

VU Research Portal

Improve the night, improve the day

van Kooten, J.A.M.C.

2020

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

van Kooten, J. A. M. C. (2020). *Improve the night, improve the day: Better sleep (measurement) in pediatric oncology*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl



Chapter 4

Psychometric properties of the Patient-Reported Outcomes Measurement Information System (PROMIS) Sleep Disturbance and Sleep-Related Impairment item banks in adolescents

Jojanneke A.M.C. van Kooten
Caroline B. Terwee
Michiel A.J. Luijten
Lindsay M.H. Steur
Sigrid Pillen
Nicole G.J. Wolters
Gertjan J.L. Kaspers
Raphaële R.L. van Litsenburg

Journal of Sleep Research 2020, in press, e13029

Summary

Sleep problems have a high prevalence and negative daytime consequences in adolescents. Current sleep measures for this age group have limitations. The Patient-Reported Outcomes Measurement Information System (PROMIS®) developed sleep item banks for adults. In a previous validation study, these item banks were adapted to a shortened version for adolescents. The current study aimed to further explore the psychometric properties of the 11-item Sleep-Related Impairment and 23-item Sleep Disturbance item bank in Dutch adolescents. We investigated structural validity by testing Item Response Theory assumptions and model fit; measurement invariance by performing Differential Item Functioning analyses; performance as a Computerized Adaptive Test; reliability by marginal reliability estimates and test-retest reliability (Intraclass Correlation Coefficients and Limits of Agreement); and construct validity by hypothesis testing. Additionally, we provide mean values for the item banks. The study sample consisted of 1046 adolescents (mean age 14.3 ± 1.6 years), including 1013 high school students and 33 sleep clinic patients. The Sleep Disturbance-23 showed lack of unidimensionality, but had sufficient test-retest reliability, and could distinguish between adolescents with and without sleep- or health issues. The Sleep-related Impairment-11 showed sufficient unidimensionality and model fit and was thus tested as a Computerized Adaptive Test, demonstrating an equal amount of reliable measures as the full item bank. Furthermore, the Sleep-related Impairment-11 could distinguish between adolescents with and without sleep or health issues and test-retest reliability was moderate. We recommend the use of both item banks in full form, and the use of the Sleep-related Impairment-11 as a Computer Adaptive Test.

Introduction

Sleep problems and sleep deprivation are common among adolescents: 20-37% of otherwise healthy adolescents struggle with sleep problems,¹⁻³ and 30-60% do not get the required 8-10 h of sleep during school nights.^{1,4,5} Compared to adults and younger children, healthy sleep is challenged by unique features during adolescence: Extrinsic factors such as social activities and academic demands - with early school start times - interact with physiological changes that cause a tendency towards later bed times and later natural wake up times.^{4,6} This increases adolescent' sleep debt during the school week and often makes for more irregular sleep patterns, with catch-up sleep during the weekends. In this phase, adolescents are also discovering autonomy regarding their sleep schedules.^{7,8} Sleep disorders can further diminish sleep duration, with delayed sleep phase disorder in 5-16% of healthy adolescents,^{6,9} and insomnia in 8-10% as the most common diagnoses.^{6,10}

Insufficient sleep during the night is associated with many daytime problems; sleepiness, behavioral problems including increased risk taking,^{3,6,9} and difficulties in emotional regulation resulting in increased irritability, anxiety, depressive symptoms and self-harm.^{2,5,6,9,11} Cognitive function and academic achievement are worse in children with poor sleep,^{5,6,9} as is physical health, illustrated by an increased presence of hypertension and obesity, and symptoms like dizziness and headaches.^{2,5,11} Sleep problems during adolescence are predictive of sleep problems later in life,¹² early recognition and treatment of sleep problems is therefore important.

Given the high prevalence of sleep problems in adolescents, the negative consequences, and the unique sleep features of this age group; it is important to have psychometrically sound measurement instruments for screening and follow-up, validated in this age group. In contrast to objective measures like actigraphy and polysomnography, sleep questionnaires are able to capture feelings and cognitions about sleep and the effects of impaired sleep.⁵ A multitude of sleep questionnaires targeting different sleep constructs is available. Previous reviews identified around 60 sleep questionnaires used in children and adolescents, but none of the generic sleep instruments were adequately validated.^{13,14} The Patient-Reported Outcomes Measurement Information System (PROMIS[®]) might offer a solution for the lack of validated sleep questionnaires for adolescents. PROMIS was initiated by six US research institutes and the National Institutes of Health (NIH) and is an international initiative aimed to standardize questionnaires measuring key health outcomes in research and clinical practice and to increase relevance of results by facilitating comparison of data. The methodological basis is the use of item response theory (IRT), enabling the creation of item banks that support fixed-length forms and



Computer Adaptive Testing (CAT). A CAT can achieve greater measurement precision with fewer items: Participants need to complete only a subset of items instead of the full set, since after the first item, the selection of subsequent items is determined by the participants' responses to the previous items.^{15,16}

PROMIS developed a Sleep Disturbance and Sleep-Related Impairment item bank for adults through factor- and IRT analyses.¹⁷ The construct validity of both the full item banks and the short forms was found to be sufficient in adults.^{17,18} We previously started validation of the PROMIS adult sleep item banks for adolescents,^{19,20} since sleep item banks for children were not yet developed.^{21,22} Since then, multiple studies used the PROMIS adult sleep item banks in young adults and adolescents.²³⁻²⁵

The Dutch-Flemish versions of the PROMIS v1.0 adult sleep item banks showed adequate content validity in adolescents,²⁰ meaning the items were considered relevant and comprehensible for adolescents, parents and sleep experts and no key issues were considered missing. Additional psychometric evaluation in a community sample of over 1000 Dutch adolescents, however, showed that the one factor models found in adults could not be replicated: Thus the items in adolescents did not reflect the same single construct measured in adults and the item banks were not unidimensional. Adaptation of the item banks to improve unidimensionality needed for IRT analyses resulted in a shortened version of the Sleep-Related Impairment item bank (11 instead of 16 items) with adequate fit (comparative fit index (CFI) 0.98) and a shortened version of the Sleep Disturbance item bank (23 instead of 27 items) with fit indices just below recommended value (CFI 0.90, recommended value > 0.95).¹⁹ The current study aims to further explore the psychometric properties of the adult-version 11-item PROMIS v1.0 Sleep-Related Impairment and 23-item PROMIS v1.0 Sleep Disturbance item bank in adolescents. We evaluated structural validity, measurement invariance, performance as a CAT, reliability and construct validity, and additionally provide mean values for these item banks in adolescents.

Methods

Participants and procedures

Test sample

A community sample of adolescents was recruited from seven randomly selected high schools in the Netherlands. Schools from all educational levels and from different regions were included. Most Dutch high school students are aged 12-18 years, but exceptions of 11 or 19 years do exist and were included in the current study.

Adolescents were asked to fill out the questionnaires during regular class hours. During this administration, author JvK was present in the classroom to supervise the procedure and provide assistance if necessary. In one school where online entry was not possible due to lack of digital resources, paper versions were distributed.

Adolescents (11-19 years) with sleep problems were recruited from four outpatient (sleep) clinics. Adolescents with any type of sleep problem were eligible. They were invited to participate during their first visit to the clinic. They received a study package containing the study questionnaires on paper, including questions on the type of sleep problems they experienced. Exclusion criteria for both samples were any impairment that precluded filling out the questionnaires independently. The study was approved by the Institutional Review Board of the VU University Medical Center Amsterdam.

Retest sample

All participants from the high school sample were invited to participate in the retest study. Participants that were interested could apply by providing their e-mail address. Two weeks after the first entry, a link to the repeat questionnaire was sent via e-mail.

Measures

Sociodemographic variables

Descriptive data were collected on gender, age, educational level and country of birth. In addition, all participants were asked to report on current use of medication and on health issues, specifically diagnosis of autism spectrum disorder (ASD) and attention deficit (hyperactivity) disorder (ADHD / ADD), as these diagnoses are common and associated with sleep problems.^{26,27}

PROMIS Item Banks

This study used the shortened versions of the adult Dutch-Flemish PROMIS Sleep Disturbance and Sleep-Related Impairment v1.0 item banks that were adapted for adolescents.¹⁹ The item banks aim to gain a general overview of the subjects' perception of sleep problems and how these problems hinder daily functioning. The original PROMIS Sleep Disturbance item bank contains 27 items that are reflective of insomnia-like symptoms. It assesses one's perception of their sleep quality and restoration associated with sleep, perceived sleep difficulties and concerns about falling and staying asleep, and perceptions of adequate and satisfactory sleep.¹⁷ Our previous study on structural validity resulted in an adapted Sleep Disturbance item bank with 23 items with better fit in adolescents. We removed items Sleep20 (I had a problem with my sleep), Sleep106 (My sleep was light), Sleep108 (My sleep was restless) and Sleep125 (I felt lousy when I woke up) from the original item bank.¹⁹

The original PROMIS Sleep-Related Impairment item bank consists of 16 items that are related to sleepiness, fatigue, and cognitive difficulties during waking hours. In addition, Sleep-Related Impairment items assess perceptions of functional impairment during waking hours that are associated with sleep problems or impaired alertness.^{17,28} Our previous study on structural validity resulted in an adapted Sleep-Related Impairment item bank with 11 items, with better fit in adolescents. We removed items Sleep4 (I had enough energy), Sleep119 (I felt alert when I woke up), Sleep120 (When I woke up I felt ready to start the day), Sleep123 (I had difficulty waking up) and Sleep124 (I still felt sleepy when I woke up) from the original item bank.¹⁹

All Sleep Disturbance and Sleep-Related Impairment items were measured on a 5-point Likert scale (1 = not at all or never; 2 = a little bit or rarely; 3 = somewhat or sometimes; 4 = quite a bit or often; 5 = very much or always). The answers were indicative of how frequently respondents had experienced problems related to sleep in the last 7 days. The official HealthMeasures scoring service tool (https://www.assessmentcenter.net/ac_scoring-service/) was used to calculate T-scores using the US calibration parameters for all participants who filled out at least one item. T-scores are anchored on the US general population, with a mean of 50 and a standard deviation of 10. Higher scores indicate more sleep disturbances or more sleep-related impairment.

Statistical analyses

Item-level descriptives

For each item the median and mean response category was calculated separately for the high school sample and the sleep clinic sample.

Structural validity

Psychometric analyses of the Sleep Disturbance-23 and the Sleep-Related Impairment-11 baseline data were conducted according to the PROMIS analyses plan.²⁹ The graded response model (GRM) was estimated with marginal maximum likelihood (MML). GRM is an IRT model for ordinal items. An IRT model requires that three assumptions are met: unidimensionality, local independence and monotonicity. Table 1 provides further explanation and criteria of investigated IRT model assumptions and fit.

Measurement invariance

Differential item functioning (DIF) between our sample and the PROMIS 2 Sleep-Wake sample was assessed, this sample was used to develop the item banks and consists of 1993 adults from the general population and 259 adults recruited from medical, psychiatric, or sleep clinics. Their mean age was 51.2 (15.9) years, range 18-88, 52% was male¹⁷. This data was available from HealthMeasures Dataverse.³⁰

Post-hoc CAT simulation

A post-hoc CAT simulation was performed using the item parameters estimated in our study sample. This is a simulation based on real responses from participants, where for each new item selected, the item that can give the most information (based on item parameters) is presented to the individual. This analysis was only performed if IRT-model assumptions were met and participants with missing data were excluded. The algorithm was set to administer a minimum of one item, and to stop administration when the reliability of the participants T-score was above 0.90 (standard error of the mean (SEM) < 0.32) or all items) had been used.³¹ The number of participants that reached a reliable score was compared between the CAT and the full-length Sleep Disturbance-23 or Sleep-Related Impairment-11.

Reliability

If IRT assumptions were met and adequate model fit was found, marginal reliability estimates were plotted, showing the standard error of theta across the scale.

Test-retest reliability was examined by calculating intraclass correlation coefficients (ICC) and limits of agreement (LoA). For the ICC a two-way random effects model for absolute agreement was used. The LoA was calculated as the mean difference between the test and retest T-score $\pm 1.96 * SD$ of this mean, 95% of differences are located between the upper and lower LoA. ICC and LoA were interpreted following the Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines: An ICC of ≥ 0.70 is considered sufficient reliability and the LoA should be smaller than the minimal important change (MIC).³²

Construct validity

To determine construct validity we assessed the difference in T-scores between groups, testing four hypotheses about the ability of the PROMIS item banks to distinguish between these groups. In line with COSMIN guidelines, construct validity is considered sufficient when $\geq 75\%$ of the results are in accordance with the hypotheses.³² We expected PROMIS T-scores to be higher (worse) in 1. the sleep clinic sample compared to the high school sample; 2. the adolescents with a high risk of sleep problems compared to healthy high school students; and 3. high school students with health issues compared to healthy high school students. Adolescents with a high risk of sleep problems included the sleep clinic sample and high school students with relevant health issues that were associated with a higher probability of sleep problems. Health issues included self-reported sleep difficulties, use of sleep medication, chronic health problems and / or use of medication associated with sleep problems (such as ADHD and ASD,^{26,27} other psychiatric conditions such as depression



(continued on following page)

Table 1 IRT-analyses

Investigated property	Explanation	Criteria for acceptable values	Sleep Disturbance-23	Sleep-Related Impairment-11
Assumptions IRT model	Unidimensionality A persons' response to an item should be accounted for by the amount of the construct measured (sleep disturbance and sleep-related impairment) and not by other factors, so all items in the item bank need to reflect the same construct. With confirmatory factor analysis (CFA) we first test how well an unidimensional model fits our data. Bi-factor analysis and Exploratory factor analysis (EFA) were performed if CFA showed poor fit. Bi-factor analysis tests if items load on randomly added factors in addition to the general factor (the construct we wish to measure). EFA tests how much of the variance in the data is explained by the first factor (the construct we wish to measure).	CFA Factor loadings > 0.50 Comparative fit index (CFI) > 0.95 Tucker-Lewis index (TLI) > 0.95 Root mean square error of approximation (RMSEA) < 0.06 Standardized root mean residuals (SRMR) < 0.08 Bi-factor analysis Factor loading general factor (G) > random factor (F) Omega-H (% explained by G) > 0.80 Explained common variance (ECV) (ratio between explained variance by G and F) > 0.80 EFA Explained variance by 1 st factor > 20% Ratio between 1 st and 2 nd factor > 4	20/23 items > 0.50 0.80 0.78 0.14 0.11 14/23 items G > F 0.68 0.54 40% 5.9	All items > 0.50 0.96 0.94 0.15 0.06 10/11 items G > F 0.86 0.79 64% 15.3
Local independence	Items should be independent of each other, when controlled for the construct measured.	Residual correlations ≤ 0.20	11 pairs (4.3%) correlation > 0.20	1 pair (1.8%) correlation > 0.20
Monotonicity	The probability of selecting an item response indicative of more problems, should increase with an increasing level of the construct. This is tested with Mokken scale analysis, which assesses scalability, the possibility of locating the item on a scale.	Scalability coefficient H (scale) > 0.50 Scalability coefficient H _i (item) > 0.30	0.37 (0.01) 20/23 items > 0.30	0.59 (0.02) All items > 0.30

IRT model fit	IRT models describe, in probabilistic terms, the relationship between a persons' response to an item and the level of construct measured by the total item bank. With $S-X^2$ the differences between observed and expected response frequencies is qualified.	$S-X^2 \geq 0.001$	All items > 0.001	All items > 0.001
Differential Item Functioning (DIF)	This questions if items perform differently across groups (measurement invariance), here Dutch adolescents and US adults. Uniform DIF means the magnitude is similar for all levels of construct, non-uniform DIF means this differs between different (higher / lower) levels of construct.	Change in McFadden $R^2 \leq 0.02$	3/23 items uniform DIF; no non-uniform DIF	No uniform or non-uniform DIF

Abbreviations: IRT = Item response theory; CFA = confirmatory factor analysis; EFA = Exploratory factor analysis; CFI = Comparative fit index; TLI = Tucker-Lewis index; RMSEA = Root mean square error of approximation; SRMR = Standardized root mean residuals; GRM = Graded response model; G = general factor; F = random factor



and anxiety,³³ medication prescribed for before mentioned conditions or strong pain medication such as opioids). A mean difference of ≥ 2 points, with a higher (worse) score in the clinical or health issues sample was considered clinically relevant.³⁴ The fourth hypothesis was that T-scores would worsen (increase) with more problems (higher response category) on item Sleep20 (I had a sleep problem). This item is included in the original PROMIS Sleep Disturbance item bank, but not in the Sleep Disturbance-23. For the fourth hypothesis we merged the last two response categories ('Quite a bit' and 'Very much'), since the separate groups were too small. Differences in T-scores between groups were evaluated using linear regression analysis, with correction for relevant demographic variables based on the results from the analyses of the mean values.

Mean T-scores for adolescents from the general population

Mean T-scores were calculated for all high school students. In addition, we compared mean T-scores between boys and girls; adolescents with low (lower general secondary education / intermediate vocational education) and high (higher general secondary education) educational level as a reflection of socioeconomic status; and younger (11-14) and older (15-19) aged adolescents, since these are factors that are associated with sleep quality and / or quantity. We expected scores to be higher (worse) in girls,^{2,35} adolescents with low educational level,³⁶ and older adolescents.^{7,36} A mean difference of ≥ 2 points was considered clinically relevant.³⁴

Analyses 3.2 to 3.5 were carried out using R, analyses 3.1 and 3.5 to 3.8 were carried out using SPSS 24.

Results

Participants

Test sample

In total 1046 adolescents provided valid data, including 1013 high school students and 33 adolescents from the sleep clinics. Sample characteristics are summarized in Table 2. Almost half of the total sample consisted of boys, this is comparable to the general population. In the general Dutch population 61% of adolescents receives high level education,³⁷ this percentage was higher in our high school sample (81%), and lower in the sleep clinic sample (38%). As expected, the percentages of adolescents with ASD and ADHD were higher in the sleep clinic sample than in the high school sample (ASD 24% versus 5% and ADHD 18% versus 5%, respectively), the percentage of adolescents in the Dutch general population with ASD is 2% and with ADHD 7%.³⁸

Table 2 Participant characteristics

Characteristic	High school sample	Healthy ^b adolescents	High school sample with health issues	Sleep clinic sample	Sleep clinic + high school sample with health issues	Retest sample
n	1013	920	93	33	126	114
Age [mean (SD); range]	14.3 (1.6); 11-19	14.3 (1.6); 11-19	14.4 (1.6); 12-19	14.8 (1.9); 11-18	14.5 (1.7); 11-19	14.7 (1.5); 11-19
Gender [% boys]	48.4	46.2	69.9	41.9	62.9	33.3
Country of birth [% Netherlands]	94.2	94.5	91.4	100	93.5	93.0
Educational level [% high] ^a	81.4	81.4	81.7	37.5	70.4	94.7
ASD [% yes]	4.4	0.0	48.8	24.4	42.1	4.4
ADHD [% yes]	4.5	0.0	49.5	18.2	41.3	1.8
T-score Sleep Disturbance-23 [mean (SD); range]	47.3 (7.0); 25.9-70.8	47.1 (6.8); 25.9 – 70.8	49.8 (8.4); 25.9-68.8	57.9 (8.8); 36.1 – 79.9	51.9 (9.2); 25.9-79.9	46.3 (7.0); 30.9-70.7
T-score Sleep- Related Impairment-11 [mean (SD); range]	48.6 (9.6); 31.1 – 82.4	48.2 (9.5); 31.1 – 82.4	51.7 (10.2); 31.1-71.7	58.7 (12.1); 31.1 – 82.4	53.5 (11.1); 31.1-82.4	47.1 (9.6); 31.1-70.6

Abbreviations: ADHD = Attention deficit hyperactivity disorder; ASD = Autism spectrum disorder

^a Low = Lower general secondary education / intermediate vocational education; high = higher / A-level general secondary education.

^b Excluding children with medical / psychiatric conditions (ASD, ADHD and other psychiatric conditions (e.g. depression, anxiety)) or medications (medication prescribed for before mentioned conditions, sleep medication and strong pain medication such as tramadol).



In the sleep clinic sample 82% had a problem with initiating and maintaining sleep, 18% had parasomnias, 18% had delayed sleep phase disorder, 12% had obstructive sleep apnea and 9% had excessive daytime sleepiness. Thirty-nine percent of the sleep clinic sample experienced multiple sleep problems at the time of inclusion.

Retest sample

Of the 1013 included high school students, 372 provided their e-mail address and 114 (11%) completed the PROMIS item banks again after 2 weeks. Compared to non-responders ($n = 899$), responders ($n = 114$) to the retest were more often girls (50% versus 66%, respectively, $P < .01$), and they more often received high level education (80% versus 93%, respectively, $P < .01$). Responders and non-responders did not differ significantly in age or baseline T-scores.

Item-level descriptives

Table 3 provides item-level descriptives. For the high school sample the median response category ranged from 1 ('Not at all' or 'Never') to 2 ('A little bit' or 'Rarely'), whereas the medians in the sleep clinic sample ranged from 1 to 4 ('Quite a bit' or 'Often'). The percentage of missing responses ranged from 0.0% to 4.7% per item in the high school sample, and from 0.0% to 9.1% per item in the sleep clinic sample. This is likely due to the fact that skipping items was not possible in the online version used in the majority of high school participants whereas sleep clinic patients filled out paper questionnaires. The two items with 9.1% missing answers in the sleep clinic sample, had 0.8% and 2.2% missings in the high school sample, thus there does not seem to be a systematic problem with these items. All 23 items of the Sleep Disturbance were filled out by 93% of adolescents and all 11 Sleep-Related Impairment items by 95%.

Structural validity

Results of analyses for IRT assumptions and fit are shown in Table 1. The Sleep Disturbance-23 did not meet the three assumptions needed to fit the IRT model. For unidimensionality exploratory factor analysis (EFA) criteria were met, but both confirmatory factor analysis (CFA) and bi-factor analysis were not satisfactory (CFI 0.80, criterion > 0.95 ; Omega-H 0.68, criterion > 0.80). IRT item fit was nevertheless satisfactory. The Sleep-Related Impairment-11 did meet the overall criteria for IRT modelling (CFI 0.96, Omega-H 0.86), IRT item fit was good.

Measurement invariance

In the Sleep Disturbance-23 three items were flagged for uniform language DIF, together they impact the total T-score by about two points, which could be relevant in the future when comparing US to Dutch scores. In the Sleep-Related Impairment-11 no items were flagged for DIF.

Post-hoc CAT simulation

A post-hoc CAT simulation was performed for the Sleep-Related Impairment-11. Out of the 1000 participants included in the simulation, 765 participants (76.5%) reached a reliable score with the full length item bank, whereas 235 participants (23.5%) could not reach a reliable score with 11 items. Of these 765, 757 also reached a reliable score using less than 11 items in CAT (mean number of items 4.5 ± 1.7).

Reliability

The Sleep-Related Impairment-11 has a reliability higher than 0.90 between a T-score of approximately 43 and 80, adolescents with lower (better) scores reach a lower reliability (Figure 1). We did not perform reliability estimates for the Sleep Disturbance-23, since assumptions needed to fit the IRT-model were not met.

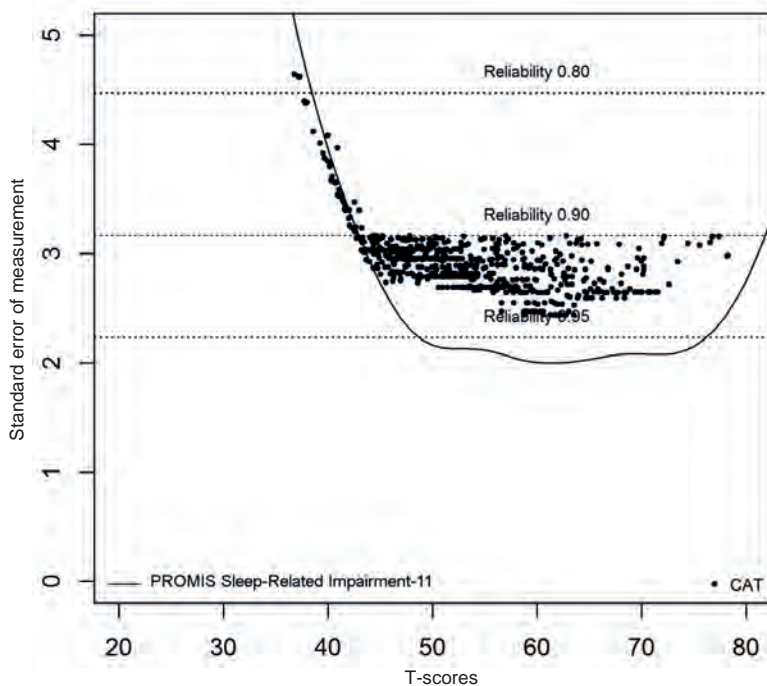


Figure 1. Standard error of measurement over the range of T-scores

Table 3 Item-level descriptive statistics

Items	High school sample (n = 1013)			Sleep clinic sample (n = 33)		
	Median	Mean (SD)	Missing (%)	Median	Mean (SD)	Missing (%)
<i>Sleep Disturbance-23</i>						
Sleep105: My sleep was restless.	2	2.4 (1.0)	0 (0.0)	4	3.8 (1.0)	0 (0.0)
Sleep107: My sleep was deep.	2	2.5 (1.1)	2 (0.2)	4	3.4 (1.4)	0 (0.0)
Sleep109: My sleep quality was ...	2	2.2 (0.8)	2 (0.2)	4	3.6 (0.8)	0 (0.0)
Sleep110: I got enough sleep.	2	2.5 (0.9)	2 (0.2)	3	3.5 (0.9)	0 (0.0)
Sleep115: I was satisfied with my sleep.	2	2.6 (1.1)	2 (0.2)	4	3.9 (1.1)	0 (0.0)
Sleep116: My sleep gave me new energy.	2	2.5 (1.1)	47 (4.7)	4	3.8 (1.0)	1 (3.0)
Sleep42: It was easy for me to fall asleep.	2	2.6 (1.1)	9 (0.9)	4	3.8 (1.4)	0 (0.0)
Sleep44: I had difficulty falling asleep.	2	2.2 (1.2)	12 (1.2)	4	3.7 (1.5)	0 (0.0)
Sleep45: I laid in bed for hours waiting to fall asleep.	2	2.1 (1.1)	10 (1.0)	4	3.4 (1.5)	1 (3.0)
Sleep50: I woke up too early and could not fall back asleep.	2	2.1 (1.1)	9 (0.9)	3	2.6 (1.3)	0 (0.0)
Sleep65: I felt physically tense at bedtime.	1	1.7 (1.0)	9 (0.9)	2	2.1 (1.4)	1 (3.0)
Sleep67: I worried about not being able to fall asleep.	1	1.5 (0.9)	8 (0.8)	2	2.2 (1.4)	0 (0.0)
Sleep68: I felt worried at bedtime.	1	1.4 (0.8)	9 (0.9)	1	1.9 (1.3)	0 (0.0)
Sleep69: I had trouble stopping my thoughts at bedtime.	1	2.0 (1.2)	8 (0.8)	2	2.5 (1.5)	3 (9.1)
Sleep70: I felt sad at bedtime.	1	1.3 (0.8)	9 (0.9)	1	1.7 (1.2)	0 (0.0)
Sleep71: I had trouble getting into a comfortable position to sleep.	1	1.8 (1.0)	9 (0.9)	2	2.3 (1.3)	0 (0.0)
Sleep72: I tried to get to sleep.	2	2.1 (1.2)	23 (2.3)	3	3.1 (1.4)	0 (0.0)
Sleep78: Stress disturbed my sleep.	1	1.7 (1.1)	22 (2.2)	2	2.2 (1.3)	0 (0.0)
Sleep86: I tossed and turned at night.	1	1.9 (1.2)	22 (2.2)	3	2.8 (1.4)	3 (9.1)
Sleep87: I had trouble staying asleep at night.	1	1.6 (0.9)	47 (4.7)	3	3.0 (1.3)	0 (0.0)
Sleep90: I had trouble sleeping.	2	1.8 (1.0)	22 (2.2)	4	3.3 (1.3)	1 (3.0)
Sleep92: I woke up and had trouble falling back to sleep.	2	2.1 (1.2)	22 (2.2)	3	3.1 (1.5)	1 (3.0)
Sleep93: I was afraid I would not get back to sleep after waking up.	1	1.5 (0.9)	22 (2.2)	2	2.2 (1.4)	2 (6.1)

(continued on following page)

<i>Sleep-related Impairment-11</i>						
Sleep6: I was sleepy during the daytime.	2	2.4 (1.1)	22 (2.2)	4	3.3 (1.2)	0 (0.0)
Sleep7: I had trouble staying awake during the day.	1	1.7 (1.0)	21 (2.1)	2	2.5 (1.3)	0 (0.0)
Sleep10: I had a hard time getting things done because I was sleepy.	1	1.7 (0.9)	36 (3.6)	2	2.5 (1.4)	0 (0.0)
Sleep11: I had a hard time concentrating because I was sleepy.	2	1.9 (1.0)	36 (3.6)	3	2.9 (1.5)	0 (0.0)
Sleep18: I felt tired.	2	2.3 (1.1)	37 (3.7)	4	3.7 (1.4)	2 (6.1)
Sleep19: I tried to sleep whenever I could.	1	1.8 (1.0)	39 (3.9)	2	2.4 (1.2)	2 (6.1)
Sleep25: I had problems during the day because of poor sleep.	1	1.6 (0.8)	37 (3.7)	2	2.7 (1.5)	2 (6.1)
Sleep27: I had a hard time concentrating because of poor sleep.	1	1.8 (1.0)	38 (3.8)	3	2.9 (1.5)	2 (6.1)
Sleep29: My daytime activities were disturbed by poor sleep.	1	1.7 (0.9)	37 (3.7)	3	2.7 (1.2)	0 (0.0)
Sleep30: I felt irritable because of poor sleep.	2	1.8 (1.0)	38 (3.8)	3	3.0 (1.3)	0 (0.0)
Sleep33: I had a hard time controlling my emotions because of poor sleep.	1	1.6 (0.9)	36 (3.6)	3	2.7 (1.4)	1 (3.0)

Table 4 Hypothesis testing Sleep Disturbance-23 and Sleep-related Impairment-11

We expected that:	Mean difference in T-score (95% confidence interval)^a	
	Sleep Disturbance-23	Sleep-related Impairment-11^b
the sleep clinic sample had higher scores than the high school students	10.6 (8.1 – 13.1)	8.6 (5.2 – 11.9)
the adolescents with sleep problems and / or relevant health issues had higher scores than healthy high school students	4.8 (3.5 – 6.2)	5.3 (3.5 – 7.1)
the high school students with relevant health issues had higher scores than healthy high school students	2.7 (1.2 – 4.2)	4.0 (2.0 – 6.0)
adolescents who answered item Sleep20 'I had a sleep problem' with a higher response category, would have higher scores		
'Not at all' versus 'A little bit'	6.1 (5.1 – 7.1)	4.5 (3.0 – 6.0)
'A little bit' versus 'Somewhat'	2.2 (0.9 – 3.4)	3.9 (1.7 – 6.1)
'Somewhat' versus 'Quite a bit / Very much'	5.2 (4.0 – 6.5)	3.2 (0.6 – 5.7)

^a A mean difference of ≥ 2 points was considered clinically relevant.

^b Corrected for age and gender.



Test-retest reliability of the Sleep Disturbance-23 was sufficient, ICC (95% confidence interval) was 0.76 (0.67-0.83). The Bland & Altman plot (Figure 2) shows a mean difference of -1.1 point, with LoA -10.3 to 8.0. Test-retest reliability of the Sleep-Related Impairment-11 was just below the recommended value, with an ICC (95% confidence interval) of 0.68 (0.57-0.77). The Bland & Altman plot (Figure 3) shows a mean difference of -1.6 point, with LoA -16.1 to 12.9. The reliability was lower for lower (better) T-scores.

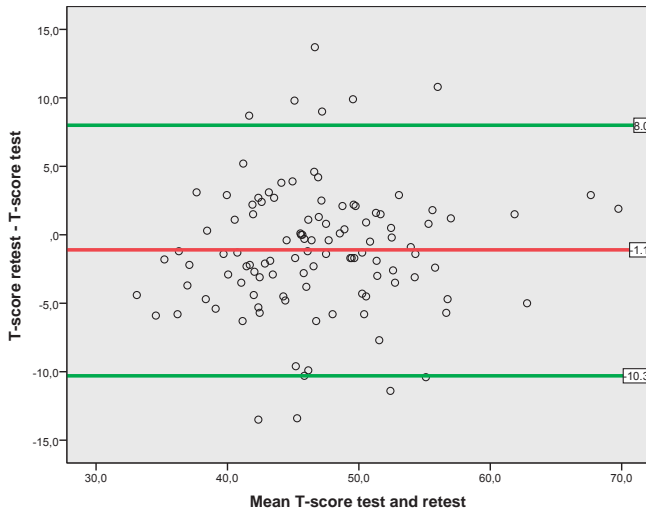


Figure 2 Test-retest reliability (Bland & Altman plot) Sleep Disturbance-23

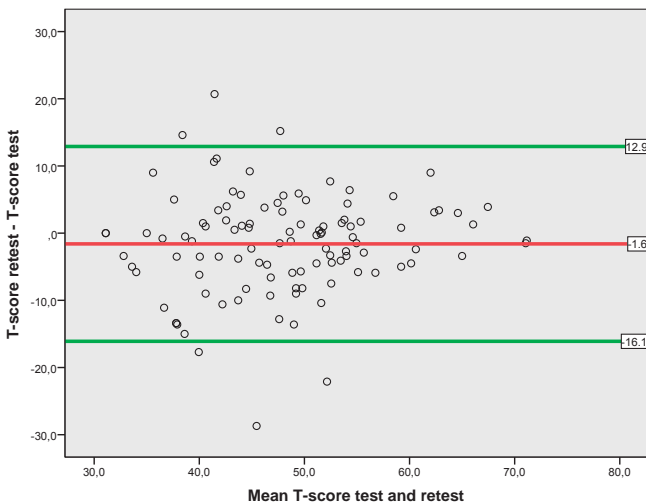


Figure 3 Test-retest reliability (Bland & Altman plot) Sleep-Related Impairment-11

Construct validity

For both the Sleep Disturbance-23 and the Sleep-Related Impairment-11 all results were in accordance with the hypotheses (Table 4). The Sleep Disturbance-23 showed differences between the different samples of 2.7 to 10.6 points and the largest difference was found between the high school sample and the sleep clinic sample. Additionally, adolescents who reported to have more sleep problems on the single item, also had worse T-scores. The Sleep-Related Impairment-11 showed differences between the different healthy and non-healthy samples of 4.0 to 8.6 points, corrected for age and gender. Here also the largest difference was found between the high school sample and the sleep clinic sample. Adolescents who reported to have more sleep problems on the single item, also had worse T-scores.

Mean T-scores for adolescents from the general population

For the Sleep Disturbance-23, the mean (SD) T-score in the high school sample was 47.3 (7.0), with a range from 25.9 to 70.8. T-scores did not differ between low and high educational level, and there were no relevant differences between boys and girls, and younger and older adolescents (Table 5). For the Sleep-Related Impairment-11, the mean (SD) T-score in the high school sample was 48.6 (9.6), with a range from 31.1-82.4. T-scores did not differ more than two points between low and high educational level. There was a relevant difference between younger and older adolescents, and between boys and girls; older adolescents and girls scored higher by 5.0 and 2.7 points, respectively, indicating more sleep-related impairment (Table 5).

Table 5 Mean T-scores high school sample

Variables		Mean (SD) Sleep Disturbance-23	Mean (SD) Sleep-Related Impairment-11
Overall		47.3 (7.0)	48.6 (9.6)
Gender	Boys	46.4 (6.9)	47.2 (9.6) ^a
	Girls	48.2 (7.0)	49.9 (9.6) ^a
Age	11-14 years	46.6 (6.9)	46.5 (9.1) ^a
	15-19 years	48.4 (7.1)	51.5 (9.6) ^a
Educational level	Low	47.7 (7.8)	49.1 (10.4)
	High	47.3 (6.8)	48.4 (9.4)

^a Clinically relevant (≥ 2 points) difference between groups.

Discussion

The field of adolescent sleep medicine is in need of sleep questionnaires with good psychometric properties. These need to be tested specifically in this age group, since adolescents are different than their younger and older peers in terms of sleep physiology and social influences on sleep. PROMIS has high potential in this field and developed item banks through IRT, enabling use as CAT, which ultimately leads to less participant burden.^{16,39} PROMIS developed sleep item banks for adults that can possibly also be used in adolescents. In this study we determined structural validity, measurement invariance, performance as CAT, reliability and construct validity of the PROMIS Sleep Disturbance-23 and Sleep-related Impairment-11 item banks, adapted for adolescents in previous research.

The Sleep Disturbance-23 did not meet the assumptions for IRT analyses due to lack of unidimensionality and is therefore not suited for use as CAT in its current form. A proper alternative for the Sleep Disturbance-23 with sufficient unidimensionality does currently not exist: Ji et al. provided an overview of sleep questionnaires used in adolescents, from 2000 to 2016. Only six generic sleep measures were validated to some extent in adolescents. In three of these questionnaires, structural validity was not assessed at all. For the Sleep Disturbance Scale for Children and the Pittsburgh Sleep Quality Index factor analyses were performed, but the results did not meet the criteria for sufficient structural validity.^{40,41} The Sleep Disorders Inventory for Students (adolescent version) did meet criteria for CFA, but did not have adequate IRT model fit.¹⁴ In contrast to structural validity, the Sleep Disturbance-23 showed sufficient test-retest reliability in terms of ICC and sufficient construct validity. Ideally test-retest reliability would also be assessed comparing the LoA to the MIC, however, the MIC is not yet determined for either of the PROMIS sleep item banks.

The Sleep-related Impairment-11 met the requirements for IRT analyses and showed good item fit. Post-hoc CAT simulations using one to ten items showed a difference of only 0.3% in amount of reliable measurements when compared to the full length item bank. There was a floor effect in reliability, meaning that in adolescents with good sleep, reliability is lower. Importantly, in the adolescents with more sleep problems, T-scores are estimated with high reliability. Test-retest reliability was just lower (0.68) than the criterion of 0.70. This might partly be explained by the large group of healthy participants in the study sample, since the Bland & Altman shows that the reliability is lower in the participants with lower (better) scores for this item bank. The Sleep-related Impairment-11 has more measurement error than the Sleep

Disturbance-23, which was shown by wider LoA. This could be explained by the lower number of items (11 versus 23). The construct validity was sufficient.

During the course of this study, PROMIS also developed a Sleep Disturbance and a Sleep-Related Impairment item bank for children aged 5-17 years. The constructs measured by the adult and pediatric item banks are similar, but the adult item banks are longer and item wording is partly different. The pediatric Sleep Disturbance item bank has 15 items, of which nine are identical in the adult item bank. The pediatric Sleep-Related Impairment item bank counts 13 items, of which six are identical to the adult item bank. The items that we deleted to fit the bank in Dutch adolescents, were both adult-only and shared items. Currently research is being conducted to evaluate the performance of the adult sleep item banks in Dutch adults and the pediatric sleep item banks in Dutch children and adolescents, since the question remains if the lack of unidimensionality we found for the Sleep Disturbance-23 is due to the adolescent age group or the construct measured. Preliminary results showed that the adult Sleep Disturbance item bank is unidimensional enough in adults (C.B. Terwee, personal communication, July 31st, 2019), but that the pediatric Sleep Disturbance item bank lacks unidimensionality in adolescents (S. Peersmann, personal communication, July 31st, 2019). This suggests that the construct of Sleep Disturbance might not be one unidimensional construct in adolescents, or that the questions developed for adults or a larger pediatric age group do not cover enough of the specific adolescent sleep issues. Regarding the latter and recalling the unique adolescent sleep features, specific questions should mostly focus on problems with sleep onset and cover the sleep-related impairments due to the tendency for later bedtimes, activities supporting this tendency, and autonomy regarding bedtime.^{7,8} Drawing upon existing measures, example questions could be 'I have trouble sleeping because I do things in bed that keep me awake (for example reading, watching TV etc.)' or 'when it is time to go to sleep, I have trouble settling down'.^{42,43} Importantly, additional content needs to be developed in accordance with PROMIS' high standards.

This study had a few limitations. First of all, our study sample was higher educated than the general Dutch adolescent population, which may limit the generalizability of these results. Secondly, regarding the analyses, we estimated DIF based on the US adult sample, ideally an adolescent sample would have been used, but this was not available. This means that we do not know if the observed DIF for the Sleep Disturbance-23 is due to language or age. Additionally, most available DIF methods can detect DIF but cannot identify the DIF items due to parameter identification issues.⁴⁴

In conclusion, the Sleep Disturbance-23 is a reliable measure (high ICC) of sleep disturbance and can properly distinguish between clinical and non-clinical groups of adolescents. Though it is not suitable as a CAT in its current form, better alternatives are currently unavailable and its use in adolescents is therefore recommended. Future research is necessary to optimize structural validity in order to enable CAT. In contrast, the Sleep-related Impairment-11 item bank has sufficient structural validity and performed well as CAT. It can properly distinguish between clinical and non-clinical groups of adolescents, but test-retest reliability was just below the recommend criterion.

References

1. Short MA, Gradisar M, Gill J, Camfferman D. Identifying adolescent sleep problems. *PLoS one*. 2013;8(9):e75301.
2. Paiva T, Gaspar T, Matos MG. Sleep deprivation in adolescents: correlations with health complaints and health-related quality of life. *Sleep medicine*. 2015;16(4):521-527.
3. Verkooijen S, de Vos N, Bakker-Camu BJW, et al. Sleep Disturbances, Psychosocial Difficulties, and Health Risk Behavior in 16,781 Dutch Adolescents. *Academic pediatrics*. 2018;18(6):655-661.
4. National Sleep Foundation. Sleep in America Poll 2014: Sleep in the modern family. <http://sleepfoundation.org/sleep-polls-data/sleep-in-america-poll/2014-sleep-in-the-modern-family/>. Accessed November 1st, 2014.
5. Paruthi S, Brooks LJ, D'Ambrosio C, et al. Recommended Amount of Sleep for Pediatric Populations: A Consensus Statement of the American Academy of Sleep Medicine. *Journal of clinical sleep medicine*. 2016;12(6):785-786.
6. Moore M, Meltzer LJ. The sleepy adolescent: causes and consequences of sleepiness in teens. *Paediatric respiratory reviews*. 2008;9(2):114-120.
7. Crowley SJ, Wolfson AR, Tarokh L, Carskadon MA. An update on adolescent sleep: New evidence informing the perfect storm model. *J Adolesc*. 2018;67:55-65.
8. Jakobsson M, Josefsson K, Hogberg K. Reasons for sleeping difficulties as perceived by adolescents: a content analysis. *Scand J Caring Sci*. 2019.
9. Carter KA, Hathaway NE, Lettieri CF. Common sleep disorders in children. *American family physician*. 2014;89(5):368-377.
10. Amaral MO, de Figueiredo Pereira CM, Silva Martins DI, de Serpa Cdo R, Sakellarides CT. Prevalence and risk factors for insomnia among Portuguese adolescents. *European journal of pediatrics*. 2013;172(10):1305-1311.
11. Chaput JP, Gray CE, Poitras VJ, et al. Systematic review of the relationships between sleep duration and health indicators in school-aged children and youth. *Applied physiology, nutrition, and metabolism*. 2016;41(6):S266-282.
12. Dregan A, Armstrong D. Adolescence sleep disturbances as predictors of adulthood sleep disturbances--a cohort study. *The Journal of adolescent health*. 2010;46(5):482-487.
13. Spruyt K, Gozal D. Pediatric sleep questionnaires as diagnostic or epidemiological tools: a review of currently available instruments. *Sleep Med Rev*. 2011;15(1):19-32.
14. Ji X, Liu J. Subjective sleep measures for adolescents: a systematic review. *Child: care, health and development*. 2016;42(6):825-839.
15. Alonso J, Bartlett SJ, Rose M, et al. The case for an international patient-reported outcomes measurement information system (PROMIS) initiative. *Health and quality of life outcomes*. 2013;11:210.
16. Cella D, Riley W, Stone A, 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. *Journal of clinical epidemiology*. 2010;63(11):1179-1194.
17. Buysse DJ, Yu L, Moul DE, et al. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. *Sleep*. 2010;33(6):781-792.
18. Yu L, Buysse DJ, Germain A, et al. Development of short forms from the PROMIS sleep disturbance and Sleep-Related Impairment item banks. *Behavioral sleep medicine*. 2011;10(1):6-24.
19. van Kooten JAM, van Litsenburg RRL, Yoder WR, Kaspers GJL, Terwee CB. Validation of the PROMIS Sleep Disturbance and Sleep-Related Impairment item banks in Dutch adolescents. *Quality of life research*. 2018;27(7):1911-1920.
20. van Kooten JA, Terwee CB, Kaspers GJ, van Litsenburg RR. Content validity of the Patient-Reported Outcomes Measurement Information System Sleep Disturbance and Sleep Related Impairment item banks in adolescents. *Health and quality of life outcomes*. 2016;14:92.
21. Bevans KB, Meltzer LJ, De La Motte A, Kratchman A, Viel D, Forrest CB. Qualitative Development and Content Validation of the PROMIS Pediatric Sleep Health Items. *Behavioral sleep medicine*. 2018:1-15.



22. Forrest CB, Meltzer LJ, Marcus CL, et al. Development and validation of the PROMIS Pediatric Sleep Disturbance and Sleep-Related Impairment item banks. *Sleep*. 2018;41(6).
23. Levenson JC, Shensa A, Sidani JE, Colditz JB, Primack BA. Social Media Use Before Bed and Sleep Disturbance Among Young Adults in the United States: A Nationally Representative Study. *Sleep*. 2017;40(9).
24. Bian J, Guo Y, Xie M, et al. Exploring the Association Between Self-Reported Asthma Impact and Fitbit-Derived Sleep Quality and Physical Activity Measures in Adolescents. *JMIR mHealth and uHealth*. 2017;5(7):e105.
25. Hanish AE, Lin-Dyken DC, Han JC. PROMIS Sleep Disturbance and Sleep-Related Impairment in Adolescents: Examining Psychometrics Using Self-Report and Actigraphy. *Nursing research*. 2017;66(3):246-251.
26. Becker SP, Langberg JM, Eadeh HM, Isaacson PA, Bourchtein E. Sleep and daytime sleepiness in adolescents with and without ADHD: differences across ratings, daily diary, and actigraphy. *Journal of child psychology and psychiatry, and allied disciplines*. 2019.
27. Richdale AL, Schreck KA. Sleep problems in autism spectrum disorders: prevalence, nature, & possible biopsychosocial aetiologies. *Sleep Med Rev*. 2009;13(6):403-411.
28. National Institutes of Health. PROMIS Domain Frameworks. <http://www.nihpromis.org/measures/domainframework1#sd>. Accessed August 30th, 2015.
29. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical care*. 2007;45(5):S22-31.
30. HealthMeasures. HealthMeasures Dataverse. <https://dataverse.harvard.edu/dataverse/HealthMeasures>. Accessed December 7th, 2018.
31. Wainer HDN, Eignor D, Flaughner R, Green BF, Mislvey RJ, Steinberg L, Thissen D. Computerized adaptive testing: A primer (2nd ed.). Mahwah, USA: Lawrence Erlbaum Associates; 2000.
32. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of life research*. 2018;27(5):1147-1157.
33. Baddam SKR, Canapari CA, van Noordt SJR, Crowley MJ. Sleep Disturbances in Child and Adolescent Mental Health Disorders: A Review of the Variability of Objective Sleep Markers. *Medical sciences*. 2018;6(2).
34. Lee AC, Driban JB, Price LL, Harvey WF, Rodday AM, Wang C. Responsiveness and Minimally Important Differences for Four Patient-Reported Outcomes Measurement Information System (PROMIS) Short Forms: Physical Function, Pain Interference, Depression, and Anxiety in Knee Osteoarthritis. *The journal of pain*. 2017;18(9):1096-1110.
35. Galland BC, Gray AR, Penno J, Smith C, Lobb C, Taylor RW. Gender differences in sleep hygiene practices and sleep quality in New Zealand adolescents aged 15 to 17 years. *Sleep Health*. 2017;3(2):77-83.
36. Moore M, Kirchner HL, Drotar D, Johnson N, Rosen C, Redline S. Correlates of adolescent sleep time and variability in sleep time: the role of individual and health related characteristics. *Sleep medicine*. 2011;12(3):239-245.
37. Dutch central bureau for statistics. StatLine: Vo; leerlingen, onderwijsoort in detail, leerjaar [high school, detailed educational level, class]. <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80040ned/table>. Accessed April 29th, 2019.
38. van Hal LBT, van Rooijen M, van der Hoff M. Een actueel perspectief op kinderen en jongeren met een chronische aandoening in Nederland: Omvang, samenstelling en participatie [A up-to-date perspective on Dutch children with a chronic condition: Extent, composition and participation]. https://www.fnozorgvoorkansen.nl/wp-content/uploads/2019/04/217026_actueel_perspectief_op_kinderen_en_WEB.pdf. Accessed May 1st, 2019.
39. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Medical care*. 2007;45(5):S3-S11.
40. Bruni O, Ottaviano S, Guidetti V, et al. The Sleep Disturbance Scale for Children (SDSC). Construction and validation of an instrument to evaluate sleep disturbances in childhood and adolescence. *Journal of sleep research*. 1996;5(4):251-261.

41. Zhou HQ, Shi WB, Wang XF, et al. An epidemiological study of sleep quality in adolescents in South China: a school-based study. *Child: care, health and development*. 2012;38(4):581-587.
42. Essner B, Noel M, Myrvik M, Palermo T. Examination of the Factor Structure of the Adolescent Sleep-Wake Scale (ASWS). *Behavioral sleep medicine*. 2015;13(4):296-307.
43. de Bruin EJ, van Kampen RK, van Kooten T, Meijer AM. Psychometric properties and clinical relevance of the adolescent sleep hygiene scale in Dutch adolescents. *Sleep medicine*. 2014;15(7):789-797.
44. Bechger TM, Maris G. A Statistical Test for Differential Item Pair Functioning. *Psychometrika*. 2015;80(2):317-340.

