks
Both the beads and the muffin task consist of four subtasks. In the beads task (Fig. 1a), the child was asked to string beads (flat discs) on a shoelace as if to feed a caterpillar. First, the child was asked to open a closed can and to grasp a disc from the can, to place the disc on the table, and to put the lid back on the can. Second, the child had to pick up an egg timer, to turn it so that the timer went off (as if to wake the caterpillar), and to place the timer back on the table. Then the child had to open a drawer that was being held back by elastic bands, and take out the shoelace. Third, the child had to pick up the disc and to string it on the shoelace. Fourth, the child had to open the drawer, put back the shoelace, and pick up the egg timer to reset it. In the muffin task (Fig. 1b), the child was asked to decorate a muffin with sweets. First, the child was asked to open a can, grasp a sweet from the can, place the sweet on a plate, and put the lid back on the can. Second, the child had to open a play oven and take out a muffin that was placed in a sieve with a handle. Third, the child had to grasp a stick that was placed upside-down in an open can and make a hole in the muffin using the stick. The child was then asked to take the sweet and to put it in the hole in the muffin. Fourth, the child had to place the sieve holding the muffin back in the oven and close the door. All subtasks were repeated four times.

The beads and the muffin task both lasted 2 to maximally 7 minutes. Participants were seated in a chair with their back supported, their forearms and hands laying on the table, and their feet placed on the floor or a footplate. In the case of a child with CP, the test instructor was one of three occupational therapists experienced in paediatric rehabilitation. A typically developing child was instructed by one of two occupational therapy students. All test instructors received a test manual and training to administer the tasks in a standardized manner. The test instructor was seated opposite the child and
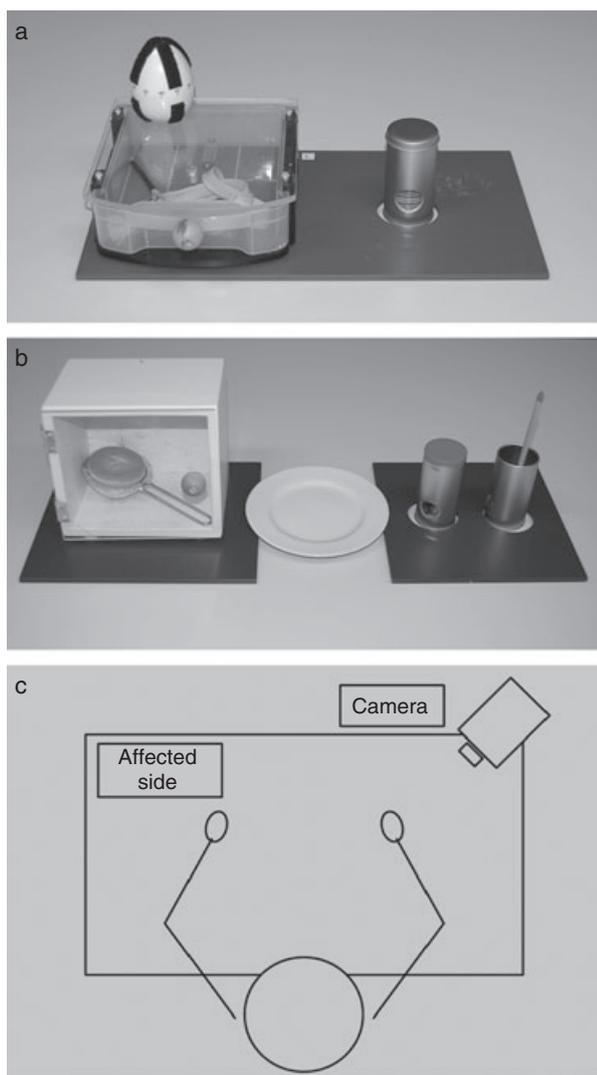
Figure 1: The (a) beads task and (b) muffin task. The materials are positioned for a child with a left-sided paresis, as observed by the child. For a child with right-sided paresis, the setting of the materials is mirrored. The camera was positioned contralateral to the child's affected side (non-dominant side for the typically developing children) at a height of 2 metres focused into the palm of the affected hand (c).

provided the instructions for each subtask, without indicating which hand had to be used. Before the child started, the test instructor demonstrated the tasks and checked whether the child had understood the instructions. The video camera was placed contralateral to the child's affected side (non-dominant side for the typically developing children) focused into the palm of the affected hand (Fig. 1c).

Scoring system

The video recordings were scored offline for the occurrence of grasping, holding, and releasing (i.e. motor behaviours) as well as the overall duration of use of the affected upper limb (non-dominant side in typically developing children). The average duration for scoring all the measurements of one

child was 30 minutes. The frequency of the three behaviours was scored irrespective of the quality (e.g. grasping with the wrist in dorsal flexion or in palmar flexion were both scored). The participant could obtain maximally one point for each of the three behaviours during each subtask, resulting in a maximum frequency score of 48 (three behaviours×four subtasks×four repetitions). Thus, the frequency measure did not take into account whether a behaviour was performed multiple times during one subtask. The observed total frequency was converted into a percentage of the maximum frequency. The frequency score during the 'demanding' beads task was termed the capacity score, whereas the frequency during the 'stimulating' muffin task was termed the performance score. When a child was unable to perform four repetitions of the task, the total frequency score was adjusted accordingly (e.g. when a child could perform only three repetitions, the maximal attainable frequency score was 36). In addition, the overall duration of use of the affected upper limb was scored for both the beads and the muffin task as a percentage of the total duration of each task. All motor behaviours related to the task performance contributed to the duration score, regardless of their success or quality. The difference in the duration of use between the beads task and the muffin task was defined as developmental disregard.

Procedure

The children with CP were assessed twice by two occupational therapists with a time interval of approximately 2 weeks, as recommended by Terwee et al.[17] The first assessments of the children with CP were scored by both raters to determine the interrater reliability. In addition, the same rater scored the first assessments twice with at least 2 weeks in between to determine the intrarater reliability. The scorings of the first and second assessments of the children with CP by the same raters were used to determine the test–retest reliability. The assessment of the typically developing children was scored by one rater (the physical therapist), and was used together with the first assessment of the children with CP to determine the construct validity.

Analysis
Participants
The characteristics of the children with CP were compared with those of the typically developing children for age (two-sided independent t-test) and sex (Mann–Whitney U test).

Validity
Construct validity was determined by comparing the scores of the children with CP with those of the typically developing children, based on the following hypotheses. Compared with the typically developing children, the children with CP were expected to score lower on capacity and performance, and higher on developmental disregard. Between-group differences (children with CP vs typically developing children) were tested with Mann–Whitney U tests and within-group differences (capacity vs performance in children with CP) with a Wilcoxon signed ranks test. Furthermore, the effects of sex

and age on the three scores were examined by testing the differences between males and females using Mann–Whitney U tests and by correlating the three scores with age using Spearman's ρ. The mean score +2 SD of the typically developing children was used as a cut-off criterion to determine developmental disregard in individual children with CP.

Reliability

The intrarater, interrater, and test–retest reliability of the capacity, performance, and developmental disregard scores were quantified with the intraclass correlation coefficient (ICC) and corresponding 95% confidence interval. An ICC less than 0.70 was considered good.[18] A two-way random model for absolute agreement was used to distinguish between random variations and 'real' differences.[18] The standard error of measurement (SEM) was used to assess the absolute agreement between the first and second assessment, according to Bland and Altman.[19] The SEM was calculated using the within-participant SD (SEM=√error variance). To determine the minimal change score in an individual that represented a real difference, the smallest detectable difference was calculated as 1.96×√2×SEM.[19]

RESULTS
Participants

The characteristics of the children with CP and the typically developing children are presented in Table I. The children with CP did not differ significantly from the typically developing children for sex (p=0.423) or age (p=0.136).

Validity

Table II shows the capacity, performance, and developmental disregard scores for the children with CP and for the typically developing children. Seven of the children with CP could only perform two or three repetitions of the subtasks within 7 minutes. Consequently, their maximally attainable score on capacity and/or performance was adjusted to 24 and 36 respectively. The typically developing children scored almost maximally on

Table I: Characteristics of the children with cerebral palsy (CP) and the typically developing children (TDC)

	CP (*n*=25)	TDC (*n*=46)
Age		
Mean (SD)	4y 9mo (1y 7mo)	5y 3mo (1y 5mo)
Range	2y 9mo–8y	2y 6mo–8y
Sex, *n*		
Male	10	23
Female	15	23
Affected side, *n*		
Right	10	–
Left	15	–
Dominant side, *n*		
Right	–	41
Left	–	5
MACS, *n*		
I	5	
II	12	
III	8	

MACS, Manual Ability Classification System.

capacity and performance, whereas the children with CP scored lower on capacity, almost reaching statistical significance (p=0.052), and significantly lower on performance (p<0.001). Their performance scores were lower than their capacity scores (p<0.001). Furthermore, children with CP scored three times higher on developmental disregard (p<0.001). There were no effects of sex on the three scores for the children with CP (p>0.428) and the typically developing children (p>0.095), nor any effects of age on the performance and developmental disregard scores (p>0.248). There was a small effect of age on the capacity score in the children with CP (ρ=0.436; p<0.05) and the typically developing children (ρ=0.758; p<0.001), indicating that older children performed better than younger children. The cut-off score for developmental disregard based on the mean scores 2SD of the typically developing children was 17.2%. Based on this value, 64% of the children with CP could be identified as having developmental disregard. All individual scores are presented in Figure 2.

Reliability

The intra- and interrater reliability of the capacity, performance, and developmental disregard scores were excellent, with ICCs ranging from 0.95 to 1.00 (Table II). The test–retest reliability was excellent for the capacity and performance scores, whereas it was good for the developmental disregard score (ICCs ranged from 0.79 to 0.99). The mean differences between the first and second assessments in children with CP were −1.2% (SD 7.1) for capacity, −1.8% (SD 6.4) for performance, and −0.3% (SD 9.7) for developmental disregard. The absolute agreement between the two assessments is presented in Figure 3. The SEMs ranged from 4.5 to 6.8%, which resulted in smallest detectable differences of between 12.5% and 19.0%.

DISCUSSION

The results of this study indicate that the three scores of the VOAA-DDD-R (i.e. capacity, performance, and developmental disregard) are both valid and reliable. The construct validity was determined by comparing the scores of children with CP with those of typically developing children, because there is no criterion standard available in the literature for the frequency and duration of use of the affected upper limb during bimanual activities. Children with CP had lower scores than typically developing children for capacity (77% vs 98%) and significantly lower scores for performance (55% vs 100%), yielding much higher scores for developmental disregard (23% vs 7%). In addition, the variability in the CP group was much higher compared with the typically developing children. Furthermore, the older children performed better than the younger ones on the capacity score, which may have been related to improvements in bimanual performance that are related to development. This finding needs to be taken into account by therapists when assessing younger children with the VOAA-DDD-R.

Based on the cut-off score for developmental disregard of typically developing children (i.e. 17%), 64% of the children

Table II: Mean scores of typically developing children (TDC; n=46) and children with cerebral palsy (CP, n=25), and reliability outcomes in children with CP

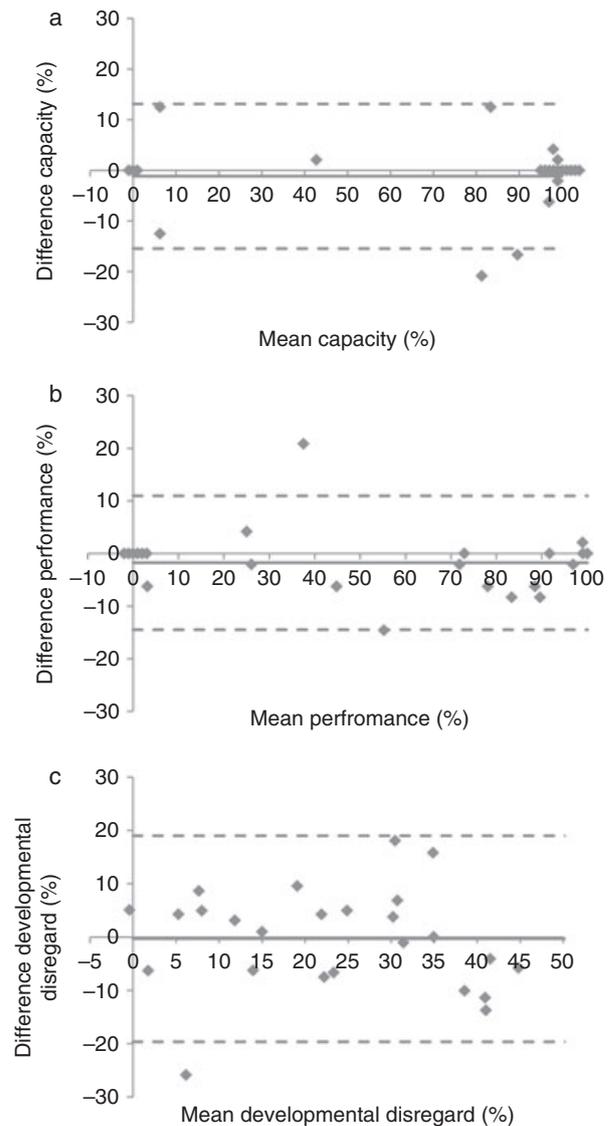| Score | Mean scores (SD) | | Mann–Whitney U test | Reliability (CP) | | | | |
	TDC (n=46)	CP (n=25)	p	Intrarater ICC (95% CI)	Interrater ICC (95% CI)	Test–retest ICC (95% CI)	SEM (%)	SDD (%)
Capacity (range 0–100), %	98.1 (3.7)	76.6 (39.8)	0.052	1.00 (1.00–1.00)	0.98 (0.95–0.99)	0.98 (0.96–0.99)	5.1	14.0
Performance (range 0–100), %	100 (0.0)	54.8 (41.1)	<0.001	1.00 (1.00–1.00)	0.99 (0.98–1.00)	0.99 (0.97–0.99)	4.5	12.5
Developmental disregard (range 0–100), %	6.6 (5.3)	23.3 (15.0)	<0.001	0.98 (0.96–0.99)	0.95 (0.90–0.98)	0.79 (0.57–0.90)	6.8	19.0

ICC, intraclass correlation coefficient; CI, confidence interval; SEM, standard error of measurement; SDD, smallest detectable difference.
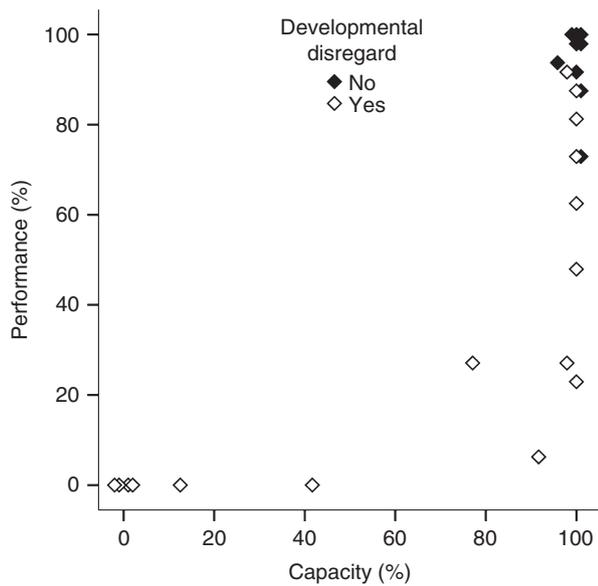
Figure 2: Individual scores of children with cerebral palsy (n=25) on capacity (x-axis) and performance (y-axis). Individuals who were identified as having developmental disregard (i.e. a developmental disregard score >17.2%; n=16) are depicted by white diamonds, whereas children without developmental disregard are depicted by black diamonds.

Figure 3: The absolute agreement between two repeated assessments of the same rater in children with cerebral palsy (n=25) according to Bland and Altman.[19] The difference score between the two assessments is plotted against the mean score for (a), capacity (b), performance and (c) developmental disregard The solid line represents the mean difference, the dotted lines represent the limits of agreement.

with CP could be identified as having developmental disregard (Fig. 2). This cut-off value is close to the cut-off value reported in our previous study on the VOAA-DDD (14%).[13] These results confirm our hypothesis that many children with CP show a discrepancy between what they can do with their affected upper limb when bimanual activity is demanded (i.e. capacity) and what they actually do when bimanual activity is merely stimulated (i.e. performance). These test scores can be used as a basis for designing an individually tailored rehabilitation intervention.[15] For instance, Figure 2 shows that the six children with a low capacity score (0–40%) scored 0% on performance. Based on these scores it is advisable that these children are primarily trained to improve their upper-limb capacity. Remarkably, even eight children with a (near) maximum capacity score of 100% showed some degree of developmental disregard, whereas nine others with a maximum capacity did not. This pattern of results suggests that a (nearly)

optimal capacity is needed to prevent the occurrence of developmental disregard, but that such a score certainly provides no guarantee for the absence of developmental disregard. Thus, these children should all be carefully monitored for signs of developmental disregard and offered appropriate training (e.g. constraint-induced movement therapy). On the other hand, one or two children with a somewhat lower performance than their optimal capacity scores did not seem to have developmental disregard based on the duration of use of their affected upper limb.

The VOAA-DDD-R showed excellent intra- and interrater reliability, as indicated by high ICCs for capacity, performance, and developmental disregard. Reliability in this context means that repeated measurements result in similar outcomes,[17,18] which are not influenced by characteristics of the instrument, differences in performance by the same rater, differences between multiple raters, or by the natural variability within an individual. The ICC values in the present study indicated that the repeated scoring of the assessments was very stable both within (intrarater reliability) and between raters (interrater reliability). This suggests that when a child is assessed twice by the same rater or by two different raters, the results are generally the same and not affected by the measurement instrument. The test–retest reliability was excellent for the capacity and performance scores and good for developmental disregard. Thus, the variability between two assessments caused by variation of the child's behaviour was larger than the variation caused by the raters. In addition, the results indicate that with repeated testing the frequency scores were more stable than the duration scores. This can be explained by the fact that for the frequency scores a child could obtain maximally one point for each behaviour per subtask, which renders the frequency scores more stable but also less sensitive to repeated behaviours within a subtask. Nevertheless, the absolute agreement between the repeated assessments was good, as indicated by SEMs between 4.5% and 6.8%. These results imply that, when two groups of children with CP are compared, a group difference of 5.1% on capacity, 4.5% on performance, and 6.8% on developmental disregard can be regarded as a real difference and not due to natural variation. For individual children, a change in the VOAA-DDD-R scores needs to be larger to be significantly different, because the smallest detectable differences ranged from 12.5 to 19.0%. These results indicate that although the VOAA-DDD-R is suitable to detect differences between groups, it needs to be further refined to be able to detect smaller changes in individual children.[19]

Until now, no reliable and valid measure of developmental disregard has been available in the literature. In this perspective, the VOAA-DDD-R is a valuable addition to the existing measures of affected upper-limb use in children with CP. Because the VOAA-DDD-R consists of common daily-life tasks that are attractive and meaningful for all children, it may also have merits for other groups of children with unilateral upper-limb disability, for instance children with peripheral nerve damage, traumatic brain injury, or stroke. A limitation of the present study is that the responsiveness (i.e. sensitivity to change) was not investigated. Thus, future studies need to examine the responsiveness of the VOAA-DDD-R to determine its usefulness and sensitivity in intervention studies. Another limitation is that one could argue that the VOAA-DDD-R is not truly a test of upper-limb performance in daily life, because it requires a standardized test situation. Yet, a drawback of real-life assessments is that they may be too subjective. For instance, self-report questionnaires[20,21] are usually completed by the child's parents or caregivers with a great influence of personal perspectives and proneness to inconsistencies. Recent developments in the use of wearable wrist activity monitors to assess actual daily-life use of the affected upper limb are promising,[22] but such monitors have only been tested during standardized activities as well. Finally, the construct validity was determined based on the assessment of typically developing children, who are expected to have no limitations in capacity and performance and show no developmental disregard. To confirm that the cut-off value for developmental disregard used in this study was indeed valid, we need to investigate other groups of children with CP with and without developmental disregard as determined, for example, by experts.

In conclusion, this study showed that the VOAA-DDD-R, using a simplified scoring system, is equally reliable, when performed 2 weeks apart, and as valid as the original VOAA-DDD when applied by trained occupational and physical therapists to children with unilateral spastic CP (2y 6mo–8y). By comparing the use of the affected upper limb during a task demanding the use of both hands compared with a task merely stimulating bimanual activity, upper-limb capacity, performance, and developmental disregard can be reliably and validly assessed offline with a computer-supported video scoring system.

ACKNOWLEDGEMENTS

REFERENCES

1. Cans C. Surveillance of cerebral palsy in Europe: a collaboration of cerebral palsy surveys and registers. *Dev Med Child Neurol* 2000; **42:** 816–24.

2. Rosenbaum P. Cerebral palsy: what parents and doctors want to know. *BMJ* 2003; **326:** 970–4.

3. Kim WH, Park EY. Causal relation between spasticity, strength, gross motor function, and functional outcome in children with cerebral palsy: a path analysis. *Dev Med Child Neurol* 2011; **53:** 68–73.

4. Beckung E, Hagberg G. Neuroimpairments, activity limitations, and participation restrictions in children with cerebral palsy. *Dev Med Child Neurol* 2002; **44:** 309–16.

5. World Health Organization. ICF: International Classification of Functioning, Disability and Health. Geneva: World Health Organization, 2001.

6. Hoare BJ, Wasiak J, Imms C, Carey L. Constraint-induced movement therapy in the treatment of the upper limb in children with hemiplegic cerebral palsy. *Cochrane Database Syst Rev* 2007; CD004149.

7. Mayston M. From 'one size fits all' to tailor-made physical intervention for cerebral palsy. *Dev Med Child Neurol* 2011; **53:** 969–70.

8. Klingels K, Jaspers E, van de Winckel A, de Cock P, Molenaers G, Feys H. A systematic review of arm activity measures for children with hemiplegic cerebral palsy. *Clin Rehabil* 2010; **24:** 887–900.

9. Gilmore R, Sakzewski L, Boyd R. Upper limb activity measures for 5- to 16-year-old children with congenital hemiplegia: a systematic review. *Dev Med Child Neurol* 2010; **52:** 14–21.

10. DeMatteo C, Law M, Russell D, Pollock N, Rosenbaum P, Walter S. QUEST: Quality of Upper Extremity Skills Test. Hamilton, ON: McMaster University, Neurodevelopmental Clinical Research Unit, 1992.

11. Johnson LM, Randall MJ, Reddihough DS, Oke LE, Byrt TA, Bach TM. Development of a clinical assessment of quality of movement for unilateral upper-limb function. *Dev Med Child Neurol* 1994; **36:** 965–73.

12. Krumlinde-Sundholm L, Holmefur M, Kottorp A, Eliasson AC. The Assisting Hand Assessment: current evidence of validity, reliability, and responsiveness to change. *Dev Med Child Neurol* 2007; **49:** 259–64.

13. Aarts PB, Jongerius PH, Geerdink YA, Geurts ACH. Validity and reliability of the VOAA-DDD to assess spontaneous hand use with a video observation tool in children with spastic unilateral cerebral palsy. *BMC Musculoskelet Disord* 2009; **10:** 145.

14. Aarts PB, Jongerius PH, Aarts MA, Van Hartingsveldt MJ, Anderson PG, Beumer A. A pilot study of the Video Observations Aarts and Aarts (VOAA): a new software program to measure motor behaviour in children with cerebral palsy. *Occup Ther Int* 2007; **14:** 113–22.

15. Aarts PB, Jongerius PH, Geerdink YA, van Limbeek J, Geurts AC. Modified constraint-induced movement therapy combined with bimanual training (mCIMT-BiT) in children with unilateral spastic cerebral palsy: how are improvements in arm-hand use established? *Res Dev Disabil* 2011; **32:** 271–9.

16. Eliasson AC, Krumlinde-Sundholm L, Rosblad B, et al. The Manual Ability Classification System (MACS) for children with cerebral palsy: scale development and evidence of validity and reliability. *Dev Med Child Neurol* 2006; **48:** 549–54.

17. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; **60:** 34–42.

18. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006; **59:** 1033–9.

19. Bland JM, Altman DG. Measurement error. *BMJ* 1996; **313:** 744.

20. Wallen M, Bundy A, Pont K, Ziviani J. Psychometric properties of the Pediatric Motor Activity Log used for children with cerebral palsy. *Dev Med Child Neurol* 2009; **51:** 200–8.

21. Arnould C, Penta M, Renders A, Thonnard JL. ABIL-HAND-Kids: a measure of manual ability in children with cerebral palsy. *Neurology* 2004; **63:** 1045–52.

22. Howcroft J, Fehlings D, Zabjek K, Fay L, Liang J, Biddiss E. Wearable wrist activity monitor as an indicator of functional hand use in children with cerebral palsy. *Dev Med Child Neurol* 2011; **53:** 1024–9.