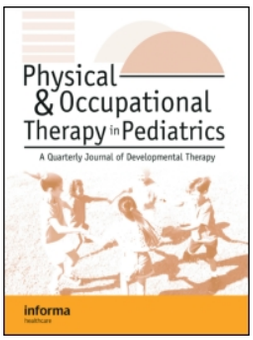

Physical & Occupational Therapy In Pediatrics

ISSN: 0194-2638 (Print) 1541-3144 (Online) Journal homepage: http://www.tandfonline.com/loi/ipop20

To Explore the Validity of Change Scores of the Children's Hand-use Experience Questionnaire (CHEQ) in Children with Unilateral Cerebral Palsy

Ulrike C Ryll, Ann-Christin Eliasson, Carolien HG Bastiaenen & Dido Green

To cite this article: Ulrike C Ryll, Ann-Christin Eliasson, Carolien HG Bastiaenen & Dido Green (2018): To Explore the Validity of Change Scores of the Children's Hand-use Experience Questionnaire (CHEQ) in Children with Unilateral Cerebral Palsy, Physical & Occupational Therapy In Pediatrics, DOI: 10.1080/01942638.2018.1438554

To link to this article: https://doi.org/10.1080/01942638.2018.1438554

Published online: 26 Feb 2018.

PHYSICAL & OCCUPATIONAL THERAPY IN PEDIATRICS
2018, VOL. 0, NO. 0, 1–13
https://doi.org/10.1080/01942638.2018.1438554

Taylor & Francis
Taylor & Francis Group

To Explore the Validity of Change Scores of the Children's Hand-use Experience Questionnaire (CHEQ) in Children with Unilateral Cerebral Palsy

Ulrike C Ryll[a], Ann-Christin Eliasson[a], Carolien HG Bastiaenen[b,c], and Dido Green[d,e]

[a]Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden; [b]Caphri Research Institute, Program Functioning and Rehabilitation, Department of Epidemiology, Maastricht University, Maastricht, The Netherlands; [c]School of Health Professions, Department of Health, Zurich University of Applied Sciences, Winterthur, Switzerland; [d]Centre for Rehabilitation, Oxford Brookes University, Oxford, UK; [e]Department of Occupational Therapy, Jönköping University, Jönköping, Sweden

ABSTRACT

Aims: To explore the validity of change scores of the Children's Hand-use Experience Questionnaire (CHEQ). *Methods*: Analysis of the CHEQ included 44 children (15 girls) between 6–16 years (median 9.0; IQR 8–11) with unilateral cerebral palsy, with baseline and post- (two-week intensive) intervention assessments using the Goal Attainment Scale (GAS) as external anchor for change. Hypotheses on the magnitude of expected change were formulated and correlation coefficients and effect sizes calculated. Receiver operating curve analysis was performed and the area under the curve (AUC) calculated to investigate the ability of CHEQ to discriminate between improvement and non-improvement according to GAS. *Results*: All hypotheses about the magnitude of change were confirmed supporting longitudinal validity of CHEQ scales to measure change in the perception of bimanual performance. AUCs for the Grasp efficacy and the Time utilization were slightly below, and for the Feeling bothered slightly above the threshold. The latter one accurately discriminating between children that improved and did not improve according to the GAS. *Conclusions*: Evidence was found that CHEQ scales capture change in bimanual performance but with limited accuracy for two out of three scales. The validity of CHEQ change scores needs to be further explored in a wider population.

ARTICLE HISTORY
Received 5 July 2017
Accepted 29 January 2018

KEYWORDS
Anchor; bimanual performance; construct approach; longitudinal validity; psychometric properties; responsiveness

The Children's Hand-use Experience Questionnaire (CHEQ) is an online questionnaire for children from 6 to 18 years with unilateral upper limb impairment. It is designed to measure children's own experiences when using the affected hand in common daily life activities requiring use of both hands. CHEQ provides insight into how the child (or proxy) perceives the efficacy in grasping objects, time taken to perform the activities, and how bothersome hand function is during performance. To accurately measure a child's performance in clinical practice and for research purposes, sufficient psychometric properties of the instrument, i.e. reliability and validity, are essential. Validity can be defined as 'the degree to which an instrument truly measures the construct(s) it purports to measure' (Vet, Terwee, Mokkink,

CONTACT Ulrike C. Ryll ulrike.ryll@ki.se Department of Women's and Children's Health, Astrid Lindgren Children's Hospital, Neuropediatric Unit, Q2:07, 171 76 Stockholm, Sweden.

& Knol, 2011). This comprises the validity of single scores in a cross-sectional setting as well as change scores in a longitudinal setting (Mokkink et al., 2010). Responsiveness is considered one aspect of validity as it concerns the validity of change scores, i.e. longitudinal validity (Hays & Hadorn, 1992; Streiner, Norman, & Cairney, 2014; Vet et al., 2011). Another aspect of validity reflects the degree to which test performance corresponds to real-world everyday performance, i.e. ecological validity (Chaytor & Schmitter-Edgecombe, 2003). According to the proposed definition by the COSMIN (COnsensus-based Standards for selection of health Measurement INstruments) group, a responsive instrument should measure changes in the construct(s) it intends to measure (Vet et al., 2011); if a participant changes on the construct of interest the measurement instrument assessing the same construct should reflect this. Yet, there is no consensus on the definition and evaluation of the responsiveness of health measurements (Beaton, 2001; Husted, Cook, Farewell, & Gladman, 2000; Terwee, Dekker, Wiersinga, Prummel, & Bossuyt, 2003; Vet et al., 2011; Wright & Young, 1997). The COSMIN criteria for investigating responsiveness are similar to those for validity; hypotheses testing against a known standard or construct (anchor-based). Hypotheses testing can be done by correlating expected changes of the instrument under study to changes in instruments measuring a similar construct (criterion approach) or a (slightly) different construct or to external variables (construct approach) (Vet et al., 2011). The criterion approach requires a gold standard in order to test changes in the construct of the measure under study. In contrast, the construct approach does not require a gold standard to measure change in the construct but considers expected correlations between change scores on the (new) measure and relating instruments. A widely used but also much criticized way to analyze responsiveness is on the basis of effect sizes (Vet et al., 2011). These have been developed to express the magnitude of change after an intervention but give no information about the validity of this change regarding the construct being measured by the instrument under study. COSMIN advises the application of effect sizes as an exception only in combination with formulating a priori hypotheses about the magnitude of expected changes to the treatment in order to be able to draw conclusions about the instruments' responsiveness. Without a priori estimation of the expected treatment effect, the instruments' ability to measure change cannot be disentangled from the treatment effect, e.g. a weak effect size may either suggest a non-responsive instrument or be a result of a non-effective intervention (Vet et al., 2011).

The purpose of this study is to investigate the validity of change scores of CHEQ in children with unilateral cerebral palsy (CP) by formulating hypotheses on the expected magnitude of change and investigating correlations and effect sizes in children with unilateral CP using change measured on the Goal Attainment Scale (GAS) as an anchor for change on the CHEQ.

Methods

Participants and Data Collection

The study sample was retrieved from a project investigating intensive bimanual training during a two-week day camp in children with unilateral CP in England. Data from some children included in the project, have been analyzed for other purposes and results have been reported elsewhere (Green et al., 2013; Schertz, Karni-Visel, Tamir, Genizi, & Roth, 2016). Inclusion criteria for the analysis of responsiveness were children aged 6 to 18 years with unilateral impairment of the upper limb who participated in the intervention with required assessments at baseline and immediately following the two-week hand-arm bimanual intensive therapy program.

Included participants were children (14 girls, 30 boys) with unilateral CP between the ages of 6.8 to 16.2 years (median 9.0; IQR 8–11) (Table 2). The children were presented with MACS level I-III and GMFCS level I-III. Since children were young, most often parents completed the CHEQ apart from two adolescents who completed the questionnaire themselves.

The intervention was assumed to be successful (Green et al., 2013) and the timespan between the two measurements sufficient for the children to change on the construct of interest. Further, improvements in bimanual hand-use in children following similar intensive training have also been reported in other studies (Geerdink, Aarts, van der Burg, Steenbergen, & Geurts, 2015; Gordon et al., 2008; Gordon, Schneider, Chinnan, & Charles, 2007). Ethical approval including study extensions was granted by the NHS Research Ethics Committee (10/H0804/40/A1M01, 10/H0804/40/A1M02).

Measures

The Children's Hand-use Experience Questionnaire (CHEQ) was developed for children aged 6–18 years with unilateral hand dysfunction to evaluate their experience with the affected hand in daily life activities typically requiring use of both hands. For children below the age of 13 years, it is recommended that parents act as proxy. CHEQ is a computer-adaptive online questionnaire consisting of 29 items; available free of charge via the website (www.cheq.se). Three scales are used to measure the grasp efficacy when both hands are involved, time utilization when performing the activity compared with peers, and experience of feeling bothered while performing the activities independently. CHEQ scales are rated on a four-point ordinal scale and raw scores can be transformed by Rasch analysis to logits and further into a 0–100 CHEQ-units (Amer, Eliasson, Peny-Dahlstrand, & Hermansson, 2016; Sköld, Hermansson, Krumlinde-Sundholm, & Eliasson, 2011). Higher scores indicate a better grasp, less time taken, and greater satisfaction. Previous investigations have shown acceptable unidimensionality and high test-retest reliability (ICC 0.87–0.91) in children with unilateral CP; results from their Rasch analysis indicates a possible ability to detect change (Amer et al., 2016; Sköld et al., 2011).

The Goal Attainment Scale (GAS) is a measure to assess the extent to which individuals' goals are achieved during intervention. Goals are assumed to be reached and are formulated individually for each child, but setting goals and scoring achievement follows a standardized procedure (Kiresuk & Sherman, 1968; McMorran et al., 2016; Turner-Stokes, 2009, 2016).

GAS has shown to be a valid method for defining the content of goals which target change in gross motor function after physical therapy (Palisano, 1993), and is assumed to be more sensitive to measure changes than standardized measures (Cusick, McIntyre, Novak, Lannin, & Lowe, 2006; Palisano, 1993; Steenbeek, Gorter, Ketelaar, Galama, & Lindeman, 2011; Steenbeek, Ketelaar, Galama, & Gorter, 2007). Good reliability has been shown when used under ideal conditions, particularly when constructed by the child's own therapist (Steenbeek, Ketelaar, Lindeman, Galama, & Gorter, 2010).

Procedure

Goals for bimanual tasks were selected by the children (usually in company of their parent/s) and scaling parameters were defined by a therapist, a parent, and the child. Goal setting was applied, setting the current state to −1 in order to account for possible deterioration (−2) and the goal achievement to 0, further improvement beyond the achieved goal was specified at +1

(better than expected) and +2 (much better than expected). Goal achievement was scored on consensus of an observation by a therapist and a report from a parent and the child.

GAS was chosen as an anchor for change, because we expect CHEQ scales and GAS goals to be related on the constructs as participants were asked to formulate goals of bimanual performance and the majority of participants formulated goals similar to CHEQ items. Further, the child and parent/guardian were involved in setting and evaluating the goals together with the therapist. This enabled a combination of standardized procedures and perceptions of children and parents regarding bimanual performance, which are measured by the construct of CHEQ. A change of two scale steps is reported as being meaningful and in resemblance with goal achievement (Kiresuk & Sherman, 1968; Steenbeek et al., 2011). Therefore, a minimum of two scale steps of change in at least two out of three GAS goals was assumed to affect the construct of the CHEQ as nearly all GAS goals were directly or indirectly reflected by CHEQ activities (Table 1). Based on prior investigations of the relationship between perceived bimanual performance measured by CHEQ and the observed bimanual performance measured by the Assisting Hand Assessment (AHA) (Ryll, Bastiaenen, & Eliasson, 2017), positive relationships between the change scores of the CHEQ scales and the GAS as anchor were hypothesized to range from r = 0.3–0.5 for the CHEQ scale Grasp efficacy and Time utilization, as both have a common focus on bimanual activity performance. A slightly weaker association (r = 0.2–0.4) was expected for the CHEQ scale Feeling bothered, because we assumed the ability in specific task performance to be less related to feelings about this performance across tasks (Ryll et al., 2017).

Data Analysis

Medians with quartiles were used to summarize characteristics of participants and CHEQ scale scores. Participants were classified into categories, improved versus not improved, according to the GAS. Classification was based on a minimum change of two scale steps of at least two out of three goals per participant. In five cases only two goals were available and the third goal was imputed based on the mean of the other participants. All analyses were performed with SPSS 24 using CHEQ-unit scores. Change scores were calculated for each GAS goal and then dichotomized according to the threshold of two scale steps of change. The sum of these scores was taken across all three goals (range 0–3) in order to determine the categories improved versus not improved.

Spearman rank correlation coefficients were calculated between change scores of the CHEQ scales and the GAS (i.e. the sum of the dichotomized change across all three GAS goals) to determine change in the construct of CHEQ and to ensure the suitability of GAS as an anchor. Change scores of CHEQ scales were not normally distributed, except for CHEQ Time utilization.

Further, we calculated effect sizes for each CHEQ scale using Cohen's $SD_{pooled}\sqrt{(SD_{baseline}{}^2 + SD_{post}{}^2)/2}$ with the pooled SD in the denominator and the standardized response mean $SRM = mean_{change} / SD_{change}{}^2$ with the SD of change scores in the denominator. We hypothesized an effect size equal to or larger than 0.5 for participants that improved and an effect size smaller than 0.5 for participants that did not improve according to the GAS. Hypotheses about expected correlations between CHEQ scales and GAS were based on information about the relationship between the Assisting Hand Assessment and the CHEQ scales ranging from low to borderline high (Pearson's r = 0.28 to 0.50) (Ryll et al., 2017). Hypotheses of expected change in CHEQ scales were based on an intervention study

Table 1. Presentation of CHEQ items and all 37 different GAS goals.

CHEQ items	GAS goals represented in CHEQ ($n = 104$)
Cut meat (or other food hard to cut up) on a plate	*Cutting meat (12)*[#]
Cut up a pancake (or other food easy to cut up) on the plate	*Cutting soft food with knife and fork, using knife and fork/cutlery (21)*[#]
Butter a slice of soft bread	Butter bread (2)
Cut on a chopping board (e.g., fruit, vegetables, bread)	Cutting slices of bread, fruit on a board, or pizza (2)
Tie shoelaces	*Tying shoelaces, manage shoes (18)*[#]
Button up the trousers	Doing buttons (9)
Pull up zipper of a jacket	Zips on coat (7)
Put on socks	Putting on socks (7)
Carry a tray (e.g., in the canteen)	Carrying tray (5)
Screw off the cap of a small, unopened softdrink bottle	Opening a can, e.g., coke (3)
Open a plastic box with a lid (e.g., an ice-cream box)	Open plastic box (1)
Open a bag (for example a bag of crisps)	Opening crisps/similar packet without using teeth (6)
Spread out glue on paper using a glue stick	Spreading glue on paper (1)
Pick money out of a purse or wallet	Taking money, Picking up coins (2)
Cut out a picture using scissors	Scissors (2)
Remove a straw from a juice box and insert it	Straw (1)
Buckle a helmet (e.g., a bike-helmet)	Helmet (2)
Pull up track suit trousers	Leggings (1)
Open up a box of milk or juice	Open juice box (1)
Handle playing-cards	Playing cards (1)
Peel an orange	
Remove the wrapping from an ice-cream	
Take off the protective plastic backing of an elastoplast	
Eat out of a small container of yoghurt	
Fasten a necklace (whilst around the neck)	
Remove the wrapping from a piece of candy	
Open the zipper on a small bag (e.g., pencil case, purse)	
Open a small box (e.g., a box of mints)	
Put toothpaste on a toothbrush	

GAS goals not represented in CHEQ ($n = 24$)
Pouring (1)
Stacking cups (1)
Strumming guitar (2)
Throwing/Catching balls (3)
Holding cup (1)
Holding paper still when writing (1)
Use a ruler (1)
Riding a scooter (1)
Self-esteem with participation in a group, confidence to approach kids (2)
Scratching top of head (1)
Hold handle bar right to release left steer (1)
Improve speed of touch typing (1)
Posturing arm (1)
Lift heavy things (2)
Putting hair (2)
Managing shampoo (1)
Putting on shirt, Managing school uniform (2)

[#]most important goals for participants out of all 37 different GAS goals, GAS - Goal Attainment Scale, CHEQ - Children's Hand-use Experience Questionnaire

by Geerdink et al. investigating bimanual intensive training in combination with constraint induced movement therapy of one week duration (40 h) in 22 mild to moderately affected similarly aged children with unilateral CP (Geerdink et al., 2015). Moderate effects for the CHEQ scale Grasp efficacy (Cohen's d = 0.50) and Time utilization (d = 0.61) were shown, and a smaller effect for the Feeling bothered (d = 0.47), four months after intervention following an intervention with shorter duration compared to our study sample. We therefore, expect Cohen's d and SRMs to be larger in our analysis for the improved group.

Table 2. Characteristics of participants ($n = 44$).

	Median (*IQR*)	Frequency *n* (%)
Number of participants		44
Age in years	9.0 (8.0, 11.0)	
Gender (girls/boys)		14 (32)/ 30 (68)
Affected side right/left		23 (52)/ 21 (48)
MACS level I/II/III		7 (16)/ 22 (50)/ 15 (34)
GMFCS level I/II/III		30 (68)/ 13 (30)/ 1 (2)
Respondent child/parent or guardian		42 (96)/ 2 (4)

MACS - Manual Ability Classification System, GMFCS - Gross Motor Function Classification System

An area under the curve (AUC) larger than 0.70 (95 % CI 0.5 to 1.0) was hypothesized to indicate good discriminative ability (Terwee et al., 2007). Receiver Operating Characteristics (ROC) Curve analysis was performed to calculate the AUC using GAS as an external anchor of change in order to determine the probability of CHEQ to correctly classify participants that demonstrated improvement versus no improvement.

Results

About 104 (81.3 %) out of 128 GAS goals were covered by the majority of CHEQ items (69.0 %) (Table 1). Nearly all children ($n = 39$; 85 %) set at least two goals that were represented by CHEQ items: 22 (50 %) children formulated three corresponding goals, 17 (39 %) children two corresponding goals, and five (11 %) children only one goal that was related to CHEQ. No floor effects were observed for any of the CHEQ scales. A ceiling effect was found for the CHEQ scale Feeling bothered at post assessment (15.9 %). Median changes across CHEQ scales ranged from 6.4 to 10.0 CHEQ-units for children that improved and from 0.4 to 1.9 CHEQ-units for children that did not improve according to the GAS (Table 3).

Spearman rank correlation coefficients showed associations between GAS and all CHEQ scales ranging from 0.34–0.38 (Grasp efficacy $\rho = 0.38$, Time utilization $\rho = 0.34$ and Feeling bothered $\rho = 0.37$) and confirmed our hypotheses that changes in the construct, i.e. bimanual performance based on children's and parents' scorings (perceptions), determined by the GAS are positively and moderately related to changes in CHEQ scales. Further, all correlations exceeded the minimum acceptable threshold of 0.3 indicating a sufficient relationship of GAS to be used as an anchor (Table 5).

Effect sizes, Cohen's d and SRM, were higher in the improved group than in the group that did not improve according to GAS as external anchor, and hence confirmed all hypotheses. For

Table 3. Median change scores with interquartile ranges of CHEQ scales for the improved ($n = 34$) and non-improved group ($n = 10$).

CHEQ scale	Pre	Post	Change
Grasp efficacy			
improved	44.6 (37.0 to 49.6)	54.4 (47.2 to 63.1)	10.0 (−1.0 to 22.0)
not improved	42.0 (34.3 to 47.4)	44.7 (37.3 to 47.7)	1.9 (− 6.7 to 9.8)
Time utilization			
improved	42.2 (33.3 to 49.0)	49.1 (40.4 to 59.6)	6.7 (0.6 to 17.0)
not improved	42.0 (26.0 to 47.1)	39.7 (35.5 to 43.6)	0.5 (− 4.8 to 8.2)
Feeling bothered			
improved	53.8 (46.1 to 58.8)	58.0 (51.1 to 79.2)	6.4 (1.2 to 17.4)
not improved	53.0 (44.4 to 56.4)	56.5 (45.4 to 61.2)	0.4 (− 9.1 to 6.1)

Table 4. Mean scores of CHEQ scales, Cohen's d and SRM for improved ($n = 34$) and non-improved group ($n = 10$).

CHEQ scale	Pre, mean (SD)	Post, mean (SD)	Mean change (95 % CI)	Cohen's d (SD_{pooled})	SRM
Grasp efficacy					
improved	44.9 (9.2)	56.0 (12.4)	11.1 (6.6 to 15.5)	1.01	0.87
not improved	40.9 (8.2)	43.4 (7.7)	2.5 (− 5.8 to 10.7)	0.31	0.22
Time utilization					
improved	42.5 (9.3)	50.6 (12.4)	8.1 (3.2 to 12.9)	0.74	0.58
not improved	38.3 (10.7)	39.2 (8.9)	0.9 (− 6.0 to 7.8)	0.09	0.10
Feeling bothered					
improved	55.2 (14.4)	65.6 (19.2)	10.4 (5.7 to 15.0)	0.61	0.77
not improved	54.1 (16.6)	52.0 (12.2)	− 2.1 (− 11.7 to 7.6)	− 0.14	− 0.15

$SD_{pooled} = \sqrt{(SD_{baseline}{}^2 + SD_{post}{}^2)/2}$, SRM - Standardized response mean

the improved group, Cohen's d ranged from 0.61–1.01 and SRM from 0.58–0.87. For the non-improved group, Cohen's d ranged from −0.14–0.31 and SRM from −0.15–0.22 (Table 4).

The AUC for the CHEQ scale Feeling bothered was the only one indicating sufficient ability (AUC = 0.73, 95 % CI 0.56 to 0.91) to classify children correctly as having improved or not improved according to the GAS. However, the AUCs for the other two CHEQ scales Grasp efficacy (AUC = 0.69, 95% CI 0.52 to 0.87) and Time utilization (AUC = 0.67, 95 % CI 0.49 to 0.84) were close to 0.70 but did not exceed the threshold; thus, indicating limited accuracy to discriminate between children with and without improvement. This was supported by limited values for specificity (true-negative value; correct classification as not having improved according to GAS; 70 %) and sensitivity (true-positive value; correct classification as having improved according to GAS; 62–71 %). The ROC analysis classified 34 (77.3 %) of the total of 44 children as having improved and 10 (22.7 %) as not having improved according to the GAS as the external anchor (Table 5, Figure 1).

Discussion

This study investigated the validity of change scores of the CHEQ anchoring to the GAS as a prospective criterion of change. Results indicated that CHEQ scales are able to capture change in bimanual performance when using GAS as an external anchor, but can only distinguish with limited accuracy between children that improved and did not improve after a short-term intensive intervention.

Hypotheses about correlations between changes on GAS and CHEQ scales showed that expectations were most accurately met for Grasp efficacy, but slightly overestimated for Time utilization. For the CHEQ scale Feeling bothered the relationship was somewhat underestimated with a correlation coefficient at the upper threshold of the expected range.

Effect sizes for children that improved were larger than for those that did not improve according to the GAS, indicating longitudinal validity of change scores after a short-term intensive intervention. However, limited accuracy in correctly classifying participants that

Table 5. Area under the curve (AUC), sensitivity and specificity, Spearman rank correlation coefficient (ρ) for GAS and CHEQ scales ($n = 44$).

CHEQ scale	AUC (95 % CI)	Sensitivity	Specificity	CHEQ value (units)	Spearman's rho (ρ)
Grasp efficacy	0.69 (0.52 to 0.87)	71 %	70 %	4.4	0.38
Time utilization	0.67 (0.49 to 0.84)	62 %	70 %	5.7	0.34
Feeling bothered	0.73 (0.56 to 0.91)	65 %	70 %	2.7	0.37

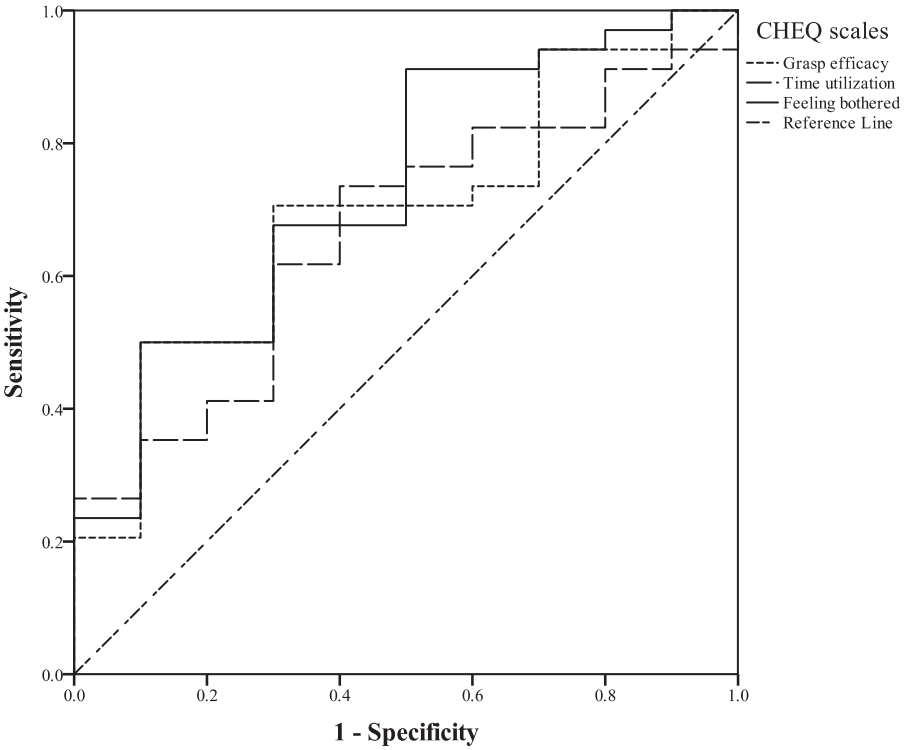

Figure 1. Receiver operating characteristic curves for change scores of CHEQ scales using the Goal Attainment Scale as external anchor for classification into improved and non-improved participants (n = 44).

demonstrated improvement versus no improvement was demonstrated by borderline AUC values for the CHEQ scales Grasp efficacy and Time utilization. The discriminant ability of the CHEQ scale Feeling bothered was higher than expected compared to the other two scales; however, the ceiling effect observed in this scale may have concealed the 'real' effect.

The choice for GAS as an anchor was supported by sufficient relationships between CHEQ scales and GAS. Hence, GAS goals formulated with a focus on bimanual performance can be linked to CHEQ items and reflect parts of the construct of the CHEQ, i.e. the aspect of bimanual performance in daily life activities and the perception and reflection on these activities from a child-parent perspective (Revicki, Hays, Cella, & Sloan, 2008; Turner et al., 2010). An advantage here was that the CHEQ most often has been completed before setting the GAS goals, which fundamentally contributed to link GAS to the construct of CHEQ. Thus, the strength of anchoring to the GAS is the similarity in perception and experience between the constructs of both instruments, i.e. the core aspect of the CHEQ; particularly as similar instruments that provide appropriate psychometric properties are not yet available. To determine change on GAS, we decided to use raw scores instead of T-scores because conceptually these are very different regarding the change they measure (Steenbeek et al., 2011). Moreover, applying GAS as a prospective anchor of change corresponding to a global rating scale seemed more appropriate and straight forward.

A challenge in investigating an instrument's ability to detect change in the construct being measured lies in the choice and psychometric properties of possible comparator instruments that may serve as an anchor. If available, information about responsiveness is often limited due to low methodological quality, inappropriate statistical methods or different populations studied, potentially reflecting the difficulties encountered when undertaking such

analysis (Gerber, Labruyère, & van Hedel, 2016; Reedman, Beagley, Sakzewski, & Boyd, 2016; Sears & Chung, 2010). A different choice for an external anchor such as the ABILHAND-Kids (Arnould, Penta, Renders, & Thonnard, 2004), a parent-reported questionnaire of children's capacity to perform daily tasks, or the Assisting Hand Assessment (AHA), a standardized observational measurement of performance of the affected hand in bimanual activities (Krumlinde-Sundholm, Holmefur, Kottorp, & Eliasson, 2007), might have led to different conclusions. However, earlier investigations showed that AHA and CHEQ also only share some parts of their constructs (Ryll et al., 2017).

Another issue may relate to limitations in the transfer of skills from the 'clinical' setting (camp) to experiencing them in every day contexts that encompass very different and varying challenges (Levac, Wishart, Missiuna, & Wright, 2009). Furthermore, as the CHEQ reflects children's feelings and experiences in daily life, this is likely to differ from the intervention setting. Likewise, parents' perspectives when acting as proxy-rater directly after intervention may be limited (Green & Wilson, 2008); especially as children may not yet have had the opportunity to demonstrate their newly acquired skills across home and community environments and may thus not be scored as having changed. This may also explain the results regarding the CHEQ scale Feeling bothered. Performing activities in the camp may not immediately change a child's perception of her/his performance until they are established in daily routines, but it may already have changed feelings towards performing these activities at the end of the training camp. Therefore, children/parents may more readily express their feelings towards the performance of activities than report perceptions related to grasping and timing as these skills may have not been implemented and observed in daily practice yet.

Regarding the interpretation of effect sizes, Cohen's thresholds (≥ 0.2 small, ≥ 0.5 moderate, ≥ 0.8 strong effect) (Cohen, 1988) are commonly applied irrespective of the formula used for calculation. These thresholds were based on Cohen's d using a pooled standard deviation in the denominator, which may over- or underestimate the magnitude of effect sizes calculated using different formulae (Middel & Van Sonderen, 2002). Cohen himself explained that these thresholds were "arbitrary, but seemed reasonable," indicating that one should also consider the importance of the effects beyond these thresholds (Cohen, 1988; Lakens, 2013; Middel, Stewart, Bouma, Sonderen, & Heuvel, 2001; Thompson, 2007). Discrepancies between Cohen's d and SRM are due to differences in the denominator, i.e. using either the pooled SD or the SD of change scores in relation to the mean change, likewise in our investigation. In our study, all corresponding effect sizes for CHEQ scales were above the same threshold.

Limitations

This analysis was performed on the former CHEQ version (which is still available and in use primarily for research and data was readily available) as the interest was to see whether the CHEQ may detect a trend for longitudinal change before a new study would be started. Change in the new CHEQ that may affect our results relate to the wording of the scale Grasp efficacy. No other changes have been made to the rating scales, thus both scales are comparable. The removal of two items is not expected to have an influence as Rasch analysis accounts for this. The main aim with the revision was to improve the response pattern of the CHEQ in order to get a more precise estimate of a person's ability, which may increase the capacity to measure longitudinal change. However, this trend found in our study needs to be confirmed for the revised CHEQ version (CHEQ 2.0) with a new data collection.

The small sample size for a validity study limited the analysis to two categories of change. Moreover, data for this analysis was collected from an intervention study, and therefore the included sample was more homogenous, which may overestimate the magnitude of effect sizes and limit the generalizability of results to children with unilateral CP. Our results need to be confirmed in a wider population using valid anchors (Revicki et al., 2006). Moreover, it is a priority to determine smallest detectable changes for CHEQ scales to relate them to the CHEQ values proposed by the ROC analysis and in the future to values of minimal important change.

Conclusions

Some evidence was shown for CHEQ scales to capture change in bimanual performance but with limited accuracy for two out of three scales. GAS can be used as an anchor to measure the construct of perception on bimanual performance when CHEQ items directly or indirectly are considered as pool for goal setting.

Acknowledgments

We would like to thank all children and their parents that contributed to this research project as well as students from Oxford Brookes University for collecting the data.

Declaration of Interest

We hereby declare that Ann-Christin Eliasson is co-developer of the Children's Hand-use Experience Questionnaire. All other authors involved in this project declare no conflict of interest.

Funding

This research was supported by funding from Stiftelsen Frimurare Barnhuset Stockholm, Guy's and St Thomas' Charity and Beit Issie Shapiro and supported by Breathe Arts Health Research.

About the Author

Ulrike C. Ryll, MSc, is a clinical epidemiologist and physiotherapist currently a PhD student at the Department of Women's and Children's Health, Karolinska Institutet, Stockholm. Her particular interests are in neuropediatric rehabilitation and health outcome measures.

Carolien H. G. Bastiaenen, PhD, is a clinical epidemiologist, physiotherapist, and Assistant Professor at the Department of Epidemiology, Maastricht University. Her expertise is in the field of health measures within the ICF framework as well as in evaluation research in the field of rehabilitation from an epidemiological perspective.

Ann-Christin Eliasson, PhD, is an occupational therapist and professor at the Department of Women's and Children's Health, Karolinska Institutet, Stockholm. Her research focus is on children with developmental disorders with her main interest in hand function in children with cerebral palsy.

Dido Green, PhD, is an occupational therapist and researcher at the Centre for Rehabilitation at Oxford Brookes University and Associate Professor at the Department of Occupational Therapy, Jönköping University. Her focus is on translational medicine and evidence based practice, emphasizing an ecological approach to understanding challenges to performance and participation in children with movement disorders.

References

Amer, A., Eliasson, A. C., Peny-Dahlstrand, M., & Hermansson, L. (2016). Validity and test-retest reliability of children's hand-use experience questionnaire in children with unilateral cerebral palsy. *Developmental Medicine & Child Neurology*, *58*(7), 743–749. doi:10.1111/dmcn.12991

Arnould, C., Penta, M., Renders, A., & Thonnard, J. L. (2004). ABILHAND-Kids: A measure of manual ability in children with cerebral palsy. *Journal of Neurology*, *63*(6), 1045–1052. doi:10.1212/01.WNL.0000138423.77640.37

Beaton, D. (2001). A taxonomy for responsiveness. *Journal of Clinical Epidemiology*, *54*(12), 1204–1217. doi:10.1016/S0895-4356(01)00407-3

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*(4), 181–197. doi:10.1023/B:NERV.0000009483.91468.fb

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cusick, A., McIntyre, S., Novak, I., Lannin, N., & Lowe, K. (2006). A comparison of goal attainment scaling and the canadian occupational performance measure for paediatric rehabilitation research. *Pediatric Rehabilitation*, *9*(2), 149–157. doi:10.1080/13638490500235581

Geerdink, Y., Aarts, P., van der Burg, J., Steenbergen, B., & Geurts, A. (2015). Intensive upper limb intervention with self-management training is feasible and promising for older children and adolescents with unilateral cerebral palsy. *Research in Developmental Disabilities*, *43*, 97–105. doi:10.1016/j.ridd.2015.06.013

Gerber, C., Labruyère, R., & van Hedel, H. (2016). Reliability and responsiveness of upper limb motor assessments for children with central neuromotor disorders a systematic review. *Neurorehabilitation*, *30*(1), 19–39. doi:10.1177/1545968315583723

Gordon, A. M., Chinnan, A., Gill, S., Petra, E., Hung, Y.-C., & Charles, J. (2008). Both constraint-induced movement therapy and bimanual training lead to improved performance of upper extremity function in children with hemiplegia. *Developmental Medicine & Child Neurology*, *50*(12), 957–958. doi:10.1111/j.1469-8749.2008.03166.x

Gordon, A. M., Schneider, J. A., Chinnan, A., & Charles, J. R. (2007). Efficacy of a hand-arm bimanual intensive therapy (HABIT) in children with hemiplegic cerebral palsy: A randomized control trial. *Developmental Medicine & Child Neurology*, *49*(11), 830–838. doi:10.1111/j.1469-8749.2007.00830.x

Green, D., Schertz, M., Gordon, A. M., Moore, A., Schejter Margalit, T., Farquharson, Y., … Fattal-Valevski, A. (2013). A multi-site study of functional outcomes following a themed approach to hand-arm bimanual intensive therapy for children with hemiplegia. *Developmental Medicine & Child Neurology*, *55*(6), 527–533. doi:10.1111/dmcn.12113

Green, D., & Wilson, B. (2008). The importance of parent and child opinion in detecting change in movement capabilities. *Canadian Journal of Occupational Therapy*, *75*(4), 208–219. doi:10.1177/000841740807500407

Hays, R. D., & Hadorn, D. (1992). Responsiveness to change: An aspect of validity, not a separate dimension. *Quality of Life Research*, *1*(1), 73–75. doi:10.1007/BF00435438

Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness. *Journal of Clinical Epidemiology*, *53*(5), 459–468. doi:10.1016/S0895-4356(99)00206-1

Kiresuk, T. J., & Sherman, R. E. (1968). Goal attainment scaling: A general method for evaluating comprehensive community mental health programs. *Community Mental Health Journal*, *4*(6), 443–453. doi:10.1007/BF01530764

Krumlinde-Sundholm, L., Holmefur, M., Kottorp, A., & Eliasson, A. C. (2007). The assisting hand assessment: Current evidence of validity, reliability, and responsiveness to change. *Developmental Medicine & Child Neurology*, *49*(4), 259–264. doi:10.1111/j.1469-8749.2007.00259.x

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. doi:10.3389/fpsyg.2013.00863

Levac, D., Wishart, L., Missiuna, C., & Wright, V. (2009). The application of motor learning strategies within functionally based interventions for children with neuromotor conditions. *Pediatric Physical Therapy*, *21*(4), 345–355. doi:10.1097/PEP.0b013e3181beb09d

McMorran, D., Robinson, L. W., Henderson, G., Herman, J., Robb, J. E., & Gaston, M. S. (2016). Using a goal attainment scale in the evaluation of outcomes in patients with diplegic cerebral palsy. *Gait & Posture*, *44*, 168–171. doi:10.1016/j.gaitpost.2015.12.003

Middel, B., Stewart, R., Bouma, J., Sonderen, E. , & Heuvel, W. J. A. (2001). How to validate clinically important change in health-related functional status. Is the magnitude of the effect size consistently related to magnitude of change as indicated by a global question rating?. *Journal of Evaluation in Clinical Practice*, *7*(4), 399–410. doi:10.1046/j.1365-2753.2001.00298.x

Middel, B., & Van Sonderen, E. (2002). Statistical significant change versus relevant or important change in (quasi) experimental design: Some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care*, *2*, 1–18. doi:10.5334/ijic.65

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., … de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745. doi:10.1016/j.jclinepi.2010.02.006

Palisano, R. J. (1993). Validity of goal attainment scaling in infants with motor delays. *Physical Therapy*, *73*(10), 651–658. doi:10.1093/ptj/73.10.651

Reedman, S. E., Beagley, S., Sakzewski, L., & Boyd, R. N. (2016). The jebsen taylor test of hand function: A pilot test–retest reliability study in typically developing children. *Physical & Occupational Therapy in Pediatrics*, *36*(3), 292–304. doi:10.3109/01942638.2015.1040576

Revicki, D. A., Cella, D., Hays, R. D., Sloan, J. A., Lenderking, W. R., & Aaronson, N. K. (2006). Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes*, *4*(1), 70. doi:10.1186/1477-7525-4-70

Revicki, D. A., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, *61*(2), 102–109. doi:10.1016/j.jclinepi.2007.03.012

Ryll, U. C., Bastiaenen, C. H. G., & Eliasson, A. C. (2017). Assisting hand assessment and children's hand-use experience questionnaire - observed versus perceived bimanual performance in children with unilateral cerebral palsy. *Physical & Occupational Therapy in Pediatrics*, *37*(2), 199–209. https://doi.org/10.1080/01942638.2016.1185498

Schertz, M., Karni-Visel, Y., Tamir, A., Genizi, J., & Roth, D. (2016). Family quality of life among families with a child who has a severe neurodevelopmental disability: Impact of family and child socio-demographic factors. *Research in Developmental Disabilities*, *53–54*, 95–106. doi:10.1016/j.ridd.2015.11.028

Sears, E. D., & Chung, K. C. (2010). Validity and responsiveness of the jebsen-taylor hand function test. *Journal of Hand Surgery*, *35*(1), 30–37. doi:10.1016/j.jhsa.2009.09.008

Sköld, A., Hermansson, L. N., Krumlinde-Sundholm, L., & Eliasson, A.-C. (2011). Development and evidence of validity for the children's hand-use experience questionnaire (CHEQ). *Developmental Medicine & Child Neurology*, *53*(5), 436–442. doi:10.1111/j.1469-8749.2010.03896.x

Steenbeek, D., Gorter, J. W., Ketelaar, M., Galama, K., & Lindeman, E. (2011). Responsiveness of goal attainment scaling in comparison to two standardized measures in outcome evaluation of children with cerebral palsy. *Clinical Rehabilitation*, *25*(12), 1128–1139. doi:10.1177/0269215511407220

Steenbeek, D., Ketelaar, M., Galama, K., & Gorter, J. W. (2007). Goal attainment scaling in paediatric rehabilitation: A critical review of the literature. *Developmental Medicine & Child Neurology*, *49*(7), 550–556. doi:10.1111/j.1469-8749.2007.00550.x

Steenbeek, D., Ketelaar, M., Lindeman, E., Galama, K., & Gorter, J. W. (2010). Interrater reliability of goal attainment scaling in rehabilitation of children with cerebral palsy. *Archives of Physical Medicine and Rehabilitation*, *91*(3), 429–435. doi:10.1016/j.apmr.2009.10.013

Streiner, D., Norman, G., & Cairney, J. (2014). *Health Measurement Scales A practical guide to their development and use* (5th ed.). Oxford, UK: Oxford University Press.

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., … de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. doi:10.1016/j.jclinepi.2006.03.012

Terwee, C. B., Dekker, F. W., Wiersinga, W. M., Prummel, M. F., & Bossuyt, P. M. M. (2003). On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. *Quality of Life Research*, *12*(4), 349–362. doi:10.1023/A:1023499322593

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools*, *44*(5), 423–432. doi:10.1002/pits.20234

Turner-Stokes, L. (2009). Goal attainment scaling (GAS) in rehabilitation: A practical guide. *Clinical Rehabilitation*, *23*(4), 362–370. doi:10.1177/0269215508101742

Turner, D., Schünemann, H. J., Griffith, L. E., Beaton, D. E., Griffiths, A. M., Critch, J. N., & Guyatt, G. H. (2010). The minimal detectable change cannot reliably replace the minimal important difference. *Journal of Clinical Epidemiology*, *63*(1), 28–36. doi:10.1016/j.jclinepi.2009.01.024

Vet, H. de, Terwee, C., Mokkink, L., & Knol, D. (2011). *Measurement in medicine: A practical guide* (1st ed.). Cambridge, Cambridge University Press.

Vroland-Nordstrand, K., Eliasson, A.-C., Jacobsson, H., Johansson, U., & Krumlinde-Sundholm, L. (2016). Can children identify and achieve goals for intervention? A randomized trial comparing two goal-setting approaches. *Developmental Medicine & Child Neurology*, *58*(6), 589–596. doi:10.1111/dmcn.12925

Wright, J. G., & Young, N. L. (1997). A comparison of different indices of responsiveness. *Journal of Clinical Epidemiology*, *50*(3), 239–246. doi:10.1016/S0895-4356(96)00373-3