Quality of Life Research (2018) 27:2165–2176
https://doi.org/10.1007/s11136-018-1873-9

Giving voice to the child perspective: psychometrics and relative precision findings for the Child Health Questionnaire self-report short form (CHQ-CF45)

Jeanne M. Landgraf[1] · Amy van Grieken[2] · Hein Raat[2]

Accepted: 27 April 2018 / Published online: 17 May 2018

Abstract

Purpose To derive and evaluate a shorter self-report Child Health Questionnaire (CHQ) legacy measure for use in research and clinical trials/care.

Methods Stepwise regression, factor analysis, and item scaling principles were used to derive and guide item selection, using data from a large general sample in the Netherlands ($n=933$). Feasibility was assessed in a school sample ($n=114$) and item internal consistency, discriminant validity, floor, and ceiling effects were evaluated using an external larger validation sample in the US ($n=1468$). Reliabilities were estimated using Cronbach's alpha. Relative precision (RP), the ability to distinguish between clinical subgroups, was computed by comparing the proportion of variance explained by the short-form scales vs. respective full-length scales.

Results The CHQ-CF was reduced from 87 to 45 items. The median alpha coefficient was 0.89. Ninety-seven to 100% scaling successes for item discriminant validity were observed. Floor effects were not observed; some ceiling effects were detected. RP estimates ranged from 0.73 to 1.37.

Conclusion The CHQ-SF45 is reliable and valid and exceeds item level scaling criteria.

Keywords Psychometrics · Children · Adolescents · Health-related Quality of life · PRO · Child-centered

Introduction

The early mid-1990s heralded an important shift in the national narrative about health-related outcomes as the child-centered perspective became increasingly integrated into the translational research agenda. Fundamental to this burgeoning perspective was the availability of conceptually sound and well-validated child/adolescent specific measures [1–5]. Historically, the parent was often selected as the child's "voice"; however, developers advocated that whenever feasible health outcomes should be elicited directly from the child [6–9].

Current research further promotes this premise by examining the specific age at which children can understand health and illness concepts and by identifying essential methods/technologies for doing so [10–13]. These efforts are fundamental to understanding the impact and potential trajectory of health beliefs across the life course and continue to inform enhancements to "legacy" measures that utilize classical test theory and newly released measures that utilize modern test theory such as the NIH-funded family relationships measures [14] and pediatric outcomes measures used in federal initiatives such as PEPR [15] and ECHO [16].

The Child Health Questionnaire (CHQ), a family of parent and child-reported general health outcomes measures, [3, 6, 7] now considered "legacy" tools, continues to be an important touchstone in furthering the advancement of PRO measurement among children and adolescents. The CHQ was constructed using classical test theory as its developmental cornerstone, although item response theory was also employed (e.g., Rasch analysis for the Physical Functioning Scale). The development of an experimental CHQ child-report short form was explored at the time the

✉ Jeanne M. Landgraf
jml@healthactchq.com

[1] HealthActCHQ, 800 Boylston Street, 16th Floor, Boston, MA 02199, USA

[2] Department of Public Health, Erasmus MC, University Medical Center, Rotterdam, The Netherlands

parent-reported CHQ was released [17]; however, it was determined that further work was warranted. To pre-empt a misleading message that the parent was the most reliable reporter for children, and to ensure that the child's own perspective was included, the full-length 87 item self-report version (CHQ-CF87) was released with a long-term objective to develop a more practical short form [17].

Despite its length (87 items), which has been cited as a potential barrier, there are more than 100 peer-reviewed publications representing several dozen countries that provide evidence of the reliability and validity of the measure across an array of conditions and settings including randomized clinical trials, clinical practice, schools, academic research, and epidemiologic studies [18]. The CHQ-CF87 has been rigorously translated and linguistically validated [19] in 34 languages. In direct response to demand, work was initiated in 2015 to derive a shorter CHQ self-report. In early 2017, CHQ-CF87 U.S.-based norms were collected and data were used to evaluate the short form. The purpose of this paper is threefold: (1) report on the analytic methods used to derive the short form (CHQ-CF45); (2) detail feasibility findings from a school-based initiative; and (3) provide evidence of the measure's reliability and validity. To further support use and data interpretation, two complementary initiatives—the estimation of U.S.-based norms and linguistic validation [19] of 34 translations—were completed as part of the CHQ-CF45 comprehensive scientific agenda.

Sample used to derive the short form

A Dutch dataset was used to empirically derive the short-form CHQ-CF45. Reliability and validity of the Dutch translation has been well-established [20–24]. This dataset was selected for many compelling reasons: (1) it provided an international perspective of previously published work [25]; (2) it included randomization of both paper–pencil and Internet administration; (3) the psychometric properties of the CHQ-CF87 were equal between the modes of administration which has been identified as a key equivalence guideline [26]; (4) the sample was sufficiently large ($N = 933$); (5) the sample was representative (87% participation rate across seven secondary schools); (6) the dataset included those with/without health conditions; and (7) feedback about the feasibility/acceptance of the CHQ-CF87 was collected.

The Medical Ethics Committee of the Erasmus University Rotterdam Medical Center approved the study protocol (reference number MEC-2007-163). Table 1 provides a brief demographic summary of the sample.

The age range was 13–17 years (mean age 15). Fifty-four percent of the sample were girls and the majority were of Dutch ethnicity (76%). Fifty-eight percent of the sample were enrolled in lower secondary/vocational education; 22% upper secondary; and 19% intermediate secondary. The CHQ-CF87 performed well in the Dutch sample. The median Cronbach's alpha coefficient was 0.83 (paper/pencil; range 0.69 physical functioning to 0.92 mental health) and 0.81 (internet; range 0.70 role/social-emotional/behavioral to 0.90 mental health). All scales demonstrated discriminant validity between adolescents with/without health conditions [25].

Table 1 Characteristics of the sample used to derive the CHQ-45 ($n = 933$)

	Total group ($n = 933$)	Internet ($n = 458$)	Paper–pencil ($n = 475$)
Mean age years (range)	15 (13–17)	15 (15–17)	15 (13–17)
Gender female (%)	54	56	51
Ethnic background %			
Dutch	76	76	76
Turkish	6	5	8
Moroccan	2	3	2
Surinamese	2	2	2
Antillean/Arubean	.4	.4	.4
Other	12	12	11
Secondary school type			
Lower secondary/vocational	58	59	58
Intermediate secondary	19	19	20
Upper secondary	22	22	22

No statistically significant differences between groups ($p \geq .05$). Some percentages may not add to 100 due to rounding

Analytic methods used to derive the short form

A multimethod approach was used to derive the short form: (1) stepwise regression; (2) exploratory factor analysis; (3) confirmatory multi-trait analysis; and (4) alpha reduction. The overall objective was to reduce each scale by 50%; but only if such a premise was empirically supported.

The multimethod approach was applied to the full sample as well as 15 unique subgroups defined by variables in the dataset: (1) mode of administration—paper/pencil or Internet; (2) child gender—boys/girls; (3) age group—13–14,15,16–17 (groupings based on data distribution); (4) smoking status—current/former smoker/never; (5) diagnostic Status—none, 1–2 conditions, ≥ 3 conditions; and (6) problems with CHQ Completion—no, yes.

The conceptual framework for the CHQ-CF87 is based on the WHO multi-dimensional model of health [27]. Scale–scale correlations (not shown but estimated as part of the short-form development) provided empirical support that each scale represented a unique construct. Thus, to maintain this conceptual/definitional structure, all scales/concepts were retained. Decision rules were comparable to those employed for previous published short-form work [28, 29].

Using the Likert-type format and a graduated response continuum that ranges from 4 to 6 levels, the CHQ measures global health; limitations with regard to daily physical activities; limitations in school and social activities due to emotional/behavioral issues and physical health issues, bodily pain, behavior, self-esteem, general health perceptions; limitations in family activities; and family cohesion. A 4-week recall is used to capture a more "reasonably stable" assessment of the average [30] rather than an acute (7-day) perspective common to more newly developed IRT measures [14].

Items are reverse scored (if appropriate) so that a higher score is better and a mean score is calculated to derive an overall scale score as is the standard for ordered categorical or Likert-type scales. For ease of interpretation, the raw scale scores are converted to a 0–100 continuum using a standard mathematical formula [31].

For each scale, items that had consistent results across all analytic methods for the full sample and 15 subgroups were retained. For the regression analysis, an adjusted R^2 between 0.80 and 0.90 was considered acceptable. The dependent variable is the full-length scale; the corresponding items for that scale are the independent variables. Each model was analyzed separately and item selection was noted where little gain was observed if the item was retained. For the item scaling analyses, items had to have an observed loading of ≥ 0.40 with their respective scale [32]. Item reliability thresholds, as measured by Cronbach's alpha [33], were set at ≥ 0.80.

Table 2 identifies the concepts, the number of items for each scale in both lengths of the CHQ, and the item content that was omitted/retained in the shorter version.

Samples used to evaluate the short form

Feasibility/ease of completion/acceptance

Since the resulting CHQ-CF45 was derived from, and is embedded within the full-length CHQ-CF87, a small unfunded feasibility study to assess the short version as an independent measure was conducted at a private school in the western part of the U.S. Table 3 provides a brief description of the convenience sample.

The school was selected because of accessibility and consistency in class size and range. The CHQ could be self-administered over a few days within the usual school flow. The small class size (approximately 13/grade; range 12–19) and the grade range (1–8) afforded a compelling opportunity to evaluate acceptability through direct observation and open-ended discussions with participants. The same researcher (JL) introduced the CHQ, oversaw its self-completion, and was available to assist if issues arose. Understanding and ease of CHQ completion were obtained from each participant. Teachers were not present during the field-testing to allow for greater anonymity and freedom with regard to open-ended discussions about health, well-being, and the CHQ items/layout/format. The project was approved by the Executive Director, Director, and classroom teachers. Parent consent and child assent was obtained. One child declined participation although the parent consented. The field test was performed in April 2014.

The school utilizes a modular learning approach, i.e., some students within each grade switch to separate classrooms for particular subjects/activities. This was an important consideration in the school's selection as it allowed for further protection of student confidentiality for those whom parents did not provide consent. A priori arrangements were made with each teacher so that non-participants could engage in a different activity in a separate room without calling attention to their absence during the CHQ administration.

Table 4 provides feasibility details. Invitations were sent to 153 parents with 114 providing consent (74% response rate). The average completion time was 11 min (range 4–35 min; high number reflects 1–3 outliers). There were negligible missing data (range 0.02–0.05%). In general, students did not have problems with CHQ content and did not feel items were difficult to complete.

Table 2 Concepts, number of items, and items omitted/retained in the CHQ-CF45

Concept/Scale Name	Number of items in CHQ-CF87	Number of items in CHQ-CF45	Items content omitted	Item content retained
Global Health item	1	1	None	My health is
Physical Functioning	9	6	Walk/climb 1 flight, eat dress bathe, get in/out of bed	Strenuous activities, moderate activities, walk blocks/climb stairs, get around school, do usual tasks, bend/lift/stoop
Role/social limitations—Emotional/Behavioral[a]	6	2	Do certain kinds of schoolwork/activities	Spend usual time on work or activities, accomplish schoolwork or activities
Role/social limitations—Physical (k = 2)	3	2	Do certain kinds of schoolwork/activities	Spend usual time on work or activities accomplish schoolwork or activities
Pain	2	2	None	Amount and frequency of pain
Getting Along/Behavior	16	6	Being alone, hard time others liking you, felt clumsy, ran away, speech problems, stole things, hard to be with others, hard time getting along	Immature, argued, inattentive, disobedient, lied/cheated, hot tempered
Global Behavior item	1	1	None	My behavior is
Mental Health	16	9	Felt like crying, lonely, nervous, bothered or upset, have fun, jittery or restless, like yourself	Sad, afraid/scared, worry, unhappy, happy, cheerful, enjoy life, trouble sleeping, headaches
Self-Esteem	14	7	Yourself, schoolwork, sports, way you most often feel, way you get along with family, way life seems to be, health in general	Friendships, things you can do, get along with others, body/looks, be a friend to others, others feel about you, talk with others
General Health	11	4	Health is excellent, sickly, do not seem to get sick, never very sick, always sick, never worry, worry more	Less healthy than others, less healthy in the future, healthy in the future, healthy now
Change in Health item	1	1	None	
Family Activities	6	3	Caused tension/conflict, source of disagreements, caused family to cancel/change plans	Limited activities, interrupted activities, limited spontaneous activities
Family Cohesion item	1	1	None	

[a]The role/social limitations-emotional and behavioral scales are combined in the CHQ-CF45

Table 3 Feasibility sample descriptors of participants ($n=114$)

Grade	Response rate CHQ'S completed/class size (%)	Gender		Health condition[a]	
		Male	Female	No	Yes
2nd	14/16 (88%)	6	8	12	2
3rd[b]	23/29 (79%)	12	11	9	14
4th	13/17 (76%)	6	7	6	7
5th[c]	25/35 (71%)	17	8	15	10
6th	11/14 (79%)	8	3	8	3
7th[d]	17/25 (68%)	8	9	9	8
8th	11/17 (65%)	7	4	5	6
Totals	114/153 (75%)	64	50	64	50

[a]Health condition: parent-reported. Full sample by gender: No male = 31, female = 33, yes male = 33, female = 17

[b]Due to the number of students in this grade, at the start of the term, the class was split into two separate rooms/teachers ($n=12/14$, $n=11/15$. Thus, response rates were 86 and 73%, respectively)

[c]Due to the number of students in this grade, at the start of the term, the class was split into two separate rooms/teachers ($n=15/19$, $n=10/16$. Thus, response rates were 79 and 62%, respectively)

[d]Due to the number of students in this grade, at the start of the term, the class was split into two separate rooms/teachers ($n=8/12$, $n=9/13$. Thus, response rates were 67 and 69%, respectively)

Table 4 Completion times and ease of completion by grade ($n=114$)

Grade	Average number of minutes to complete (range)	Difficult to complete (n)[a,b]		Difficult to understand (n)[a,b]	
		No	Yes	No	Yes
2nd	17 (12–30)[c]	8/11	3/11	9/11	2/11
3rd	17 (6–35)[c]	17/22	5/22	9/22	13/22
4th	11 (5–19)	8/13	5/13	5/13	8/13
5th	9 (4–14)	19/23	4/23	17/21	4/21
6th	9 (6–12)	11/11	0/11	11/11	0/11
7th	7 (5–15)	16/16	0/16	15/16	1/16
8th	8 (6–9)	10/11	1/11	10/11	1/11

[a]Indicates the number of students in this grade who answered the two open-ended questions regarding CHQ difficulty/ease of understanding. The number of students who completed the CHQ for each grade is provided in Table 3

[b]Those who responded "yes" to Q1 or Q2 indicated that "only a little bit" or "some" items were more difficult to complete/understand than others. Probing revealed that in general, ≤ 3 items were problematic

[c]Only 1 participant in this grade required the lengthier time as noted to complete the CHQ

There were some initial issues with layout in Grade 2 but once the format was explained these younger children had no problems completing the CHQ. Grades 3 and 4 students reported that the CHQ was not difficult to complete. However, they did seem to have more challenges than other grades but when items were clarified (e.g., stoop, tasks, body/looks, family) they were able to answer. It was later disclosed that learning and comprehension challenges were not unique to the CHQ for these grades. Interestingly, a 3rd grader commented, "It was hard not because I did not understand but because it was tough deciding how I really felt." A 5th grader noted, "I am going through a few things so I would prefer not to answer some questions." A 4th grader commented, "This is not what I expected. I thought you'd just ask about the usual stuff but it was not just about that, but my body, my family. That's important." Others commented "this was fun and easy" (Grade 4), "it's cool" (Grade 3), "I love this. It's fun and makes me feel like a grown-up (Grade 2)." Grades 7 and 8 felt it complemented what they were learning in health/science. One of the most compelling comments came from a 3rd grader, who emphatically voiced, "It's about time people started to ask US how we feel instead of our parents. We do know about these things, you know!."

Internal consistency reliability for the nine CHQ-CF45 multi-item scales was evaluated using Cronbach's alpha coefficient [33]. Due to small class sizes, coefficients were estimated for the entire sample ($n=114$). Change in Health and Family Cohesion was retained as independent single items in the CHQ-CF45 but test/retest was not possible; therefore, they were not included in the estimations. Table 5 provides a summary for the CHQ scales. All coefficients were ≥ 0.71; the median coefficient was 0.82 (range 0.71–0.87; Role/social-Physical, and Mental Health,

Table 5 CHQ-CF45 alpha coefficients for the feasibility sample ($n=114$)

Scale name	Alpha coefficient
Physical Functioning (k = 6)[a]	0.81
Role/social limitations—Emotional/Behavioral (k = 2)	0.85
Role/social limitations—Physical (k = 2)	0.71
Bodily Pain (k = 2)	0.82
Getting along/Behavior (k = 7)	0.80
Mental Health (k = 9)	0.87
Self-Esteem (k = 7)	0.85
General Health (k = 5)	0.80
Family Activities (k = 3)	0.86

Participants had to complete all items within the respective scale to be included in the analysis. This eliminated nine subjects for BE and SE only ($n=105$)

[a]k number of items in the scale. Global health item and global behavior item are included in their respective scales (General Health and Getting Alon/Behavior). Change in Health and Family Cohesion are single items. Test–retest not performed so internal consistency reliability not analyzed

respectively) suggesting that the CHQ performed well in the feasibility sample.

National U.S. sample

The CHQ-CF45 was validated in a 3rd independent project using a U.S.-based representative sample [34, 35]. Responses from children aged 8–18 years were collected between November 2015–January 2016 using an online research panel (GfK KnowledgePanel) representative of the U.S. population. All eligible respondents completed the CHQ-CF87.

The GfK recruitment protocol relies on probability-based sampling of all U.S. addresses from the latest Delivery Sequence File (DSF) of the U.S. Postal Service and a stratification plan to boost representation of subgroups with higher rates of attrition (e.g., Hispanic/Non-Hispanic, ages 18–29) and other hard to reach adults (e.g., no Internet). To further improve the demographic composition, GfK makes adjustments to account for oversampling and every quarter, invites adults from randomly selected addresses to join the panel. Thus, the parent sample, from which children/teens were recruited, generally represents an equal probability selection method sample [34]. The merits/disadvantages of probability-based and convenience Internet panels have been addressed by others [36]. Both parental consent and child assent was obtained.

Table 6 provides a summary of the demographic characteristics of the U.S. representative sample. The sample ($N=1500$) comprised an equal number of males/females ($N=750$) overall and within each of the targeted age groups which were created based on development parameters and distribution within the GfK panel (8–9, 10–11, 12–13, 14–15, 16–18; $n=150$, respectively). The mean sample age was 13 years. The least represented age was 18 ($N=73$). The majority of the sample attended public school, with a little more than 8% attending private schools and almost 6% ($N=88$) being home-schooled.

More than half of the sample was White Non-Hispanic (59%), 21% were Hispanic with the remaining identifying as Black, Other non-Hispanic, and 2+ Races, non-Hispanic. Regional representation was varied (17–36%) with slightly more representation from the South and Midwest U.S. An overwhelming majority (85%) lived in attached/detached one-family homes with married working parents (80; 73%, respectively—data not shown).

Few child conditions (e.g., asthma, ADHD) were reported by parents (mean 1.5; range 0–12; not shown). Table 7 presents the conditions with/without comorbidities. Due to small samples within discrete conditions, even with comorbidities, children were re-classified for analytic purposes as having no condition (51%), one condition (24%), two conditions (12%), or ≥ 3 conditions (12%).

Table 6 Background descriptors for the validation sample ($N=1500$)

Characteristic	Mean	Range	*N*	%[a]
Gender				
Male	–		750	50
Female	–		750	50
Age[b]	12.57	8–18	–	–
Grade	8.17	k-College	–	–
School type	–	–		
Public			1255	84
Private			129	9
Home-schooled			88	6
Vocational/technical			5	0.3
Not in school			19	1
Not answered/missing			4	0.3
Ethnicity	–	–		
White, Non-Hispanic			888	59
Black, Non-Hispanic			132	9
Other, Non-Hispanic			70	5
Hispanic			321	21
2+ Races, Non-Hispanic			89	6
Region	–	–		
Northeast			257	17
Midwest			390	26
South			541	36
West			312	21
Type of housing	–	–		
Detached one-family house			1174	78
Attached one-family house			101	7
Building with ≤ 2 apartments			169	11
Mobile home			54	4
Boat, RV, van, etc			2	0.1

Reproduced with permission [35]

– Not applicable

[a]Percentages may not equal 100 due to rounding

[b]There were 300 (150 boys, 150 girls) in the following age categories 8–9, 10–11, 12–13, 14–15, 16–18

Methods used to evaluate the short form

Multi-item scaling

Scaling criteria and confirmatory factor analysis [32] were used to evaluate the psychometrics and the a priori hypothesized structure of the CHQ-CF45. A subtype of structural equation modeling, MAPR is recognized as a hallmark classical test methodology for evaluating legacy tools that utilize a multi-item Likert-type approach to scale construction. Unlike traditional factor analysis, MapR allows for the direct test of a priori structures (i.e., item "fit") and can be used to estimate "error free" constructs. MAPR extends traditional tests of internal consistency by testing item discrimination

Table 7 Prevalence of parent-reported conditions and comorbidities in the CHQ-CF45 U.S. normative sample: ages 8–18 ($N = 1500$)

Condition*	*N*	%**
Acute illness		
Yes	115	8
Anxiety only		
Yes	19	1
With another condition	147	10
Asthma only		
Yes	73	5
With another condition	219	15
Attention/executive functioning		
Yes	32	2
With another condition	158	11
Autism only		
Yes	8	.5
With another condition	61	4
Behavioral problems only		
Yes	6	.4
With another condition	100	7
Cardiovascular with another condition	9	.6
Cerebral Palsy with another condition	1	.1
Chronic allergies/sinus only		
Yes	6	.4
With another condition	169	12
Chronic bone/joint with another condition	6	.4
Chronic respiratory (not asthma) with another condition	7	.5
Depression only		
Yes	2	.1
With another condition	56	4
Developmental delay/learning problems only		
Yes	7	.5
With another condition	107	7
Diabetes with another condition	11	.7
Epilepsy with another condition	12	.8
Hearing impairment/deafness with another condition	17	1
Sleep issues only		
Yes	3	.2
With another condition	58	4
Speech problems		
Yes	13	.9
With another condition	74	5
Vision problems		
Yes	85	6
With another condition	228	16
Other		
Yes	23	2
With another condition	69	5

Reproduced with permisison [35]

*Exclusivity calculated and presented only for conditions with comorbidities $N > 50$

**% based on actual number of responses for that diagnosis/condition. Missing ranged from $N = 8$ (Acute) to $N = 41$ (Depression) and the median was $N = 32$ ("Other "category not counted in the "missing" calculation since this item did not warrant a response from all respondents)

across other item sets in the matrix (i.e., how well an item represents a given construct relative to other constructs). Correlations between items are corrected for indicator overlap so that estimates are not spurious [30, 32].

Specific examination using MAPR included (1) item convergent validity: correlations between 0.30 and 0.40 for items within their hypothesized scale; (2) item discriminant validity: correlations one to two standard errors higher than correlations with other scales; (3) floor and ceiling effects; percentage of respondents at the lowest and highest possible scale score; (4) internal consistency reliability as measured by Cronbach's alpha coefficient [33, 37]. Strict analytic selection criteria were used; completion of at least half of the items for all nine scales was required. The final sample for analysis was 1468.

Relative precision in clinical subgroups

The validity of the short and full-length scales to detect differences among clinical subgroups was estimated by computing the ratio of pair-wise F-statistics [28, 29, 35, 38, 39]. The F-statistic represents the ratio of systematic variance (between groups) versus error variance (within group). The larger the F value, the better the concept distinguishes across the different classifications. Following published convention, the full-length CHQ-CF87 scales were considered the "standard" (i.e., denominator was set to 1.0) against which the shorter scales were compared. Precision estimates greater than 1.0 indicate that the short-form scale is superior in detecting differences relative to the corresponding full-length version. Estimates under 1.0 provide evidence, that although respondent burden is reduced, some loss of precision is observed. Parametric methods are routinely used with summated rating scales, however, for the precision assessment, the Kruskal–Wallis non-parametric test was also performed. No differences were found. Thus, per convention, data are reported using parametric methods (ANOVA). All analyses were performed using SPSS Version 23.

Results

Short-form scale evaluation

Multi-trait item scaling results are summarized in Table 8 for the full sample, and by gender, age group, and condition group (none, 1, 2, or ≥ 3 conditions). For ease of data presentation, item convergent and discriminant validity findings are summarized in the row "Scaling Success." In the full sample and across all 10 subgroups, all item-scale correlations (not shown) were ≥ 0.40 for all items with the exception of 3 items—global health (General Health Scale) (0.34) in the 8–9 year-old group and headache (Mental Health) in

Table 8 Item scaling success rates, floor/ceiling effects, and alpha coefficients for the CHQ-SF45 Scales

	All ($n=1468$)	Gender		Age				Number of health conditions[a]			
		M ($n=736$)	F ($n=732$)	8–9 ($n=295$)	10–12 ($n=429$)	13–15 ($n=452$)	16–18 ($n=292$)	0 ($n=751$)	1 ($n=362$)	2 ($n=182$)	≥3 ($n=172$)
PF[b] (k = 6)											
% Scaling success[c]	100	100	100	100	100	100	100	100	100	100	100
% Floor/ceiling effects[d]	0.2/71	0.3/71	0.1/70	0.3/71	0.0/70	0.4/66	0.0/77	0.0/79	0.3/70	0.0/60	1/44
Alpha coefficient[e]	0.86	0.86	0.86	0.85	0.86	0.86	0.87	0.83	0.87	0.83	0.87
REB[a] (k = 2)											
% Scaling success	100	100	100	100	100	100	100	100	100	100	100
% Floor/ceiling effects	0.5/84	0.8/84	0.3/85	1/84	0.5/86	0.0/83	1/85	0.1/91	0.3/85	0.0/82	4/56
Alpha coefficient	0.89	0.88	0.89	0.85	0.89	0.85	0.95	0.85	0.91	0.84	0.92
RP[a] (k = 2)											
% Scaling success	100	100	100	100	100	100	100	100	100	100	100
% Floor/ceiling effects	0.6/91	0.7/91	0.5/92	0.3/91	0.5/93	0.4/89	1/92	0.3/94	0.0/92	0.5/89	3/79
Alpha coefficient	0.92	0.90	0.94	0.89	0.88	0.93	0.95	0.89	0.94	0.82	0.98
BP[a] (k = 2)											
% Scaling success	100	100	100	100	100	100	100	100	100	100	100
% Floor/ceiling effects	0.3/44	0.3/45	0.3/43	0.7/51	0.2/48	0.0/35	0.3/44	0.1/53	0.3/41	0.0/30	1/23
Alpha coefficient	0.89	0.88	0.91	0.89	0.87	0.88	0.93	0.88	0.88	0.91	0.89
BE[a] (k = 7)											
% Scaling success	100	100	100	100	100	100	100	100	100	100	100
% Floor/ceiling effects	0.2/5	0.0/4	0.4/6	0.0/4	0.2/4	0.2/6	0.3/8	0.1/8	0.0/3	0.5/1	0.6/2
Alpha coefficient	0.87	0.87	0.86	0.85	0.86	0.88	0.86	0.86	0.81	0.85	0.87
MH[a] (k = 9)											
% Scaling success	100	97	100	99	99	100	100	99	100	99	99
% Floor/ceiling effects	0.0/4	0.0/3	0.0/5	0.0/6	0.0/5	0.0/4	0.0/2	0.0/7	0.0/3	0.0/2	0.0/0.6
Alpha coefficient	0.84	0.79	0.87	0.81	0.79	0.86	0.86	0.80	0.84	0.84	0.84
SE[a] (k = 7)											
% Scaling success	100	100	100	100	100	100	100	100	100	100	100
% Floor/ceiling effects	0.3/26	0.4/24	0.1/27	1/31	0.2/28	0.0/22	0.0/23	0.3/34	0.0/20	0.0/17	1/11
Alpha coefficient	0.93	0.93	0.92	0.92	0.93	0.92	0.94	0.92	0.92	0.92	0.93
GH[a] (k = 5)											
% Scaling success	100	100	100	93	100	100	100	100	100	100	100
% Floor/ceiling effects	0.0/16	0.0/15	0.0/17	0.0/20	0.0/18	0.0/14	0.0/11	0.0/22	0.0/12	0.0/9	0.0/4
Alpha coefficient	0.76	0.78	0.75	0.67	0.77	0.79	0.78	0.73	0.75	0.75	0.76
FA[a] (k = 3)											
% Scaling success	100	100	100	100	100	100	100	100	100	100	100

Table 8 (continued)

	All ($n=1468$)	Gender		Age				Number of health conditions[a]			
		M ($n=736$)	F ($n=732$)	8–9 ($n=295$)	10–12 ($n=429$)	13–15 ($n=452$)	16–18 ($n=292$)	0 ($n=751$)	1 ($n=362$)	2 ($n=182$)	≥ 3 ($n=172$)
% Floor/ceiling effects	0.5/55	0.5/55	0.5/55	1/52	0.5/57	0.4/51	0.3/62	0.4/66	0.0/54	0.0/45	3/20
Alpha coefficient	0.90	0.90	0.91	0.90	0.90	0.90	0.93	0.92	0.88	0.88	0.87

[a]Reported independently by the child's parent prior to the child self-completion of the CHQ

[b]*k* number of items. Scale abbreviations: *PF* Physical Functioning, *REB* role/social limitations-emotional/behavioral, *RP* role/social-physical limitations, *BP* bodily pain, *BE* getting along/behavior, *MH* general well-being/mental health, *SE* self-esteem, *GH* general health perceptions, *FA* family activities

[c]Item discriminant validity. Correlations one to two standard errors higher than correlations with other scales

[d]Percentage of respondents at the lowest and highest possible scale score, respectively

[e]Cronbach's Alpha. Internal consistency reliability. All reliabilities from MapR except REB and RP in DX2 and RP in the full sample which were uninterpretable and therefore estimated in SPSS

the same age group and those with 1 or 2 reported conditions (0.39 and 0.37, respectively) and trouble sleeping (Mental Health) in 8–9-year-olds. 100% item scaling success rates were observed for all nine scales across the full sample and 10 subgroups with the exception of Mental Health, which although not perfect, were still quite high (range 97%–99%). Given the reported good health of the sample, minimal floor effects were observed indicating that few if any respondents scored at the lowest possible end of the response continuums across all items (range 0.0 Mental Health and General Health to 0.6% Role Physical). However, not surprisingly, ceiling effects—the percentage of respondents scoring at the highest/best end of the response options—were detected. High ceiling effects included the Role Physical, Role Emotional/Behavioral and Physical Functioning scales (91, 84, and 71%, respectively). Moderate ceiling effects were observed for the Family Activities (55%), Bodily Pain (44%), and Self-Esteem (26%) scales. Negligible/no ceiling effects were observed for the Behavior and General Health Perceptions scales. Ceiling effects were much smaller for those in the condition groups, especially those with ≥ 3. Overall, findings indicate that the CHQ-CF45 short-form scales exceeded item level scaling criteria even for the youngest ages (8/9 year-olds).

Table 8 also presents alpha coefficients for the full sample and the subgroups. In the full sample, with the exception of General Health (0.76), alpha coefficients for the remaining eight scales were ≥ 0.84 with the highest reliabilities observed for Self-Esteem (SE; 0.93) and Family Activities (FA; 0.90) followed by bodily Pain (0.89) and Behavior (0.87) and Physical Functioning (0.86). Similar alpha coefficients were observed across all 10 subgroups. These findings suggest that the CHQ-CF45 is reliable even in the younger age groups and across children with/without reported health conditions. The reliability estimates are remarkable considering that several of the CHQ scales comprised only 2 or 3 items (REB, RP, BP, FA) with the remaining scales defined by ≤ 9 items.

Relative precision

Table 9 provides group means, *p* values, f-statistics, and relative precision (RP) estimates. As noted, the groups were created by classifying children based on the number of parent-reported conditions (0, 1, 2, ≥ 3). The two item Bodily Pain scale, and the two single items (Change in Health, and Family Cohesion) were excluded from the RP estimates since they are identical in both lengths of the CHQ.

Statistically significant differences ($p = .000$) were observed for all scales in the full-length CHQ-CF87 and the derived short-form CHQ-CF45. Overall, lower scores were observed for the more "psychosocial" scales (e.g., Mental Health, Behavior, Self-Esteem) relative to Physical

Table 9 Relative precision of the CHQ-CF45 for children with and without reported health conditions

Scales	Number of health conditions				*p* value	F-statistic	Relative precision
	0 ($n=737$)[a] Mean (SD)	1 ($n=358$) Mean (SD)	2 ($n=176$) Mean (SD)	≥3 ($n=167$) Mean (SD)			
PF[b] (k=9)	97.37 (7.81)	95.68 (11.34)	94.20 (11.30)	88.50 (17.32)	.000	32.43	1.00
PF[b] (k=6)	96.64 (9.26)	94.77 (12.53)	92.59 (13.42)	86.03 (19.80)	.000	35.48	1.09
RE[b] (k=3)	94.42 (14.57)	91.00 (19.06)	90.34 (17.62)	75.85 (30.42)	.000	45.51	1.00
RB[b] (k=3)	96.77 (11.24)	94.23 (14.94)	94.00 (13.59)	78.51 (29.60)	.000	62.20	1.00
REB[b] (k=2)	95.70 (11.71)	92.50 (16.00)	92.19 (14.40)	77.44 (26.84)	.000	62.36	1.37 vs. RE 1.00 vs. RB
RP[b] (k=3)	97.83 (10.39)	96.71 (11.80)	96.15 (11.81)	89.55 (23.48)	.000	18.28	1.00
RP[b] (k=2)	97.74 (10.68)	96.79 (12.34)	96.59 (11.73)	89.52 (23.75)	.000	17.33	0.95
BE[b] (k=17)	83.38 (13.25)	79.77 (13.37)	75.86 (15.75)	65.11 (17.85)	.000	79.38	1.00
BE[b] (k=7)	76.46 (16.78)	71.68 (16.17)	67.62 (18.91)	56.56 (21.73)	.000	62.68	0.79
MH[b] (k=16)	81.98 (12.69)	79.20 (13.70)	75.18 (15.34)	66.82 (17.19)	.000	58.52	1.00
MH[b] (k=9)	81.18 (12.71)	78.33 (13.54)	74.34 (15.20)	66.56 (16.92)	.000	55.63	0.95
SE[b] (k=14)	85.41 (15.22)	80.77 (17.09)	77.29 (16.78)	67.89 (22.24)	.000	53.47	1.00
SE[b] (k=7)	85.85 (15.92)	81.02 (18.17)	78.48 (18.36)	68.26 (23.39)	.000	47.20	0.88
GH[b] (k=12)	79.67 (13.46)	74.19 (14.56)	72.76 (15.87)	63.42 (17.21)	.000	61.83	1.00
GH[b] (k=5)	83.23 (15.00)	78.16 (16.40)	77.76 (15.76)	67.86 (18.57)	.000	44.96	0.73
FA[b] (k=6)	89.28 (17.43)	85.94 (17.64)	81.30 (21.37)	65.72 (25.16)	.000	71.92	1.00
FA[b] (k=3)	89.55 (18.90)	86.69 (18.85)	82.58 (21.90)	66.82 (26.64)	.000	58.48	0.81

[a]Sample held constant. Respondents had to have completed half the items for each scale in order to be included

[b]*k* number of items in the CF87 and CF45 multi-item scales, respectively. Scale abbreviations: *PF* Physical Functioning, *RE* role/social limitations-emotional, *RB* role/social limitations-behavioral, *REB* role/social limitations-emotional/behavioral reflects combined items and counted as two concepts, *RP* role/social-physical limitations, *BP* bodily pain, *BE* getting along/behavior, *MH* mental health/general well-being, *SE* self-esteem, *GH* general health perceptions, *FA* family activities. Both lengths contain the same single Change in Health (CH) and Family Cohesion (FC) items and two identical pain items (BP). Therefore, ANOVA not run for CH, FC, or BP

Functioning, Role Physical. As might be expected, the lowest scores were observed for the ≥3 conditions group. Precision estimates for the short-form scales that met/exceeded the standardized estimate for the corresponding full-length versions included Role/social-Emotional Behavioral (REB = 1.37 vs. RE), Role/social-Emotional Behavioral (REB = 1.00 vs. RB), and Physical Functioning (PF = 1.09). Precision estimates for all remaining scales (except GH) were high: RP and MH (0.95), SE (0.88), and FA (0.81). A lower, but still quite modest estimate was observed for General Health (GH 0.73) which is noteworthy given that the short-form scale was reduced by 42%. RP estimates are remarkable given that overall the CHQ was reduced by 52%. The reduction for individual scales ranged from 3 to 9 items (41–66%) with the single exception being the role/social limitations-physical (RP) which was reduced in length by only one item.

Discussion

Shorter, well-validated, child-centered PRO measures, irrespective of methodological development (i.e., legacy/classical or modern test theory) are needed to keep pace with emergent innovative technologies (e.g., geographic information systems, fit-for purpose activity trackers/"wearables") that are increasingly being used to assess clinical/biomedical markers in concert with PROs [15, 16]. The objective of the present work was to derive and psychometrically evaluate a more practical-length version of a comprehensive, multidimensional child-focused PRO "legacy" measure that has been rigorously translated and used across an array of conditions and countries/cultures for many years.

An important strength of the present investigation is that the CHQ short form was empirically derived from a large Dutch representative sample using a multi-dimensional strategy but "externally" validated in a second larger U.S. representative sample that included 8-year-olds. Even though Flesch–Kincaid readability statistics [40] places the reading level for the CHQ-45 at Grade 5, with a reading ease of 74%, both the feasibility sample and the U.S. sample provide evidence that the CHQ-CF45 can be used with confidence in younger primary grades (e.g., 2nd–4th).

Items in the CHQ-CF45 exceeded all scaling criteria in the full U.S. sample and across all 10 subgroups (gender, age, number of health conditions) including item convergent and discriminant validity. All nine scales demonstrated strong internal consistency (alpha's ≥ 0.73) in the full sample

and across all subgroups with the exception of GH in ages 8–9 (0.67). RP estimates ranged from 0.73 to 1.37 providing strong evidence that using a multimethod approach, it was possible to achieve the objective of a 50% reduction in items with only a small loss in precision (range 5–27 RP/MH and GH, respectively). RP estimates for two short-form scales (PF, REB) were higher than the full-length versions (1.09, 1.37, and 1.00, respectively). Overall, these findings are encouraging and support the structural integrity of the CHQ-CF45 and provide evidence that it is reliable and valid and can be used with confidence.

A few study limitations should be noted. First, the U.S. representative sample was healthy with about 50% for whom no condition was reported. Thus, not unexpectedly, high ceiling effects were observed for the "physical scales" (PF RP BP) but less so for "psychosocial" scales (MH BE SE). Although the sample was representative of U.S. children, it is possible that the parent panel from which the sample was drawn did not truly represent children with chronic conditions and this resulted in small numbers for common "exclusive" conditions (e.g., asthma, ADHD). Although systematically recruited, it is possible that parents of children with more debilitating conditions may be less likely to participate in the panel because of time/personal limitations as their children may be less independent and require more attention/care. It is important to note, however, that for roughly 24% of the sample parents did report ≥ 2 chronic conditions. Due to small samples for discrete conditions, it became necessary to classify children for the RP estimates based on number of conditions that included both physical, mental, and/or behavioral diagnoses. Further work is needed to extend the current findings across specific condition groups to better determine potential ceiling effects and tradeoffs in precision.

Second, although item scaling results were more than satisfactory for ages 8–9, findings in the feasibility study suggest that younger ages or those with learning challenges may require some guidance as to the CHQ-CF45 layout/format. This could just be an artifact of the paper–pencil methodology since students may be accustomed to a certain style in which questions and sub-items are posed in school-related work. Further study may be warranted.

Finally, it will be important to assess the performance of the CHQ-CF45 across other countries/cultures and across different data collection platforms (e.g., social networks, smartphones, OLED-based displays, Google Glass, and other AR-enabled wearables/devices). It is important to reiterate that data for the U.S. sample were Internet-based which allowed for privacy and flexibility of questionnaire completion. In both the feasibility and the U.S. sample, respondents were allowed to skip items but little missing data were observed which suggests that children felt comfortable answering the CHQ using either print or electronic administration.

The use of family-centered child-reported PROs across an array of integrated settings is not new. More than 20 years have passed since the release of well-validated measures. Currently, the pressing need is for shorter yet equally robust child self-report measures that capture the physical, emotional, and social well-being of the child and his/her family. The integration of PRO content (such as the CHQ-CF45) with emergent technologies being used in classrooms, clinical care, and academic settings presents as an exciting evolutionary "next" step. Central to these initiatives is the child's own voice. It is essential that he/she be heard if we are to truly advance child-focused PROs as a viable and important source for meaningful and actionable data that can enhance communication and better inform decision-making.

Acknowledgements The principal author would like to thank Suzanne Fitch, D. Ed., Mr. Mike Collins, P., teachers, and parents for their overwhelming support for this work and most especially, the students for their enthusiastic participation and insights.

Compliance with ethical standards

Conflict of interest None of the authors received financial compensation for this work nor do they have a conflict of interest to report. Ms. Landgraf is the Vice President and Chief Scientific Officer at HealthActCHQ, Inc., which owns the intellectual property rights to the CHQ-CF45.

References

1. Starfield, B., Bergner, M., Ensminger, M., et al. (1993). Adolescent health status measurement: development of the Child Health and Illness Profile. *Pediatrics, 91*(2), 430–435.
2. Feeny, D., Furlong, W., Boyle, M., & Torrance, G. W. (1995). Multi-attribute health status classification systems: Health Utilities Index. *Pharmacoeconomics, 7*, 490–502.
3. Landgraf, J. M., Maunsell, E., Nixon-Speechley, K. N., Bullinger, M., Campbell, S., Abetz, L., & Ware, J. E. (1998). Canadian-French, German and United Kingdom versions of the Child Health Questionnaire (CHQ-PF50): Methodology and preliminary item scaling results. *Quality of Life Research, 7*(5), 433–445.
4. Drotar, D. (1998). *Measuring health-related quality of life in children and adolescents: Implications of research and practice*. New Jersey: Lawrence Erlbaum Associates.
5. Eiser, C., & Morse, R. (2001). Quality-of-life measures in chronic diseases of childhood. *Health Technology Assessment, 5*(4), 1–157.
6. Kurtin, P., Landgraf, J. M., & Abetz, L. (1994). Patient-based health status measures in pediatric dialysis: Expanding the assessment of outcome. *American Journal of Kidney Diseases, 24*(2), 376–382.
7. Landgraf, J. M., & Abetz, L. (1997). Functional Status and well-being of children representing three cultural groups: Initial self-reports using the CHQ-CF87. *Journal of Psychology and Health, 12*, 839–854.
8. Landgraf, J. M. & Abetz, L. (1996). Measuring health outcomes in pediatric populations: Issues in psychometrics and application. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (pp. 793–802). Philadelphia: Lipincott-Raven.

9. Landgraf, J. M. (2005). Practical considerations in the measurement of health-related quality of life in child/adolescent clinical trials. In P. Fayers & R. D. Hays (Eds.), *Eds.), Assessing the quality of life in clinical trials* (2nd edn., pp. 339–356). Oxford: Oxford University Press.
10. Riley, A. (2004). Evidence that school-age children can self-report on their health. *Ambulatory Pediatrics, 4*(Suppl), 371–376.
11. Olson, L. M., Radecki, L., Frintner, M. P., Weiss, K. B., Korfmacher, J., & Siegel, R. M. (2007). At what age can children report dependably on their asthma health status. *Pediatrics, 119*(1), e93–e102.
12. Bevans, K. A., Riley, A. W., Moon, J., & Forrest, C. B. (2010). Conceptual and methodological advances in child-reported outcomes measurement. *Expert Review of Pharmaeconomics & Outcomes Research, 10*(4), 385–396. https://doi.org/10.1586/erp/10.52/.
13. Forrest, C. B., & Riley, A. W. (2004). Childhood origins of adult health: A basis for life-course health policy. *Health Affairs (Millwood), 23*(5), 155–164.
14. Bevans, K. B., Riley, A. W., Landgraf, J. M., Carle, A. C., Teneralli, R. E., Fiese, B. H., Meltzer, L. J., Ettinger, A. K., Becker, B. D., & Forrest, C. B. (2017). Children's family experiences: Development of the PROMIS® pediatric family relationships measures. *Quality of Life Research, 26*, 3011. https://doi.org/10.1007/s11136-017-1629-y. (**Epub ahead of print**).
15. Pediatric Patient Reported Outcomes in Chronic Diseases (PEPR) https://www.peprconsortium.org/. Accessed 10 Jan 2018.
16. Environmental Influences on Child Health Outcomes (ECHO) Program https://www.nih.gov/echo. Accessed 10 Jan 2018.
17. Landgraf, J. M., Abetz, L., & Ware, J. E. (1996). *The Child Health Questionnaire (CHQ): A user's manual*. Boston, MA: The Health Institute, New England Medical Center.
18. https://www.healthactchq.com/bibliographies.php. Accessed 10 Jan 2018.
19. Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health, 8*(2):94–104.
20. Raat, H., Landgraf, J. M., Bonsel, G. J., Gemke, R. J., & Essink-Bot, M. L. (2002). Reliability and validity of the Child Health Questionnaire-Child Form (CHQ-CF87) in a Dutch adolescent population. *Quality of Life Research, 11*(6), 575–581.
21. de Wit, M., Delemarre-van de Waal, H. A., Bokma, J. A., Haasnoot, K., Houdijk, M. C., Gemke, R. J., & Snoek, F. J. (2007). Self-report and parent-report of physical and psychosocial well-being in Dutch adolescents with type 1 diabetes in relation to glycemic control. *Health Quality of Life Outcomes, 5*, 10.
22. Buysse, C. M., Raat, H., Hazelzet, J. A., Hop, W. C., Maliepaard, M., & Joosten, K. F. (2007). Long-term health-related quality of life in survivors of meningococcal septic shock in childhood and their parents. *Quality of Life Research, 16*, 1567–1576. https://doi.org/10.1007/s11136-007-9271-8.
23. de Wit, M., Delemarre-van de Waal, H. A., Pouwer, F., Gemke, R. J., & Snoek, F. J. (2007). Monitoring health related quality of life in adolescents with diabetes: A review of measures. *Archieves of Disease in Childhood, 92*(5), 434–439.
24. Hosli, E. J., Detmar, S. B., Raat, H., Bruil, J., Vogels, T. G. C., & Verrips, E. H. W. (2007). Reliability, validity, and structure of the self-report form of the Child Health Questionnaire (CHQ-CF87) in a Dutch adolescent population. *Expert Review of Pharmacoeconomics & Outcomes Research, 7*(4), 393–401. (https://doi.org/10.1586/14737167.7.4.393).
25. Raat, H., Mangunkusumo, R. T., Landgraf, J. M., Kloek, G., & Brug, J. (2007). Feasibility, reliability and validity of adolescent health status measurement by the Child Health Questionnaire Child Form (CHQ-CF87): Internet administration compared with the standard paper version. *Quality of Life Research, 16*, 675–685.
26. Coons, S. J., Gwaltney, C. J., Hays, R. D., Lundy, J. J., Sloan, J. A., Revicki, D. A., Lenderkinbg, W. R., Cella, D., & Basch, E. on behalf of the ISPOR ePRO Task Force. (2009). Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient reported outcome (PRO) measures: ISPOR ePRP good research practices task force report. *Value in Health 12* (4):419–429.
27. World Health Organization. (1948). *Constitution of the World Health Organization*. Geneva: WHO Basic Documents.
28. Landgraf, J. M., Vogel, I., Oostenbrink, R., van Baar, M. E., & Raat, H. (2013). Parent-reported health outcomes in infants/toddlers: Measurement properties and clinical validity of the ITQOL-SF47. *Quality of Life Research, 22*(3), 635–646. https://doi.org/10.1007/s11136-012-0177-8. (**Epub 2012 April**).
29. Landgraf, J. M. (2007). Precision and sensitivity of the short-form pediatric enuresis module to assess quality of life (PEMQOL). *Journal of Pediatric Urology, 3*(2), 109–117. https://doi.org/10.1016/j.jpurol.2006.04.004.
30. Stewart, A., & Ware, J. E. (1992). *Measuring functioning and well-being the medical outcomes study approach*. Durham: Duke University Press.
31. HealthActCHQ. (2018). *CHQ-CF45 scoring manual*. Boston, MA: HealthActCHQ.
32. Ware, J. E., Harris, W. J., Gandek, B., Rogers, B. W., & Reese, P. R. (1997). *MAPR for windows version 1.1*. Boston, MA: Health Assessment Lab.
33. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.
34. The GfK Group Confidential Proprietary Project Report for the HealthActCHQ (HACHQ) Norming Surveys (2015). GfK: East Hanover, NJ; HACHQ: Boston, MA.
35. HealthActCHQ Inc (2018). CHQ-CF45 US Norms. Boston, MA, USA: HealthActCHQ, Inc.
36. Hays, R. D., Honghu, L., & Kapteyn, A. (2015). Use of Internet panels to conduct surveys. *Behaviour Research Methods*. https://doi.org/10.3758/s13428-015-0617-9. (**Published Online 14 July**)
37. Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd edn.). New York: McGraw-Hill.
38. Buckstein, D. A., McGrath, M. M., Buchner, D. A., Landgraf, J. M., & Goss, T. P. (2000). Evaluation of a short-form for measuring health related quality of life among pediatric wheezing illness patients. *Journal of Allergy Clinical Immunology 105*, 245–251.
39. McHorney, C. A., Ware, J. E., Rogers, W., Racek, A. E., & Lu, J. F. R. (1992). The validity and relative precision of MOS short-and-long-form health status scales and Dartmouth COOP Charts. *Medical Care, 30*, 253–265.
40. Lapp, D., Flood, J., Brock, C. H., & Fisher, D. (2007). *Teaching reading to every child* (4th edn.). New York: Routledge Taylor and Francis Group.