SLEEPJ, 2018, 1–13

doi: 10.1093/sleep/zsy054
Advance Access Publication Date: 13 March 2018
Original Article

ORIGINAL ARTICLE

Development and validation of the PROMIS Pediatric Sleep Disturbance and Sleep-Related Impairment item banks

Christopher B. Forrest[1,2,*], Lisa J. Meltzer[3], Carole L. Marcus[2], Anna de la Motte[1], Amy Kratchman[1], Daniel J. Buysse[4], Paul A. Pilkonis[4], Brandon D. Becker[1] and Katherine B. Bevans[5]

[1]Project Performance Site: Applied Clinical Research Center, Children's Hospital of Philadelphia, Philadelphia, PA, [2]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, [3]Department of Pediatrics, National Jewish Health, Denver, CO, [4]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA and [5]College of Public Health, Temple University, Philadelphia, PA

*Corresponding author. Christopher B. Forrest, Applied Clinical Research Center, Children's Hospital of Philadelphia, 2716 South Street, Room 11-473, Philadelphia, PA 19146. Email: forrestc@email.chop.edu.

Abstract

Study Objectives: To develop and evaluate the measurement properties of child-report and parent-proxy versions of the Patient-Reported Outcomes Measurement Information System (PROMIS) Pediatric Sleep Disturbance and Sleep-Related Impairment item banks.

Methods: A national sample of 1104 children (8–17 years old) and 1477 parents of children 5–17 years old was recruited from an internet panel to evaluate the psychometric properties of 43 sleep health items. A convenience sample of children and parents recruited from a pediatric sleep clinic was obtained to provide evidence of the measures' validity; polysomnography data were collected from a subgroup of these children.

Results: Factor analyses suggested two dimensions: sleep disturbance and daytime sleep-related impairment. The final item banks included 15 items for *Sleep Disturbance* and 13 for *Sleep-Related Impairment*. Items were calibrated using the graded response model from item-response theory. Of the 28 items, 16 are included in the parallel PROMIS adult sleep health measures. Reliability of the measures exceeded 0.90. Validity was supported by correlations with existing measures of pediatric sleep health and higher sleep disturbance and sleep-related impairment scores for children with sleep problems and those with chronic and neurodevelopmental disorders. The sleep health measures were not correlated with results from polysomnography.

Conclusions: The PROMIS Pediatric Sleep Disturbance and Sleep-Related Impairment item banks provide subjective assessments of child's difficulty falling and staying asleep as well as daytime sleepiness and its impact on functioning. They may prove useful in the future for clinical research and practice. Future research should evaluate their responsiveness to clinical change in diverse patient populations.

Statement of Significance

This article describes the development of two patient-reported outcome measures of children's subjective experiences of their sleep using state-of-the-art methods developed by the National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS) program of research. The Sleep Disturbance measure assesses problems with falling and staying asleep and sleep quality. The Sleep-Related Impairment measure provides assessments of children's daytime sleepiness, difficulties waking up, and the impact of sleepiness on thinking, mood, behavior, and daily activities. Both measures provide excellent precision from low to high levels of severity, and both include a child self-report version that can be used for children 8–17 years old and a parent-proxy version that can be used for children 5–17 years old.

Key words: child; sleep health; sleep disturbance; sleep-related impairment; self-report; PROMIS; item-response theory

Submitted: 14 October, 2017; **Revised:** 14 February, 2018

Introduction

Sleep is an essential health asset for living beings [1–3], sustaining life, maintaining and restoring health, and enabling optimal functioning [4, 5]. By the age of 18, 40 per cent of a child's life should have been spent sleeping [6]. Sleeping well is important for children's cognitive functioning [7, 8], emotion regulation [9–11], energy management [12–14], mood [15, 16], immune function [17], and school performance [18]. Sleep problems affect a large share of youth [19], with higher rates among those with chronic disease and neurodevelopmental disorders [20–22]. Although sleep–wake function can be measured using polysomnography and actigraphy, these methods are not practical for research on large populations or in routine clinical settings, nor do they capture an individual's lived experience of sleep and its effects. Person-reported outcome measures fill this gap. Self-reported measures of sleep health are questionnaires that provide an individual's assessment of sleep-related experiences and the impact of those experiences on functional status.

Over 300 person-reported measures have been developed to assess the subjective experiences of children's sleep [23]. Most are parent-proxy forms rather than child self-reported. Based on a comprehensive review of these measures, Spruyt concluded that just two (Sleep Disturbance Scale for Children [24] and Sleep Disorders Inventory for Students [25]) have undergone adequate evaluation of their reliability and validity [26]. Only about a third of existing measures were developed using factor analytic methods and none was created using item-response theory (IRT). IRT methods provide information about a measure's precision for individuals with varying levels of severity. Thus, there is a need for new measures of pediatric patient-reported sleep health that are practical to use, informed by the experiences of children, parents, and sleep health experts, and developed in accordance with rigorous measurement science.

In 2004, the National Institutes of Health launched a program of research called the Patient-Reported Outcomes Measurement Information System (PROMIS) [27]. The goal of PROMIS was to provide clinicians and researchers access to efficient, precise, and valid patient-reported measures of health. PROMIS has produced over 40-item banks for adults and 20 for children. PROMIS measures are freely available and are being used in clinical and behavioral research, epidemiology and population surveillance, and clinical practice. The PROMIS suite of measures includes two sleep health instruments for adults: Sleep Disturbance (measures sleep quality and difficulties obtaining restful sleep) and Sleep-Related Impairment (measures daytime wakefulness and sleepiness) [28]. One of the major gaps in the PROMIS portfolio is the absence of sleep health item banks for children.

The purpose of this article is to describe the development and evaluation of the measurement properties of pediatric versions of the PROMIS Sleep Disturbance and Sleep-Related Impairment item banks. We started with a pool items that we previously developed using qualitative methods [23]. That process began with items from the PROMIS adult sleep health item banks [28]. Additional items were elicited through interviews with sleep medicine experts, children aged 8–17 years old, parents of children 5–17 years old, and a systematic literature review of all existing pediatric sleep health instruments. The items were evaluated for their understandability in cognitive interviews with children and parents, and poorly understood items were deleted. This content validation process resulted in 43 items that were expressions of all facets of children's lived experiences of sleep disturbance and sleep-related impairments and were well understood by children themselves [23]. Demonstrating that a self-reported outcome measures is content valid is critical for ensuring that the items measure the intended concept and that they cover the full range of experiences of individuals. The Food and Drug Administration has produced guidance for developing patient-reported outcomes that are useful for medical product studies, emphasizing the critical importance of extensive content validation [29].

Methods

The Children's Hospital of Philadelphia's (CHOP) Institutional Review Board reviewed and approved study procedures (protocol numbers 15-012503 and 16-013083). Parents provided informed consent for children, and children gave their assent.

PROMIS approach to instrument development

PROMIS uses items (i.e. questions) aggregated into item banks to measure a health domain. An item bank includes items that measure various manifestations of the domain concept from low to high levels of severity. The PROMIS approach for creating an item bank includes six major steps: (1) conceptual specification of the domain; (2) item pool construction; (3) content validation; (4) survey administration to a sample representative of the full range of severity for the domain; (5) psychometric evaluation using classical test and IRT approaches; and (6) clinical evaluation of the final products. This study focused on steps 4–6 (Figure 1); steps 1–3 are described in a separate article [23].

To develop the item banks from the item pool, we adhered to the PROMIS methodological standards for patient-reported outcome measure development [30, 31]. PROMIS uses a domain-specific approach that generates self-reported health measures that are relevant to all persons. A domain is a clinically coherent health attribute, such as sleep disturbance, that is applicable to all persons regardless of disease status. This approach contrasts with disease-attributed measures that require respondents to relate their health state to their disease (e.g. "In the past 7 days I felt sleepy *because of* my asthma"). The validity of health-state attribution is questionable because patients' responses to health surveys are influenced by many physical, psychological, environmental, and cultural factors in addition to their underlying disease(s) [32].

Item pool

We started with an item pool composed of 43 items, previously developed and evaluated for their content validity [23], representing six facets of sleep disturbance (i.e. sleep onset, sleep continuity, dreams, breathing, parasomnias, and sleep quality) and six facets of sleep-related impairment (daytime sleepiness, energy, sleep offset, and the impact of sleep on cognitive functioning, affect and behavior, and daily activities). Each facet included multiple items for relevant concepts. The items used a 7 day recall period. Response categories were frequency-based (1: never, 2: almost never, 3: sometimes, 4: almost always, and 5: always). Parent-proxy versions replaced the pronoun "I" with "my child."

General population survey sample

We recruited a sample of 1104 children aged 8–17 years old and 1477 parents of children aged 5–17 years old from the GfK Knowledge Panel, an existing dual-frame (random-digit dial and address-based) online, national panel of participants [33, 34]. We generated sampling weights to render the final sample representative of the US population; weights were used in the IRT calibration. The initial weights adjusted for oversampling of individuals living in minority communities and Spanish-language dominant areas and nonresponse. The weights were iteratively adjusted until the sample's marginal distributions of gender, census region (Northeast, Midwest, South, West), metropolitan area (Metro, Non-Metro), and household income (under $25,000, $25,000–$49,999, $50,000–$74,999, $75,000+) matched those for children aged 5–17 years in the 2015 Current Population Survey [35].

General population survey administration

Parents of 5–17 years old children participating in the internet panel were emailed an invitation to join the study. Those accepting the invitation completed a few questions to ensure eligibility (presence of a child in the correct age range, able to read English, and child did not have an intellectual or development delay). Eligible parents were asked to complete an informed consent and a sleep questionnaire. When done with this, parents asked their children to complete their portion, which started with an assent form followed by the sleep questions. Because we evaluated the two sleep health item pools plus a set of sleep practice items, we created four questionnaire forms to reduce the burden on any single child respondent and administered each to an equal number of respondents: form 1 (43 items) with questions on sleep practices and sleep disturbance; form 2 (36 items) with sleep practices and sleep-related impairment questions; form 3 (43 items) with sleep disturbance and sleep-related impairment questions; and form 4 (61 items) with sleep practices, sleep disturbance, and sleep-related impairment questions. (Note that the data for the sleep practices items will be presented in a forthcoming article.) The parents were also administered one of the four forms (range in number of items 56–81); each parent took the same content as their children plus 20 sociodemographic, clinical, and health status questions. The median time for children to complete their portion of the survey was 15 min and for parents 11 min.

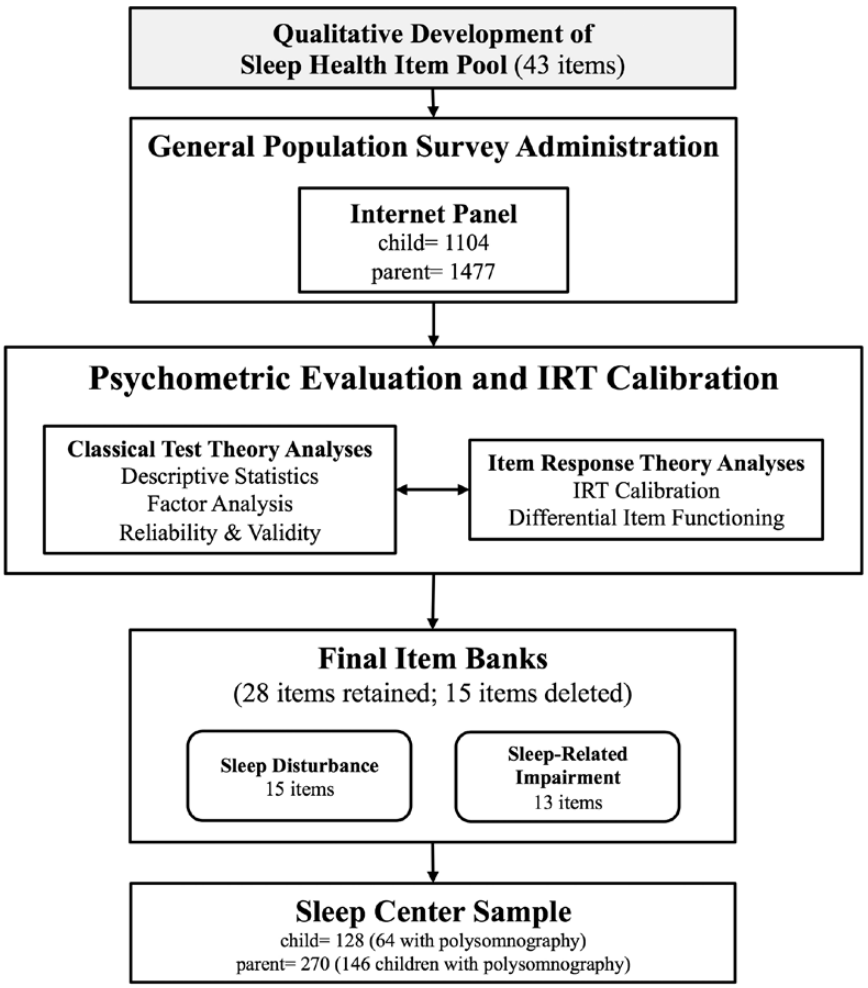

Figure 1. Psychometric evaluation and IRT calibration of the item banks. The flow chart illustrates the major steps in the development and evaluation of the PROMIS Pediatric Sleep Disturbance and Sleep-Related Impairment Item Banks, and items at each stage of development. The first box is grey shaded to indicate that the qualitative development of the item pool preceded this work.

Classical test theory analyses

At the scale level, we examined the range of the IRT-based scores (see below for scoring details) and the percentage of individuals at the floor and ceiling of these scores. Reliability was evaluated with an IRT-based estimate of Cronbach's α called marginal reliability [36].

Using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), we evaluated the structural validity of the item pool testing the assumption of two factors (i.e. sleep disturbances and sleep-related impairments). Both EFA and CFA were done with the weighted least squares means and variance adjusted estimator and Promax rotation using Mplus 6.1. Following the guidance of the PROMIS scientific standards [31], we examined the comparative fit index (CFI > 0.95 for good fit), the Tucker-Lewis index (TLI > 0.95 for good fit), and root mean square error of approximation (RMSEA < 0.06 for good fit). To test the assumption of two factors, we fit a CFA with the full-item pool (expected poor fit), a two-factor EFA (items expected to load on their hypothesized factor), and CFAs for items expected to measure sleep disturbance and sleep-related impairment [23]. Items were considered locally dependent if their residual correlations (the correlation between items that remains after removing their common correlation with the underlying latent variable) were ≥0.20 in the single-factor CFA model [37], and when present, one of the correlated items was removed. Graphs of item–mean scores conditional on the total test scale score minus the item score were examined to confirm item monotonicity (i.e. the probability of item endorsement should increase as the measured trait increases), and non-monotonic items, if found, were removed.

Differential item functioning

Differential item functioning (DIF) refers to the possibility that two individuals with equivalent levels of a given latent variable (e.g. sleep disturbance) answer questions differently as a function of another variable (e.g. age) assumed to be extraneous to the testing situation. DIF can bias scale scores and lead to spurious study conclusions. To identify potential item bias by child age, sex, race, and ethnicity, DIF analyses for these variables were done using the Lordif package in R [38]. Lordif regresses an item's ordered responses on an IRT-derived scale score, the indicator variable (such as age or sex) and the interaction between the two to test for both uniform and nonuniform DIF. Items that showed a 1 per cent change in the McFadden pseudo R^2 measure were considered to demonstrate DIF [39] and, if present, were removed from the item pool.

IRT analysis

The final item pool was calibrated using Samejima's Graded Response Model [40]. In IRT, calibration refers to estimating discrimination and threshold parameters for each item using a sample's item responses. The discrimination parameter (or slope, designated *a*) measures the ability of an item to differentiate respondents by severity—i.e. higher versus lower levels of sleep disturbance or sleep-related impairment. The IRT model also produces threshold parameters (referred to as item difficulty, designated *b*), which correspond to the *difficulty* of endorsing item responses. Thresholds indicate the point on the latent variable where a respondent is more likely to respond to a higher versus lower response option—e.g. "almost never" versus "never." For an item with five response options, the IRT model produces four item threshold parameters. A major advantage of IRT calibration of item pools to form item banks is that the estimation of parameters is sample independent, which is in contrast to scales that are formed with classical test theory analyses only [41]. In recruiting samples for IRT calibration, it is important that the sample has a wide range of distribution of the latent variable (from low to high severity levels).

Because we calibrated items using a weighted representative sample of the general US pediatric population, scale scores can be interpreted relative to the general US pediatric population aged 5–17 years old. We implemented these analyses using Samejima's cumulative logit–graded response model in Stata 14.2 with maximum likelihood estimation [42]. The parameters were estimated with mean and variance adaptive Gauss–Hermite quadrature using seven quadrature points to compute the log likelihood.

Scoring

PROMIS measures are scored in the direction of their concept's name, so higher sleep disturbance and sleep-related impairment scores indicate poorer sleep health. After finalizing item parameters, we used Firestar v1.3.2, an R-based simulation software [43], and estimated full bank and short form scores using Bayesian Expected A Posteriori estimation [44]. Expected A Posteriori scoring uses an individual's pattern of responses and the model's parameters to estimate an individual's score, called theta, which is set to a mean of 0 with a standard deviation of 1. The PROMIS convention is to transform θ scores to T-scores by multiplying by 10 and adding 50, producing a T-score distribution with a mean of 50 and standard deviation of 10. Thus, a score of 40 is 1 standard deviation below the national average, and vice versa for a score of 60. This type of scoring provides an intuitive and easy-to-use metric and allows scores to be converted to percentile ranks using T-distribution tables. Once we constructed the T-scores, we evaluated the distribution of scores in the general population sample to ensure that we had a representative range of severity of sleep disturbance and sleep-related impairment.

To aid in the interpretation and clinical utility of sleep disturbance and sleep-related impairment scores, we produced conversion tables that map total sum score to Expected A Posteriori T-scores using the Lord and Wingersky algorithm [45, 46]. The conversion tables allow a simple total sum score to be placed on the PROMIS T-score metric. In situations where the full instrument (full bank, 8-item short form, or 4-item short form) has missing item responses, the conversion tables can still be used by generating a pro-rated total sum score using the following formula: raw score × number of items on the instrument)/number of items answered by the participant. A minimum of six items for the 8-item short form and three for the 4-item form are required to generate a score using this look-up table method. (See Supplementary Material for the conversion tables (Appendix A1) and scoring manual (Appendices A2 and A3).)

Short forms

Fixed-length short forms can be constructed from an item bank. Regardless of the specific items administered, the scores from

any short form are comparable, because they are calibrated on the same IRT-based metric. For each domain, PROMIS has produced recommended short forms that minimize respondent burden, but maintain measurement precision for a wide range of severity of the measured concept [47]. Items from item banks may also be given as a computerized adaptive test, which use algorithms to choose which items to administer based on the respondent's prior responses [48, 49]. We selected short form items based on the IRT item discrimination and threshold parameter estimates and ensured that the content validity of the measure was maintained by assuring that items selected represented the domain's conceptual facets. We gave priority to items that were also included in the PROMIS adult sleep health item banks. For the 4-item short form, items with high levels of item discrimination were preferred, because this parameter is reflective of their ability to differentiate among individuals at different levels of sleep health.

Sleep center sample

We collected data from a sample of children seen at the Children's Hospital of Philadelphia (CHOP) Sleep Center, which provides diagnostic and management services for children with sleep disorders. We reviewed the appointment schedule up to 7 days before a scheduled visit or polysomnography evaluation to identify eligible children. To confirm eligibility, using the same criteria as the general population sample, we also reviewed charts of scheduled children. Parents of eligible children were either recruited during their child's scheduled appointment or sent an email including study information and a link to the online questionnaire. Those joining the study as a result of the online invitation completed the questionnaire on their home computers using the same procedures described for the general population survey. Within the clinic, we used iPads to administer the questionnaires. Median time to complete the questionnaires was 12 min for parents and 10 min for children. Parents completed sociodemographic, health condition, and sleep medication use questions along with parent-proxy versions of the 8-item sleep disturbance, 8-item sleep-related impairment, and 4-item PROMIS Pediatric fatigue short forms. When done with their part of the questionnaire, parents asked their children to complete their portion of the survey; children completed the same content as parents excluding the health condition checklist and sleep medication questions.

Polysomnography

The children recruited from the sleep laboratory underwent polysomnography primarily to rule-out obstructive sleep apnea. The laboratory used a Rembrandt polysomnography system (Embla, Broomfield, CO), which recorded the following parameters: electroencephalogram (C3/A2, C4/A1, F_3A_2, F_4A_1, O1/A2, O2/A1), left and right electrooculograms, submental electromyogram (EMG), chest and abdominal wall motion using respiratory inductance plethysmography, heart rate by electrocardiogram, arterial oxygen saturation (SpO_2) by pulse oximetry; end-tidal PCO_2 measured by infrared capnometry (Novametrix Medical System, Inc., Wallingford, CT), airflow using a 3-pronged thermistor (Pro-Tech Services, Inc., Mukilteo, WA), nasal pressure by a pressure transducer (Pro-Tech Services, Inc., Walnut Cove, NC), and bilateral tibialis anterior EMG. Participants were continuously observed by a polysomnography technician and were recorded on video with the use of an infrared video camera. Studies were scored using standard pediatric sleep scoring criteria [50]. We evaluated eight polysomnography variables: total sleep time, sleep latency, sleep efficiency, awakenings after sleep onset, periodic limb movement index, obstructive apnea–hypopnea index, minimum oxygen saturation, and the arousal index. The questionnaire data were collected within 7 days of the polysomnography tests.

Preliminary validity evidence

In the general population, we expected poorer sleep health scores to be associated with older children [51], parent-reported sleep problem [52], parent-reported poor general health status [53], and presence of chronic and neurodevelopmental conditions [20–22]. Children from the sleep center were expected to have poorer sleep health scores than counterparts from the general population. These known-group validity analyses were done using Cohen's *d* statistic to produce an effect size estimate [54], computed as the difference in means divided by the standard deviation of the combined sample. We considered Cohen's *d* values of 0.2 to <0.5 to be small effects, 0.5 to <0.8 moderately sized, and 0.8 and above large effects.

To assess convergent validity, we evaluated the correlations between the two new sleep measures and existing measures of sleep problems, daytime sleepiness, and fatigue in the sleep center sample. After our review of over 300 sleep health measures [23], we concluded that none provided valid assessments of the same sleep health concepts that our item banks measured; thus, we expected moderate correlations only between the legacy and the newly developed measures. Both legacy measures we gave included items related to sleep disturbance and sleep-related impairment. The Children's Sleep Habits Questionnaire Daytime Sleepiness Subscale (minus the open response question regarding child wake time) is a 9-item scale that assesses *parent-reported* sleep awakening time, difficulties waking in the morning (which can be a sign of poor quality or deficient sleep), and daytime sleepiness during specific activities (i.e. watching TV, riding in car) [55]. We also gave the School Sleep Habits Survey, which is a self-reported measure of sleep patterns, including descriptions of school- and weekend-night sleep habits and daytime sleepiness. To assess fatigue, we used a 4-item subset of the Pediatric Fatigue Short Form 10a items. The items addressed general fatigue experiences (I felt weak, I got tired easily) and fatigue impact (I was so tired it was hard for me to pay attention, being tired made it hard for me to play or go out with friends), and were expected to correlate more strongly with *Sleep-related Impairment* than *Sleep Disturbance* [56, 57]. We anticipated higher correlations between child-reported measures than between parent-reported and child-reported measures.

Finally, we computed Spearman rank correlation coefficients between the polysomnography variables and the sleep health scales. Based on prior research [58], our hypothesis was that the sleep health measures, which are subjective assessments of sleep experiences and sleep-related impairments, would be weakly or not associated with the polysomnography variables which evaluate physiological sleep parameters.

Results

For the general population sample, email invitations were sent to 4419 parent participants in the internet panel who also had a child between the ages of 5 and 17. Of those, 2461 parents accepted the invitation to learn more about the study. A total of 1614 parents were deemed eligible for the study. Ineligible parents included those who did not provide consent (n = 59), did not enter the survey after consenting (n = 81), were not able to participate because the quota for their child's age had already been met (n = 344), or did not meet eligibility criteria (38 reported a cognitive limitation or developmental delay that would prevent their child from completing a survey, and 111 parents did not provide permission for their child to participate). Of the 1614 eligible parents, 1477 completed the parent survey (92%). Of the 1234 eligible children (had an eligible parent, and were between the ages of 8 and 17), 1104 completed the child survey (89%).

For the sleep center sample, 1214 parents were sent an invitation by email. Of those, 399 clicked on a link to learn more about the study; 89 did not complete the consent form and 1 parent indicated that they did not speak or read English well enough to complete a survey. Of the 309 eligible parents, 270 completed the parent survey (87%). Of the 160 eligible children (had an eligible parent, and were between the ages of 8 and 17), 128 completed the child survey (80%). Polysomnography data were collected for 146 children in the parent-report sample and 64 in the child-report sample. Sociodemographic and clinical characteristics for the general population and clinical samples are shown in Table 1. The former sample can be considered a national sample of children in the general population while the latter a convenience sample of children with sleep problems.

Item deletion

Item bank development began with a pool of 43 items that were previously produced using qualitative methods [23]. Although we expected that the pool would have two factors, one for sleep disturbance and another for sleep-related impairment items, initial analysis examined these items as a single underlying dimension. Internal consistency reliability was high, consistent with the large number of items (child-report 0.97; parent-proxy 0.96). However, CFA indicated marginal model fit when treating all 43 items as a single scale (child-report CFI 0.90, TLI 0.89, RMSEA 0.09; parent-proxy CFI 0.89, TLI 0.88, RMSEA 0.07).

In a two-factor EFA, we identified six items that were positively worded and did not load on their hypothesized factor; each measured satisfaction with sleep: *got enough sleep*; *felt refreshed when woke up*; *satisfied with sleep*; *happy with sleep*; *slept well;* and *woke up feeling ready to start the day*. These six items were excluded. All other items were reflective of a disturbance in sleep, daytime sleepiness, or an impairment resulting from poor sleep. When we repeated the two-factor EFA, the remaining 37 items (20 sleep disturbance items and 17 sleep-Related impairment items) loaded on their hypothesized factor.

We deleted additional sleep disturbance items because of low factor loadings (<0.60) from single factor CFAs: *wet the bed* (0.27) and *take medicine to fall asleep* (0.58). Other items were deleted because they had high residual correlation (>0.20) with one or more items: *scary dreams, bad dreams*, and *hard time falling asleep because felt upset*. These item deletions resulted in a final set of 15 sleep disturbance items, which had adequate model fit: child-report, CFI 0.96, TLI 0.95, RMSEA 0.12; parent-proxy, CFI 0.95, TLI 0.95, and RMSEA 0.12. Of these 15 items, nine are also in the PROMIS Adult *Sleep Disturbance* Item Bank.

Two sleep-related impairment items were excluded before psychometric analysis because they are included in the PROMIS Pediatric Fatigue item bank: *felt tired* and *had enough energy*. The item *difficulty waking up* had a high residual correlation with *felt sleepy when woke up*, so the former was deleted. Finally, *slept during the day* had nonmonotonic thresholds and low factor loading (0.58), so it was deleted, leaving a total of 13 items. Model fit for these 13 items was adequate: child-report, CFI 0.97, TLI 0.96, and RMSEA 0.12; parent-proxy, CFI 0.98, TLI 0.97, and RMSEA 0.11. Of these 13 items, seven are also in the PROMIS Adult *Sleep-Related Impairment* Item Bank.

Results from the final CFA analyses for the item banks, child-report and parent-proxy versions, item-level descriptive statistics, and a two-factor EFA done in the final set of items are shown in Supplementary Material (Appendices A4–A6).

IRT item calibration

The graded response model was fitted to each of the item pools for child-report and parent-proxy versions separately (Tables 2 and 3). The item calibrations produce the *a* parameter (higher values indicate greater item information) and four threshold parameters, *b1-b4*, for item-category responses (lower thresholds detect lower severity levels and vice versa). The *Sleep Disturbance* threshold parameters ranged for child-report from –0.7 to 3.8 (child-report) and –0.5 to 4.6 (parent-proxy), a span of 4.5 and 5.1 standard deviations. The *Sleep-Related Impairment* threshold parameters ranged for child-report from –1.2 to 3.1 (child-report) and –0.8 to 3.9, a span of 4.3 and 4.7 standard deviations. Tables 2 and 3 also indicate which items were selected for the short forms, and which of the pediatric items are also included in the parallel PROMIS adult item banks. The final set of items in both item banks, child and parent versions, is shown in Supplementary Material (Appendix A7).

Descriptive statistics and reliability

The *Sleep Disturbance* child-reported scale score range was 33.8 to 84.9 (5.1 standard deviations), 10.0 per cent of the sample had the lowest score (floor effect) and 0.1 per cent had the highest score (ceiling effect). The *Sleep-Related Impairment* child-reported scale score range was 34.7 to 76.1 (4.1 standard deviations), 13.0 per cent of the sample had the lowest score, and 0.1 per cent had the highest score. The wide range of scores indicates that the sample was representative of the full range of the latent variable and 87%–90% of children experience some degree of sleep disturbance and sleep-related impairment.

The marginal reliabilities for the item banks and their short forms were high (Table 4). The marginal reliability by T-score estimates (from low-to-high severity of sleep health) is shown for the child-report editions of both item banks in Figure 2. The precision of the item bank estimates was high (marginal reliability > 0.90) from T-scores in the range of 40 to a high of 85. Reliability declined more precipitously at the tails for the SF4 compared with the SF8.

Table 1. Parent-reported sociodemographic and clinical characteristics for the general population and sleep center child-report and parent-report samples

Characteristic	General population		Sleep center	
	Child-report*n* (%)	Parent-report*n* (%)	Child-report*n* (%)	Parent-report*n* (%)
Overall sample	1104 (100)	1477 (100)	128 (100)	270 (100)
Sociodemographics				
Child age, years				
5–7	—	373 (23)	—	98 (36)
8–12	604 (48)	604 (37)	78 (61)	108 (40)
13–17	500 (52)	500 (40)	50 (39)	64 (24)
Child female gender	540 (49)	720 (49)	72 (56)	136 (50)
Child race				
White	833 (71)	1,118 (70)	60 (47)	144 (53)
African American/Black	103 (15)	137 (15)	49 (38)	86 (32)
Asian	38 (3)	51 (3)	4 (3)	10 (4)
Other/more than one	130 (11)	171 (12)	15 (12)	30 (11)
Child Hispanic ethnicity	220 (24)	293 (24)	13 (10)	24 (9)
Parental education attainment				
High school	382 (36)	463 (33)	25 (20)	44 (16)
Some college	341 (29)	448 (28)	39 (31)	90 (33)
College or higher	381 (35)	566 (39)	64 (50)	136 (50)
Family financial strain: somewhat/very hard to pay for food, housing, medical care and heating	383 (36)	511 (36)	69 (54)	138 (51)
Somewhat/very hard to pay for food, housing, medical care and heating				
Family food insecurity: often/sometimes not enough to eat	106 (10)	147 (11)	10 (8)	28 (10)
Often/sometimes not enough to eat				
Health conditions				
Asthma	154 (14)	188 (14)	64 (50)	117 (43)
Attention deficit hyperactivity disorder	150 (13)	176 (12)	36 (28)	73 (27)
Obstructive sleep apnea	9 (1)	14 (1)	39 (31)	90 (33)
Health status				
Poor/fair general health status	33 (3)	40 (3)	36 (28)	68 (25)
Health, mental health, or developmental condition lasting 12 months or more	185 (16)	238 (16)	77 (60)	169 (63)
Has a sleep problem	87 (8)	108 (7)	107 (84)	216 (80)
Obese[†]	161 (16)	226 (17)	53 (41)	109 (40)
Medications currently taken for sleep				
Melatonin	67 (6)	84 (6)	25 (20)	50 (19)
Other medicine	24 (2)	27(2)	12 (9)	22 (8)

The general population sample is weighted to be representative of the 2015 US Current Population Survey distributions of children's gender, census region, metropolitan area, and household income.

[†]Defined as ≥95 percentile (CDC 2000 growth charts) for age and sex calculated from parent-reported height, weight, and child age.

Preliminary validation evidence

For the clinical sample, we converted T-scores to percentile ranks using T-distribution tables. The *Sleep Disturbance* means for the sleep center clinical sample were substantively higher than the general population: child-report 58.4 (80th percentile) and parent-proxy 62.4 (89th percentile). The *Sleep-Related Impairment* means were child-report 56.9 (75th percentile) and parent-proxy 61.8 (88th percentile).

In the general population sample, we evaluated sociodemographic and clinical known-group differences (Table 5). Sleep disturbance and sleep-related impairment increased with age and financial strain, and, as expected, was no different by gender, race, or ethnicity. There were moderate effects observed for parent-reported chronic condition (Does your child have a health, mental health, or developmental condition that is expected to last 12 months or more?) and poor/fair general health status, and large effects observed for melatonin use (Does your child currently take melatonin to help him/her sleep), use of other sleep medications (Does your child currently take any other medication to help him/her sleep), and presence of a sleep problem (Do you think your child has a sleep problem?).

The correlations between child-reported *Sleep Disturbance* and *Sleep-Related Impairment* and other measures of sleep health and fatigue are shown in Table 6. The two PROMIS pediatric sleep health measures were moderately correlated (0.57). Child-reported *Fatigue* correlated more strongly with *Sleep-Related Impairment* than *Sleep Disturbance*. The parent-reported Children's Sleep Habits Questionnaire Morning Waking/Daytime Sleepiness subscale was weakly correlated with the *Sleep Disturbance* and moderately correlated with *Sleep-Related Impairment*. Finally, the child-reported School Sleep Habits Survey was moderately correlated with both *Sleep Disturbance* and *Sleep-Related Impairment*.

For the polysomnography and sleep health scale correlations, we evaluated 32 comparisons: two child-report scales and two

Table 2. Item calibrations for the PROMIS Pediatric Sleep Disturbance Item Banks, Child-Report and Parent-Proxy Versions, and Assignments to the 8- and 4-item Short Forms (SF8 and S4)

Items	Facet	SF8	SF4	a	b1	b2	b3	b4
Difficulty falling asleep†	Sleep onset	X	X					
Child-report				3.30	−0.34	0.57	1.48	2.02
Parent-proxy				3.92	0.27	1.10	1.97	2.55
Easy to fall asleep†	Sleep onset							
Child-report				2.25	−0.69	0.55	1.55	2.31
Parent-proxy				2.11	−0.49	0.92	1.82	2.55
Worried about not being able to fall asleep†	Sleep onset	X						
Child-report				2.24	0.55	1.31	2.49	3.31
Parent-proxy				2.47	0.89	1.69	2.63	3.49
Trouble falling asleep because worried	Sleep onset							
Child-report				2.02	0.41	1.29	2.60	3.81
Parent-proxy				2.72	0.66	1.51	2.83	3.55
Trouble stopping thoughts at bedtime†	Sleep onset							
Child-report				2.02	−0.35	0.41	1.55	2.28
Parent-proxy				1.98	−0.07	0.87	2.07	2.76
Trouble falling asleep could not get comfortable†	Sleep onset							
Child-report				2.21	−0.13	0.63	1.89	2.66
Parent-proxy				2.32	0.40	1.33	2.70	3.93
Took long time to fall asleep	Sleep onset	X						
Child-report				3.47	−0.38	0.43	1.45	2.03
Parent-proxy				2.79	−0.12	0.88	1.92	2.75
Slept through the night	Sleep continuity	X	X					
Child-report				2.05	−0.02	1.20	2.09	2.72
Parent-proxy				1.71	0.35	1.75	2.47	2.80
Woke up at night and worried about not going back to sleep	Sleep continuity							
Child-report				2.17	0.57	1.33	2.59	3.28
Parent-proxy				2.26	0.91	1.90	3.30	4.62
Woke up at night and had trouble falling back to sleep†	Sleep continuity	X						
Child-report				2.38	0.17	0.92	2.07	2.84
Parent-proxy				2.18	0.41	1.52	2.84	4.34
Woke up too early and could not fall back asleep	Sleep continuity							
Child-report				1.68	−0.08	0.84	2.33	3.39
Parent-proxy				1.57	0.18	1.45	3.22	4.41
Tossed and turned at night†	Sleep quality	X						
Child-report				2.10	−0.18	0.68	1.70	2.32
Parent-proxy				1.69	0.10	1.15	2.39	3.25
Slept poorly	Sleep quality							
Child-report				3.46	0.02	0.89	1.73	2.47
Parent-proxy				3.26	0.21	1.26	2.21	2.86
Problem with sleep†	Sleep quality	X	X					
Child-report				3.47	0.21	0.95	1.91	2.54
Parent-proxy				3.61	0.32	1.21	2.09	2.54
Trouble sleeping†	Sleep quality	X	X					
Child-report				4.48	0.04	0.78	1.72	2.42
Parent-proxy				6.25	0.34	1.17	1.97	2.54

IRT-graded response model item calibrations produce the *a* parameter (higher values indicate greater item information) and four threshold parameters, *b1*-*b4*, for item-category responses (lower thresholds detect lower severity levels and vice versa).
†Indicates that the item is also included in the PROMIS Sleep Disturbance item bank for adults.

parent-report scales correlated with eight polysomnography measures. None was statistically significant; the average *p*-value for these comparisons was 0.49 and the range was 0.06–0.97.

Discussion

This article describes the development of the PROMIS Pediatric *Sleep Disturbance* and *Sleep-Related Impairment* item banks. The qualitative development of the item pool is described elsewhere [23]; it involved extensive content validation to ensure that the items measure the intended concepts and that the item pool covers the full range of sleep disturbance and sleep-related impairment experiences. The pool of 43 items that we used in this study underwent extensive psychometric testing, IRT item calibration, and evaluation of the measures' validity. The work was done in general population (1100 children and 1400 parents) and sleep referral clinic (134 children and 286 parents) samples. The final *Sleep Disturbance* item bank includes 15 items that assess difficulties with sleep onset, sleep continuity, and sleep quality. The *Sleep-Related Impairment* item bank includes 13 items that assess daytime sleepiness, sleep offset, and the impact of sleepiness on cognitive functioning, affect and behaviors, and daily activities.

Table 3. Item calibrations for the PROMIS Pediatric Sleep-Related Impairment item banks, Child-Report and Parent-Proxy Editions, and Assignments to the 8- and 4-item Short Forms (SF8 and S4)

Items	Facet	SF8	SF4	a	b1	b2	b3	b4
Sleepy during the daytime†	Daytime sleepiness	X	X					
Child-report				2.67	–0.56	0.21	1.60	2.15
Parent-proxy				2.32	–0.47	0.71	2.38	3.30
Trouble staying awake during the day†	Daytime sleepiness	X						
Child-report				3.33	0.12	1.02	1.98	2.59
Parent-proxy				3.26	0.68	1.55	2.69	na‡
Hard time concentrating because sleepy†	Impact: Cognitive	X	X					
Child-report				2.87	–0.11	0.68	2.00	2.83
Parent-proxy				4.21	0.39	1.22	2.42	3.40
Bad mood because sleepy	Impact: Affect & Behavior	X						
Child-report				2.83	0.00	0.79	2.08	2.58
Parent-proxy				1.94	–0.40	0.68	2.47	3.91
Got mad easily because sleepy	Impact: Affect & Behavior							
Child-report				2.62	0.07	0.82	1.97	2.53
Parent-proxy				2.57	0.06	0.97	2.49	3.00
Trouble controlling feelings because sleepy	Impact: Affect & Behavior							
Child-report				2.52	0.20	0.99	2.20	3.12
Parent-proxy				2.53	0.24	1.23	2.67	na‡
Still felt sleepy when woke up†	Sleep Offset							
Child-report				1.71	–1.23	–0.30	1.18	2.03
Parent-proxy				1.46	–0.77	0.44	2.05	3.28
Hard time getting things done because sleepy†	Impact: Activities	X	X					
Child-report				3.88	0.04	0.87	1.80	2.69
Parent-proxy				4.37	0.40	1.33	2.46	na‡
Problems during the day because of poor sleep†	Impact: Activities	X	X					
Child-report				5.01	0.34	1.01	2.00	na‡
Parent-proxy				3.95	0.39	1.34	2.30	2.93
Hard to have fun because sleepy	Impact: Activities	X						
Child-report				3.48	0.51	1.24	2.50	3.15
Parent-proxy				3.24	0.67	1.68	2.81	na‡
Trouble sitting still because sleepy								
Child-report				3.05	0.61	1.37	2.23	2.85
Parent-proxy				2.89	0.74	1.76	2.73	3.76
Could not keep eyes open during the day	Daytime sleepiness	X						
Child-report				3.13	0.46	1.25	2.39	3.05
Parent-proxy				2.64	0.59	1.78	3.10	3.59
Daytime activities disturbed by poor sleep†	Impact: Activities							
Child-report				4.26	0.38	1.11	2.08	2.51
Parent-proxy				5.96	0.62	1.40	2.64	na‡

IRT-graded response model item calibrations produce the *a* parameter (higher values indicate greater item information) and four threshold parameters, *b1*-*b4*, for item-category responses (lower thresholds detect lower severity levels and vice versa).

†Indicates that the item is also included in the PROMIS Sleep Disturbance item bank for adults.

‡The b4 threshold is not reported because of sparse cell size for the fifth (highest) response category.

Table 4. Marginal reliabilities for the item banks and short forms.

	Marginal reliabilities	
Version	Sleep disturbance	Sleep-related impairment
Child-report		
Item bank	.93	.92
SF8	.91	.92
SF4	.88	.88
Parent-proxy		
Item bank	.92	.91
SF8	.90	.90
SF4	.87	.87

For each item bank, we created 8- and 4-item short forms. Both provide excellent precision across a large range of the sleep health continua, but the SF4 has more error associated with scores at the tails of the distribution. In addition, both banks include a child self-report version that can be used for children aged 8–17 years old and a parent-proxy version that can be used for children aged 5–17 years old.

The validity of the measures is supported in a variety of ways. Extensive content validation, involving elicitation of lived experiences from children and parents, drafting of the theoretical construct by experts, and cognitive debriefing to ensure that the items are understandable was done on all the items evaluated in this study. The structural validity of the measures was demonstrated in factor analyses that showed that indeed they measured

Table 5. Known-group validity of the Child-Reported Sleep Disturbance and Sleep-Related Impairment Scales

Characteristic [2]	Sleep Disturbance (mean)	Sleep-Related Impairment (mean)
Sociodemographics		
Age		
13–17 years	51.1	51.5
8–12 years	48.8	48.0
Effect size	0.24	0.38
p-Value	< 0.001	< 0.001
Financial strain		
Yes	51.3	51.0
No	49.0	48.9
Effect size	0.25	0.23
p-Value	< 0.001	0.001
Health status		
Has sleep problem		
Yes	55.6	54.3
No	45.7	46.4
Effect size	1.21	0.92
p-Value	< 0.001	< 0.001
General health		
Poor/Fair	55.8	56.4
Good/Very Good/Excellent	49.7	49.4
Effect size	0.65	0.75
p–Value	0.007	0.004
Chronic health condition		
Yes	54.0	52.7
No	49.0	49.0
Effect size	0.53	0.39
p-Value	< 0.001	< 0.001
Attention deficit hyperactivity disorder		
Yes	52.8	52.1
No	49.4	49.2
Effect size	0.37	0.31
p-Value	< 0.001	0.002
Medications for insomnia		
Uses melatonin for sleep		
Yes	58.2	55.6
No	49.3	49.2
Effect size	0.97	0.69
p-Value	< 0.001	< 0.001
Uses other medication for sleep		
Yes	59.7	55.9
No	49.7	49.5
Effect size	1.07	0.69
p-Value	< 0.001	0.007

Data are from the general population sample (n = 940).
Only those comparisons with an effect size of >0.20 and a p-value of 0.01 or smaller are shown in the table. Variables not meeting these criteria are gender, race, ethnicity, food insecurity, obesity, obstructive sleep apnea, and asthma.
†All characteristics, except general health status, are obtained from parental report, whereas the sleep health scale scores are derived from the child-reported questionnaire.
‡NS refers to not significant.

single and distinct concepts. The measures were shown to have excellent reliability and precision across a wide range of the latent variables. We found that the measures differed significantly between groups hypothesized to have differences in sleep. Convergent validity was supported by correlations of the new measures with existing scales of fatigue and sleep health, and by correlations between child- and parent-proxy reports on the same scale. Although the PROMIS sleep health item banks have excellent precision across a wide range of severity, we have not evaluated their responsiveness to clinical change. Longitudinal studies that examine sleep health and its relationship to clinical change in diverse populations a direction for future research.

Of the final 28 items in the two item banks, 16 are also included in the parallel adult PROMIS item banks. This overlap provides further support for the validity of the pediatric item banks. Furthermore, the pediatric short forms have substantial overlap with their adult counterparts (Tables 2 and 3). Because the pediatric and adult item banks are conceptually harmonized and share several of the same items, it will be possible in the future to statistically link the pediatric and adult versions, creating a common metric (both item banks scored on the same scale) and a life course sleep health measurement system from ages 5 to 85 years. Methods for linking alternative forms (i.e. instruments) for the same test (i.e. sleep health construct) are well established in the field of educational measurement [59–64], where, for example, they have been used to equate different forms of standardized tests across grade levels. More recently linkage methods have been applied to patient self-report health status measures, where, for instance, alternative measures for depressive symptoms are linked to the same common metric [65].

The PROMIS Pediatric sleep health measures assess a child's lived experiences of sleeping–wake function. That is, they assess children's subjective experiences of falling and staying asleep, waking up, and the effects of being sleepy on daytime activities. They do not measure the timing of sleep, which can be evaluated with sleep diaries or actigraphy, and are not intended to be a screening or diagnostic tool for sleep disorders. In fact, items representing the conceptual facets of dreams, sleep-related breathing problems, and parasomnias were deleted from the final versions of item banks because they did not adhere to IRT model assumptions regarding unidimensionality and monotonicity. Furthermore, we found no statistically significant associations between the sleep health measures and polysomnography, further emphasizing that the self-report measures assess unique attributes of sleep–wake function—i.e. the lived experiences of sleep well/poorly and the impact of poor sleep/sleepiness on daytime functioning—rather than physiological sleep parameters.

The general population sample enrolled participants from a large, national internet panel. Advantages of internet panels include the efficiency with which large amounts of data can be collected, the accessibility of diverse populations, and standardization of the data collection process [66]. Because not all individuals and families have access to home computers and participants of internet panels tend to have a higher socioeconomic status than the general US population [67], there are limitations to the representativeness of the sample; nonetheless, it is fair to say that the PROMIS Pediatric sleep health measures have been standardized to a national, highly diverse sample.

In conclusion, the PROMIS Pediatric *Sleep Disturbance* and *Sleep-Related Impairment* item banks provide assessments of a child's difficulties falling and staying asleep as well as daytime sleepiness and its impact on daytime functioning. The scales have excellent precision across a wide range of the latent variables and initial evidence for their validity. The child-report edition can be used for children aged 8–17 years old, and a Parent-Proxy edition is available for children aged 5–7 years old. They may prove useful in the future for clinical research and practice.

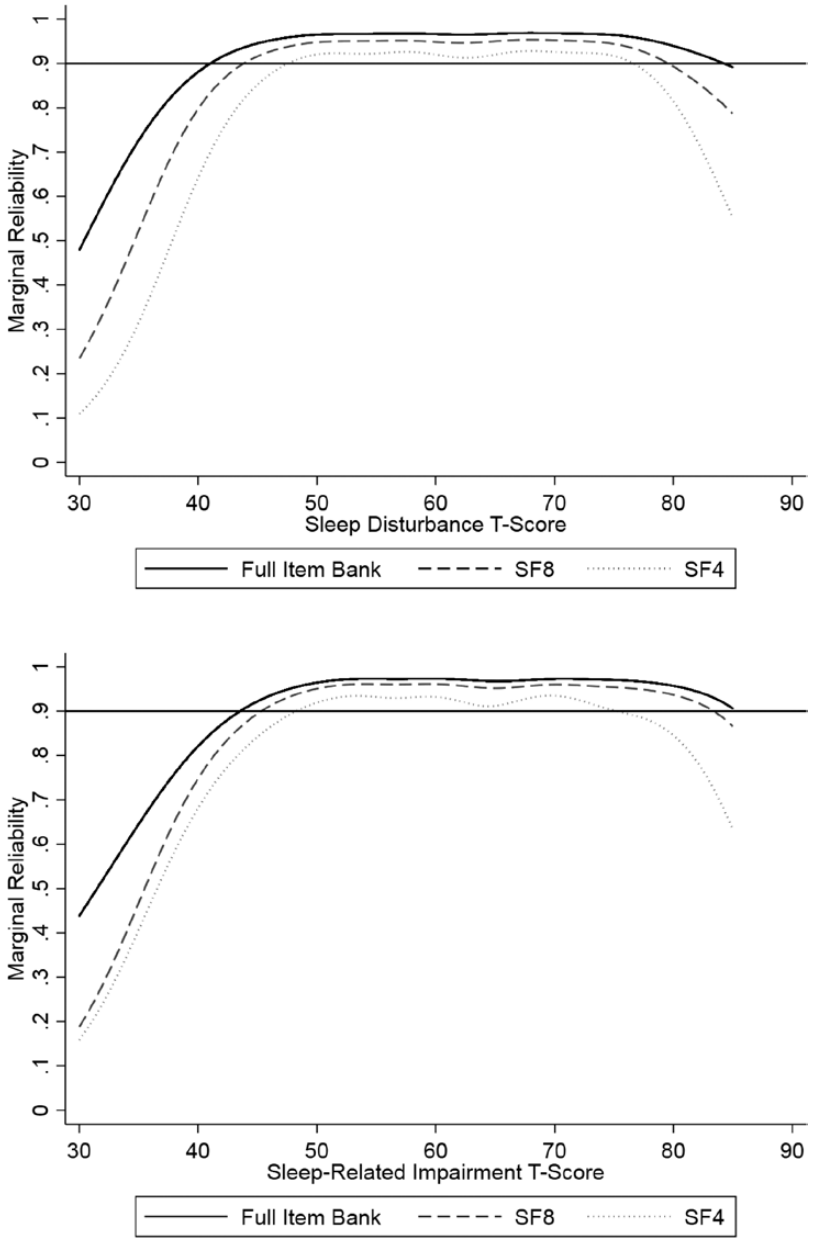

Figure 2. Marginal Reliabilities of the Sleep Health Item Banks, Child-Report Versions.For each of the two item banks and their short forms, the figures show the reliability of the measure (y-axis) across the full range of severity (x-axis). The horizontal line indicates a reliability of 0.90.

Table 6. Correlations of the Child-Reported Sleep Disturbance and Sleep-Related Impairment scales with other measures. Data are from the Sleep Referral Center Clinical Sample that included 128 children and 270 parents

	Child-Reported	
Measures	Sleep Disturbance	Sleep-Related Impairment
Child-Reported		
PROMIS Pediatric Sleep Disturbance	—	0.57
PROMIS Pediatric Sleep-Related Impairment	0.57	—
PROMIS Pediatric Fatigue	0.46	0.77
School Sleep Habits Survey	0.49	0.59
Parent-Reported		
PROMIS Pediatric Sleep Disturbance	0.77	0.42
PROMIS Pediatric Sleep-Related Impairment	0.46	0.71
PROMIS Pediatric Fatigue	0.42	0.67
Sleep Habit Questionnaire Daytime Sleepiness Sub-Scale	0.20	0.40

Supplementary Material

Supplementary material is available at *SLEEP* online.

Funding

This work was funded by a grant from the Patient-Centered Outcomes Research Institute (PCORI-D-17-00187; PI Christopher Forrest).

Note

Conflict of interest statement: None declared.

References

1. Forrest CB. A living systems perspective on health. *Med Hypotheses*. 2014;**82**(2):209–214.
2. Cirelli C, *et al*. Is sleep essential? *PLoS Biol*. 2008;**6**(8):e216.
3. Robles TF, *et al*. Restorative biological processes and health. *Soc Personal Psychol Compass*. 2011;**5**(8):518–537.
4. Mignot E. Why we sleep: the temporal organization of recovery. *PLoS Biol*. 2008;**6**(4):e106.
5. Frank MG. The mystery of sleep function: current perspectives and future directions. *Rev Neurosci*. 2006;**17**(4):375–392.
6. Mindell JA, *et al*. *A Clinical Guide to Pediatric Sleep: diagnosis and Management of Sleep Problems*. Vol 2nd. Philadelphia, PA: Lippincott, Willians, & Wilkins; 2010.
7. Touchette E, *et al*. Associations between sleep duration patterns and behavioral/cognitive functioning at school entry. *Sleep*. 2007;**30**(9):1213–1219.
8. Paavonen EJ, *et al*. Sleep quality and cognitive performance in 8-year-old children. *Sleep Med*. 2010;**11**(4):386–392.
9. Gregory AM, *et al*. Sleep, emotional and behavioral difficulties in children and adolescents. Sleep Med Rev. 2012;16(2):129-136.
10. Berger RH, *et al*. Acute sleep restriction effects on emotion responses in 30- to 36-month-old children. *Journal of sleep research*. 2011:no–no.
11. Carvalho Bos S, *et al*. Sleep and behavioral/emotional problems in children: a population-based study. *Sleep Med*. 2009;**10**(1):66–74.
12. Javaheri S, *et al*. Association of short and long sleep durations with insulin sensitivity in adolescents. *J Pediatr*. 2011;**158**(4):617–623.
13. Van Cauter E, *et al*. Sleep and the epidemic of obesity in children and adults. *Eur J Endocrinol*. 2008;**159** (Suppl 1):S59–S66.
14. Javaheri S, *et al*. Sleep quality and elevated blood pressure in adolescents. *Circulation*. 2008;**118**(10):1034–1040.
15. Carskadon MA, *et al*. Regulation of adolescent sleep: implications for behavior. *Ann N Y Acad Sci*. 2004;**1021**:276–291.
16. Roberts RE, *et al*. Chronic insomnia and its negative consequences for health and functioning of adolescents: a 12-month prospective study. *J Adolesc Health*. 2008;**42**(3):294–302.
17. Cohen S, *et al*. Sleep habits and susceptibility to the common cold. *Arch Intern Med*. 2009;**169**(1):62–67.
18. Dewald JF, *et al*. The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: a meta-analytic review. *Sleep Med Rev*. 2010;**14**(3):179–189.
19. Owens JA, *et al*. Sleep disturbance and injury risk in young children. *Behav Sleep Med*. 2005;**3**(1):18–31.
20. Konofal E, *et al*. Sleep and ADHD. *Sleep Med*. 2010;**11**(7):652–658.
21. Lewandowski AS, *et al*. Sleep problems in children and adolescents with common medical conditions. *Pediatr Clin North Am*. 2011;**58**(3):699–713.
22. Richdale AL, *et al*. Sleep problems in autism spectrum disorders: prevalence, nature, & possible biopsychosocial aetiologies. *Sleep Med Rev*. 2009;**13**(6):403–411.
23. Bevans KB, et al. Qualitative Development and content validation of the PROMIS Pediatric Sleep Health Items. Behav Sleep Med. in press.
24. Bruni O, *et al*. The Sleep Disturbance Scale for Children (SDSC). Construction and validation of an instrument to evaluate sleep disturbances in childhood and adolescence. *J Sleep Res*. 1996;**5**(4):251–261.
25. Luginbuehl M, *et al*. Pediatric sleep disorders: validation of the sleep disorders inventory for students. *Sch Psych Rev*. 2008;**37**(3):409e431.
26. Spruyt K, *et al*. Pediatric sleep questionnaires as diagnostic or epidemiological tools: a review of currently available instruments. *Sleep Med Rev*. 2011;**15**(1):19–32.
27. Cella D, *et al*.; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care*. 2007;**45**(5 Suppl 1):S3–S11.
28. Buysse DJ, *et al*. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. *Sleep*. 2010;**33**(6):781–792.
29. US Food & Drug Administration. Guidance for Industry—Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009; https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf. Accessed February 10, 2018.
30. Forrest CB, *et al*. Commentary: the patient-reported outcome measurement information system (PROMIS®) for children and youth: application to pediatric psychology. *J Pediatr Psychol*. 2012;**37**(6):614–621.
31. HealthMeasures. Measure Development & Research. 2017; http://www.healthmeasures.net/explore-measurement-systems/promis/measure-development-research. Accessed June 5, 2017.
32. Siegel K, *et al*. Strategies adopted by late middle-age and older adults with HIV/AIDS to explain their physical symptoms. *Psychol Health*. 2011;**26** (Suppl 1):41–62.
33. DiSogra C, *et al*. On the quality of ancillary data available for address-based sampling. *Proceedings of the American Statistical Association, Section on Survey Research Methods*. 2010:4174–4183.
34. Dennis JM. KnowledgePanel®: Processes & Procedures Contributing to Sample Representativeness & Tests for Self-Selection Bias. 2010; http://www.knowledgenetworks.com/ganp/docs/KnowledgePanelR-Statistical-Methods-Note.pdf. Accessed June 6, 2016.
35. Lohr S. *Sampling: design and analysis*. Scarborough, ON: Nelson Education; 2009.
36. Green BF, *et al*. Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*. 1984;**21**(4):347–360.
37. Edelen MO, *et al*. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res*. 2007;**16** (Suppl 1):5–18.
38. Choi SW, *et al*. 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**(8):1–30.
39. Crane PK, *et al*. Rapid detection of differential item functioning in assessments of health-related quality of life: the functional assessment of cancer therapy. *Qual Life Res*. 2007;**16**(1):101–114.
40. Samejima F. Graded response model. In: van der Linden WJ, Hambleton RK, eds. *Handbook of Modern Item Response Theory*. New York, NY: Springer; 1997:85–100.
41. Hambleton RK, *et al*. *Fundementals of Item Response Theory*. Newbury Park, CA: Sage; 1991.
42. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*. 1969;**34**:100.
43. Choi SW. Firestar: computerized adaptive testing simulation program for polytomous item response theory models. *Applied Psychological Measurement*. 2009;**33**(8):644–645.
44. Bock RD, *et al*. Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*. 1981;**46**:443–459.
45. Cai L. Lord-wingersky algorithm version 2.0 for hierarchical item factor models with applications in test scoring, scale alignment, and model fit testing. *Psychometrika*. 2015;**80**(2):535–559.
46. Thissen D, *et al*. Item response theory for scores on tests including polytomous items with ordered responses. *Applied Psychological Measurement*. 1995;**19**(1):39–49.
47. Cella D, *et al*. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res*. 2007;**16** (Suppl 1):133–141.
48. Bjorner JB, *et al*. Developing tailored instruments: item banking and computerized adaptive assessment. *Qual Life Res*. 2007;**16** (Suppl 1):95–108.
49. Cella D, *et al*. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res*. 2007;**16** (Suppl 1):133–141.
50. Berry RB, Brooks R, Gamaldo CE. *The AASM Manual for the Scoring of Sleep and Associated Events: rules, Terminology and Technical Specifications, Version 2.0*. Darien, IL :American Academy of Sleep Medicine; 2012.
51. Sadeh A, *et al*. Sleep patterns and sleep disruptions in school-age children. *Dev Psychol*. 2000;**36**(3):291–301.
52. Meltzer LJ, *et al*. The Children's report of sleep patterns–sleepiness scale: a self-report measure for school-aged children. *Sleep Med*. 2012;**13**(4):385–389.
53. Smaldone A, *et al*. Sleepless in America: inadequate sleep and relationships to health and well-being of our nation's children. *Pediatrics*. 2007;**119** (Suppl 1):S29–S37.
54. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates; 1988.
55. Owens JA, *et al*. The Children's sleep habits questionnaire (CSHQ): psychometric properties of a survey instrument for school-aged children. *Sleep*. 2000;**23**(8):1043–1051.
56. Wolfson AR, *et al*. Evidence for the validity of a sleep habits survey for adolescents. *Sleep*. 2003;**26**(2):213–216.
57. Lai JS, *et al*. Development and psychometric properties of the PROMIS(®) pediatric fatigue item banks. *Qual Life Res*. 2013;**22**(9):2417–2427.
58. Markovich AN, *et al*. Validating the children's sleep habits questionnaire against polysomnography and actigraphy in school-aged children. *Front Psychiatry*. 2014;**5**:188.
59. Holland PW, *et al*. Linking and equating test scores. In: Brennan RL, ed. *Educational Measurement*. 4th ed. Westport, CT: Praeger Publishers; 2007:187–220.
60. von Davier AA, *et al*. *The Kernel Method of Equating*. New York, NY: Springer; 2004.
61. Kolen MJ, *et al*. *Test equating, linking, and scaling: methods and practices*. 2nd ed. New York, NY: Springer-Verlag; 2004.
62. Petersen NS, *et al*. Scaling, norming and equating. In: Linn RL, ed. *Educational Measurement*. 3rd ed. New York, NY: Macmillan; 1989:221–262.
63. Holland PW, *et al*. eds. *Test Equating*. New York, NY: Academic Press; 1982.
64. Angoff WH. Scales, norms and equivalent scores. In: Thorndike RL, ed. *Educational Measurement*. 2nd ed. Washington, DC: American Council on Education; 1971:508–600.
65. Choi SW, *et al*. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. *Psychol Assess*. 2014;**26**(2):513–527.
66. Hays RD, *et al*. Use of Internet panels to conduct surveys. *Behav Res Methods*. 2015;**47**(3):685–690.
67. Craig BM, *et al*. Comparison of US panel vendors for online surveys. *J Med Internet Res*. 2013;**15**(11):e260.