SLEEP DISTURBANCE

A brief guide to the PROMIS® Sleep Disturbance instruments:

ADULT	PEDIATRIC	PARENT PROXY
PROMIS Bank v1.0 – Sleep Disturbance PROMIS Short Form v1.0 – Sleep Disturbance 4a PROMIS Short Form v1.0 – Sleep Disturbance 6a PROMIS Short Form v1.0 – Sleep Disturbance 8a PROMIS Short Form v1.0 – Sleep Disturbance 8b	PROMIS Pediatric Bank v1.0 – Sleep Disturbance PROMIS Pediatric Short Form v1.0 – Sleep Disturbance 4a PROMIS Pediatric Short Form v1.0 – Sleep Disturbance 8a	PROMIS Parent Proxy Bank v1.0 – Sleep Disturbance PROMIS Parent Proxy Short Form v1.0 – Sleep Disturbance 4a PROMIS Parent Proxy Short Form v1.0 – Sleep Disturbance 8a

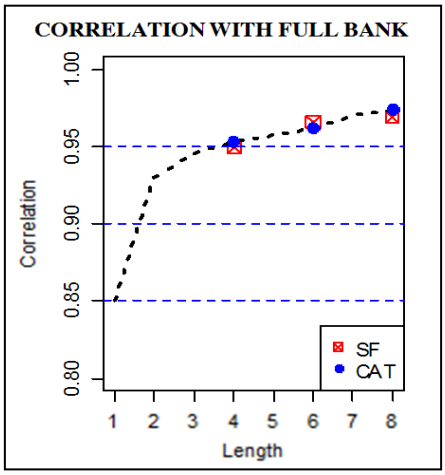
ABOUT SLEEP DISTURBANCE

The PROMIS Sleep Disturbance instruments assess self-reported perceptions of sleep quality, sleep depth, and restoration associated with sleep. This includes perceived difficulties and concerns with getting to sleep or staying asleep, as well as perceptions of the adequacy of and satisfaction with sleep. Sleep Disturbance does not focus on symptoms of specific sleep disorders, nor does it provide subjective estimates of sleep quantities (total amount of sleep, time to fall asleep, amount of wakefulness during sleep). The Sleep Disturbance short form is universal rather than disease-specific. It assesses sleep disturbance over the past seven days.

Sleep Disturbance instruments are available for adults (ages 18+), pediatric self-report (ages 8-17) and for parents serving as proxy reporters for their child (youth ages 5-17).

INTRODUCTION TO ASSESSMENT OPTIONS

There are two administration options for assessing Sleep Disturbance: short forms and computerized adaptive test (CAT). When administering a short form, instruct participants to answer all of the items (i.e., questions or statements) presented. With CAT, participant responses guide the system's choice of subsequent items from the full item bank (27 items in total). Although items differ across respondents taking CAT, scores are comparable across participants. Some administrators may prefer to ask the same question of all respondents or of the same respondent over time, to enable a more direct comparability across people or time. In these cases, or when paper administration is preferred, a short form would be more desirable than CAT. This guide provides information on all Sleep Disturbance short form and CAT instruments.



Whether one uses a short form or CAT, the score metric is Item Response Theory (IRT), a family of statistical models that link individual questions to a presumed underlying trait or concept of sleep disturbance represented by all items in the item bank. When choosing between CAT and a short form, it is useful to consider the demands of computer-based assessment, and the psychological, physical, and cognitive burden placed on respondents as a result of the number of questions asked.

Figure 1 illustrates the correlations (strength of relationship) of the full

Figure 1

bank with CAT and with short forms of varying length. The correlation of CAT scores with the full bank score is greater than a short form of any length. A longer CAT or longer short form offers greater correlation, as well as greater precision. When evaluating precision, not all questions are equally informative. The flexibility of CAT to choose more informative questions offers more precision.

SHORT FORM DIFFERENCES

Adult Profile Short Forms

You will notice that there are 4 Sleep Disturbance short forms for adults. Items in the 4a, 6a, and 8a short forms were selected based on rankings using two psychometric criteria: (1) maximum interval information; and 2) CAT simulations. Item rankings were similar for both criteria. For the maximum interval criterion, each item information function was integrated (without weighting) for the interval from the mean to 2 SDs worse than the mean. For the CAT simulations, responses to all items in each bank were generated using a random sample of 1,000 simulees drawn separately for each bank (centered on 1.0 SD worse than the general population mean). Items were rank ordered based on their average administration rank over the simulees. Content experts reviewed the items and rankings and made cuts of 8, 6, and 4 items. For each domain, 4-item, 6-item and 8-items have been selected so that the items are nested/overlap (e.g., the 8-item form is the 6-item form plus two additional items). The 4a, 6a, and 8a short forms can be administered with short forms of similar length from other domains (physical function, anxiety, pain interference, fatigue, depression, and ability to participate in social roles and activities v2.0) as part of a PROMIS Profile (see PROMIS-29, 43 or 57 Profile v2.0), though they can also be administered individually.

Other Adult Short Forms

The original adult short form (8b) was constructed by the domain team with a focus on representing the range of the trait and also representing the content of the item bank. Domain experts reviewed short forms to give input on the relevance of each item. Each domain group worked independently and the original short forms are 6-10 items long depending on the domain. Psychometric properties and clinical input were both used and likely varied in importance across domains.

Pediatric and Parent Proxy Short Forms

There are 2 Pediatric and 2 Parent Proxy short forms. Items were selected based on content and psychometric characteristics.

Selecting a Short Form

In selecting between short forms, the difference is instrument length. The reliability and precision of the short forms within a domain is highly similar. If you are working with a sample in which you want the most precise measure, select the longest short form. If you have little room for additional measures but really wanted to capture something as a secondary outcome, select one of the shorter instruments (e.g., 4-item short form).

SELECTING A PEDIATRIC OR PARENT PROXY INSTRUMENT

In selecting whether to use the pediatric or parent proxy instrument for this domain, it is important to consider both the population and the domain which you are studying. Pediatric self-report should be considered the standard for measuring patient-reported outcomes among children. However, circumstances exist when the child is too young, cognitively impaired, or too ill to complete a patient-reported outcome instrument. While

information derived from self-report and proxy-report is not equivalent, it is optimal to assess both the child and the parent since their perspectives may be independently related to healthcare utilization, risk factors, and quality of care.

SCORING THE INSTRUMENT

Short Forms: PROMIS instruments are scored using item-level calibrations. This means that the most accurate way to score a PROMIS instrument is to use the HealthMeasures Scoring Service (https://www.assessmentcenter.net/ac_scoringservice) or a data collection tool that automatically calculates scores (e.g., Assessment Center, REDCap auto-score). This method of scoring uses responses to each item for each participant. We refer to this as "response pattern scoring." Because response pattern scoring is more accurate than the use of raw score/scale score look up tables included in this manual, it is preferred. Response pattern scoring is especially useful when there is missing data (i.e., a respondent skipped an item), different groups of participants responded to different items, or you have created a new questionnaire using a subset of questions from a PROMIS item bank.

To use the scoring tables in this manual, calculate a summed score. Each question usually has five response options ranging in value from one to five. To find the total raw score for a short form with all questions answered, sum the values of the response to each question. For example, for the v2.0 adult 4-item form, the lowest possible raw score is 4; the highest possible raw score is 16 (see all short form scoring tables in Appendix 1). **All questions must be answered in order to produce a valid score using the scoring tables**. If a participant has skipped a question, use the HealthMeasures Scoring Service (https://www.assessmentcenter.net/ac_scoringservice) to generate a final score.

With the total raw score for a measure, locate the applicable score conversion table in Appendix 1 and use this table to translate the total raw score into a T-score for each participant. The T-score rescales the raw score into a standardized score with a mean of 50 and a standard deviation (SD) of 10. Therefore a person with a T-score of 40 is one SD below the mean.

For the adult PROMIS Sleep Disturbance 8a short form v1.0, a raw score of 10 converts to a T-score of 35.9 with a standard error (SE) of 3.3 (see scoring table for the 4a v1.0 short form in the Appendix). Thus, the 95% confidence interval around the observed score ranges from 29.4 to 42.4 (T-score $\pm$ (1.96*SE) or 35.9 $\pm$ (1.96*3.3).

CAT: A minimum number of items (4 for adult and adult cancer CATs and 5 for peds and parent proxy CATs) must be answered in order to receive a score for Sleep Disturbance CAT. The response to the first item will guide the system's choice of the next item for the participant. The participant's response to the second item will dictate the selection of the following question, and so on. As additional items are administered, the potential for error is reduced and confidence in the respondent's score increases. CAT will continue until either the standard error drops below a specified level (on the T-score metric 3.0 for adult and adult cancer CATs), or the participant has answered the maximum number of questions (12), whichever occurs first.

For most PROMIS instruments, a score of 50 is the average for the United States general population with a standard deviation of 10 because calibration testing was performed on a large sample of the general population. You can read more about the calibration and centering samples on HealthMeasures.net (http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis). The T-score is provided with an error term (Standard Error or SE). The Standard Error is a statistical measure of variance and represents the "margin of error" for the T-score.

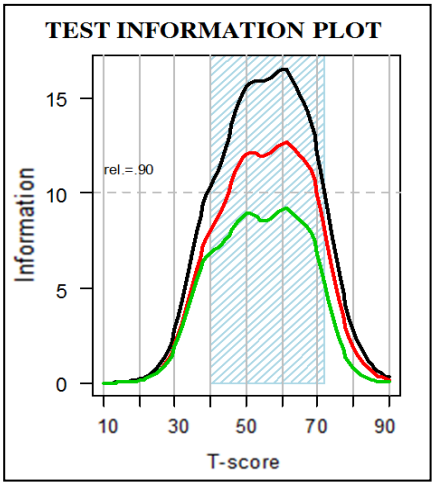


Figure 2

Important: *A higher PROMIS T-score represents more of the concept being measured.* For negatively-worded concepts like Sleep Disturbance, a T-score of 60 is one SD worse than average. By comparison, a Sleep Disturbance T-score of 40 is one SD better than average.

STATISTICAL CHARACTERISTICS

There are four key features of the score for Sleep Disturbance:

- **Reliability**: The degree to which a measure is free of error. It can be estimated by the internal consistency of the responses to the measure, or by correlating total scores on the measure from two time points when there has been no true change in what is being measured (for z-scores, reliability = $1 - SE^2$).
- **Precision**: The consistency of the estimated score (reciprocal of error variance).
- **Information:** The precision of an item or multiple items at different levels of the underlying continuum (for z-scores, information = $1/SE^2$).
- **Standard Error (SE):** The possible range of the actual final score based upon the scaled T-score. For example, with a T-score of 52 and a SE of 2, the 95% confidence interval around the actual final score ranges from 48.1 to 55.9 (T-score $\pm$ (1.96*SE) = 52 $\pm$ 3.9 = 48.1 to 55.9).

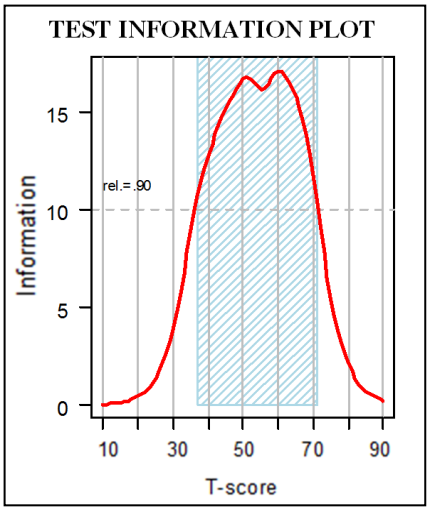


Figure 3

The final score is represented by the T-score, a standardized score with a mean of 50 and a standard deviation (SD) of 10.

In Figure 2 (adult 8b short form), the dotted horizontal line represents a degree of internal consistency reliability (i.e., .90 or .95) typically regarded as sufficient for an accurate individual score. The shaded blue region marks the range of the scale where measurement precision is comparable to the reliability of .90 for the eight-item form. Figure 2 also tells us where on the scale the form is most informative based upon the T-score. This form would typically be more informative than a Sleep Disturbance form with fewer items.

Figure 3 (adult 4a, 6a & 8a short forms) also tells us where on the scale the form is most informative based upon the T-score: the 8-item form is more informative than the 6-item form, which is more informative than the 4-item form.

Figure 4 is a sample of the statistical information available in Assessment Center for the Sleep Disturbance CAT.

More information is available at HealthMeasures.net.

Scaling Model Used For Calibration	Graded Response Model (GRM)
Total Number of Items	27

Sample	N	Alpha Reliability
Sleep Calibration	2252	0.97

Score Distributions									
	Mean	SD	P5	P10	P25	P50	P75	P90	P95
Raw	60.73	23.88	30.00	33.00	40.00	56.00	78.00	96.00	105.00
Scale	49.78	10.30	33.44	36.49	42.09	49.42	57.22	63.40	66.51

										Min	Max
Scale Score	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0	10.0	90.0
SE	2.30	.90	.40	.20	.20	.20	.20	.30	.60		
Reliability	.00	.16	.87	.96	.97	.97	.97	.92	.68		

Figure 4

PREVIEW OF SAMPLE ITEM

Figure 5 is an excerpt from the paper version of the eight-item short form. This is the paper version format used for all Sleep Disturbance instruments. It is important to note, CAT is not available for paper administration.

	In the past 7 days…	Not at all	A little bit	Somewhat	Quite a bit	Very much
Sleep116 2	My sleep was refreshing.....	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1
Sleep20 3	I had a problem with my sleep	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Sleep44 4	I had difficulty falling asleep	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Figure 5

FREQUENTLY ASKED QUESTIONS (FAQ)

Q: I am interested in learning more. Where can I do that?

Review the HealthMeasures website at www.healthmeasures.net.

Q: Do I need to register with PROMIS to use these instruments?

No.

Q: Are these instruments available in other languages?

Yes! Look at the HealthMeasures website (www.healthmeasures.net) for current information on PROMIS translations.

Q: Can I make my own short form?

Yes, custom short forms can be made by selecting any items from an item bank. This can be scored using the Scoring Service (https://www.assessmentcenter.net/ac_scoringservice).

Q: How do I handle multiple responses when administering a short form on paper?

Guidelines on how to deal with multiple responses have been established. Resolution depends on the responses noted by the research participant.

- If two or more responses are marked by the respondent, and they are next to one another, then a data entry specialist will be responsible for randomly selecting one of them to be entered and will write down on the form which answer was selected. Note: To randomly select one of two responses, the data entry specialist will flip a coin (heads - higher number will be entered; tails – lower number will be entered).To randomly select one of three (or more) responses, a table of random numbers should be used with a statistician's assistance.
- If two or more responses are marked, and they are NOT all next to one another, the response will be considered missing.

Q: What is the minimum change on a PROMIS instrument that represents a clinically meaningful difference?

To learn more about research on the meaning of a change in scores, we suggest conducting a literature review to identify the most current information. The HealthMeasures website (http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis) has additional information on interpreting scores.

APPENDIX-SCORING TABLES

Sleep Disturbance 8b
Short Form Conversion Table

Raw Score	T-score	SE*
8	28.9	4.8
9	33.1	3.7
10	35.9	3.3
11	38.0	3.0
12	39.8	2.9
13	41.4	2.8
14	42.9	2.7
15	44.2	2.7
16	45.5	2.6
17	46.7	2.6
18	47.9	2.6
19	49.0	2.6
20	50.1	2.5
21	51.2	2.5
22	52.2	2.5
23	53.3	2.5
24	54.3	2.5
25	55.3	2.5
26	56.3	2.5
27	57.3	2.5
28	58.3	2.5
29	59.4	2.5
30	60.4	2.5
31	61.5	2.5
32	62.6	2.5
33	63.7	2.6
34	64.9	2.6
35	66.1	2.7
36	67.5	2.8
37	69.0	3.0
38	70.8	3.2
39	73.0	3.5
40	76.5	4.4

*SE = Standard Error on T-score metric

Sleep Disturbance 4a
Short Form Conversion Table

Raw Score	T-score	SE*
4	32.0	5.2
5	37.5	4.0
6	41.1	3.7
7	43.8	3.5
8	46.2	3.5
9	48.4	3.4
10	50.5	3.4
11	52.4	3.4
12	54.3	3.4
13	56.1	3.4
14	57.9	3.3
15	59.8	3.3
16	61.7	3.3
17	63.8	3.4
18	66.0	3.4
19	68.8	3.7
20	73.3	4.6

*SE = Standard Error

Sleep Disturbance 6a
Short Form Conversion Table

Raw Score	T-score	SE*
6	31.7	5.1
7	36.9	3.9
8	40.1	3.5
9	42.5	3.3
10	44.6	3.2
11	46.4	3.1
12	48.0	3.0
13	49.5	3.0
14	50.9	3.0
15	52.3	2.9
16	53.6	2.9
17	54.8	2.9
18	56.1	2.9
19	57.3	2.9
20	58.5	2.9
21	59.7	2.9
22	61.0	2.9
23	62.3	2.9
24	63.6	2.9
25	65.0	2.9
26	66.5	3.0
27	68.1	3.1
28	70.0	3.3
29	72.4	3.6
30	76.1	4.4

*SE = Standard Error

Adult v1.0

Adult v1.0 - Sleep Disturbance 8a		
Short Form Conversion Table		
Raw Score	**T-Score**	**SE***
8	30.5	4.9
9	35.3	3.7
10	38.1	3.3
11	40.4	3.1
12	42.2	3
13	43.9	2.9
14	45.3	2.8
15	46.7	2.7
16	47.9	2.7
17	49.1	2.6
18	50.2	2.6
19	51.3	2.6
20	52.4	2.6
21	53.4	2.6
22	54.3	2.5
23	55.3	2.5
24	56.2	2.5
25	57.2	2.5
26	58.1	2.5
27	59.1	2.5
28	60	2.5
29	61	2.5
30	62	2.6
31	63	2.6
32	64	2.6
33	65.1	2.6
34	66.2	2.7
35	67.4	2.8
36	68.7	2.9
37	70.2	3
38	72	3.2
39	74.1	3.5
40	77.5	4.2
*Standard Error on T-score metric		

Note: The Sleep Disturbance 8a table was revised on 5/22/2014. Instruments scored prior to this date should be re-scored using this table.

Conversion table applies only when ALL items on the short form have been answered. T-score metric is a linear transformation from the IRT theta scale: T-score=10*theta+50.

Pediatric v1.0 - Sleep Disturbance 4a			
Short Form Conversion Table			
Raw Summed Score	**T Score**	**SE***	**Theta Score**
4	38.8	6	-1.12
5	45.4	4	-0.46
6	48.8	3.4	-0.12
7	51.5	3.1	0.15
8	53.7	3	0.37
9	55.8	3	0.58
10	57.9	3.1	0.79
11	60	3.1	1
12	61.9	3.1	1.19
13	63.7	3.2	1.37
14	65.5	3.2	1.55
15	67.5	3.1	1.75
16	69.3	3.1	1.93
17	71.2	3.1	2.12
18	73.3	3.3	2.33
19	75.5	3.4	2.55
20	79.1	3.8	2.91
*SE = Standard Error on T-score metric			

Pediatric v1.0 - Sleep Disturbance 8a			
Short Form Conversion Table			
Raw Summed Score	**T Score**	**SE***	**Theta Score**
8	36.6	5.6	-1.34
9	42.1	3.8	-0.79
10	44.8	3.4	-0.52
11	46.8	3	-0.32
12	48.5	2.8	-0.15
13	50	2.6	0
14	51.3	2.5	0.13
15	52.5	2.5	0.25
16	53.7	2.4	0.37
17	54.9	2.4	0.49
18	56	2.4	0.6
19	57.1	2.5	0.71
20	58.2	2.5	0.82
21	59.3	2.5	0.93
22	60.3	2.5	1.03
23	61.4	2.5	1.14
24	62.4	2.5	1.24
25	63.5	2.5	1.35
26	64.5	2.5	1.45
27	65.6	2.5	1.56
28	66.6	2.4	1.66
29	67.6	2.4	1.76
30	68.7	2.4	1.87
31	69.7	2.4	1.97
32	70.7	2.4	2.07
33	71.8	2.5	2.18
34	72.9	2.5	2.29
35	74.1	2.6	2.41
36	75.4	2.7	2.54
37	76.8	2.9	2.68
38	78.5	3.1	2.85
39	80.3	3.3	3.03
40	82.7	3.5	3.27
*SE = Standard Error on T-score metric			

Parent Proxy v1.0 – Sleep Disturbance 4a			
Short Form Conversion Table			
Raw Score	**T Score**	**T Score SE***	**Theta Score**
4	41.4	6.4	-0.86
5	48.2	4.4	-0.18
6	52.1	3.5	0.21
7	55.0	3.1	0.5
8	56.9	3.3	0.69
9	59.1	3.1	0.91
10	61.3	3.2	1.13
11	63.3	3.3	1.33
12	65.0	3.3	1.5
13	66.6	3.4	1.66
14	68.1	3.3	1.81
15	70.1	3.1	2.01
16	71.8	3.1	2.18
17	73.6	3.2	2.36
18	75.3	3.2	2.53
19	76.9	3.0	2.69
20	80.2	3.5	3.02
*Standard Error on T-score metric			

Parent Proxy v1.0 – Sleep Disturbance 8a			
Short Form Conversion Table			
Raw Score	**T Score**	**T Score SE***	**Theta Score**
8	38.7	6.1	-1.13
9	44.4	4.4	-0.56
10	47.3	3.8	-0.27
11	49.7	3.3	-0.03
12	51.5	3.1	0.15
13	53.2	2.8	0.32
14	54.7	2.7	0.47
15	56.0	2.6	0.6
16	57.2	2.6	0.72
17	58.4	2.6	0.84
18	59.6	2.6	0.96
19	60.8	2.6	1.08
20	62.0	2.6	1.2
21	63.1	2.6	1.31
22	64.2	2.6	1.42
23	65.3	2.6	1.53
24	66.3	2.6	1.63
25	67.3	2.6	1.73
26	68.4	2.6	1.84
27	69.4	2.6	1.94
28	70.5	2.5	2.05
29	71.5	2.5	2.15
30	72.5	2.5	2.25
31	73.5	2.5	2.35
32	74.6	2.5	2.46
33	75.7	2.6	2.57
34	76.8	2.7	2.68
35	78.0	2.8	2.8
36	79.3	2.9	2.93
37	80.7	3.0	3.07
38	82.3	3.1	3.23
39	84.1	3.1	3.41
40	85.6	2.9	3.56
*Standard Error on T-score metric			