Journal of Clinical Epidemiology 184 (2025) 111855

Journal of Clinical Epidemiology

ORIGINAL RESEARCH

# Psychometric properties and reference values of the Patient-Reported Outcomes Measurement Information System (PROMIS) pediatric item banks Mobility, Upper Extremity, and Pain Interference in the Dutch population

Dorinde L. Korteling[a,b,c,d], Marjolijn Ketelaar[e,f], Selina Limmen[a,b,c,g], Caroline B. Terwee[d,h], Manon A.T. Bloemen[i], Eugene A.A. Rameckers[j,k,l], Raoul H.H. Engelbert[m,n], Hedy A. van Oers[a,b,c,o], Lotte Haverman[a,b,c,p], Michiel A.J. Luijten[a,b,c,d,h,*]

[a]*Amsterdam UMC location University of Amsterdam, Emma Children's Hospital, Child and Adolescent Psychiatry & Psychosocial Care, Meibergdreef 9, Amsterdam, The Netherlands*
[b]*Amsterdam Reproduction and Development, Child development, Amsterdam, The Netherlands*
[c]*Amsterdam Public Health, Mental health, Amsterdam, The Netherlands*
[d]*Amsterdam Public Health, Methodology, Amsterdam, The Netherlands*
[e]*UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht, The Netherlands*
[f]*De Hoogstraat Rehabilitation, Utrecht, Center of Excellence for Rehabilitation Medicine Utrecht, Utrecht, The Netherlands*
[g]*Amsterdam Public Health, Personalized Medicine, Amsterdam, The Netherlands*
[h]*Amsterdam UMC, Vrije Universiteit, Department of Epidemiology and Data Science, Amsterdam, The Netherlands*
[i]*Research Group Moving, Growing and Thriving Together, HU University of Applied Sciences Utrecht, Utrecht, The Netherlands*
[j]*CAPHRI, Maastricht University, Maastricht, The Netherlands*
[k]*Centre of Expertise, Adelante Rehabilitation Centre, Valkenburg, The Netherlands*
[l]*Rehabilitation Science and Physiotherapy, REVAL, Hasselt University, Hasselt, Belgium*
[m]*Department of Rehabilitation Medicine, Amsterdam Movement Sciences, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands*
[n]*Centre of Expertise Urban Vitality, Faculty of Health, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands*
[o]*Amsterdam Public Health, Quality of Care, Amsterdam, The Netherlands*
[p]*Amsterdam Public Health, Digital Health, Amsterdam, The Netherlands*

Accepted 22 May 2025; Published online 29 May 2025

## Abstract

**Objectives:** This study investigated psychometric properties and reference values of the Patient-Reported Outcomes Measurement Information System (PROMIS) pediatric v2.0 Mobility, Upper Extremity, and Pain Interference item banks, short forms and computerized adaptive tests (CATs) in the Dutch general population, supplemented with a clinical sample to improve low-end item parameter estimates.

**Study Design and Setting:** Children (aged 8-18 years) completed PROMIS item banks and legacy instruments (Pediatric Quality of Life Inventory 4.0 subdomain Physical Health, Numeric Pain Rating Scale). Structural validity of item banks was evaluated by fitting a graded response model and inspecting item-fit statistics. Reliability of item banks, short forms, and post-hoc CATs was expressed as standard error of measurement/theta. To compare measurement efficiency of instruments, relative efficiency was calculated. Construct validity was assessed by correlating item banks with legacy instruments. Differential item functioning between Dutch and US samples was evaluated.

**Results:** Seven hundred eighty three children participated: 555 children from the general population and 228 children receiving physical therapy. Structural validity was sufficient for all banks. PROMIS Pain Interference was reliable at the sample mean (standard error of theta $< 0.32$) and up to 2 standard deviations in the clinically relevant direction (indicating worse health). PROMIS Mobility and Upper

Funding: This work was supported by the ZonMW (grant number: 10270032130005) and Zorginstituut Nederland (grant number: 20220254492022). They had no role in the conceptualization, design, decision to publish, or preparation of the manuscript.

* Corresponding author. Department of Child and Adolescent Psychiatry & Psychosocial Care, G8-136, Emma Children's Hospital, Amsterdam UMC, Postbox 22660, 1100 DD, Amsterdam, The Netherlands.

*E-mail address:* m.luijten@amsterdamumc.nl (M.A.J. Luijten).

https://doi.org/10.1016/j.jclinepi.2025.111855

Extremity scales were reliable in the clinically relevant direction, but less so within the normal range. CAT outperformed other assessment methods in efficiency. Construct validity was sufficient. No items displayed differential item functioning.

**Conclusion:** The PROMIS v2.0 pediatric Mobility, Upper Extremity, and Pain Interference item banks displayed sufficient validity in the Dutch general population and sufficient reliability in the clinically relevant direction. 

*Keywords:* Computerized adaptive testing; Reliability; Validity; Psychometrics; Physical functioning; Patient-reported outcome measures

## 1. Introduction

Patient-reported outcomes (PROs) have become increasingly important in pediatric healthcare, strengthening patient−physician communication; facilitating the monitoring of daily functioning, quality of life, and symptoms; and potentially supporting shared decision-making [1,2]. Patient-reported outcome measures (PROMs) are tools for assessing PROs and are gaining widespread implementation in medical settings [2,3]. However, PROMs often differ in content, psychometric properties, and scoring methods when measuring the same PROs [4]. Consequently, scores across instruments may be incomparable, and score interpretation is unstandardized [5]. Moreover, children often struggle with completing PROMs due to questionnaire length and the presence of irrelevant questions [2,3]. This complicates implementation and may contribute to relatively low response rates [2,3].

To address these challenges, the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative developed item banks for children (and adults), covering generic, relevant domains of physical, social, and mental health [4,6−8]. These item banks consist of collections of items measuring the same domain across various levels of functioning and were constructed using item-response theory (IRT) modeling [9]. Through IRT modeling, items are ranked by their difficulty and discriminative ability, facilitating the use of computerized adaptive tests (CATs). CATs select items from an item bank based on responses to previous items. This can drastically reduce questionnaire length and the presence of irrelevant questions, by selecting questions tailored to the child's functioning level [7,10].

In the Netherlands, efforts have been made to implement pediatric PROMIS in research and daily clinical care [4,11−15]. However, the psychometric properties of the PROMIS pediatric Mobility, Upper Extremity, and Pain Interference item banks have not been explored on population-level outside the United States despite prior evidence of relevant differences in PROM performance between the United States and European countries, highlighting the need for validation to ensure responsible use [15]. Reliable and valid measurements, such as defined by the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) initiative [16], are essential to avoid incorrect conclusions and potentially poor intervention decisions [17]. Therefore, this study aims to evaluate the psychometric properties and provide Dutch reference values of the PROMIS pediatric Mobility, Upper Extremity, and Pain Interference item banks short forms and CATs in a representative sample of the Dutch children from the general population, supplemented with a clinical sample.

## 2. Methods

The study design followed the "COSMIN Study Design checklist for PROMs" [18] and the "PROMIS Psychometric Evaluation and Calibration Plan" [19] and was reported according to the "COSMIN Reporting guideline for studies on measurement properties of PROMs" [20].

### 2.1. Procedure and participants

Between December 2017 and April 2018, children/adolescents aged 8 to 18 years, recruited through marketing agency Kantar Public, completed PROMs as part of this research. This data collection aimed to obtain representative data for multiple Dutch PROMIS pediatric item banks, including the Mobility, Upper Extremity, and Pain Interference item banks, to calculate reference values. This data collection has previously been described [13]. However, preliminary (unpublished) analyses revealed a positive skew in PROMIS responses, likely caused by limited inclusion of lower-functioning participants, which may affect IRT parameter estimation. Therefore, data collection resumed from June 2023 to April 2024 to supplement data from the general Dutch population with a clinical sample, ensuring adequate representation of children with lower physical functioning. Data were collected of children/adolescents (aged 8-18 years) with difficulties in daily functioning due to physical complaints, who received (pediatric) physical therapy (within the last year). Participants were recruited through 18 pediatric physical therapy (PPT) practices, 11 sport clubs, 3 rehabilitation centers, 8 patient associations, 2 professional associations, and 1 academic hospital and marketing agency Panel Inzicht. Prior to participation via Panel Inzicht, a participant eligibility screening procedure occurred (child's date of birth and child's contact with medical professionals in the last year). Participants were compensated by the marketing agency or received a €5 gift card. We aimed to include > 500

**What is new?**

**Key findings**

- This study examined the psychometric properties and reference values of Patient-Reported Outcomes Measurement Information System (PROMIS) v2.0 pediatric Mobility, Upper Extremity, and Pain Interference item banks short forms and computerized adaptive tests in the Dutch general population, with additional data from a clinical sample. The item banks showed sufficient performance for use in both Dutch general and clinical populations.

**What this adds to what is known?**

- In the Netherlands, pediatric PROMIS has been introduced in research and clinical care. However, the psychometric properties of the Mobility, Upper Extremity, and Pain Interference item banks had not yet been evaluated at the population level outside the United States.
- The examination of the psychometric properties and reference values of PROMIS v2.0 pediatric Mobility, Upper Extremity, and Pain Interference instruments supports their responsible use in the Netherlands and offers insights into the interpretation of scores in both general and clinical populations.

**What is the implication and what should change now?**

- Future research is needed to improve the reliability of the PROMIS Mobility and Upper Extremity item banks for children with average to high physical functioning.

participants, as recommended for IRT analyses by the ''COSMIN Study Design checklist for PROMs'' [18].

The data collection was approved by the Medical Ethics Committee of the Amsterdam UMC, location AMC [W20_136 # 20.175 and W23_069 # 23.092].

## 2.2. Measures

### 2.2.1. Sociodemographic questionnaire

Parents completed a sociodemographic questionnaire about themselves (including age, country of birth, and educational level) and their child (including date of birth, gender, and education). For the clinical sample, the Functional Mobility Scale [21] was incorporated into the sociodemographic questionnaire. This scale classifies children's functional mobility based on their use of mobility aids.

### 2.2.2. PROMIS pediatric item banks v2.0—Mobility, Upper Extremity, and Pain Interference

Children completed 3 full Dutch-Flemish PROMIS, version 2.0, self-reported pediatric Mobility (24 items), Upper Extremity (34 items), and Pain Interference (20 items) item banks. PROMIS item banks are based on a reflective model, where all items reflect the same underlying construct and are strongly correlated, whereas formative models define the construct through the combination of distinct items [18]. Eight-item short forms were extracted from the full bank data. The item banks use a 7-day recall period and a 5-point Likert score [22]. For the Mobility and Upper Extremity items, the scale ranges from 1 (''With no trouble'') to 5 (''Not able to do''). For the Pain Interference items, the scale ranges between 1 (''Never'') and 5 (''Almost always''). For the clinical sample, we added a sixth response option (''Not able to do'') to 5 items, as the original response options could be confusing for children using mobility aids (Supplement A). The additional response category was treated as missing data in the analyses, as they were not original to the instrument and participants unable to perform an activity may be unable to meaningfully assess its pain interference.

T-scores were calculated using the HealthMeasures Scoring Service, which uses the US model parameters. A T-score of 50 indicates average levels of functioning as defined by the US calibration sample (standard deviation [SD] = 10), which consists of healthy and chronically ill children (~33%) [23]. Higher scores on the Mobility and Upper Extremity item banks indicate better functioning, while higher scores on the Pain Interference item bank indicate more pain interference.

### 2.2.3. Pediatric quality of life inventory (4.0), domain 'Physical Health'

Children completed the 'Physical Health' domain of the Pediatric Quality of Life inventory (PedsQL) 4.0 questionnaire (8 items), which served as a legacy instrument for the 3 PROMIS item banks. The PedsQL measures health-related quality of life of children aged 8-18 years and uses a recall period of 1 week [24]. Questions are scored on a 5-point Likert scale, ranging from 1 (''Never a problem'') to 5 (''Almost always a problem''). Responses are transformed to a 0-100 scale, with a higher score indicating better function. The PedsQL has been validated in the Netherlands [25,26].

### 2.2.4. PROMIS pediatric numeric rating scale v.2.0—Pain Intensity

The clinical sample completed the Dutch-Flemish PROMIS v2.0 Pediatric Numeric Rating Scale for Pain Intensity, which served as a legacy instrument for the PROMIS Pain Interference item bank regarding construct validity. It consists of a single question on self-reported averaged pain intensity over the past 7 days on a scale from

0 to 10, where 0 indicates no pain and 10 represents the worst imaginable pain [27].

### 2.3. Statistical analyses

To preserve quality, data-cleaning occurred prior to analyses. Data from the clinical sample with a response time of $\leq 5$ minutes for all questions while showing no variation in responses were removed. To investigate whether the data had sufficient heterogeneity for IRT analyses, distributions of the used response options of PROMIS item banks were examined. Validity and reliability analyses were performed on the combined (general plus clinical) data.

#### 2.3.1. Structural validity

To evaluate the structural validity of the PROMIS item banks, a graded response model (GRM) was fitted, using the Expectation−Maximization algorithm within the R-package "mirt (v1.29)" [28]. A GRM is contingent upon meeting the following assumptions: unidimensionality, local independence, and monotonicity.

Unidimensionality was assessed via confirmatory factor analysis (CFA), using the weighted least squares mean-adjusted and variance-adjusted estimator with the R-package "lavaan (v0.6−3)" [29]. The following criteria were used to assess CFA fit acceptability: standardized root mean square residual value $< 0.08$, scaled Comparative Fit Index value $> 0.95$, Tucker−Lewis Index value $> 0.95$, and a root mean square error of approximation value $< 0.08$ [16,30].

Local independence was evaluated by examining residual correlations within the CFA model, with an item pair considered locally independent if the residual correlation was $< 0.20$ [19]. Monotonicity was assessed with Mokken scaling [31,32], considering the assumption met if item H values were $\geq 0.30$ and the overall scale H value was $\geq 0.50$.

Upon meeting the assumptions, a GRM model was fitted to estimate item discrimination and threshold parameters. Differences between observed and expected responses were assessed using the S-$X^2$ statistic to assess item fit [33]. Items were considered misfits if the $P$ value of the S-$X^2$ statistic was $< .001$ [19]. In addition, the range of discrimination and threshold parameters was investigated.

#### 2.3.2. Reliability

In IRT, every response pattern corresponds to a different level of functioning, theta (θ). These different levels of functioning are associated with a standard error of theta (SE(θ)), indicating the reliability of the score. To estimate the reliability of the PROMIS item banks and short forms, θ estimates and SE(θ) were calculated using the GRM model fitted in the study population and the Expected A Posteriori estimator. To investigate the reliability of CATs, post-hoc CAT simulations were performed using the R-package "catR (v3.16)" [34]. CAT performance with model parameters estimated in the study population was evaluated using maximum posterior weighted information selection criterion and the Expected A Posteriori estimator. The CAT started with the item providing the most information at the sample mean ($\theta = 0$). Administration stopped after a minimum of 4 items and a maximum of 12, or earlier if SE(θ) $< 0.32$ [35]. To compare reliability of the PROMIS measurements to the PedsQL, a GRM model was fit to the PedsQL responses. θ estimates and SE(θ) were presented in a reliability plot. We deemed an administration mode sufficiently reliable if the SE(θ) was $< 0.32$ (indicating a reliability $> 0.90$ as SE(θ) $= SD\sqrt{1 - reliability}$; criteria in line with previous PROMIS research in the Dutch general pediatric population [12,14]) at the sample mean of 0 and covered at least 2 SD of θ in the clinically relevant direction (lower θ for Mobility and Upper Extremity; higher θ for Pain Interference). Because test information ($Information(\theta) = 1/(1 - Reliability(\theta))$) [36] increases with item count, administration modes were compared on efficiency ($(1/SE(\theta))^2/n_{items}$; calculated per individual and averaged across respondents), to assess information relative to the number of items administered. Assessment methods with broader measurement ranges or higher reliability result in a higher efficiency. As a sensitivity analysis, reliability and efficiency of the instruments were also assessed after removing data of participants showing a ceiling/floor effect (ie, answered the most positive response category on all items from an item bank).

#### 2.3.3. Construct validity

To assess construct validity of the PROMIS item banks, the T-scores, based on United States metrics, were correlated with the PedsQL Physical Health subscale scores. A strong correlation (Spearman's rho $> 0.70$) was hypothesized between the PedsQL subscale and the Mobility item bank [12,37]. A moderate correlation (Spearman's rho $> 0.50$) was hypothesized between the PedsQL subscale and the Upper Extremity and Pain Interference item banks [38]. The T-scores of the Pain Interference item bank were correlated with the Numeric Pain Rating Scale, with a hypothesized correlation of Spearman's rho $> 0.50$ [39]. Lower correlations ($\Delta r > 0.10$) were hypothesized between the Pain Interference and Mobility item bank and the Pain Interference and Upper Extremity item bank. Construct validity was considered sufficient if 4 of 5 (80%) hypotheses were met.

#### 2.3.4. Cross-cultural validity

To investigate cross-cultural validity, our sample was compared to the calibration sample from the United States, sourced from the Harvard Dataverse (mean $n = 1742$) [23]. Uniform and nonuniform differential item functioning (DIF) between the Dutch and United States samples was evaluated using the R-package "lordif (v0.3−3)" [40],

employing McFadden's pseudo $R^2$, where $R^2 \geq 0.02$ indicated possible DIF [41].

#### 2.3.5. Dutch reference values

We calculated referential PROMIS T-scores for the Dutch general population sample using the HealthMeasures Scoring Service, which uses the United States model parameters. Cut-off scores were calculated based on the 75th and 95th percentile of the T-scores.

## 3. Results

Seven hundred eighty three children completed the PROMIS Mobility, Upper Extremity, and Pain Interference item banks, of which 555 from the general population and 228 from the clinical sample. The combined data showed sufficient heterogeneity to perform IRT analyses (Fig 1; Supplement B-C), with the θ range of the full sample (Mobility: −2.63 to 1.21; Upper Extremity: −3.57 to 0.79; Pain Interference: 1.89-2.57) extending beyond that of the general population sample (Mobility: −1.89 to 1.21; Upper Extremity: −3.02 to 0.79; Pain Interference: 1.89-2.26). The full sample was largely representative of the Dutch population based on age, gender, and country of birth of parents (Supplement D). In the clinical sample, 22 participants (9.6%) used the added response option "Not able to do" for 5 PROMIS Pain Interference items, with 8 of these individuals using a mobility aid (Supplement A). Sociodemographic information is provided in Table 1.

A large percentage of participants showed a ceiling/floor effect (Mobility: 53.3%; Upper Extremity: 67.8%; Pain Interference: 45.1%; PedsQL: 40.4%; See Supplement B for details on general population and clinical sample). All PROMIS item banks contained items for which the worst response option was unused (Mobility: 4 items; Upper Extremity: 4 items; Pain Interference: 1 item; Supplement C). No response options were collapsed and all items were included for IRT analyses.

### 3.1. Structural validity

Data of PROMIS item banks were unidimensional according to the CFA (Comparative Fit Index $> 0.95$, Tucker−Lewis Index $> 0.95$, root mean square error of approximation $< 0.08$, standardized root mean square residual $< 0.08$; Supplement E). Local dependencies were found between 4-item pairs of the Mobility item bank (residual correlations: 0.20-0.23; Supplement F). As the percentage of locally dependent items was low (1.5%), no items were removed for subsequent analyses. The assumption of monotonicity was met for all items and item banks. The PROMs item discrimination parameters ranged between 2.01 and 10.16 (Supplement G). One item showed a discrimination parameter above 10 ("I could zip up my clothes"; $\alpha = 10.16$). No items showed item misfit ($S\text{-}X^2 < 0.001$).

### 3.2. Reliability

In the model based on item parameters estimated on the full study population, the Mobility and Upper Extremity item banks provided reliable measurements ($SE(\theta) < 0.32$) for θ in the clinically relevant direction. They were less reliable for measurements at the sample mean ($\theta = 0$). The Pain Interference item bank showed reliable measurements at the sample mean and up to 2 SD in the clinically relevant direction. The Pain Interference short form and CAT results are similar for SE(θ) and number of items, as participants with ceiling/floor effect are administered additional items (to a maximum of 12) by CAT that do not provide additional information, resulting in lower efficiency for this subgroup. The reliability of measurements across the range of θ is presented in Figure 2 and Table 2. The relative measurement efficiency of the CAT

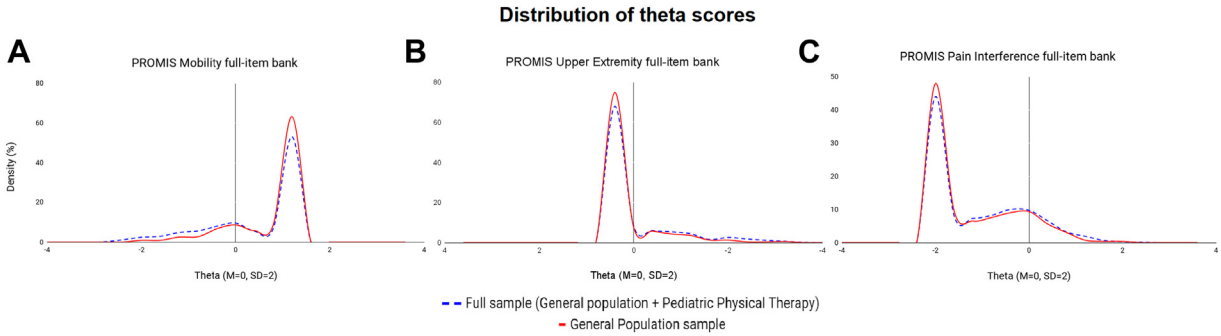


**Figure 1.** Distribution of theta scores of the full item bank of the (A) Patient-Reported Outcomes Measurement Information System (PROMIS) Mobility item bank, (B) Upper Extremity item bank, and (C) Pain Interference item bank as calculated for the Dutch general population sample ($n = 555$) and the full study sample (Dutch general population sample and clinical sample, $n = 783$).

**Table 1.** Sociodemographic information of all participants ($n$ = 783)

| Sociodemographic information | Participants from general population ($n$ = 555) | Clinical sample (pediatric physical therapy participants) ($n$ = 228)[a] | All participants ($n$ = 783)[a] |
|---|---|---|---|
| Age in years | | | |
| Mean [SD] | 13.7 [3.1] | 11.8 [2.8] | 13.2 [3.2] |
| Range | 8-18 | 8-18 | 8-18 |
| Gender | | | |
| Boy | 284 (51.2%) | 125 (54.8%) | 409 (52.2%) |
| Girl | 271 (48.8%) | 94 (41.2%) | 365 (46.6%) |
| Different/prefer not to say | 0 (0.0%) | 1 (0.4%) | 1 (0.1%) |
| Unknown | 0 (0.0%) | 8 (3.5%) | 8 (1.0%) |
| Country of birth of parents | | | |
| Both parents born in the Netherlands | 447 (80.5%) | 200 (87.7%) | 647 (82.6%) |
| One parent not born in the Netherlands | 90 (16.2%) | 12 (5.3%) | 102 (13.0%) |
| Both parents not born in the Netherlands | 18 (3.2%) | 7 (3.1%) | 25 (3.2%) |
| Unknown | 0 (0.0%) | 9 (3.9%) | 9 (1.1%) |
| Educational level of parent[b] | | | |
| Low | 67 (12.1%) | 3 (1.2%) | 70 (8.9%) |
| Middle | 267 (48.1%) | 176 (77.2%) | 443 (56.6%) |
| High | 221 (39.8%) | 41 (18.0%) | 262 (33.5%) |
| Unknown | 0 (0.0%) | 8 (3.5%) | 8 (1.0%) |

SD, standard deviation.
[a] 3 participants did not fill out the PROMIS upper extremity and pain interference item banks.
[b] Low: no education, primary education, and prevocational secondary education. Middle: secondary vocational education, senior general secondary education, and preuniversity education. High: higher professional education, university education, and higher.

outperformed the PROMIS full item bank and short form (Table 3).

### 3.3. Construct validity

The Mobility item bank T-scores correlated strongly with the PedsQL Physical Health scores ($r_s$ = 0.73). The Upper Extremity and Pain Interference item banks correlated moderately with the PedsQL Physical Health scores ($r_s$ = 0.50 and $r_s$ = 0.64, respectively; Table 4). The Pain Interference item bank correlated moderately with the Numeric Pain Rating Scale ($r_s$ = 0.67). PROMs measuring different health domains showed weaker correlations; for instance, the Pain Interference item bank correlated moderately with the Mobility item bank ($r_s$ = −0.54) and weakly with the Upper Extremity item bank

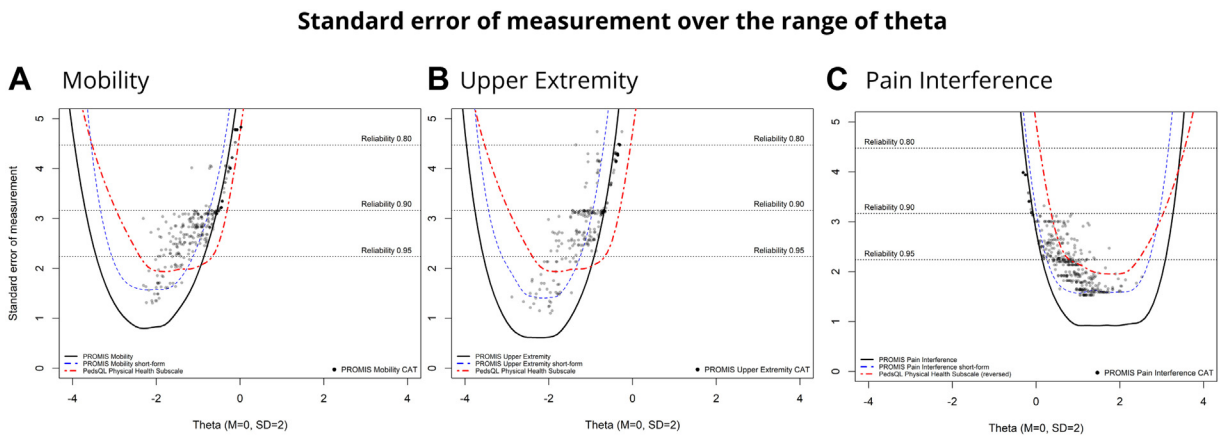


**Figure 2.** Standard error of measurement (SE(θ)) of the full item bank, short form, and computerized adaptive test (CAT) of the (A) Patient-Reported Outcomes Measurement Information System (PROMIS) Mobility item bank ($n$ = 783), (B) Upper Extremity item bank ($n$ = 780), and (C) Pain Interference item bank ($n$ = 780), and the Pediatric Quality of Life inventory (PedsQL) Physical Health domain, using the model parameters estimated in the study population.

**Table 2.** Reliability of measurement of the PROMIS Mobility, Upper Extremity, and Pain Interference pediatric item banks in the full sample ($n$ = 783)

| | Full-length item bank | | | Short form | | | CAT | | |
|---|---|---|---|---|---|---|---|---|---|
| **PROMIS scale** | **Mean SE(θ)** | **SE(θ) < 0.32[a]** | **No. of items** | **Mean SE(θ)** | **SE(θ) < 0.32[a]** | **No. of items** | **Mean SE(θ)** | **SE(θ) < 0.32[a]** | **Mean no. of items** |
| Mobility | | | | | | | | | |
| All participants ($N$ = 783) | 0.502 | 226 (28.9%) | 24 | 0.634 | 134 (17.1%) | 8 | 0.529 | 217 (27.7%) | 10.1 |
| Excluding participants with ceiling effect ($N$ = 367) | 0.274 | 226 (61.2%) | 24 | 0.573 | 192 (52.3%) | 8 | 0.331 | 217 (59.1%) | 7.9 |
| Upper Extremity | | | | | | | | | |
| All participants ($N$ = 780) | 0.558 | 199 (25.5%) | 34 | 0.674 | 131 (16.8%) | 8 | 0.592 | 179 (22.9%) | 10.5 |
| Excluding participants with ceiling effect ($N$ = 273) | 0.230 | 199 (72.8%) | 34 | 0.284 | 131 (48.0%) | 8 | 0.318 | 179 (65.6%) | 7.8 |
| Pain Interference | | | | | | | | | |
| All participants ($N$ = 758[b]) | 0.368 | 377 (49.7%) | 20 | 0.412 | 355 (46.8%) | 8 | 0.414 | 368 (48.5%) | 8.4 |
| Excluding participants with floor effect ($N$ = 417[b]) | 0.153 | 377 (90.4%) | 20 | 0.210 | 355 (85.1%) | 8 | 0.235 | 368 (88.3%) | 5.4 |

PROMIS, Patient-Reported Outcomes Measurement Information Systems; SE(θ), standard error of theta.
[a] Number/percentage of participants with an SE(θ) < 0.32. An SE(θ) < 0.32 equals a reliability of 0.90.
[b] After removing participants who used the added sixth response category, which was counted as missing data.

($r_s$ = −0.45). As 80% of hypotheses were met, construct validity was considered sufficient.

### 3.4. Cross-cultural validity

No PROMIS items showed DIF ($R^2 \geq 0.02$) between the Dutch and US samples (Supplement G).

### 3.5. Dutch reference values

Dutch PROM reference values can be found in Table 5.

## 4. Discussion

This is the first study to analyze the psychometric properties of the PROMIS pediatric Mobility, Upper Extremity, and Pain Interference item banks in a population outside of the United States. All PROMIS item banks measure reliably across a broad spectrum of levels of functioning in the clinically relevant direction from the mean. Structural and construct validity of all PROMIS item banks showed to be sufficient. No DIF was found between the Dutch and US samples. The CAT outperformed the full-length item banks and short forms based on relative efficiency.

For investigation of the psychometric properties, sufficient heterogeneity of the sample had to be ensured. As such, we included additional children with difficulties in daily functioning due to physical complaints. However, the full study population, comprising of participants from the Dutch general population and PPT participants, still contained many average to high functioning participants.

**Table 3.** Relative efficiency of the PROMIS Mobility ($n$ = 783), Upper Extremity ($n$ = 780), and Pain Interference ($n$ = 780) full item banks short forms and computerized adaptive test (CAT) compared to the PedsQL Physical Health domain ($n$ = 725)

| | PROMIS Mobility | | | PROMIS Upper Extremity | | | PROMIS Pain Interference | | |
|---|---|---|---|---|---|---|---|---|---|
| **Assessment Method** | **FL** | **SF** | **CAT** | **FL** | **SF** | **CAT** | **FL** | **SF** | **CAT** |
| All participants | | | | | | | | | |
| PedsQL Physical Health domain[a] | 0.66 | 0.82 | 1.11 | 0.59 | 0.65 | 0.95 | 2.25 | 2.30 | 3.25 |
| Full length | | 1.24 | 1.69 | | 1.13 | 1.64 | | 0.98 | 1.46 |
| Short form | | | 1.36 | | | 1.45 | | | 1.44 |
| Excluding participants with ceiling/floor effect | | | | | | | | | |
| PedsQL Physical Health domain[b] | 0.56 | 0.69 | 0.94 | 0.43 | 0.46 | 0.87 | 1.19 | 1.16 | 2.12 |
| Full length | | 1.23 | 1.69 | | 1.08 | 2.02 | | 0.94 | 1.79 |
| Short form | | | 1.37 | | | 1.89 | | | 1.90 |

Here, efficiency is defined as the amount of information provided per item. A value greater than 1 indicates that the column outperforms the row, providing, on average, that many times more information per item.
CAT, computerized adaptive test; FL, full length; PedsQL, Pediatric Quality of Life inventory; PROMIS, Patient-Reported Outcomes Measurement Information System; SF, Short form 8 items.
[a] Based on $n$ = 725.
[b] Based on $n$ = 344.

**Table 4.** Spearman's rank correlation between scores of the PROMIS Mobility ($n$ = 783), Upper Extremity ($n$ = 780), and Pain Interference ($n$ = 780) full item banks, the PedsQL Physical Health domain ($n$ = 725) and the NPRS ($n$ = 225)

| PROM | PROMIS Mobility | PROMIS Upper Extremity | PROMIS Pain Interference | PedsQL Physical Health domain |
|---|---|---|---|---|
| PROMIS Upper Extremity | 0.44 [0.37-0.51] | - | | |
| PROMIS Pain Interference | −0.54 [−0.60 to 0.49] | −0.45 [−0.50 to 0.38] | | |
| PedsQL Physical Health domain | 0.73 [0.69-0.77] | 0.50 [0.44-0.56] | 0.64 [0.58-0.68] | |
| NPRS[a] | −0.48 [−0.59 to 0.37] | −0.18 [−0.30 to 0.05] | 0.67 [0.57-0.74] | 0.49 [0.38-0.60] |

PedsQL, Pediatric Quality of Life inventory; PROM, Patient-Reported Outcome Measure; PROMIS, Patient-Reported Outcomes Measurement Information System; NPRS, Numeric Pain Rating Scale.

[a] Completed only by the clinical sample.

This is especially true for the Upper Extremity items (Supplement B), as children with severe upper extremity issues are less often seen by physical therapists. This might explain the Upper Extremity item displaying a discrimination parameter above 10 ("I could zip up my clothes"; $\alpha$ = 10.16). Dutch discrimination parameters are generally higher than US parameters [15,43], which may inflate item information and reliability estimates. Ideally, Dutch and US models should be compared using more comparable samples (eg, bilingual or bicultural). If differences persist, Dutch parameters could be justified to optimize CAT efficiency. However, given the lack of DIF in our analysis, we consider US parameters suitable for use in the Dutch population and recommend their application.

Our results show that the PROMIS Mobility and Upper Extremity item banks are reliable measures for children with clinically relevant deviations from the mean ($\theta$ of 0). As PROMIS focuses on measuring PROs for improving patient care and research [13,44], their PROMs function optimally for populations with potentially clinically relevant deviations (eg, low mobility, poor upper extremity function, or high pain interference). Our findings indicate that, currently, the PROMIS Mobility and Upper Extremity item banks do not measure sufficiently reliable around the mean of the (general plus clinical) population. Similar results were found in clinical samples [12,45]. To reliably measure the entire population, we agree with previous research [37] that future work is needed to improve the reliability of the PROMIS Mobility and Upper Extremity item banks for children with average to high levels of physical functioning. The item banks might require additional high-end differentiating items to appropriately measure healthy or high-functioning individuals. This was done for the PROMIS Physical Function (v2.0) item bank for adults, which now includes 165 items to cover all levels of physical function (PROMIS pediatric Mobility v2.0 contains 24 items and Upper Extremity v2.0 34 items) [46].

Unlike PROMIS item banks, the PedsQL Physical Health subdomain combines several PRO domains into a single score. This approach reduces the likelihood of ceiling/floor effects, as participants are less likely to report the highest scores across multiple domains simultaneously. This resulted in a higher ceiling/floor effect for the PROMIS item banks (53.3%, 67.8%, and 45.1%) compared to the PedsQL Physical Health subdomain (40.4%). However, combining multiple PRO domains into a single score complicates interpretability of the score [47]. Generally, all investigated PROMs showed high levels of ceiling/floor effect, mirroring previous research [48−51] and indicating an inability to detect variability at the high end of the scale.

Although this study focuses on v2.0 of the PROMIS pediatric Mobility, Upper Extremity, and Pain Interference item banks, v3.0 has recently been released [52,53], and our results do not apply to v3.0. A major change in v3.0

**Table 5.** Dutch reference values of the PROMIS Mobility, Upper Extremity, and Pain Interference T-scores and the PedsQL Physical Health domain scores ($n$ = 555)

| PROM | Mean [SD][b] | Range | Cut-off[c] | |
|---|---|---|---|---|
| | | | Moderate | Severe |
| PROMIS Mobility T-score | 58.2 [6.6] | 31.1-62.1 | ≤ 55.9 | ≤ 44.8 |
| PROMIS Upper Extremity T-score | 54.5 [6.9] | 19.8-57.9 | ≤ 52.8 | ≤ 39.4 |
| PROMIS Pain Interference T-score | 39.3 [9.8] | 31.1-72.6 | ≥ 47.5 | ≥ 56.7 |
| PedsQL Physical Health domain sum score[a] | 92.0 [12.9] | 12.5-100 | - | - |

Age-specific average values for the PROMIS instruments can be found in Supplement H.

PedsQL, Pediatric Quality of Life inventory; PROM, Patient-Reported Outcome Measure; PROMIS, Patient-Reported Outcomes Measurement Information System; SD, standard deviation.

[a] 58 participants did not fill out the PedsQL Physical Health domain.

[b] A T-score of 50 and standard deviation of 10 corresponds to the average as defined by the US calibration sample.

[c] The cut-off are based on the 75th percentile (moderate) and the 95th percentile (severe) of the T-scores in the Dutch general population [42].

is the collapsing of response categories among the clinical side of the Likert scale (eg, ''Not able to do'' and ''With a lot of trouble''). However, our results show that nearly all response categories from v.2.0 were used (Supplement C). This indicates a potential loss of nuance in the v3.0 item banks, which may affect reliability and responsiveness. Moreover, 9.6% of individuals presented with the added 'Not able to do' response option in the Pain Interference item bank selected it. This suggests a reason to consider adding this option to the PROMIS item bank to better accommodate lower-functioning individuals and improve validity for this subgroup.

Our study presents Dutch reference T-scores of the PROMIS v2.0 Mobility, Upper Extremity, and Pain Interference item banks. The mean T-scores for the 3 domains differ from the US calibration sample average of 50, especially for Pain Interference (39.3; Table 5). This is potentially due to the US sample including approximately 33% chronically ill participants, a demographic not explicitly represented in our general population sample. We recommend Dutch PROMIS users to use the T-scores presented in Table 5 as reference values for the Dutch general population. Future research will explore reference T-scores for PPT participants.

In conclusion, the PROMIS v2.0 pediatric Mobility, Upper Extremity, and Pain Interference item banks perform sufficiently in the Dutch general and clinical population. These measures are distributed by the Dutch-Flemish PROMIS National Center (www.dutchflemishpromis.nl). Future research is needed to improve the reliability of the PROMIS Mobility and Upper Extremity item banks for children with average to high physical functioning.

## Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGTP in order to draft and improve the readability of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

## CRediT authorship contribution statement

**Dorinde L. Korteling:** Writing − review & editing, Writing − original draft, Visualization, Formal analysis, Data curation, Conceptualization. **Marjolijn Ketelaar:** Writing − review & editing, Supervision, Conceptualization. **Selina Limmen:** Writing − review & editing. **Caroline B. Terwee:** Writing − review & editing. **Manon A.T. Bloemen:** Writing − review & editing. **Eugene A.A. Rameckers:** Writing − review & editing. **Raoul H.H. Engelbert:** Writing − review & editing, Supervision, Conceptualization. **Hedy A. van Oers:** Writing − review & editing, Funding acquisition. **Lotte Haverman:** Writing − review & editing, Supervision, Funding acquisition, Conceptualization. **Michiel A.J. Luijten:** Writing − review & editing, Supervision, Formal analysis, Conceptualization.

## Declaration of competing interest

Authors L. Haverman, M. Luijten, and C.B. Terwee are part of the Dutch-Flemish PROMIS National Center. All other authors report no conflict of interest.

## Acknowledgments

The authors would like to thank all individuals and organizations that helped in recruiting participants for this study. Furthermore, the authors would like to thank all children, adolescents, and parents for participating in this study. The authors would like to thank Biomedia for building and maintaining the research website.

## Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2025.111855.

## Data availability

Data will be made available on request.

## References

[1] Damman OC, Jani A, de Jong BA, Becker A, Metz MJ, de Bruijne MC, et al. The use of PROMs and shared decision-making in medical encounters with patients: an opportunity to deliver value-based health care to patients. J Eval Clin Pract 2020;26:524−40.

[2] van Muilekom MM, Teela L, van Oers HA, van Goudoever JB, Grootenhuis MA, Haverman L. Patients' and parents' perspective on the implementation of Patient Reported Outcome Measures in pediatric clinical practice using the KLIK PROM portal. Qual Life Res 2022;31:241−54.

[3] Teela L, van Muilekom MM, Kooij LH, Gathier AW, van Goudoever JB, Grootenhuis MA, et al. Clinicians' perspective on the implemented KLIK PROM portal in clinical practice. Qual Life Res 2021;30:3267−77.

[4] Terwee CB, Ahmed S, Alhasani R, Alonso J, Bartlett SJ, Chaplin JE, et al. Comparable real-world patient-reported outcomes data across health conditions, settings, and countries: the PROMIS international collaboration. NEJM Catalyst 2024;5:CAT.24.0045.

[5] Terwee CB, Zuidgeest M, Vonkeman HE, Cella D, Haverman L, Roorda LD. Common patient-reported outcomes across ICHOM Standard Sets: the potential contribution of PROMIS(R). BMC Med Inform Decis Mak 2021;21:259.

[6] Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care 2007;45:S3−11.

[7] Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010;63:1179−94.
[8] Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. Qual Life Res 2007;16:133−41.
[9] Fries JF, Witter J, Rose M, Cella D, Khanna D, Morgan-DeWitt E. Item response theory, computerized adaptive testing, and PROMIS: assessment of physical function. J Rheumatol 2014;41:153−8.
[10] Choi SW, Reise SP, Pilkonis PA, Hays RD, Cella D. Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. Qual Life Res 2010;19:125−36.
[11] Peersmann SHM, Luijten MAJ, Haverman L, Terwee CB, Grootenhuis MA, van Litsenburg RRL. Psychometric properties and CAT performance of the PROMIS pediatric sleep disturbance, sleep-related impairment, and fatigue item banks in Dutch children and adolescents. Psychol Assess 2022;34:860−9.
[12] Luijten MA, Terwee CB, van Oers HA, Joosten MM, van den Berg JM, Schonenberg-Meinema D, et al. Psychometric properties of the pediatric Patient-Reported Outcomes Measurement Information System item banks in a Dutch clinical sample of children with juvenile idiopathic arthritis. Arthritis Care Res 2020;72:1780−9.
[13] Luijten MAJ, van Litsenburg RRL, Terwee CB, Grootenhuis MA, Haverman L. Psychometric properties of the Patient-Reported Outcomes Measurement Information System (PROMIS(R)) pediatric item bank peer relationships in the Dutch general population. Qual Life Res 2021;30:2061−70.
[14] van Muilekom MM, Luijten MAJ, van Litsenburg RRL, Grootenhuis MA, Terwee CB, Haverman L. Psychometric properties of the patient-reported outcomes measurement information system (PROMIS(R)) pediatric anger scale in the Dutch general population. Psychol Assess 2021;33:1261−6.
[15] Klaufus LH, Luijten MAJ, Verlinden E, van der Wal MF, Haverman L, Cuijpers P, et al. Psychometric properties of the Dutch-Flemish PROMIS((R)) pediatric item banks Anxiety and Depressive Symptoms in a general population. Qual Life Res 2021;30:2683−95.
[16] Mokkink LB, Elsman EBM, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. Qual Life Res 2024.
[17] de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A Practical Guide: Introduction. Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press; 2011:1−6.
[18] Mokkink LB, Prinsen C, Patrick DL, Alonso J, Bouter LM, De Vet H, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. Amsterdam, The Netherlands: COSMIN initiative; 2019:1−32.
[19] Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 2007;45:S22−31.
[20] Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res 2021;30:2197−218.
[21] Graham HK, Harvey A, Rodda J, Nattrass GR, Pirpiris M. The functional mobility scale (FMS). J Pediatr Orthop 2004;24:514−20.
[22] Abma IL, Butje BJD, Ten Klooster PM, van der Wees PJ. Measurement properties of the Dutch-Flemish patient-reported outcomes measurement information system (PROMIS) physical function item bank and instruments: a systematic review. Health Qual Life Outcomes 2021;19:62.
[23] DeWalt D. PROMIS 1 pediatric supplement. V1 ed.: Harvard Dataverse. 2016. Available at: https://doi.org/10.7910/DVN/IBWSUD. Accessed August 19, 2024.
[24] Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. Ambul Pediatr 2003;3:329−41.
[25] Engelen V, Haentjens MM, Detmar SB, Koopman HM, Grootenhuis MA. Health related quality of life of Dutch children: psychometric properties of the PedsQL in The Netherlands. BMC Pediatr 2009;9:68.
[26] van Muilekom MM, Luijten MAJ, van Oers HA, Conijn T, Maurice-Stam H, van Goudoever JB, et al. Paediatric patients report lower health-related quality of life in daily clinical practice compared to new normative PedsQL(TM) data. Acta Paediatr 2021;110:2267−79.
[27] Health Measures. Patient-Reported Outcomes Measurement Information System®; Pain intensity scoring manual; A brief guide to scoring the PROMIS® pain intensity instruments. Evanston, IL: Health Measures; 2021.
[28] Mirt CRP. A multidimensional item response theory package for the R environment. J Stat Softw 2012;48:1−29.
[29] Lavaan RY. An R package for structural equation modeling. J Stat Softw 2012;48:1−36.
[30] Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Methods Psychol Res Online 2003;8:23−74.
[31] Mokken RJ. A theory and procedure of scale analysis: With applications in political research. Berlin, Germany: Walter de Gruyter; 2011.
[32] van der Ark LA. Mokken scale analysis in R. J Stat Softw 2007;20:1−19.
[33] Kang T, Chen TT. Performance of the generalized S-X2 item fit index for polytomous IRT models. J Educ Meas 2008;45:391−406.
[34] Magis D, Raîche G. catR: an R package for computerized adaptive testing. Appl Psychol Meas 2011;35:576−7.
[35] Wainer H, Dorans NJ, Flaugher R, Green BF, Mislevy RJ. Computerized adaptive testing, a primer. New York, NY: Routledge; 2000.
[36] Petrillo J, Cano SJ, McLeod LD, Coon CD. Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples (vol 18, pg 25, 2015). Value in Health 2015;18:547.
[37] Kashikar-Zuck S, Carle A, Barnett K, Goldschneider KR, Sherry DD, Mara CA, et al. Longitudinal evaluation of patient-reported outcomes measurement information systems measures in pediatric chronic pain. Pain 2016;157:339−47.
[38] Jones JT, Wootton J, Ying J, Liberio B, Lee J, Carle A, et al. Validation of patient-reported outcomes measurement information system (PROMISr) modules for use in childhood-onset. Lupus Arthritis Rheumatol 2015;67:3151.
[39] Bernstein DN, Kelly M, Houck JR, Ketz JP, Flemister AS, DiGiovanni BF, et al. PROMIS pain interference is superior vs numeric pain rating scale for pain assessment in foot and ankle patients. Foot Ankle Int 2019;40:139−44.
[40] Choi SW, Gibbons LE, Crane PK. Lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. J Stat Softw 2011;39:1−30.
[41] Crins MHP, Terwee CB, Ogreden O, Schuller W, Dekker P, Flens G, et al. Differential item functioning of the PROMIS physical function, pain interference, and pain behavior item banks across patients with different musculoskeletal disorders and persons from the general population. Qual Life Res 2019;28:1231−43.
[42] Carle AC, Bevans KB, Tucker CA, Forrest CB. Using nationally representative percentiles to interpret PROMIS pediatric measures. Qual Life Res 2021;30:997−1004.
[43] DeWitt EM, Stucky BD, Thissen D, Irwin DE, Langer M, Varni JW, et al. Construction of the eight-item patient-reported outcomes

measurement information system pediatric physical function scales: built using item response theory. J Clin Epidemiol 2011;64:794−804.
[44] HealthMeasures. PROMIS (Patient-Reported Outcomes Measurement Information System). HealthMeasures. Published 2023. Available at: https://www.healthmeasures.net/explore-measurement-systems/promis. Accessed February 1, 2025.
[45] Brandon TG, Becker BD, Bevans KB, Weiss PF. Patient-reported outcomes measurement information system tools for collecting patient-reported outcomes in children with juvenile arthritis. Arthritis Care Res (Hoboken) 2017;69:393−402.
[46] Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. J Clin Epidemiol 2014;67:516−26.
[47] van der Willik EM, Terwee CB, Bos WJW, Hemmelder MH, Jager KJ, Zoccali C, et al. Patient-reported outcome measures (PROMs): making sense of individual PROM scores and changes in PROM scores over time. Nephrology (Carlton) 2021;26:391−9.
[48] Varni JW, Magnus B, Stucky BD, Liu Y, Quinn H, Thissen D, et al. Psychometric properties of the PROMIS (R) pediatric scales: precision, stability, and comparison of different scoring and administration options. Qual Life Res 2014;23:1233−43.
[49] Kratz AL, Slavin MD, Mulcahey MJ, Jette AM, Tulsky DS, Haley SM. An examination of the PROMIS pediatric instruments to assess mobility in children with cerebral palsy. Qual Life Res 2013;22:2865−76.
[50] Singh A, Dasgupta M, Retherford D, Fiallo-Scharer R, Simpson PM, Panepinto JA. Measurement properties of patient reported outcomes measurement information system domains for children with type 1 diabetes. Pediatr Diabetes 2021;22:335−44.
[51] Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. Med Care 2001;39:800−12.
[52] Lai JS, Tang X, Schalet BD, Kallen MA, Kaat AJ, D C, PROMIS® Health Organization (PHO) 2022 Conference Abstracts. The PROMIS pediatric item banks norming project. J Patient-Reported Outcomes 2023;7:23.
[53] HealthMeasures. PROMIS Physical Function Measure Differences. Published July 10, 2024. Available at: https://www.healthmeasures.net/images/PROMIS/Differences_Between_PROMIS_Measures/PROMIS_Physical_Function_Measure_Differences_10July2024.pdfPatient-Reported. Accessed October 7, 2024. Outcomes Measurement Information System®. Physical function measure differences. Health Measures; 2024.