

# Psychometric Evaluation of 2 New Upper Extremity Functional Strength Tests in Children With Cerebral Palsy

Koen J.F.M. Dekkers, Rob J.E.M. Smeets, Yvonne J.M. Janssen-Potten, Andrew M. Gordon, Lucianne A.W.M. Speth, Eugene A.A. Rameckers

**Background.** For children with unilateral spastic cerebral palsy (USCP), reduced muscle strength can lead to activity limitations. However, none of the existing measures of upper extremity strength measure strength in the context of functional activities in which strength must be maintained for several seconds.

**Objective.** The objective of this study was to evaluate the psychometric properties of 2 newly developed functional hand and upper extremity muscle-strength tests (Cup-Task and Box-Task) in children aged 7 to 12 years with USCP.

**Design.** A longitudinal study design was used.

**Methods.** A standardized protocol with detailed descriptions of all procedures and measurements was used to determine test-retest reliability, interrater reliability, and criterion validity.

**Results.** A total of 86 children (53 males, 33 females, mean age = 9.3 years) with USCP participated in this study, with a subset performing each measurement. Only the results of children who were able to perform the measurement were included for analysis. Excellent test-retest reliability (intraclass correlation coefficients = 0.887–0.944; 95% confidence intervals = 0.713–0.969) and interrater reliability (intraclass correlation coefficients = 0.896–0.960; 95% confidence intervals = 0.813–0.980) were observed. The Cup-Task Affected-Hand and Box-Task were moderately correlated with maximum isometric grip strength. The Cup-Task Nonaffected-Hand had a low correlation with maximum isometric grip strength.

**Limitations.** Age, sex, and manual ability were not normally distributed, which could have influenced the results.

**Conclusions.** For children with USCP who can perform the tasks, the Cup-Task and Box-Task are reliable and valid instruments for measuring functional upper extremity muscle strength.

K.J.F.M. Dekkers, PT, Rehabilitation Centres, Revant, Brabantlaan 1, 4817 JW Breda, the Netherlands; Department of Rehabilitation Medicine, Research School CAPHRI, Maastricht University, Maastricht, the Netherlands; and University for Professionals for Pediatric Physical Therapy, AVANSplus, Breda, the Netherlands. Address all correspondence to Mr Dekkers at: k.dekkers@revant.nl.

R.J.E.M. Smeets, MD, PhD, Department of Rehabilitation Medicine, Research School CAPHRI, Maastricht University; and CIR Revalidatie, Eindhoven, the Netherlands.

Y.J.M. Janssen-Potten, PhD, Department of Rehabilitation Medicine, Research School CAPHRI, Maastricht University; and Adelante Centre of Expertise in Rehabilitation and Audiology, Hoensbroek, the Netherlands.

A.M. Gordon, PhD, Department of Biobehavioral Sciences, Teachers College, Columbia University, New York, New York.

L.A.W.M. Speth, MD, PhD, Department of Rehabilitation Medicine, Research School CAPHRI, Maastricht University; and Adelante Centre of Expertise in Rehabilitation and Audiology.

E.A.A. Rameckers, PT, PhD, Department of Rehabilitation Medicine, Research School CAPHRI, Maastricht University; University for Professionals for Pediatric Physical Therapy, AVANSplus; Adelante Centre of Expertise in Rehabilitation and Audiology; and Pediatric Rehabilitation, Biomed, Faculty of Medicine and Health Science, Hasselt University, Diepenbeek, Belgium.

[Dekkers KJFM, Smeets RJEM, Janssen-Potten YJM, Gordon AM, Speth LAWM, Rameckers EAA. Psychometric evaluation of 2 new upper extremity functional strength tests in children with cerebral palsy. *Phys Ther*. 2019;99:1107–1115.]

© 2019 American Physical Therapy Association

Published Ahead of Print:

February 04, 2019

Accepted: January 9, 2019

Submitted: March 8, 2018



Post a comment for this article at:

<https://academic.oup.com/ptj>

Impaired performance of upper extremity activities is reported in ~ 50% of children with unilateral spastic cerebral palsy (USCP), and these impairments affect the children's independence and quality of life.<sup>1,2</sup> Brændvik et al concluded that muscle strength strongly correlates with activity and that reduced strength can result in activity limitations.<sup>3</sup> Upper extremity activities are assessed using a number of clinical tests and questionnaires. The National Institutes of Health Common Data Elements recommends a variety of supplemental tests for assessing upper extremity motor function for children with cerebral palsy.<sup>4</sup> These tests include the ABILHAND-kids,<sup>5</sup> the Assisting Hand Assessment,<sup>6</sup> the Melbourne Assessment-v2,<sup>7</sup> the Quality of Upper Extremity Skills Test,<sup>8</sup> and the Shriners Hospital Upper Extremity Evaluation.<sup>9</sup> Recommended exploratory measures of dexterity include the Box and Blocks Test of Manual Dexterity,<sup>10</sup> the Jebsen-Taylor Test of Hand Function,<sup>11</sup> the Nine-Hole Peg Test,<sup>12</sup> and the Purdue Pegboard.<sup>13</sup> All of these instruments measure function (active and passive range of motion, tone, segmental alignment of the extremity) or quantify the capacity (upper extremity use in a standardized, controlled environment), capability (upper extremity function in the daily environment), or performance (actual use of the upper extremity in the daily environment) of the upper extremity. However, none of these measures directly consider the role of strength in upper extremity activities.

For upper extremity strength measurements, the National Institutes of Health Common Data Elements recommends maximum voluntary isometric contraction testing using hand-held dynamometers and a grip/pinch dynamometer, and manual muscle testing using the Medical Research Council Muscle Grading Scale<sup>14</sup> to determine different grades of muscle strength. Within manual muscle testing, grades 4 and 5 seem insufficiently sensitive to assess muscle strength or to detect small to moderate increases of strength.<sup>15</sup>

In most manual activities, not only is a certain amount of muscle strength required, but also the ability to maintain that strength for a short time (2–5 seconds), for example, during carrying/moving a heavy box. Therefore, a certain level of coordination is also necessary to perform the tests adequately. Measuring muscle strength during a functional task in which it is expected that muscle strength plays a major role (ie, “functional strength” measurements) enables measurement not only of the strength of different simultaneously working muscles but also the task-specific generation of the strength.

For children with USCP, various *functional* muscle-strength tests for the lower extremity are available, including the “Lateral Step-up,” “Sit-to-Stand,” and the “Attain stand through half kneel, without using arms” tests.<sup>16</sup> For the upper extremities, the “functional strength measurement” test is feasible in children with CP.<sup>17</sup> With this test, maximum explosive muscle strength and

30-second repetitive measurements are recorded. Unfortunately, the ability to maintain the functional strength in a sustained contraction is not measured with the currently available muscle-strength tests. Thus, there are no functional upper extremity strength measures that quantify strength when sustained contractions are required. Therefore, we have developed 2 specific functional muscle-strength tests in the context of unimanual and bimanual activity: the “Cup-Task” for determining maximal functional unimanual upper extremity strength, and the “Box-Task” for determining maximal functional bimanual upper extremity strength. Both tests measure a combination of functional grip and arm strength which must be sustained for 5 seconds. In a pilot study, both tests were found to be feasible in children with USCP.<sup>18</sup>

We used the COSMIN (Consensus-based Standards for the Selection of health Measurement INstruments) checklist ([www.cosmin.nl](http://www.cosmin.nl)) as guidance for designing and reporting our study on the clinimetric properties of this new instrument. The COSMIN checklist is a consensus-based checklist that can be used for selecting a measurement instrument, peer reviewing a manuscript, and designing or reporting a study on measurement properties.<sup>19</sup> With the checklist, the methodological quality of a study can be classified as poor, moderate, good, or excellent.<sup>20</sup>

The objective of our study was to investigate test-retest and interrater reliability of the Cup-Task and Box-Task for children aged 7 to 12 years with USCP. A secondary objective was to assess the criterion validity of the Cup-Task and Box-Task.

## Methods

### Study Design

A longitudinal study design was used. Ethics approval was granted by the Medical Ethical Board of the Maastricht University Medical Center and Maastricht University (METC azM/UM, trial number NL45430.068.1) in the Netherlands and at Teachers College, Columbia University, New York City, New York (USA). Data were collected in the Netherlands and United States from 2009 to 2016.

### Participants

In the Netherlands, children were recruited from 4 rehabilitation centers and related special education schools, namely the Adelante Rehabilitation Centre in Valkenburg, Libra Rehabilitation and Audiology in Tilburg, Revant Rehabilitation Centers in Breda and Goes, and Tolbrug Rehabilitation Centre in Den Bosch. In the United States, children were recruited while they were participating in ongoing intensive upper extremity studies at Teachers College, Columbia University.

To be included in this study, children were required to be diagnosed with USCP and aged between 7 and 12 years.

Furthermore, children had to be classified as GMFCS level I to II (Gross Motor Function Classification System<sup>21</sup>), MACS-level I, II, or III (Manual Ability Classification System<sup>2</sup>), and be able to follow simple instructions. Children were excluded if within the past 6 months they had undergone surgery or received botulinum toxin-A treatment of the upper extremity. Children were also excluded for the test-retest reliability measurements if they were participating in an intensive upper extremity training program between the 2 measurements. The aim was to include more than 50 children per measurement.<sup>20</sup>

## Procedure

Measurements took place at the rehabilitation center where the children were being treated. A standardized protocol with detailed descriptions of all procedures and measurements was used. Prior to testing, the participants' body weight and MACS and GMFCS levels were determined.

One set of strength measurements consisted of the Cup-Task, the Box-Task, and the E-LINK system. The sequence of the tasks was randomized. For the Cup-Task and E-LINK system, both the nonaffected hand (NAH) and the affected hand (AH) were measured.

To evaluate test-retest reliability, the set of strength measurements was conducted twice by the same assessor with a 2-week interval. This time interval was chosen because during normal development (without intensive upper extremity training), no loss or gain in muscle strength was expected and memory of the first results would be limited. For the second session, test conditions were kept identical. For some children, the results of the second set of strength measurements were used for the premeasurements of intensive upper extremity studies, and the time span was sometimes extended to 4 weeks for these children due to scheduling difficulties.

To evaluate interrater reliability, 2 sets of strength measurements were conducted by 2 different assessors on the same day. There was at least 30 minutes of rest between each assessment. This amount of rest time was judged sufficient for the child to recover and limited the possibility for personal and environmental factors to change. If a child was measured for test-retest reliability and interrater reliability, the results of the first interrater reliability measurements were used for the test-retest reliability.

To evaluate criterion validity, a maximum isometric grip strength (IGS) measurement was performed with the Biometrics E-LINK evaluation and exercise hand kit (Biometrics Ltd, Newport, Gwent, UK). The IGS measurement was performed 3 consecutive times, and the mean of the 3 measurements was calculated. Thus, variation in muscle strength, due to variations in grasping/handling of the measurement instrument, was

minimized. Between each measurement, the child had at least 30 seconds of rest to allow recovery.

Test scores were determined by the assessor and registered by the same assessor on a test form. In case of IGS measurements, test scores were also stored on the E-LINK evaluation system computer.

All measurements were performed by 8 different assessors, all (pediatric) physical therapists, who had no direct professional connection with the participants. Two of the assessors (K.D. and E.R.) are experienced pediatric research physical therapists and were involved during the entire project. The other 6 assessors were physical therapists who are studying for their advanced degree in pediatric physical therapy. They were involved in the study for 6 consecutive months. Each assessor received 4 hours of training from K.D. or E.R. regarding the use of the standardized protocol in children with USCP. Before each measurement, each assessor had to read/practice the measurements protocol.

## Measures

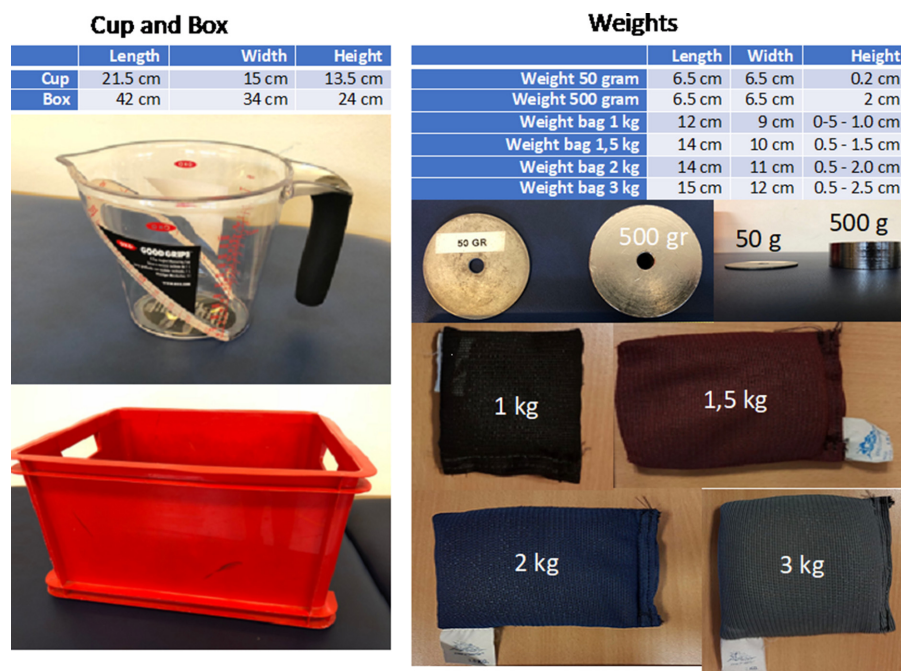
Both tasks were developed based on an expert's opinion and the identification of the most frequently reported needs of the children with USCP in our Bimanual Intensive Movement Treatment. Goals involving lifting a box, tablet, or cup were most frequently reported by the children.

### Cup-Task: unimanual functional muscle-strength test.

The goal of the Cup-Task was to test maximum unilateral functional muscle strength by determining the weight (in grams) the child could lift and hold for 5 seconds with 1 hand, using a measuring cup filled with adjustable weight. The equipment included an adjustable table, small weights, a water jug filled with 1000 cc water, and a measuring cup (322 g) with a maximum content of 1000 cc and a handle that could be held with the cylinder grip (see Fig. 1). The table was set at the height of the iliac crest of the child. The measuring cup was placed on the table. The NAH was tested prior to the AH. Between each attempt, 90 seconds of rest was provided. The number of attempts needed to determine the maximal weight could vary.

The child was instructed to lift the cup horizontally by flexing the elbow and fixating the wrist with the cup stable in the horizontal plane for 5 seconds without lowering it. The wrist was in the neutral position or in slight radial deviation. Ulnar deviation was only allowed when there was no other possible way to lift the cup and to keep it horizontal. After 5 seconds, the cup was replaced on the table.

*Nonaffected hand.* Each participant started the Cup-Task NAH with a starting weight of 500 g. After each successful attempt, a weight of 100 to 500 g was added until an



**Figure 1.**

Cup and box used for the Cup-Task and Box-Task. cm = centimeter; g = gram; kg = kilograms.

attempt was unsuccessful. The assessor was instructed to gradually build up to the maximum weight. If an attempt was unsuccessful, the weight was reduced in increments of 100 g until the child could perform the task as described.

**Affected hand.** Because the abilities of the AH differ considerably between children with USCP, it is very likely that functional muscle strength of the AH also differs between children. Therefore, before the Cup-Task AH started, the starting weight was determined using water. First, the child was instructed to lift and hold the empty cup with a flexed elbow and fixated wrist and to keep it steady in the horizontal plane. Next, the assessor filled the cup with water, using a fluent movement, until the child could no longer hold the measuring cup horizontally. The amount of water was measured, and this amount was used as the starting weight for the first attempt. When, during the first attempt, the task was not performed as described, the weight was reduced by 100 cc of water for the second attempt. When the task was performed as described, in each attempt the weight was increased by increments of 100 cc until an attempt was no longer successful. Above 1000 cc the water was replaced by weights.

#### **Box-Task: bimanual functional muscle-strength test.**

The goal of the Box-Task was to test maximum bilateral functional muscle strength by determining the weight (in kilograms) a child can lift and hold for 5 seconds with

both hands using a box filled with weight bags. The equipment included an adjustable table and a plastic box (0.8 kg) with handles (Fig. 1).

The box was situated on a table with the height adjusted so that the top of the box was at the height of the child's iliac crest. The child was instructed to lift and hold the box horizontally for 5 seconds by flexing the elbows and fixating the wrists. After 5 seconds, the box was replaced on the table. The first attempt involved lifting the box without weight. If the weight of the empty box could be held according to the criteria, 0.5 to 2 kg (depending on how easy it was for the child to lift the box) of weight was supplemented. The assessor was instructed to gradually build up to the maximum weight. Between each attempt, 90 seconds of rest was provided. The number of attempts needed to determine the maximal weight could vary.

**Isometric grip strength.** Maximum IGS strength was measured with the Biometrics E-LINK evaluation and exercise hand kit, both in the AH and NAH. The E-LINK evaluation system is a calibrated, computerized system incorporating a modified (digitalized) grip dynamometer. The applied force was measured in 100-g increments. Children were seated in an upright position in a chair with back support and armrests. The initial posture was a neutral position (0°) of the wrist joint and 90° flexion of the elbow joint. The handle position of the E-LINK handgrip was adapted to the child's hand size, according

to E-LINK guidelines. The child was also asked whether the position of the handle felt comfortable. When there was doubt, other handle positions were attempted.

In children with USCP, a previous study showed excellent test-retest reliability (intraclass correlation coefficient [ICC] values of 0.948 (95% confidence interval [CI] = 0.914–0.968) for the AH, and 0.942 (95% CI = 0.904–0.964) for the NAH. Interrater reliability was excellent for the AH (ICC = 0.976; 95% CI = 0.959–0.986) and for the NAH (ICC = 0.960; 95% CI = 0.932–0.977).<sup>22</sup>

### Statistical Analysis

Before determining reliability and validity of the Cup-Task and Box-Task, the percentage of participants who could not perform each test was determined. The measurements of these participants were excluded for the reliability and validity analyses.

ICC, model 2-way random, type absolute agreement, with 95% CI were used to assess test-retest reliability and interrater reliability. An ICC greater than 0.80 reflects excellent reliability, whereas ICCs from 0.70 to 0.79 reflect good reliability.<sup>23</sup>

The standard error of measurement (SEM) agreement was calculated as the square root of the error variance (including the systematic error).<sup>24</sup> The smallest detectable change (SDC) was computed as 1.96 multiplied by the square root of 2, multiplied by the SEM ( $SDC = 1.96 \times \sqrt{2} \times SEM$ ).<sup>24</sup>

A simple and widely used method to interpret the SDC values is the Bland-Altman limits of agreement.<sup>25</sup> An assumption of the limits of agreement is that the differences between 2 measurements are normally distributed. When differences are not normally distributed, log transformation can be attempted.<sup>26</sup> However, in log-transformed data, the antilog of the difference between 2 values on a log scale is a dimensionless ratio.<sup>25</sup> For that reason, we only calculated limits of agreement when all differences were normally distributed (determined by the Shapiro-Wilk test).

To assess criterion validity, a Pearson correlation coefficient between the functional upper extremity strength measurements and the maximum IGS was calculated. For the Cup-Task AH and the Box-Task, a comparison with the IGS AH was made. For the Cup-Task NAH, a comparison with the IGS NAH was made. In all measurements, a significant moderate Pearson correlation coefficient<sup>27</sup> between 0.50 and 0.70 was hypothesized, because muscle strength seems to be the most important component of the functional strength measurement, but coordination and some endurance are also important components. The hypothesized values are also in line with previously reported validity values for functional strength measurement in children with CP.<sup>17</sup>

Values were considered statistically significant at  $P < .05$ .

### Role of the Funding Source

This study was funded by the Forward for Children With Disabilities Foundation, Valkenburg, the Netherlands; the Revant Innovation Foundation, Breda, the Netherlands; and the Johanna Children's Foundation, Arnhem, the Netherlands. The funders had no role in the design, data collection, analysis, interpretation, or reporting of this work, or the decision to submit the work for publication.

### Results

A total of 86 children with USCP participated in this study. Because not every child performed every measurement (due to unavailability and/or measurement instrument), a different number of children was available for each test. For the statistical analyses of the reliability and validity values, only the children who could perform the specific measurement were included. See Figure 2 for a detailed description of the participant characteristics.

The results of the reliability studies for all tasks at different time points are shown in Table 1 (outcomes of the measurements) and Table 2 (reliability and validity).

#### Cup-Task AH

**Test-retest reliability.** Of the 54 children tested, 9 children (16.9%) could not (adequately) perform 1 or more measurements with the empty cup. See Figure 2 for details.

Test-retest reliability was excellent ( $N = 45$ ; ICC = 0.887; 95% CI = 0.713–0.948), with an SEM value of 284 g and an SDC value of 787 g.

**Interrater reliability.** Of the 54 children, 11 (20.3%) could not (adequately) perform the measurement with the empty cup in one of the attempts. See Figure 2 for details.

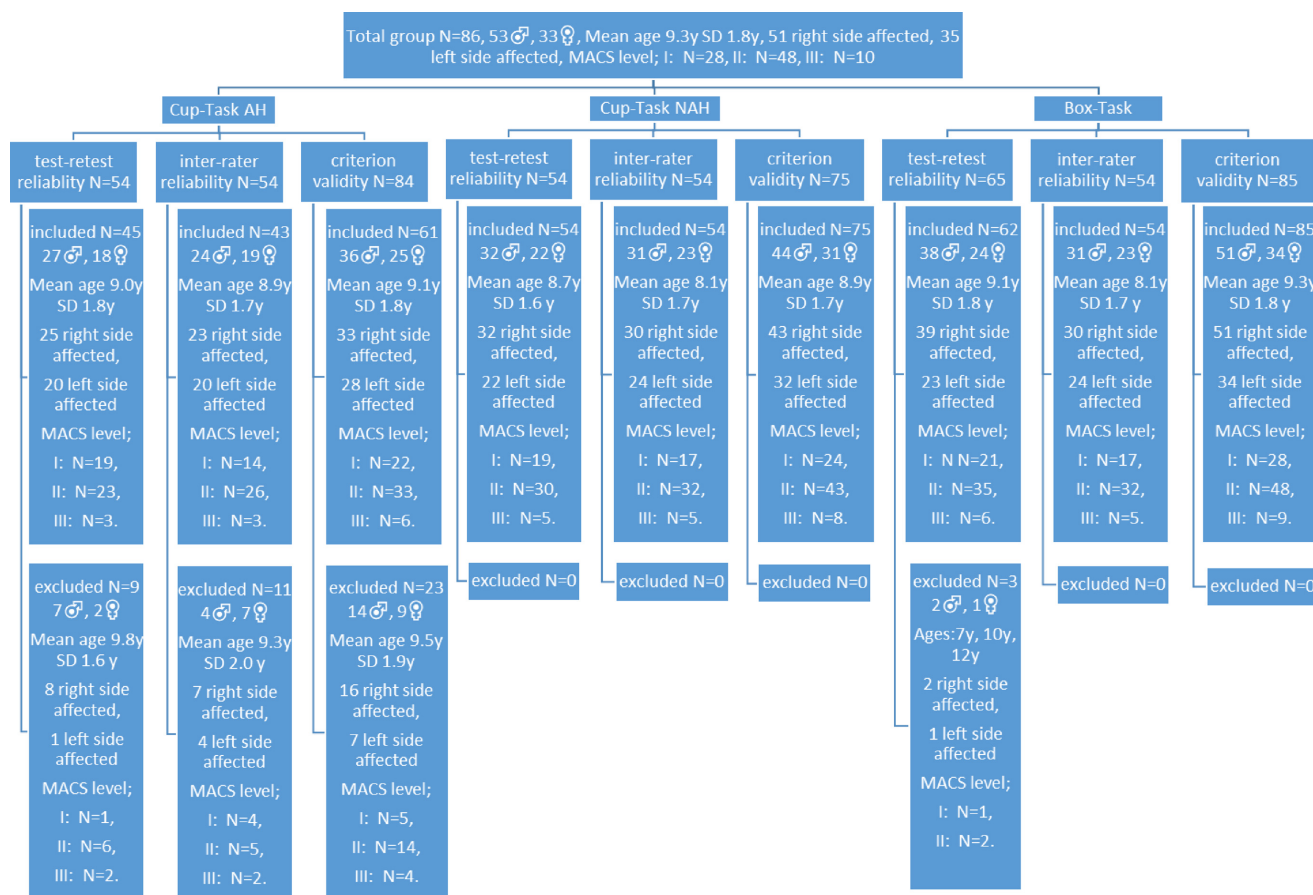
Interrater reliability was excellent ( $N = 43$ ; ICC = 0.960; 95% CI = 0.918–0.980), with an SEM value of 142 g and an SDC value of 393 g.

**Criterion validity.** Of the 84 participants, 23 children (27.4%) could not perform the measurement with an empty cup and the IGS measurement. See Figure 2 for details. The Pearson correlation coefficient ( $N = 61$ ) between the Cup-Task AH and IGS of the AH was moderate ( $r = 0.638$ ;  $P \leq .001$ ).

#### Cup-Task NAH

**Test-retest reliability.** All 54 children were able to perform the measurements on repeated occasions (Fig. 2). Test-retest reliability was excellent (ICC = 0.944; 95% CI = 0.895–0.969), with an SEM value of 272 g and an SDC value of 755 g. As seen in the Bland-Altman plot (Fig. 3), there were several outliers, but these are included in the calculations.

## Functional UE Strength Measurement in Pediatric CP



**Figure 2.** Participant characteristics. AH = affected hand; MACS = Manual Ability Classification System; NAH = nonaffected hand.

**Table 1.**  
Outcomes<sup>a</sup>

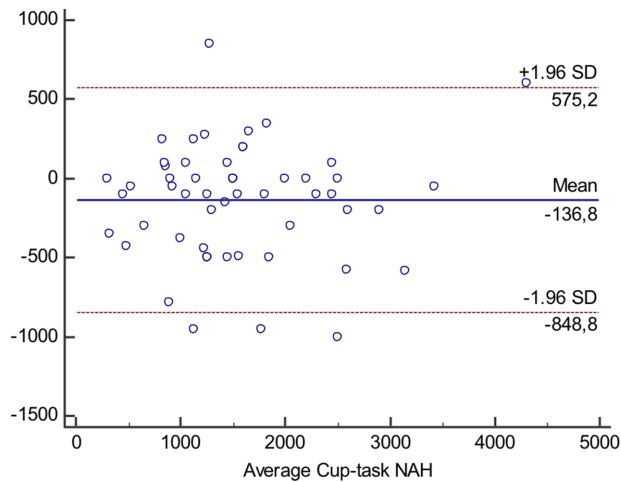
Task	Test-Retest Reliability				Interrater Reliability			
	Mean T0 (SD)	Mean T1 (SD)	Difference Between T0 and T1	Shapiro-Wilk Test (P value)	Mean T0 (SD)	Mean T1 (SD)	Difference Between T0 and T1	Shapiro-Wilk Test (P value)
Cup-Task AH	568 g (535 g)	787 g (687 g)	210 g (342 g)	≤ .001	602 g (475 g)	679 g (545 g)	77.3 g (187 g)	≤ .001
Cup-Task NAH	1500 g (836 g)	1637 g (823 g)	137 g (363 g)	.09	1622 g (1068 g)	1580 g (881 g)	42.7 g (599 g)	.01
Box-Task	5.79 kg (3.74 kg)	6.53 kg (3.99 kg)	0.74 kg (1.81 kg)	≤ .001	5.42 kg (4.03 kg)	6.26 kg (4.30 kg)	0.84 kg (2.46 kg)	≤ .001

<sup>a</sup>AH = affected hand; Diff = difference; NAH = nonaffected hand.

**Table 2.**  
Psychometric Properties of the Measurement Instruments<sup>a</sup>

Movement	Test-Retest Reliability					Interrater Reliability					Criterion Validity		
	N	ICC	95% CI	SEM	SDC	N	ICC	95% CI	SEM	SDC	N	Pcc with IGS	P value
Cup-Task AH	45	0.887	0.713–0.948	284 g	787 g	43	0.960	0.918–0.980	142 g	393 g	61	0.638	≤ .001
Cup-Task NAH	54	0.944	0.895–0.969	272 g	755 g	54	0.898	0.825–0.941	421 g	1166 g	75	0.489	≤ .001
Box-Task	62	0.934	0.875–0.963	1.38 kg	3.82 kg	54	0.896	0.813–0.941	1.82 kg	5.05 kg	85	0.555	≤ .001

<sup>a</sup>AH = affected hand; CI = confidence interval; ICC = intraclass correlation coefficient; IGS = isometric grip strength; NAH = nonaffected hand; Pcc = Pearson correlation coefficient; SDC = smallest detectable change; SEM = standard error of measurement.



**Figure 3.** Limits of agreement: Cup-Task NAH test-retest reliability. NAH = nonaffected hand.

**Interrater reliability.** All 54 children were able to perform the measurements during both tests (Fig. 2). Interrater reliability was excellent ( $ICC = 0.898$ ; 95%  $CI = 0.825$ – $0.941$ ), with an SEM value of 421 g and an SDC value of 1166 g.

**Criterion validity.** All 75 participants were able to perform both measurements. The Pearson correlation coefficient between the Cup-Task NAH and IGS NAH was low ( $r = 0.489$ ;  $P \leq .001$ ).

### Box-Task

**Test-retest reliability.** Sixty-five children performed both measurements, and 3 children (4.8%) were not able to perform 1 or 2 measurements adequately. Test-retest reliability was excellent ( $N = 62$ ;  $ICC = 0.934$ ; 95%  $CI = 0.875$ – $0.963$ ), with an SEM value of 1.38 kg and an SDC value of 3.82 kg.

**Interrater reliability.** All 54 children who participated in this part of the study were able adequately to perform the measurements twice. Interrater reliability was excellent ( $ICC = 0.896$ ; 95%  $CI = 0.813$ – $0.941$ ), with an SEM value of 1.82 kg and an SDC value of 5.05 kg.

**Criterion validity.** All 85 participants could lift the box and also perform the IGS adequately. For 1 child, an assessor determined that the empty box was the maximum weight capable of being lifted. The Pearson correlation coefficient between the Box-Task and maximum IGS AH was moderate ( $r = 0.555$ ;  $P \leq .001$ ).

## Discussion

The purpose of this study was to evaluate the psychometric properties of 2 newly developed functional

hand and upper extremity muscle-strength tests in children aged 7 to 12 years with USCP.

### Reliability

Not all the children were able to perform the Cup-Task AH sufficiently. Therefore, the number of included children for the reliability part of the study was lower than intended. For the Cup-Task NAH and Box-Task, the target population size was achieved.

All test-retest reliability ICC values and all interrater reliability ICC values for all measurement instruments showed excellent reliability. These results are in line with the previously performed pilot study.<sup>18</sup> Because of the skewed distribution of the data of the Cup-Task AH and Box-Task, only the limits of agreement for the test-retest reliability of the Cup-Task NAH could be calculated. The Bland-Altman plot showed good agreement, as reflected by the narrow limits of agreement.

Because of the excellent test-retest reliability and excellent interrater reliability, all instruments can reliably be used for cross-sectional measurements, for example, to determine the muscle strength of the AH and NAH in children with USCP.

Beside test-retest reliability and interrater reliability, the SEM and SDC values are important components of reliability.<sup>24</sup> SEM and SDC values for these measurement instruments have not been previously reported. With the SEM and SDC values, the usability in patients in clinical practice can be determined, certainly when these instruments are used to determine changes over time in individual patients. By using these values, it can be determined whether the changes in functional strength after a strength training program are larger than the changes that can occur due to variability between 2 measurements. Unfortunately, no information is available on how much functional upper extremity strength improvement a child with USCP can achieve after such a training program. Therefore, it is not possible to draw clear conclusions about the usability of the measurement instruments to measure real and clinically important changes.

### Criterion Validity

In more than 25% of the participants, the assessor judged that the child was not able to perform the Cup-Task AH. Therefore, this measurement instrument is not suitable for a substantial proportion of children with USCP. Subanalyses of the participants who could not perform the task showed a wide distribution of age, sex, and MACS level; they included 5 children with MACS level I, 14 children with MACS level II, and 4 children with MACS level III. Comparing these MACS levels with those of the children included in the validity study (MACS level I,  $n = 22$ ; level II,  $n = 33$ ; level III,  $n = 6$ ; see Fig. 2), it might be concluded that this task is more difficult to

perform for children with MACS level III. Within the MACS level III group, most children were unable to lift the cup horizontally with the AH because they could not grasp the handle of the cup. For the children who were able to perform the Cup-Task AH, the Pearson correlation coefficient met the expectations. Therefore, it can be concluded that the Cup-Task AH is valid for the children who are able to perform the task. For the children who could not perform the Cup-Task AH, another measurement instrument is needed to measure functional hand and upper extremity muscle strength.

The Cup-Task NAH and Box-Task were feasible for all participants, irrespective of the MACS level. The Pearson correlation coefficient of the Box-Task also met the expectations. However, the Pearson correlation coefficient of the Cup-Task NAH ( $r = 0.489$ ) was just below the expected range (0.50–0.70).

Because these functional strength instruments have not been studied previously with children with USCP, no direct comparison with other studies can be made. Of the 11 supplemental tests recommended by the National Institutes of Health Common Data Elements for assessing upper extremity motor function in children with CP, only maximum voluntary isometric contraction testing and manual muscle testing measure muscle strength, and none measure upper extremity functional strength.<sup>4</sup> Thus the current tests have the potential to expand the number of reliable and valid tests for this population, and fill an important gap in our understanding of strength in the context of function.

### Strengths and Limitations

The strength of this study is its methodological quality. The COSMIN criteria were important for the design of this study, which resulted in scientifically valuable results.

Unfortunately, age, sex, and MACS level were not ideally distributed, which could have influenced the results. Because this study focused on the age range of 7 to 12 years, caution is advised when testing children in other age groups. The population size was too small to calculate separate ICC/SEM/SDC values for each MACS level. The small number of children included in this study with MACS level III means that caution is advised when testing such children. Moreover, the large number of assessors could be a limitation of the study because this could have resulted in more measurement errors. However, the large number of assessors more closely resembles clinical practice.

### Conclusion

The Cup-Task and Box-Task are reliable and valid measurement instruments for measuring functional hand and upper extremity muscle strength in children with USCP who can perform such tasks. However, in most cases, for children with USCP and MACS level III the

Cup-Task AH will not be suitable. To determine the usability of both instruments in children with USCP in longitudinal measurements, more research on the effects of increasing functional upper extremity strength in children with USCP is recommended.

### Author Contributions and Acknowledgments

Concept/idea/research design: K.J.F.M. Dekkers, R.J.E.M. Smeets, Y.J.M. Janssen-Potten, A.M. Gordon, L.A.W.M. Speth, E.A.A. Rameckers  
Writing: K.J.F.M. Dekkers, R.J.E.M. Smeets, Y.J.M. Janssen-Potten, A.M. Gordon, L.A.W.M. Speth, E.A.A. Rameckers  
Data collection: K.J.F.M. Dekkers, E.A.A. Rameckers  
Data analysis: K.J.F.M. Dekkers, R.J.E.M. Smeets, Y.J.M. Janssen-Potten, E.A.A. Rameckers  
Project management: K.J.F.M. Dekkers, R.J.E.M. Smeets, Y.J.M. Janssen-Potten, E.A.A. Rameckers  
Fund procurement: K.J.F.M. Dekkers, R.J.E.M. Smeets, Y.J.M. Janssen-Potten, E.A.A. Rameckers  
Providing participants: K.J.F.M. Dekkers, A.M. Gordon, L.A.W.M. Speth, E.A.A. Rameckers  
Providing facilities/equipment: Y.J.M. Janssen-Potten, A.M. Gordon, E.A.A. Rameckers  
Providing institutional liaisons: K.J.F.M. Dekkers, R.J.E.M. Smeets, Y.J.M. Janssen-Potten, A.M. Gordon, E.A.A. Rameckers  
Clerical/secretarial support: R.J.E.M. Smeets

Special thanks to all children who participated in this study. Also, special thanks to the schools and/or rehabilitation centres for their cooperation.

### Ethics Approval

Ethics approval was granted by the Medical Ethical Board of Maastricht University Medical Center and Maastricht University (METC azM/UM, ref. no. NL45430.068.1), Maastricht, the Netherlands, and Teachers College, Columbia University, New York City, New York, USA. Data were collected in the Netherlands and United States from 2009 to 2016.

### Funding

This study was funded by the Forward for Children With Disabilities Foundation, Valkenburg, the Netherlands; the Revant Innovation Foundation, Breda, the Netherlands; and the Johanna Children's Foundation, Arnhem, the Netherlands.

### Study Registration

This study was registered at METC azM/UM (ref. no. NL45430.068.1).

### Disclosures

The authors completed the ICJME Form for Disclosure of Potential Conflicts of Interest and reported no conflicts of interest.

DOI: 10.1093/ptj/pzz019

### References

- 1 Uvebrant P. Hemiplegic cerebral palsy aetiology and outcome. *Acta Paediatrica Scand Suppl.* 1988;345:1–100.
- 2 Eliasson AC, Krumlinde-Sundholm L, Rösblad 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;48:549–554.
- 3 Brændvik SM, Elvrum AK, Vereijken B, Roeleveld K. Relationship between neuromuscular body functions and upper extremity activity in children with cerebral palsy. *Dev Med Child Neurol.* 2010;52:29–34.

- 4 National Institutes of Health. Common data elements. Cerebral palsy. [https://www.commondataelements.ninds.nih.gov/CP.aspx#tab=Data\\_Standards](https://www.commondataelements.ninds.nih.gov/CP.aspx#tab=Data_Standards). Updated December 13, 2018. Accessed January 30, 2019.
- 5 Arnould C, Penta M, Renders A, Thonnard JL. ABILHAND-Kids: a measure of manual ability in children with cerebral palsy. *Neurology*. 2004;63:1045–1052.
- 6 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;49:259–264.
- 7 Randall M, Johnson L, Reddihough DS. *The Melbourne Assessment of Unilateral Upper Limb Function*. Melbourne, Australia: Occupational Therapy Department, Royal Children's Hospital; 1999.
- 8 DeMatteo C, Law M, Russell D, Pollock N, Rosenbaum P, Walter S. *Quality of Upper Extremity Skills Test Manual*. Hamilton, Ontario: Canchild, McMaster University; 1992.
- 9 Davids JR, Peace LC, Wagner LV, Gidewall MA, Blackhurst DW, Roberson WM. Validation of the Shriners Hospital for Children Upper Extremity Evaluation (SHUEE) for children with hemiplegic cerebral palsy. *J Bone Joint Surg Am*. 2006;88:326–333.
- 10 Mathiowetz V, Federman S, Wiemer D. Box and block test of manual dexterity: norms for 6–19 year olds. *Can J Occup Ther*. 1985;52:241–245.
- 11 Jebson RH, Taylor N, Trieschmann R, Trotter MJ, Howard LA. An objective and standardized test of hand function. *Arch Phys Med Rehabil*. 1969;50:311–319.
- 12 Smith YA, Hong E, Presson C. Normative and validation studies of the Nine-hole Peg Test with children. *Percept Mot Skills*. 2000;90:823–843.
- 13 Gardner RA, Broman M. The Purdue Pegboard: normative data on 1334 school children. *J Clin Child Adolesc Psychol*. 1979;8:156–162.
- 14 Hislop HJ, Montgomery J. *Daniels and Worthingham's Muscle Testing: Techniques of Manual Examination*. 6th ed. Philadelphia, PA: W.B. Saunders; 1995.
- 15 Noreau L, Vachon J. Comparison of three methods to assess muscular strength in individuals with spinal cord injury. *Spinal Cord*. 1998;36:716.
- 16 Verschuren O, Ketelaar M, Takken T, van Brussel M, Helders PJ, Gorter JW. Reliability of hand-held dynamometry and functional strength tests for the lower extremity in children with cerebral palsy. *Disabil Rehabil*. 2008;30:1358–1366.
- 17 Aertssen W, Smulders E, Smits-Engelsman B, Rameckers E. Functional strength measurement in cerebral palsy: feasibility, test–retest reliability, and construct validity. *Dev Neurorehabil*. 2018;9:1–9.
- 18 Rameckers E. Strength training in bimanual tasks for children with cerebral palsy. <http://www.trialregister.nl/trial/NL4533>. 2014. Accessed January 30, 2019.
- 19 Mokkink LB, Terwee CB, Patrick DL et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19:539–549.
- 20 Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2011;21:651–657.
- 21 Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol*. 1997;39:214–223.
- 22 Dekkers KJ, Janssen-Potten YJ, Gordon AM, Speth LA, Smeets RJ, Rameckers E. Reliability of maximum isometric arm, grip and pinch strength measurements in children (7–12 years) with unilateral spastic cerebral palsy. *Disabil Rehabil*. 2019;1:1–6.
- 23 Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. Upper Saddle River, NJ: Prentice Hall; 2000.
- 24 De Vet HC, Terwee CB, Mokkink LB. *Measurement in Medicine: A Practical Guide*. Cambridge, United Kingdom: Cambridge University Press; 2011.
- 25 Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327:307–310.
- 26 Giavarina D. Understanding Bland-Altman analysis. *Biochem Med (Zagreb)*. 2015;25:141–151.
- 27 Hinkle DE, Wiersma W, Jurs SG. *Applied Statistics for the Behavioral Sciences*. 5th ed. Boston, MA: Houghton Mifflin; 2003.