Validity and test–retest reliability of Children's Hand-use Experience Questionnaire in children with unilateral cerebral palsy

AHMED AMER[1,2] | ANN-CHRISTIN ELIASSON[3] | MARIE PENY-DAHLSTRAND[4,5] | LISELOTTE HERMANSSON[1,6]

1 Faculty of Medicine and Health, School of Health and Medical Sciences, Örebro University, Örebro; **2** University Health Care Research Center, Region Örebro County; **3** Department of Women's and Children's Health, Karolinska Institute, Stockholm; **4** Regional Rehabilitation Centre, Queen Silvia Children's Hospital, Gothenburg; **5** Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg; **6** Department of Prosthetics and Orthotics, Region Örebro County, Sweden.

Correspondence to Ahmed Amer, University Health Care Research Center (UFC), Region Örebro County, Box 1613, S-701 16 Örebro, Sweden. E-mail: ahmed.amer@regionorebrolan.se

PUBLICATION DATA
Accepted for publication 19th October 2015.
Published online

ABBREVIATIONS

CHEQ	Children's Hand-use Experience Questionnaire
ICC	Intraclass correlation coefficient
OBPP	Obstetric brachial plexus palsy
ULRD	Upper limb reduction deficiency

AIM To investigate the validity of the internet-based version of the Children's Hand-use Experience Questionnaire (CHEQ) by testing the new four-category rating scale, internal structure, and test–retest reliability.

METHOD Data were collected for 242 children with unilateral cerebral palsy (CP) (137 males and 105 females; mean age 9y 10mo, SD 3y 5mo, range 6–18y). Twenty children from the study sample (mean age 11y 8mo, SD 3y 10mo) participated in a retest within 7 to 14 days. Validity was tested by Rasch analysis based on a rating scale model and test–retest reliability by Kappa analysis and intraclass correlation coefficient (ICC).

RESULTS The four-category rating scale was within recommended criteria for rating scale structure. One item was removed because of misfit. CHEQ showed good scale structure according to the criteria. The effective operational range was >90% for two of the CHEQ scales. Test–retest reliability for the three CHEQ scales was: grasp efficacy, ICC=0.91; time taken, ICC=0.88; and feeling bothered, ICC=0.91.

INTERPRETATION The internet-based CHEQ with a four-category rating scale is valid and reliable for use in children with unilateral CP. Further studies are needed to investigate the validity of the internet-based version of CHEQ for children with upper limb reduction deficiency or obstetric brachial plexus palsy and the validity of the recommended improvements to the current version.

Children with unilateral hand dysfunction frequently encounter difficulties in daily activities, in particular related to performance of bimanual activities.[1] The Children's Hand-use Experience Questionnaire (CHEQ) has been developed to capture children's perceived quality of performance when using the affected hand in these situations.[2] The CHEQ has a unique feature in that it is an assessment of both the use of the affected hand in bimanual activities and the children's experience of their performance during these activities. The importance of capturing perceived performance in the children's own environment has resulted in the development of several instruments.[3] The review by Wallen and Stewart also highlighted that several of the available instruments need to be further clarified according to their item selection, validation, test–retest reliability, and sensitivity to change.[3]

The CHEQ was developed for children aged 6 to 18 years with unilateral hand impairment caused by unilateral cerebral palsy (CP), upper limb reduction deficiency (ULRD), or obstetric brachial plexus palsy (OBPP). The quality aspects of hand function are evaluated in CHEQ on three scales: grasp efficacy; time taken to perform the activity; and feeling bothered by the affected hand during performance of the activity. The CHEQ is recommended to be answered as self-report by the child from about 12 years of age, together with parent or guardian before this age or later if needed, or by parents or guardians as proxy.[2] The questionnaire is now internet-based, easily accessible, free to use, and has been translated into 12 different languages (www.cheq.se).

In a previous article describing the development and validation of CHEQ, the appropriateness of the included activities was confirmed by their reported relevance and bimanual nature, and the internal structure of the scales was confirmed by Rasch analysis.[2] The first version was paper-administered with 10 category rating scales (1–10), later collapsed into four-category rating scales (1–4) as a result of the Rasch analysis.[2] Although previous results showed that the 10-category version of CHEQ is a valid instrument for the assessment of experience of bimanual

activities in children and adolescents with unilateral hand dysfunction,[2] there is a need to investigate the validity of the new four-category rating scale. In addition, since a change in the rating scale may have consequences for the validity of the overall instrument, new analyses of internal validity are required. Further, the test–retest reliability needs to be investigated to increase the validity of future score interpretations.[4]

Thus, the aim of this study was to evaluate the psychometric properties of the internet-based version of CHEQ in terms of the four-category rating scale, test content and internal structure of the internet-based version, and the test–retest reliability of CHEQ in children with unilateral CP.

What this paper adds

- CHEQ can be used to consistently measure the experience of dysfunctional hand use.
- CHEQ four-category rating scale provides a valid estimation of children's hand use experience.
- Test–retest reliability of CHEQ is high for children with unilateral CP.

METHOD

A cross-sectional online survey design was chosen. Data were collected in six different countries: Australia, Israel, Italy, the Netherlands, Sweden, and the UK. For each language version of CHEQ, the translation was performed in the respective country. For the retest, a subgroup of the Swedish children was used.

Participants and data collection

A total of 242 children with unilateral CP participated in this study. This sample size is sufficient to achieve stable item calibration within a standard deviation (SD) of 0.5 logits with a 99% confidence interval (CI).[5] The age range was 6 to 18 years old (mean age 9y 10mo, SD 3y 5mo). More descriptive data are presented in Table I. Twenty of the Swedish children (mean age 11y 7mo, SD 3y 8mo) participated in the test–retest assessment after an interval of 7 to 14 days.

The participants represent a convenience sample: some children were participants in other studies where CHEQ was included in the data collection, other children were recruited specifically for this study, and the remainder answered the questionnaire for clinical purposes and gave consent to the data being used for research. One hundred and eighty-six (77%) parents answered CHEQ on behalf of their child and, for a further three children (1%), the CHEQ was answered by another proxy. Forty-three participants (18%, all >12y) completed the CHEQ by self-report, and 10 participants (4%) completed the CHEQ together with their caregiver. For the retest, 20 children participated: 10 children with parents as proxy, and 10 children answered the CHEQ themselves. The same person completed the CHEQ on both occasions.

The study was approved by the Regional Ethical Review board of Stockholm, Sweden, and was performed in accordance with the ethical principles of the Declaration of Helsinki.[6] Informed consent was given by the carers of the participating children.

Instrumentation

CHEQ is an internet-based questionnaire containing 29 items (bimanual activities) presented one by one in random order. It can be answered as self-report by the child, by parents, or others answering CHEQ together with the child, or by parents or other caregivers as proxy. For children aged 12 years or younger, it is recommended that parents answer CHEQ together with the child or as a proxy, whereas adolescents are expected to answer CHEQ by self-report. For each CHEQ item there are two opening questions. The first question reads: 'Is this something you usually do independently?', and has the response options: 'yes', 'no', 'I get help/avoid doing it', or 'not applicable'. If the answer is 'no' or 'not applicable', the item is scored as missing and the respondent moves to the next item. If the answer is 'yes', the second opening question appears: 'Do you use one hand or both hands together?', with the response options: 'one hand', 'both hands', 'with the involved hand supporting but not holding', and 'both hands, with the involved hand holding the object'.

Next, the respondent's experience is evaluated by three questions rated on four-category scales with verbal anchors on each end, constituting three dimensions of hand use: grasp efficacy, indicating how effective the grasp is perceived, where 1 is 'ineffective' and 4 is 'effective'; time

Table I: Descriptive information on the participating children

Country	*n*	Age group in years		Mean age (SD)	Sex female/male	Affected side right/left
		≤12	>12			
Australia	33	32	1	8y 4mo (2y 2mo)	11/22	[a]
Netherlands	61	56	5	8y 4mo (2y 11mo)[b]	25/36	36/25
Israel	28	24	4	8y 11mo (3y)	13/15	21/7
Italy	46	31	15	10y 6mo (3y 7mo)	19/27	32/14
Sweden	47	26	21	11y 8mo (3y 8mo)	26/21	19/28
UK	27	17	10	11y 5mo (3y)	11/16	19/8
Total	242	186	56	9y 10mo (3y 5mo)[b]	105/137	127/82

[a]Missing information for 33 children from Australia. [b]Missing information for four children from the Netherlands.

taken, indicating the time taken to perform the activity compared with peers, where 1 is 'considerably longer' and 2 is 'equally long'; and feeling bothered, indicating whether the child feels irritated, sad, or uncomfortable when doing the activity, where 1 is 'it bothers me a lot' and 4 is 'it does not bother me at all'. Approximate time to answer CHEQ is 30 minutes. When this is finished, the system automatically generates the results in a PDF file, including the scores on each question and the average value of each scale.

Statistical analysis

To provide evidence of validity based on test content, the opening questions were reported as descriptive statistics. Rasch analysis was used to assess the internal structure, and to convert the ordinal data into interval data. More information about Rasch analysis can be found elsewhere.[7] Each scale was analyzed separately using the rating scale model.[2,8] Evidence of validity was sought for: rating scale functioning according to guidelines described by Linacre;[9] internal structure according to Fisher;[10] and item-fit statistics where information-weighted (infit) mean square values between 0.71 and 1.40 logits, in combination with an infit z-standardized value between −2.0 and 2.0, were considered acceptable[10,11] (Table II). At least 95% of items (28/29) should be within this range.[12] Targeting between the CHEQ items and children's responses was assessed by effective operational range of CHEQ, difference between item and person mean, and ceiling and floor effects. The effective operational range indicates the proportion of participants covered by the instrument was based on 50% cumulative thresholds.[13] For this study, if >90% of the children were within this range, targeting would be regarded as high (Table II).

To assess test–retest reliability on the opening questions, Kappa analysis was used. The strength of the agreement was interpreted according to Fleiss's guidelines:[14] poor agreement for $\kappa \leq 0.40$; fair to good agreement for $k>0.40$ and <0.75; and excellent agreement for $\kappa \geq 0.75$. For the three scales representing the child's experience, a two-way mixed effects model was used to generate an intraclass correlation coefficient (ICC [2,1]) with 95% CI.[15,16] Data were normally distributed (Shapiro–Wilk test, p=0.352–0.969) and ICCs were calculated on the logit measure for each person on each of the three scales. A reliability coefficient above 0.80 is desirable for group-level comparisons and above 0.90 for individual comparisons.[17] The Winsteps version 3.80.1 Rasch analysis software (Winsteps, Beaverton, OR, USA) was used for analyzing the data.[18] SPSS version 22 (IBM Corp., Armonk, NY, USA) was used for ICC and Kappa analyses.

RESULTS

The activities were performed independently by 26% to 95% of the children (Table III). Generally, the easier activities were performed by most of the children, whereas a large proportion of the difficult activities were rated 'not

Table II: Recommended and observed values of the validity criteria for the Children's Hand-use Experience Questionnaire

	Criterion	Recommended value	Observed value		
			Grasp efficacy	Time taken	Feeling bothered
Rating scale functioning	Number of observations (1, 2, 3, 4)[a]	≥10 observations/category	141, 887, 1673, 1366	365, 1131, 1777, 1319	270, 923, 1362, 1931
	Average measure (1, 2, 3, 4)[a,b]	Increased monotonically	−0.87, −0.13, 1.38, 3.30	−1.28, −0.32, 0.97, 2.93	−1.13, −0.25, 0.89, 3.20
	Step difficulties (1, 2, 3, 4)[a,b]	Increased monotonically	None, −2.57, 0.02, 2.55	None, −2.11, −0.11, 2.22	None, −1.95, −0.03, 1.99
	Distance between adjacent steps[b]	1.4–5.0 logits	2.5, 2.5	2.0, 2.3	1.9, 2.0
CHEQ as a rating scale instrument	Difference in mean (SE)[b]	−1.0 to 1.0 logits[c]	**1.22** (0.52)	0.67 (0.46)	**1.55** (0.63)
	Effective operational range	≥90%[c]	92	90	**73**
	Ceiling effect	≤2%	**2.5**	**2.1**	**12.8**
	Floor effect	≤2%	0.4	1.2	0
	Person reliability	>0.80	0.91	0.92	0.89
	Item reliability	>0.80	0.92	0.92	0.87
	Person separation	≥2	3.1	3.5	2.9
	Item separation	≥2	3.3	3.5	2.6
	Explained variance by measure	>60%	**52.5**	**53.8**	**56.9**
	Unexplained variance by first contrast	<5%	3.3	3.3	3.1
	Value	<2.0 eigenvalue	2.0	2.0	2.0

[a]1, 2, 3, 4 are the rating scale categories. [b]Measures in logits. Recommended values for rating scale functioning are based on Linacre,[9] and for CHEQ as a rating scale instrument are based on Fisher,[10] except [c]where no criteria are available – cut-off values are based on a number where most participants will be covered by the instrument. Emboldened values are deviations from the criteria. CHEQ, Children's Hand-use Experience Questionnaire; SE, standard error.

Table III: Item fit statistics for the online version of Childrens Hand-use Experience Questionnaire in children with unilateral cerebral palsy

No.	Item	Independently performed (%)	With two hands (%)	Not applicable (*n*=242)	Grasp efficacy				Time taken				Feeling bothered			
					Measure	SE	MnSq	ZSTD	Measure	SE	MnSq	ZSTD	Measure	SE	MnSq	ZSTD
1	Pull up tracksuit trousers	94	61	4	−0.27	0.14	1.04	0.37	−0.36	0.11	1.13	1.34	−0.34	0.12	1.15	1.33
2	Remove a straw from the front of a juice box and insert it	81	72	19	−0.37	0.13	0.93	−0.59	−0.31	0.12	0.93	−0.71	−0.35	0.13	0.89	−0.98
3	Put on socks	85	58	4	0.12	0.14	1.26	2.09	0.25	0.11	1.25	2.42	0.13	0.12	1.26	2.27
4	Eat out of a small container of yoghurt	95	81	6	−0.60	0.13	1.10	0.97	−1.00	0.11	1.08	0.89	−0.49	0.12	1.20	1.87
5	Spread out glue on paper using a glue stick	91	88	3	−0.32	0.12	0.76	−2.60	−0.40	0.11	0.92	−0.78	−0.20	0.12	0.92	−0.76
6	Cut out a picture using scissors	85	83	6	0.12	0.12	0.90	−1.00	0.43	0.11	0.84	−1.67	0.57	0.12	1.01	0.15
7	Butter a slice of soft bread	59	52	25	0.31	0.15	1.11	0.89	0.31	0.14	0.69	−2.91	0.34	0.15	0.76	−2.04
8	Open a small box (e.g. a box of mints)	88	79	7	−0.56	0.13	0.82	−1.84	−0.63	0.12	0.75	−2.72	−0.48	0.12	0.80	−1.91
9	Cut up a pancake (or other food easy to cut up) on the plate	69	51	8	0.25	0.15	1.18	1.41	0.01	0.12	0.92	−0.72	0.11	0.14	1.16	1.27
10	Put toothpaste on a toothbrush	86	67	7	−1.05	0.15	0.85	−1.36	−0.66	0.12	0.98	−0.21	−0.72	0.13	0.73	−2.57
11	Open the zipper on a small bag (e.g. pencil case or purse)	95	88	5	−0.85	0.13	0.84	−1.69	−0.69	0.11	0.87	−1.44	−0.57	0.12	0.77	−2.34
12	Remove the wrapping from a pack of sweets	86	69	6	0.63	0.15	1.38	2.9	−0.35	0.12	0.89	−1.11	−0.39	0.13	0.92	−0.73
13	Handle playing cards	68	54	28	−0.23	0.13	0.97	−0.20	0.55	0.13	1.35	2.93	**0.74**	**0.14**	**1.64**	**4.47**
14	Pick money out of a purse or wallet	84	75	19	0.63	0.15	1.38	2.86	−0.24	0.12	0.86	−1.39	−0.20	0.12	0.81	−1.83
15	Carry a tray (e.g. in the canteen)	67	61	40	−0.42	0.13	0.82	−1.83	−0.07	0.13	1.22	1.87	0.10	0.14	1.25	1.92
16	Pull up the zipper of a jacket	76	67	4	0.00	0.15	1.28	2.21	0.18	0.12	1.33	2.96	0.09	0.13	0.91	−0.80
17	Fasten the button on trousers/pants	46	33	25	−0.20	0.14	1.09	0.87	0.37	0.16	1.27	1.84	0.24	0.17	0.93	−0.42
18	Open a plastic box with a lid (e.g. an ice-cream box)	83	75	7	0.72	0.19	0.89	−0.67	−0.28	0.12	0.76	−2.53	−0.21	0.13	0.76	−2.32
19	Screw off the cap of a small unopened soft drink bottle	72	67	6	−0.13	0.13	0.85	−1.45	−0.63	0.13	0.84	−1.48	−0.32	0.14	0.94	−0.46
20	Remove the wrapping from an ice cream	62	53	9	−0.56	0.14	1.08	0.72	0.01	0.14	0.83	−1.46	0.09	0.15	0.84	−1.27
21	Buckle a helmet (e.g. a bike helmet)	33	28	76	0.08	0.16	0.95	−0.36	−0.05	0.19	1.27	1.56	−0.37	0.21	0.73	−1.60
22	Cut on a chopping board (e.g. fruit, vegetables, bread)	52	47	44	−0.53	0.23	0.84	−0.90	0.56	0.14	1.02	0.22	0.27	0.16	1.05	0.38
23	Peel an orange	30	28	59	0.46	0.16	0.97	−0.18	0.93	0.20	1.02	0.15	0.59	0.22	0.94	−0.24
24	Open a bag (e.g. a bag of crisps)	64	57	5	1.08	0.21	0.93	−0.36	−0.03	0.14	1.22	1.83	−0.08	0.15	1.23	1.64
25	Take off the protective plastic backing of an Elastoplast	49	45	52	0.13	0.15	1.32	2.47	0.09	0.16	0.99	−0.03	−0.11	0.18	0.89	−0.67
26	Cut meat (or other food hard to cut up) on a plate	33	30	13	0.26	0.17	1.12	0.89	1.42	0.19	1.04	0.31	**1.38**	**0.20**	**1.45**	**2.38**
27	Open up a carton of milk or juice	39	36	57	1.73	0.20	0.76	−1.60	−0.22	0.18	0.67	−2.40	−0.17	0.19	1.02	0.15
28	Tie shoelaces	26	24	52	−0.10	0.20	0.87	−0.85	0.80	0.22	1.35	1.74	**0.34**	**0.24**	**1.70**	**2.80**

Misfit items are presented in bold. Measure unit: logits. SE, standard error; MnSq, mean square value; ZSTD, z-standardized value.

applicable' for many of the children. This is shown by the number of missing values for each scale (the maximum is 6776, i.e. 242 children×28 items): grasp efficacy, 2709 (39%); time taken, 2184 (32%); and feeling bothered, 2290 (34%). The applicable activities were performed using both hands by 24% to 88% of the children, while the remainder used one hand (Table III). This gives an indication of the relevance of the test content. An initial item-fit analysis demonstrated that the item 'Fasten a necklace' misfits all three scales. This item was subsequently removed and further analyses were based on 28 items.

Rating scale functioning

Average measures and step difficulties were ordered and increased monotonically, and the distances between adjacent thresholds were within the recommended range (Table II). The number of observations for rating scale category 1 was acceptable but the proportion was low compared to categories 2 to 4 (grasp efficacy 3%, time taken 8%, and feeling bothered 6%).

Validity based on internal structure

The CHEQ scales showed good unidimensionality, although the explained variances were below the recommended value. The standardized residuals of the items were distributed randomly in the first principal component analysis of residuals which meant the secondary dimension could not affect the main dimension (Table II). Item and person reliability were high, and the separation showed that CHEQ was able to distinguish between participants based on grasp efficacy, time taken, and feeling bothered. In contrast, the opening questions were not able to distinguish between participants based on what activities they performed (separation 1.89) or whether they used the affected hand for grasp or support (separation 1.74).

The effective operational range in grasp efficacy and time taken was ≥90%, whereas in feeling bothered about 27% of the children were outside the operational range. This was confirmed by the ceiling effect for the scale (Table II).

Table IV: Test–retest reliability for the online version of Children's Hand-use Experience Questionnaire in children with unilateral cerebral palsy

	n	ICC	95% CI
Whole group			
Grasp efficacy	19	0.89	0.73–0.96
Time taken	20	0.87	0.69–0.94
Feeling bothered	20	0.91	0.79–0.96
Parent proxy-report			
Grasp efficacy	10	0.85	0.50–0.96
Time taken	10	0.93	0.74–0.98
Feeling bothered	10	0.93	0.75–0.98
Child self-report			
Grasp efficacy	9	0.92	0.35–0.98
Time taken	10	0.74	0.22–0.93
Feeling bothered	10	0.83	0.45–0.95

ICC, intraclass correlation coefficient; CI, confidence interval.

Item fit statistics showed acceptable fit for the scales grasp efficacy and time taken, but in feeling bothered three items (11%) were misfit (Table III).

Test–retest reliability

In the retest with 20 children, the group-level Kappa analysis for the opening questions showed fair to good agreement for performing the activity independently (average κ 0.63) and for using the affected hand as support or to grasp (average κ 0.57). The ICC was high for all three CHEQ scales (average ICC 0.87–0.91), sufficient for group-level comparisons, and near sufficient for individual comparisons (Table IV). The results were similar when analyzing the data separately for parent proxy-reports and child self-reports, except for the scale time taken, where child self-reports showed a lower reliability.

DISCUSSION

The results show that the CHEQ is a valid tool for use in children with CP: the four-category rating scale is used as expected, the internal structure of the CHEQ scales was acceptable, and the test–retest reliability was excellent, with the ICC values providing a strong indication that results from CHEQ assessments are reliable.

There has been a lack of valid and reliable questionnaires to measure children's perspective of using their hands in daily activities.[19,20] In today's health care, emphasis is placed on client-centred practice where the children's perspective and their preferences when deciding on interventions are important. CHEQ could be a useful tool for this purpose. CHEQ provides a different perspective on hand function compared to tests measuring observed ability, such as the Assisting Hand Assessment.[12] CHEQ is also different from ABILHAND-Kids, a questionnaire that has a broader perspective on activities of daily living not specifically related to problems apparent in children with unilateral hand impairment.[21]

The high ICC values in this study demonstrate that CHEQ can discriminate well between different ability levels in the scale, supporting the valid use of CHEQ both for individual decision-making and group comparison in research.[17] An important finding is that there were only minor differences, although larger variation, between respondent groups, namely parents acting as proxies for the younger children compared to the older children's self-reports. This indicates that parent proxy-reports on time taken and feeling bothered are more stable compared to children's self-report, whereas children's report on grasp efficacy is more stable than the parents' proxy-report. This finding is important when using CHEQ for pre- and post-testing. The reason for this difference between respondent groups is not evident and requires more research to understand. Further studies with larger samples are needed to confirm this and to calculate the smallest detectable difference for CHEQ in both proxy and self-rating versions.

Although parents can serve as proxy, it is important to recognize that the ratings are neither perceived experience

(as scored by the children themselves) nor perceived experience (as experienced by parents) but merely the parents' perception of their child's experience, and the scorings are not directly comparable. An earlier study by Ylimainen et al.[22] found that there was a discrepancy between children's and parents' ratings. It would be interesting to compare CHEQ ratings made by parents to the ratings of their child to further explore the parents' understanding of their child's experience of the hand function in activities of daily living.

The results confirmed that the CHEQ items are able to differentiate between children's experience of hand use when performing the selected activities. Overall, the items demonstrated good fit to the construct and targeting of the children's experience. However, the results also showed a ceiling effect, especially on the scale feeling bothered. Furthermore, the most difficult activities (e.g. 'Open a bag' or 'Cut meat') had a high number of 'not applicable' and 'no' responses to the question about performing the activity independently, and were thus rated by a very low number of respondents. This suggests that the difficulty of the activity may be the reason for the low response rate for that item, as the children may typically ask for help to perform these activities. Furthermore, in this scale there were three items that misfit, indicating that they gave unexpected responses, two of which ('Handle playing cards' and 'Cut meat') were difficult activities with a low number of responses.

When comparing the results from this study to the previous validation study on CHEQ,[2] the number of items with an unacceptable rate of 'not applicable' (more than 10%) responses is high: previously three items and in this study 11 items (Table III). Furthermore, the three items in the previous study have an increased proportion of 'not applicable' responses in this study: 'Cut on chopping board' (18%); 'Peel an orange' (24%); and 'Fasten a necklace' (45%). The difference between the studies (multicultural sample, single type of dysfunction) may explain the discrepancy, but this needs to be verified in future studies on children with ULRD or OBPP in different countries.

Limitations

In order to get a large enough sample for the analyses, children from different countries were included in this study. This could be a potential limitation. Children living in different countries with different cultures, may experience the CHEQ items differently. This may explain the difference in the proportion of 'not applicable' items and some of the misfits of items on the CHEQ. However, as the major results are in favour of this version of CHEQ, the use of a varied sample could in fact strengthen the validity of CHEQ for use in different countries. Several language versions of CHEQ were created and, although the translations were made in a rigorous way,[23] there may nonetheless be some differences between the understanding of the items in the different language versions. Besides the need for a sufficiently large sample, test evaluation also needs a large enough number of responses to each item. Another problem in this study, induced by the test itself, is the high number of missing scores. The missing data is a fundamental problem in the CHEQ instrument. One interesting aspect of a test is to investigate how the items function on different subgroups, so-called differential item functioning. The recommended size of each subgroup for differential item functioning analyses is 200 to 300 participants.[24] Hence, it was not possible to perform country-based differential item functioning analysis in this study, but this is something that we recommend for future studies.

Finally, CHEQ was developed for children with OBPP, ULRD, or unilateral CP, but only the latter were included in this study. Hence, the results of this study are limited to children with unilateral CP and we need similar studies on children with OBPP and ULRD to confirm the results.

Recommended improvements to the CHEQ instrument

Based on the current results, a new version of CHEQ is recommended in which the number of items is reduced by omitting the misfit item and possibly by dropping some of the items that received a large number of 'not applicable' responses. The results for the opening questions (whether the activity is performed independently and how the affected hand is used) show that they are not useful in distinguishing children based on their performance and should not be used as a measure. One way to improve the CHEQ is to remove the two opening questions from the instrument. This action will also resolve the problem with the ceiling effect, since the responses 'not applicable' and 'no' are no longer available – instead, the rating category 1 for the grasp efficacy scale will be adjusted to cover the reported difficulty of the activity. This means that the respondent will give a score on the three scales for all items, which may improve the targeting of the scale and give a truer picture of the respondents' experience of using the affected hand in everyday activities and their level of frustration if they are unable to perform the activity independently. The practical advantage of this is that it will be possible to have the raw scores directly converted online to a 0 to 100 scale based on logits from the Rasch analysis. Also, by taking away these questions, the time to complete the CHEQ will be reduced. This might be of clinical relevance since feedback from CHEQ respondents indicates that they find it time-consuming to complete the questionnaire.

CONCLUSION

In conclusion, the results demonstrate the validity of the internet-based CHEQ with a four-category rating scale in children with CP. The test–retest reliability is high, allowing for group level comparisons and near sufficient for individual comparisons. This means that CHEQ provides a unique perspective of children's experience of using the affected hand in daily activities and can be used as a complement to other tests measuring aspects of capacity and performance. Given that the development of a test is an ongoing process, further studies are needed to investigate

the validity of the internet-based version of CHEQ for children with ULRD or OBPP and the recommended improvements to the current version.

ACKNOWLEDGEMENTS

We would like to thank all participating units for helping us with the data collection. We would especially like to thank Leanne Sakzewski, PhD, University of Queensland, South Brisbane, Australia; Pauline Aarts, PhD, Sint Maartenskliniek, Nijmegen, the Netherlands; Dido Green, PhD, Oxford Brookes University, Oxford, UK; Elisa Siccola, Department of Developmental Neuroscience, IRCCS Fondazione Stella Maris, Pisa, Italy; and Marilyn Cohen, OTR MSc, Alyn Pediatric and Adolescent Rehabilitation Hospital, Jerusalem, Israel. This study has been financially supported by Stiftelsen Frimurarna Barnhuset, Swedish Research Council (grant nos 521-211-2655 and 521-2011-456) and Centre for Rehabilitation Research, Region Örebro County. The authors have stated that they had no interests that could be perceived as posing a conflict or bias.

REFERENCES

1. Skold A, Josephsson S, Eliasson AC. Performing bimanual activities: the experiences of young persons with hemiplegic cerebral palsy. *Am J Occup Ther* 2004; **58**: 416–25.
2. Skold A, Hermansson LN, Krumlinde-Sundholm L, Eliasson AC. Development and evidence of validity for the Children's Hand-use Experience Questionnaire (CHEQ). *Dev Med Child Neurol* 2011; **53**: 436–42.
3. Wallen M, Stewart K. Upper limb function in everyday life of children with cerebral palsy: description and review of parent report measures. *Disabil Rehabil* 2015; **37**: 1353–61.
4. American Educational Research Association, American Psychological Association, National Council on Measurement in Education editors. The Standards for Educational and Psychological Testing. Washington: American Educational Research Association, 2014.
5. Linacre J. Sample size and item calibration stability. *Rasch Meas Trans* 1994; **7**: 1.
6. World Medical Association. Declaration of Helsinki: ethical principles for medical research involving human subjects. *J Postgrad Med* 2002; **48**: 206–8.
7. Bond T, Fox C. Applying the Rasch model. Mahwah, New Jersey: Lawrence Erlbaum Associates Inc, 2001.
8. Wright B, Mok M. An Overview of the Family of Rasch Measurement Models. In: Smith E, Smith R, editors. Introduction to Rasch Measurement. Maple Grove, Minnesota: JAM Press, 2004.
9. Linacre J. Optimizing Rating Scale Category Effectiveness. In Smith E, Smith R, editors. Introduction to Rasch Measurement. Maple Grove, Minnesota: JAM Press, 2004: 258–78.
10. Fisher W. Rating scale instrument quality criteria. *Rasch Meas Trans* 2007; **21**: 1.
11. Wright B, Linacre M. Reasonable mean-square fit values. *Rasch Meas Trans* 1994; **8**: 370.
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. Linacre J. Table 1 Wright maps of the latent variable. Available from: http://www.winsteps.com/winman/table1.htm (accessed 24 March 2015).
14. Fleiss J, Levin B, Cho P. Statistical Methods for Rates and Proportions. New York: Wiley NY, 2003: 598–626.
15. Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; **86**: 420–8.
16. Weir J. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005; **19**: 231–40.
17. Polit D, Beck CT. Nursing Research. Generating and Assessing Evidence for Nursing Practice. Philadelphia: Lippincott, 2008.
18. Linacre J. Winsteps Rasch measurement computer program. 3.81.0 edn. Beaverton, Oregon: Winsteps.com, 2014.
19. 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.
20. 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.
21. Arnould C, Penta M, Renders A, Thonnard J. ABILHAND-Kids: a measure of manual ability in children with cerebral palsy. *Neurology* 2004; **63**: 1045–52.
22. Ylimainen K, Nachemson A, Sommerstein K, Stockselius A, Norling Hermansson L. Health-related quality of life in Swedish children and adolescents with limb reduction deficiency. *Acta Paediatr* 2010; **99**: 1550–5.
23. Regmi K, Naidoo J, Pilkimgton P. Understanding the processes of translation and transliteration in qualitative research. *Int J Qual Methods* 2010; **9**: 16–26.
24. Linacre J. Differential item functioning DIF sample size nomogram. *Rasch Meas Trans* 2013; **26**: 1.