SLEEP DISTURBANCE: CHILD REPORT
A brief guide to the PROMIS Sleep Disturbance Pediatric Report instruments:

Pediatric
PROMIS Pediatric Item Bank v1.0 Sleep Disturbance
PROMIS Pediatric Short Form v1.0 Sleep Disturbance 8
PROMIS Pediatric Short Form v1.0 Sleep Disturbance 4

ABOUT SLEEP DISTURBANCE

The PROMIS Pediatric Sleep Disturbance item bank assesses reported thoughts of one's sleep quality, and perceived difficulties with getting to sleep or staying asleep. High levels of sleep disturbance result from endorsement of poor sleep quality, and difficulty falling and staying asleep. Conceptual facets include sleep quality, sleep onset, and sleep continuity. The Pediatric Sleep Disturbance item bank uses a 7-day reporting period and has 15 items.

Sleep Disturbance instruments are available for pediatric self-report (ages 8-17) and for parents serving as proxy reporters for their child (youth ages 5-17).
(For complete definition see http://nihpromis.org/measures/domainframework2*)*

INTRODUCTION TO ASSESSMENT OPTIONS

There are two administration options for assessing Sleep Disturbance: <u>short forms</u> and <u>computerized adaptive test (CAT)</u>. When administering a short form, instruct participants to answer all of the items (.e., questions or statements) presented. With CAT, participant responses guide the system's choice of subsequent items from the full item bank (15 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
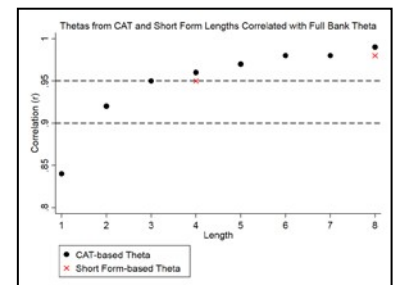

Figure 1

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 bank with CAT and with short forms of varying 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.

Some PROMIS domains have multiple versions of instruments (i.e. Item Banks/Computerized Adaptive Tests (CATs) and Short Forms). Generally, it is recommended that you use the most recent version available, which can be identified as the instruments with the highest version number.

In most cases, an instrument that has a decimal increase (v1.0 to v1.1 or v1.2) retains the same item-level parameters as well as instrument reliability and validity. In cases where a version number increases by a whole number (e.g., v1.0 to v2.0), the changes to the instrument are more substantial. For example, the PROMIS Bank v1.0 – Satisfaction with Participation in Social Roles is a small item bank comprised of 14 role function items. While the PROMIS Bank v2.0 – Satisfaction with Social Roles and Activities instrument includes a broader range of item content. Version 2.0 was re-calibrated independent of version 1.0, and is a larger (44 items), superior item bank. Details on the specific differences between instrument versions can be found in scoring manuals available on the Assessment Center homepage.

SHORT FORM DIFFERENCES

There are 2 Sleep Disturbance short forms for pediatric report: an 8-item form (SF-8) and a 4-item form (SF-4). Short form items were selected based on analyses conducted using data from the child-report calibration sample. Items in the 8 and 4 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, items were rank ordered based on their average administration rank over the simulees. The domain team constructed the short forms 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. Domain experts reviewed the items and rankings to select items for the SF-4 and SF-8. For each domain, 4-item and 8-items have been selected so that the items are nested/overlap (i the 8-item form is the 4-item form plus four additional items). The longer version of the short form will provide a more precise score with less error than the shorter short form. However, this does increase the respondent burden.

In selecting between short forms, the difference is instrument length, content coverage given fewer items, as well as possible minor differences in the reliability and precision of the short form scores. The reliability and precision of the short forms within a domain are similar. If you are

working with a sample in which you wanted the most precise measure, select the SF-8. If you are working in a sample in which you expect substantial variability in a domain area or wanted diverse content covered, you should select the SF-8. If you had little room for additional measures but really wanted to capture this domain as a secondary outcome, you should use the SF-4.

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 that 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. Since 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 utilize scoring tools, such as those found in Assessment Center, that look at responses to each item for each participant. We refer to this as "response pattern scoring." Response pattern scoring tools within Assessment Center can be used even if data was collected on paper or in another software package. Because response pattern scoring is more accurate than the use of raw score/scale score look up tables, it is preferred. However, if you aren't able to use response pattern scoring, you can use the instructions below, which rely on raw score/scale score look-up tables.

For the pediatric form, each question 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 pediatric 8-item form, the lowest possible raw score is 8; the highest possible raw score is 40 (see all short form scoring tables in Appendix 1). All questions must be answered in order to produce the most valid summed score on a short form.

A score can be approximated if a participant skips a question. If items are missing, first check how many items were answered. For the SF-8, confirm that 4 or 50% of items were answered. The SF-8 can be scored if 4 items were answered. The SF-4 form can only be scored with complete data. After confirming that enough responses were provided, sum the response scores from the items that were answered (not including any screening question). Multiply this sum by the total number of items in the short form. Finally, divide by the number of items that were answered. For example, if a respondent answered 5 of 8 questions and answered all items with the second lowest response option (2), you would sum all responses (10), multiply by the number of items in the short form (8) and divide by the number of items that were

answered (5). Here (10x8)/5=16. If the result is a fraction, round up to the nearest whole number. This is a pro-rated raw score.

Again, the formula is:
(Raw sum x number of items on the short form) Number of items that
were actually answered

Locate the applicable score conversion table in Appendix 1 and use this table to translate the total raw score or pro-rated 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. It is important to note that Assessment Center will convert a participant's pattern of responses to a standardized T-score after they have finished a CAT. The standardized T-score is reported as the final score for each participant.

For the pediatric PROMIS Sleep Disturbance SF-8, a raw score of 10 converts to a T-score of 44.8 with a standard error (SE) of 3.4 (see scoring table for the SF-8 short form in appendix). Thus, the 95% confidence interval around the observed score ranges from 38.1 to 51.5 (T-score $\pm$ (1.96*SE) or 44.8 $\pm$ (1.96*3.4).
For pro-rated scores, this calculation assumes that responses are missing at random. This isn't always true. Therefore, use caution when interpreting the final pro-rated T-score.

CAT: A minimum number of items (4 for Pediatric and Parent Proxy CATs) must be answered in order to receive a score for Sleep Disturbance CAT. The first item is selected because it provides the most information about Sleep Disturbance for the U.S. general population. The response to this item will guide the system's choice of the next item for the participant. The participant's response to this 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, or the participant has answered the maximum number of questions (12), whichever occurs first.

For most PROMIS instruments including Pediatric Sleep Disturbance, 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. This instrument was calibrated on data combined from two samples: 1) a large sample that answered all items and was comprised of participants from school, clinic and internet panel settings and 2) an additional large sample that answered only the 8-item short form and was comprised of participants from a nationally representative internet panel. 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.

e1

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 = $1/SE^2$).
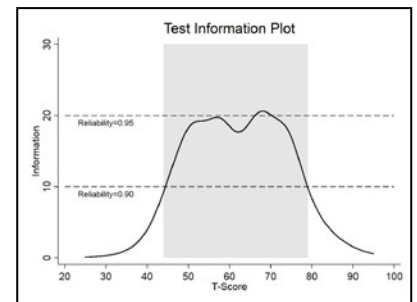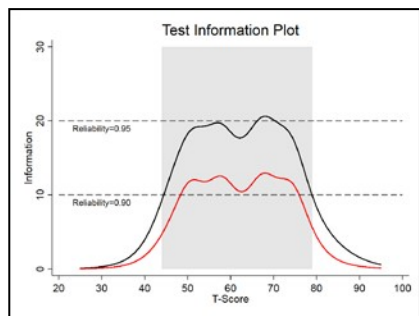

Figure 2

- **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 $^+_-$(1.96*SE) = 52 $^+_-$ 3.9 = 48.1 to 55.9).

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

Figure 3

In Figure 2 (Pediatric 8-item short form), the two dotted horizontal lines each represent 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 the 4-item Sleep Disturbance short form.

Figure 3 (Pediatric 4- and 8-item short forms) also tells us where on the scale the form is most informative based upon the T-score: the 8-item form (black) is more informative than the 4-item form (red).

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

More information is available online via Assessment Center (assessmentcenter.net).

Scaling Model Used For Calibration	Graded Response Model (GRM)
Total Number of Items	28

Sample	N	Alpha Reliability
PROMIS Wave 1 Full Bank ▾	782	0.98

Score Distributions									
	Mean	SD	P5	P10	P25	P50	P75	P90	P95
Raw	49.16	22.58	28.00	29.00	31.00	41.00	60.00	83.00	96.95
Scale	49.31	9.63	33.49	37.66	42.42	48.62	55.25	62.01	66.00

									Min	Max
Scale Score	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0	0.0	0.0
SE		13.90	4.00	1.10	.30	.10	.10	.10	.20	.50	
Reliability		.00	.00	.00	.92	.98	.99	.99	.97	.75	

Figure 4

PREVIEW OF SAMPLE ITEM

Figure 5 shows a Pediatric Sleep Disturbance item from the full item bank as it would appear to a study participant during data collection in Assessment Center. Several formats for presenting the items are available for computer-based administration through Assessment Center (see FAQ section).

Figure 5

Figure 6 is an excerpt from the paper version of the Pediatric 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…		Never	Almost Never	Sometimes	Almost Always	Always
sq005c	I had difficulty falling asleep.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
sq010c	I worried about not being able to fall asleep.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Figure 6

FREQUENTLY ASKED QUESTIONS (FAQ)

Q: *I am interested in learning more. Where can I do that?*
All instruments are available on the PROMIS website through Assessment Center, which houses all PROMIS instruments for each domain. Assessment Center is a free online data collection tool. It enables researchers to create study-specific websites for capturing participant data securely. Studies can include measures within the Assessment Center library, as well as custom instruments created or entered by the researcher. PROMIS instruments (short forms, CATs, profiles) are a central feature of the instrument library within Assessment Center. Any PROMIS measure can be included in an online study or downloaded for administration on paper. Detailed statistical information and development history about PROMIS items and instruments are available for review at nihpromis.org or assessmentcenter.net. To learn more, contact help@assessmentcenter.net.

Q: *Do I need to register with PROMIS to use these instruments?*
PROMIS instruments require endorsement of Terms and Conditions. You can click on the Request PDF button on www.assessmentcenter.net. You can also get access to instruments within Assessment Center, REDCap, and other data collection platforms (eg, Epic).

Q: *Are these instruments available in other languages?*
Many of these instruments are currently available in Spanish in Assessment Center. The PROMIS group is also working to translate this form into other languages. Information on available translations is updated periodically at http://nihpromis.org/measures/translations.

91

Q: Can I make my own short form?
Yes, custom Sleep Disturbance short forms can be made by selecting any items from the item bank. Instructions for creating a custom short form in Assessment Center can be found in the Assessment Center User Manual https://www.assessmentcenter.net/UserManuals.aspx.

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?***
This question is related to an area of active research in the PROMIS network, namely the determination of the "minimally important difference" or "MID" for a PROMIS instrument. A manuscript in the *Journal of Clinical Epidemiology* outlines the process for MIDs for adult PROMIS measures and estimates the MIDs for six PROMIS-Cancer scales: Yost, K. J., Eton, D. T., Garcia, S. F., & Cella, D. (2011). Minimally important differences were estimated for six PROMIS-Cancer scales in advanced-stage cancer patients. *Journal of Clinical Epidemiology, 64(5)*, 507-16. As described in that manuscript, the MID is a tool to enhance the interpretability of patient-reported outcomes and is often defined as the "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management" (Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials* 1989; 10(4):407-415).

APPENDIX-SCORING TABLES

SF8 Sum Score to T-Score Conversion Table				SF4 Sum Score to T-Score Conversion Table		
Raw Score	T Score	T Score SE		Raw Score	T Score	T Score SE
8	36.6	5.6		4	38.8	6
9	42.1	3.8		5	45.4	4
10	44.8	3.4		6	48.8	3.4
11	46.8	3		7	51.5	3.1
12	48.5	2.8		8	53.7	3
13	50	2.6		9	55.8	3
14	51.3	2.5		10	57.9	3.1
15	52.5	2.5		11	60	3.1
16	53.7	2.4		12	61.9	3.1
17	54.9	2.4		13	63.7	3.2
18	56	2.4		14	65.5	3.2
19	57.1	2.5		15	67.5	3.1
20	58.2	2.5		16	69.3	3.1
21	59.3	2.5		17	71.2	3.1
22	60.3	2.5		18	73.3	3.3
23	61.4	2.5		19	75.5	3.4
24	62.4	2.5		20	79.1	3.8
25	63.5	2.5				
26	64.5	2.5				
27	65.6	2.5				
28	66.6	2.4				
29	67.6	2.4				
30	68.7	2.4				
31	69.7	2.4				
32	70.7	2.4				
33	71.8	2.5				
34	72.9	2.5				
35	74.1	2.6				
36	75.4	2.7				
37	76.8	2.9				
38	78.5	3.1				
39	80.3	3.3				
40	82.7	3.5				

<div align="center">**SLEEP DISTURBANCE: PARENT PROXY REPORT**</div>
<div align="center">A brief guide to the PROMIS Sleep Disturbance Parent Proxy Report instruments:</div>

Proxy
PROMIS Proxy Item Bank v1.0 Sleep Disturbance
PROMIS Proxy Short Form v1.0 Sleep Disturbance 8
PROMIS Proxy Short Form v1.0 Sleep Disturbance 4

ABOUT SLEEP DISTURBANCE

The PROMIS Parent Proxy Sleep Disturbance item bank assesses evaluations of a child's sleep quality, and his/her perceived difficulties with getting to sleep or staying asleep. High levels of sleep disturbance result from endorsement of a child's poor sleep quality, and his/her difficulty falling and staying asleep. Conceptual facets include sleep quality, sleep onset, and sleep continuity. The Parent Proxy Sleep Disturbance item bank uses a 7-day reporting period and has 15 items.

Sleep Disturbance instruments are available for pediatric self-report (ages 8-17) and for parents serving as proxy reporters for their child (youth ages 5-17).
(For complete definition see http://nihpromis.org/measures/domainframework2*)*

INTRODUCTION TO ASSESSMENT OPTIONS

There are two administration options for assessing Sleep Disturbance: <u>short forms</u> and <u>computerized adaptive test (CAT)</u>. When administering a short form, instruct participants to answer all of the items (ie, questions or statements) presented. With CAT, participant responses guide the system's choice of subsequent items from the full item bank (15 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.

Figure 7

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.
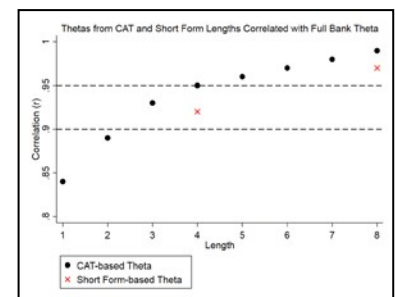
Figure 1 illustrates the correlations (strength of relationship) of the full bank with CAT and with short forms of varying length. A longer CAT or longer short form offers greater correlation, as well as greater precision. When evaluating precision, not all questions are equally informative. The flexibility of CAT to choose more informative questions offers more precision.

Some PROMIS domains have multiple versions of instruments (i.e. Item Banks/Computerized Adaptive Tests (CATs) and Short Forms). Generally, it is recommended that you use the most recent version available, which can be identified as the instruments with the highest version number.

In most cases, an instrument that has a decimal increase (v1.0 to v1.1 or v1.2) retains the same item-level parameters as well as instrument reliability and validity. In cases where a version number increases by a whole number (eg, v1.0 to v2.0), the changes to the instrument are more substantial. For example, the PROMIS Bank v1.0 – Satisfaction with Participation in Social Roles is a small item bank comprised of 14 role function items. While the PROMIS Bank v2.0 – Satisfaction with Social Roles and Activities instrument includes a broader range of item content. Version 2.0 was re-calibrated independent of version 1.0, and is a larger (44 items), superior item bank. Details on the specific differences between instrument versions can be found in scoring manuals available on the Assessment Center homepage.

SHORT FORM DIFFERENCES

There are 2 Sleep Disturbance short forms for parent proxy report: an 8-item form (SF-8) and a 4-item form (SF-4). Short form items were selected based on analyses conducted using data from the proxy-report calibration sample. Items in the 8 and 4 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, items were rank ordered based on their average administration rank over the simulees. The domain team constructed the short forms 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. Domain experts reviewed the items and rankings to select items for the SF-4 and SF-8. For each domain, 4-item and 8-items have been selected so that the items are nested/overlap (eg, the 8-item form is the 4-item form plus four additional items). The longer version of the short form will provide a more precise score with less error than the shorter short form. However, this does increase the respondent burden.

In selecting between short forms, the difference is instrument length, content coverage given fewer items, as well as possible minor differences in the reliability and precision of the short form scores. The reliability and precision of the short forms within a domain are similar. If you are working with a sample in which you wanted the most precise measure, select the SF-8. If you are working in a sample in which you expect substantial variability in a domain area or wanted diverse

content covered, you should select the SF-8. If you had little room for additional measures but really wanted to capture this domain as a secondary outcome, you should use the SF-4.

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 that 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. Since 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 utilize scoring tools, such as those found in Assessment Center, that look at responses to each item for each participant. We refer to this as "response pattern scoring." Response pattern scoring tools within Assessment Center can be used even if data was collected on paper or in another software package. Because response pattern scoring is more accurate than the use of raw score/scale score look up tables, it is preferred. However, if you aren't able to use response pattern scoring, you can use the instructions below which rely on raw score/scale score look-up tables.

For the parent proxy form, each question 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 parent proxy 8-item form, the lowest possible raw score is 8; the highest possible raw score is 40 (see all short form scoring tables in Appendix 1). All questions must be answered in order to produce the most valid summed score on a short form.

A score can be approximated if a participant skips a question. If items are missing, first check how many items were answered. For the SF-8, confirm that 4 or 50% of items were answered. The SF-8 can be scored if 4 items were answered. The SF-4 form can only be scored with complete data. After confirming that enough responses were provided, sum the response scores from the items that were answered (not including any screening question). Multiply this sum by the total number of items in the short form. Finally, divide by the number of items that were answered. For example, if a respondent answered 5 of 8 questions and answered all items with the second lowest response option (2), you would sum all responses (10), multiply by the number of items in the short form (8) and divide by the number of items that were answered (5). Here (10x8)/5=16. If the result is a fraction, round up to the nearest whole number. This is a pro-rated raw score.

Again, the formula is:
<u>(Raw sum x number of items on the short form)</u> Number of items that
were actually answered

Locate the applicable score conversion table in Appendix 1 and use this table to translate the total raw score or pro-rated 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. It is important to note that Assessment Center will convert a participant's pattern of responses to a standardized T-score after they have finished a CAT. The standardized T-score is reported as the final score for each participant.

For the parent proxy PROMIS Sleep Disturbance SF-8, a raw score of 10 converts to a T-score of 47.3 with a standard error (SE) of 3.8 (see scoring table for the SF-8 short form in appendix). Thus, the 95% confidence interval around the observed score ranges from 39.9 to 54.7 (T-score $\pm$ (1.96*SE) or 47.3 $\pm$ (1.96*3.8).

For pro-rated scores, this calculation assumes that responses are missing at random. This isn't always true. Therefore, use caution when interpreting the final pro-rated T-score.

<u>CAT</u>: A minimum number of items (4 for Pediatric and Parent Proxy CATs) must be answered in order to receive a score for Sleep Disturbance CAT. The first item is selected because it provides the most information about Sleep Disturbance for the U.S. general population. The response to this item will guide the system's choice of the next item for the participant. The participant's response to this 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, or the participant has answered the maximum number of questions (12), whichever occurs first.

For most PROMIS instruments including Parent Proxy Sleep Disturbance, 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. This instrument was calibrated on data combined from two samples: 1) a large sample that answered all items and was comprised of participants from school, clinic and internet panel settings and 2) an additional large sample that answered only the 8-item short form and was comprised of participants from a nationally representative internet panel. 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.

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 = $1 - SE^2$).

Figure 8

- **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).
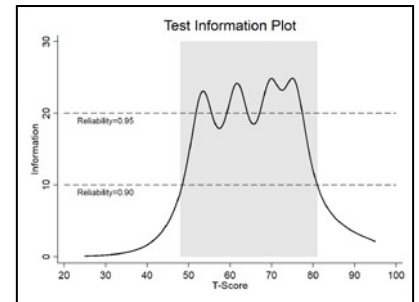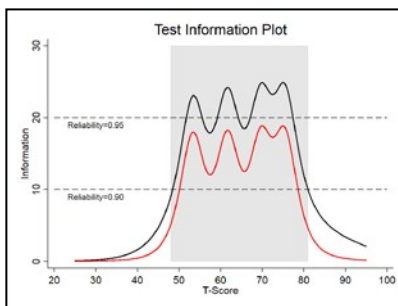
Figure 9

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 (Parent Proxy 8-item short form), the two dotted horizontal lines each represent a degree of internal consistency reliability (ie, .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 the 4-item Sleep Disturbance short form.

Figure 3 (Parent Proxy 4- and 8-item short forms) also tells us where on the scale the form is most informative based upon the T-score: the 8-item form (black) is more informative than the 4-item form (red).

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

More information is available online via Assessment Center (assessmentcenter.net).

Scaling Model Used For Calibration	Graded Response Model (GRM)
Total Number of Items	28

Sample	N	Alpha Reliability
PROMIS Wave 1 Full Bank	782	0.98

Score Distributions									
	Mean	SD	P5	P10	P25	P50	P75	P90	P95
Raw	49.16	22.58	28.00	29.00	31.00	41.00	60.00	83.00	96.95
Scale	49.31	9.63	33.49	37.66	42.42	48.62	55.25	62.01	66.00

										Min	Max
Scale Score	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0	0.0	0.0
SE	13.90	4.00	1.10	.30	.10	.10	.10	.20	.50		
Reliability	.00	.00	.00	.92	.98	.99	.99	.97	.75		

Figure 10

PREVIEW OF SAMPLE ITEM

Figure 5 shows a Parent Proxy Sleep Disturbance item from the full item bank as it would appear to a study participant during data collection in Assessment Center. Several formats for presenting the items are available for computer-based administration through Assessment Center (see FAQ section).

Figure 6 is an excerpt from the paper version of the Parent Proxy 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,

My child had a problem with his/her sleep.

○ Never
○ Almost Never
○ Sometimes
○ Almost Always
○ Always

Figure 11

In the past 7 days…	Never	Almost Never	Sometimes	Almost Always	Always
sq005p My child had difficulty falling asleep.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
sq010p My child worried about not being able to fall asleep.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Figure 6

FREQUENTLY ASKED QUESTIONS (FAQ)

Q: *I am interested in learning more. Where can I do that?*
All instruments are available on the PROMIS website through Assessment Center, which houses all PROMIS instruments for each domain. Assessment Center is a free online data collection tool. It enables researchers to create study-specific websites for capturing participant data securely. Studies can include measures within the Assessment Center library, as well as custom instruments created or entered by the researcher. PROMIS instruments (short forms, CATs, profiles) are a central feature of the instrument library within Assessment Center. Any PROMIS measure can be included in an online study or downloaded for administration on paper. Detailed statistical information and development history about PROMIS items and instruments are available for review at nihpromis.org or assessmentcenter.net. To learn more, contact help@assessmentcenter.net.

Q: *Do I need to register with PROMIS to use these instruments?*
PROMIS instruments require endorsement of Terms and Conditions. You can click on the Request PDF button on www.assessmentcenter.net. You can also get access to instruments within Assessment Center, REDCap, and other data collection platforms (e.g., Epic).

Q: *Are these instruments available in other languages?*
Many of these instruments are currently available in Spanish in Assessment Center. The PROMIS group is also working to translate this form into other languages. Information on available translations is updated periodically at http://nihpromis.org/measures/translations.

Q: Can I make my own short form?
Yes, custom Sleep Disturbance short forms can be made by selecting any items from the item bank. Instructions for creating a custom short form in Assessment Center can be found in the Assessment Center User Manual https://www.assessmentcenter.net/UserManuals.aspx.

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?***
This question is related to an area of active research in the PROMIS network, namely the determination of the "minimally important difference" or "MID" for a PROMIS instrument. A manuscript in the *Journal of Clinical Epidemiology* outlines the process for MIDs for adult PROMIS measures and estimates the MIDs for six PROMIS-Cancer scales: Yost, K. J., Eton, D. T., Garcia, S. F., & Cella, D. (2011). Minimally important differences were estimated for six PROMIS-Cancer scales in advanced-stage cancer patients. *Journal of Clinical Epidemiology, 64(5)*, 507-16. As described in that manuscript, the MID is a tool to enhance the interpretability of patient-reported outcomes and is often defined as the "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management" (Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials* 1989; 10(4):407-415).

PPENDIX-SCORING TABLES

SF8 Sum Score to T-Score	SF4 Sum Score to T-Score

	Conversion Table			**Conversion Table**	
Raw Score	T Score	T Score SE	Raw Score	T Score	T Score SE
8	38.7	6.1	4	41.4	6.4
9	44.4	4.4	5	48.2	4.4
10	47.3	3.8	6	52.1	3.5
11	49.7	3.3	7	55	3.1
12	51.5	3.1	8	56.9	3.3
13	53.2	2.8	9	59.1	3.1
14	54.7	2.7	10	61.3	3.2
15	56	2.6	11	63.3	3.3
16	57.2	2.6	12	65	3.3
17	58.4	2.6	13	66.6	3.4
18	59.6	2.6	14	68.1	3.3
19	60.8	2.6	15	70.1	3.1
20	62	2.6	16	71.8	3.1
21	63.1	2.6	17	73.6	3.2
22	64.2	2.6	18	75.3	3.2
23	65.3	2.6	19	76.9	3
24	66.3	2.6	20	80.2	3.5
25	67.3	2.6			
26	68.4	2.6			
27	69.4	2.6			
28	70.5	2.5			
29	71.5	2.5			
30	72.5	2.5			
31	73.5	2.5			
32	74.6	2.5			
33	75.7	2.6			
34	76.8	2.7			
35	78	2.8			
36	79.3	2.9			
37	80.7	3			
38	82.3	3.1			
39	84.1	3.1			
40	85.6	2.9			

SLEEP-RELATED IMPAIRMENT: CHILD REPORT

A brief guide to the PROMIS Sleep-related Impairment Pediatric Report instruments:

Pediatric
PROMIS Pediatric Item Bank v1.0 Sleep-Related Impairment
PROMIS Pediatric Short Form v1.0 Sleep-Related Impairment 8
PROMIS Pediatric Short Form v1.0 Sleep-Related Impairment 4

ABOUT SLEEP-RELATED IMPAIRMENT

The PROMIS Pediatric Sleep-related Impairment item bank assesses perceptions of sleepiness and tiredness during usual awake hours and reported impairments during the day associated with sleep problems or ability to function to one's full potential. High levels of Sleep-related impairment result from endorsement of daytime sleepiness, difficulty waking up, and problems during the day as a result of poor sleep. Conceptual facets include daytime sleepiness, sleep offset, impact: cognitive, impact: activities, and impact: emotional. The Pediatric Sleep-related Impairment item bank uses a 7-day reporting period and has 13 items.

Sleep-related Impairment instruments are available for pediatric self-report (ages 8-17) and for parents serving as proxy reporters for their child (youth ages 5-17).
(For complete definition see http://nihpromis.org/measures/domainframework2*)*

INTRODUCTION TO ASSESSMENT OPTIONS

There are two administration options for assessing Sleep-related Impairment: <u>short forms</u> and <u>computerized adaptive test (CAT)</u>. 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 (12 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-related Impairment 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-related Impairment 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,
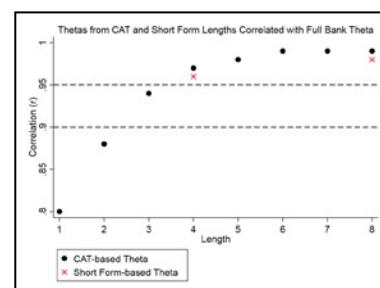
Figure 12

102

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 bank with CAT and with short forms of varying 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.

Some PROMIS domains have multiple versions of instruments (i.e. Item Banks/Computerized Adaptive Tests (CATs) and Short Forms). Generally, it is recommended that you use the most recent version available, which can be identified as the instruments with the highest version number.

In most cases, an instrument that has a decimal increase (v1.0 to v1.1 or v1.2) retains the same item-level parameters as well as instrument reliability and validity. In cases where a version number increases by a whole number (e.g., v1.0 to v2.0), the changes to the instrument are more substantial. For example, the PROMIS Bank v1.0 – Satisfaction with Participation in Social Roles is a small item bank comprised of 14 role function items. While the PROMIS Bank v2.0 – Satisfaction with Social Roles and Activities instrument includes a broader range of item content. Version 2.0 was re-calibrated independent of version 1.0, and is a larger (44 items), superior item bank. Details on the specific differences between instrument versions can be found in scoring manuals available on the Assessment Center homepage.

SHORT FORM DIFFERENCES

There are 2 Sleep-related Impairment short forms for pediatric report: an 8-item form (SF-8) and a 4-item form (SF-4). Short form items were selected based on analyses conducted using data from the child-report calibration sample. Items in the 8 and 4 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, items were rank ordered based on their average administration rank over the simulees. The domain team constructed the short forms 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. Domain experts reviewed the items and rankings to select items for the SF-4 and SF-8. For each domain, 4-item and 8-items have been selected so that the items are nested/overlap (e.g., the 8-item form is the 4-item form plus four additional items). The longer version of the short form will provide a more precise score with less error than the shorter short form. However, this does increase the respondent burden.

In selecting between short forms, the difference is instrument length, content coverage given fewer items, as well as possible minor differences in the reliability and precision of the short form scores. The reliability and precision of the short forms within a domain are similar. If you are

working with a sample in which you wanted the most precise measure, select the SF-8. If you are working in a sample in which you expect substantial variability in a domain area or wanted diverse content covered, you should select the SF-8. If you had little room for additional measures but really wanted to capture this domain as a secondary outcome, you should use the SF-4.

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 that 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. Since 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 utilize scoring tools, such as those found in Assessment Center, that look at responses to each item for each participant. We refer to this as "response pattern scoring." Response pattern scoring tools within Assessment Center can be used even if data was collected on paper or in another software package. Because response pattern scoring is more accurate than the use of raw score/scale score look up tables, it is preferred. However, if you aren't able to use response pattern scoring, you can use the instructions below, which rely on raw score/scale score look-up tables.

For the pediatric form, each question 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 pediatric 8-item form, the lowest possible raw score is 8; the highest possible raw score is 40 (see all short form scoring tables in Appendix 1). All questions must be answered in order to produce the most valid summed score on a short form.

A score can be approximated if a participant skips a question. If items are missing, first check how many items were answered. For the SF-8, confirm that 4 or 50% of items were answered. The SF-8 can be scored if 4 items were answered. The SF-4 form can only be scored with complete data. After confirming that enough responses were provided, sum the response scores from the items that were answered (not including any screening question). Multiply this sum by the total number of items in the short form. Finally, divide by the number of items that were answered. For example, if a respondent answered 5 of 8 questions and answered all items with the second lowest response option (2), you would sum all responses (10), multiply by the number of items in the short form (8) and divide by the number of items that were

answered (5). Here (10x8)/5=16. If the result is a fraction, round up to the nearest whole number. This is a pro-rated raw score.

Again, the formula is:
(Raw sum x number of items on the short form) Number of items that
were actually answered

Locate the applicable score conversion table in Appendix 1 and use this table to translate the total raw score or pro-rated 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. It is important to note that Assessment Center will convert a participant's pattern of responses to a standardized T-score after they have finished a CAT. The standardized T-score is reported as the final score for each participant.

For the pediatric PROMIS Sleep-Related Impairment SF-8, a raw score of 10 converts to a T-score of 45.9 with a standard error (SE) of 3.3 (see scoring table for the SF-8 short form in appendix). Thus, the 95% confidence interval around the observed score ranges from 39.4 to 52.4 (T-score $\pm$ (1.96*SE) or 45.9 $\pm$ (1.96*3.3).

For pro-rated scores, this calculation assumes that responses are missing at random. This isn't always true. Therefore, use caution when interpreting the final pro-rated T-score.

CAT: A minimum number of items (4 for Pediatric and Parent Proxy CATs) must be answered in order to receive a score for Sleep-related Impairment CAT. The first item is selected because it provides the most information about Sleep-related Impairment for the U.S. general population. The response to this item will guide the system's choice of the next item for the participant. The participant's response to this 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, or the participant has answered the maximum number of questions (12), whichever occurs first.

For most PROMIS instruments including Pediatric Sleep-related Impairment, 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. This instrument was calibrated on data combined from two samples: 1) a large sample that answered all items and was comprised of participants from school, clinic and internet panel settings and 2) an additional large sample that answered only the 8-item short form and was comprised of participants from a nationally representative internet panel. 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.

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-related Impairment:

- **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 = $1/SE^2$).
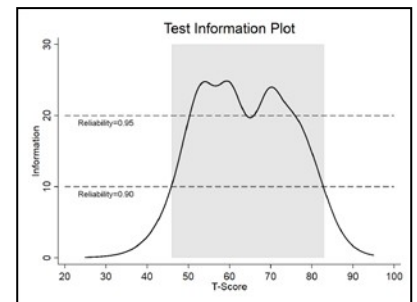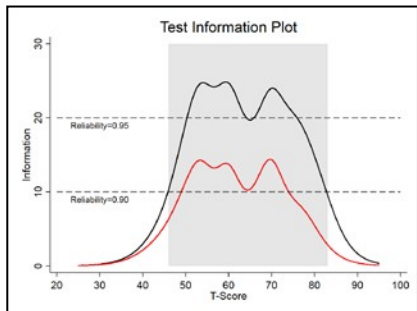

Figure 13

- **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 $^+$ (1.96*SE) = 52 $^+$ 3.9 = 48.1 to 55.9).

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

Figure 14 In Figure 2 (Pediatric 8-item short form), the two dotted horizontal lines each represent a degree of internal consistency reliability (ie, .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 the 4-item Sleep-related Impairment short form.

Figure 3 (Pediatric 4- and 8-item short forms) also tells us where on the scale the form is most informative based upon the T-score: the 8-item form (black) is more informative than the 4-item form (red).

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

More information is available online via Assessment Center (assessmentcenter.net).

Scaling Model Used For Calibration	Graded Response Model (GRM)
Total Number of Items	28

Sample	N	Alpha Reliability
PROMIS Wave 1 Full Bank	782	0.98

Score Distributions									
	Mean	SD	P5	P10	P25	P50	P75	P90	P95
Raw	49.16	22.58	28.00	29.00	31.00	41.00	60.00	83.00	96.95
Scale	49.31	9.63	33.49	37.66	42.42	48.62	55.25	62.01	66.00

										Min	Max
Scale Score	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0	0.0	0.0
SE	13.90	4.00	1.10	.30	.10	.10	.10	.20	.50		
Reliability	.00	.00	.00	.92	.98	.99	.99	.97	.75		

Figure 15

REVIEW OF SAMPLE ITEM

Figure 5 shows a Pediatric Sleep-related Impairment item from the full item bank as it would appear to a study participant during data collection in Assessment Center. Several formats for presenting the items are available for computer-based administration through Assessment Center (see FAQ section).
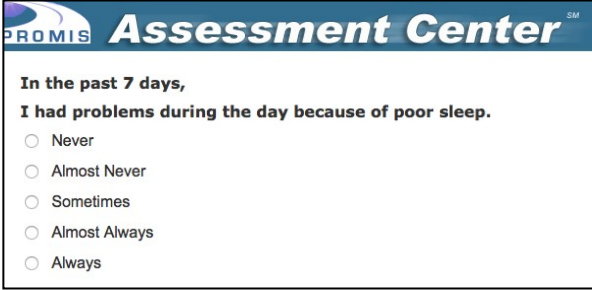

Figure 6 is an excerpt from the paper version of the Pediatric eight-item short form. This is the paper version format used for all Sleep-related Impairment instruments. It is important to note, CAT is not available for paper administration.

In the past 7 days…		Never	Almost Never	Sometimes	Almost Always	Always
w001c	I was sleepy during the daytime.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
w002c	I had trouble staying awake during the day.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Figure 6

FREQUENTLY ASKED QUESTIONS (FAQ)

Q: *I am interested in learning more. Where can I do that?*
All instruments are available on the PROMIS website through Assessment Center, which houses all PROMIS instruments for each domain. Assessment Center is a free online data collection tool. It enables researchers to create study-specific websites for capturing participant data securely. Studies can include measures within the Assessment Center library, as well as custom instruments created or entered by the researcher. PROMIS instruments (short forms, CATs, profiles) are a central feature of the instrument library within Assessment Center. Any PROMIS measure can be included in an online study or downloaded for administration on paper. Detailed statistical information and development history about PROMIS items and instruments are available for review at nihpromis.org or assessmentcenter.net. To learn more, contact help@assessmentcenter.net.

Q: *Do I need to register with PROMIS to use these instruments?*
PROMIS instruments require endorsement of Terms and Conditions. You can click on the Request PDF button on www.assessmentcenter.net. You can also get access to instruments within Assessment Center, REDCap, and other data collection platforms (e.g., Epic).

Q: *Are these instruments available in other languages?*

Many of these instruments are currently available in Spanish in Assessment Center. The PROMIS group is also working to translate this form into other languages. Information on available translations is updated periodically at http://nihpromis.org/measures/translations.

Q: Can I make my own short form?
Yes, custom Sleep-related Impairment short forms can be made by selecting any items from the item bank. Instructions for creating a custom short form in Assessment Center can be found in the Assessment Center User Manual https://www.assessmentcenter.net/UserManuals.aspx.

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?*
This question is related to an area of active research in the PROMIS network, namely the determination of the "minimally important difference" or "MID" for a PROMIS instrument. A manuscript in the *Journal of Clinical Epidemiology* outlines the process for MIDs for adult PROMIS measures and estimates the MIDs for six PROMIS-Cancer scales: Yost, K. J., Eton, D. T., Garcia, S. F., & Cella, D. (2011). Minimally important differences were estimated for six PROMIS-Cancer scales in advanced-stage cancer patients. *Journal of Clinical Epidemiology, 64(5)*, 507-16. As described in that manuscript, the MID is a tool to enhance the interpretability of patient-reported outcomes and is often defined as the "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management" (Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials* 1989; 10(4):407-415).

PPENDIX-SCORING TABLES

SF8 Sum Score to T-Score Conversion Table				SF4 Sum Score to T-Score Conversion Table		
Raw Score	T Score	T Score SE		Raw Score	T Score	T Score SE
8	37.4	5.7		4	38.3	6
9	43.3	3.8		5	44.7	4.1
10	45.9	3.3		6	47.8	3.7
11	48.1	2.8		7	50.6	3.1
12	49.7	2.5		8	52.9	2.9
13	51.2	2.3		9	55.1	2.8
14	52.5	2.2		10	57.2	2.8
15	53.7	2.1		11	59.5	2.9
16	54.8	2.1		12	61.9	2.9
17	55.9	2.1		13	64.1	3
18	57	2.1		14	66.1	3.1
19	58.1	2.1		15	68.4	3
20	59.2	2.1		16	70.7	3
21	60.3	2.1		17	73	3.1
22	61.5	2.2		18	75.7	3.3
23	62.7	2.2		19	79.6	3.9
24	63.9	2.2				
25	65.1	2.3				
26	66.2	2.3				
27	67.4	2.3				
28	68.6	2.2				
29	69.7	2.2				
30	70.9	2.2				
31	72	2.2				
32	73.1	2.2				
33	74.2	2.2				
34	75.4	2.3				
35	76.6	2.3				
36	78	2.4				
37	79.6	2.6				
38	81.4	2.8				
39	84	2.9				

SLEEP-RELATED IMPAIRMENT: PARENT PROXY REPORT
A brief guide to the PROMIS Sleep-related Impairment Parent Proxy Report instruments:

Proxy
PROMIS Proxy Item Bank v1.0
Sleep-related Impairment
PROMIS Proxy Short Form v1.0
Sleep-related Impairment 8
PROMIS Proxy Short Form v1.0
Sleep-related Impairment 4

ABOUT SLEEP-RELATED IMPAIRMENT

The PROMIS Parent Proxy Sleep-related Impairment item bank assesses evaluations of a child's sleepiness and tiredness during usual awake hours and reported impairments during the day associated with sleep problems or ability to function to a child's full potential. High levels of Sleep-related impairment result from endorsement of a child's daytime sleepiness, difficulty waking up, and problems during the day as a result of poor sleep. Conceptual facets include daytime sleepiness, sleep offset, impact: cognitive, impact: activities, and impact: emotional. The Parent Proxy Sleep-related Impairment item bank uses a 7-day reporting period and has 13 items.

Sleep-related Impairment instruments are available for pediatric self-report (ages 8-17) and for parents serving as proxy reporters for their child (youth ages 5-17).
(For complete definition see http://nihpromis.org/measures/domainframework2*)*

INTRODUCTION TO ASSESSMENT OPTIONS

There are two administration options for assessing Sleep-related Impairment: <u>short forms</u> and <u>computerized adaptive test (CAT)</u>. 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 (12 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-related Impairment 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-related Impairment 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
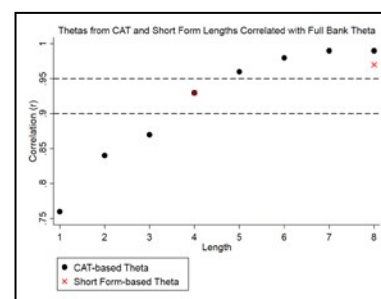
Figure 17

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 bank with CAT and with short forms of varying 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.

Some PROMIS domains have multiple versions of instruments (i.e. Item Banks/Computerized Adaptive Tests (CATs) and Short Forms). Generally, it is recommended that you use the most recent version available, which can be identified as the instruments with the highest version number.

In most cases, an instrument that has a decimal increase (v1.0 to v1.1 or v1.2) retains the same item-level parameters as well as instrument reliability and validity. In cases where a version number increases by a whole number (e.g., v1.0 to v2.0), the changes to the instrument are more substantial. For example, the PROMIS Bank v1.0 – Satisfaction with Participation in Social Roles is a small item bank comprised of 14 role function items. While the PROMIS Bank v2.0 – Satisfaction with Social Roles and Activities instrument includes a broader range of item content. Version 2.0 was re-calibrated independent of version 1.0, and is a larger (44 items), superior item bank. Details on the specific differences between instrument versions can be found in scoring manuals available on the Assessment Center homepage.

SHORT FORM DIFFERENCES

There are 2 Sleep-related Impairment short forms for parent proxy report: an 8-item form (SF-8) and a 4-item form (SF-4). Short form items were selected based on analyses conducted using data from the proxy-report calibration sample. Items in the 8 and 4 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, items were rank ordered based on their average administration rank over the simulees. The domain team constructed the short forms 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. Domain experts reviewed the items and rankings to select items for the SF-4 and SF-8. For each domain, 4-item and 8-items have been selected so that the items are nested/overlap (e.g., the 8-item form is the 4-item form plus four additional items). The longer version of the short form will provide a more precise score with less error than the shorter short form. However, this does increase the respondent burden.

In selecting between short forms, the difference is instrument length, content coverage given fewer items, as well as possible minor differences in the reliability and precision of the short form scores. The reliability and precision of the short forms within a domain are similar. If you are working with a sample in which you wanted the most precise measure, select the SF-8. If you are

working in a sample in which you expect substantial variability in a domain area or wanted diverse content covered, you should select the SF-8. If you had little room for additional measures but really wanted to capture this domain as a secondary outcome, you should use the SF-4.

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 that 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. Since 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 utilize scoring tools, such as those found in Assessment Center, that look at responses to each item for each participant. We refer to this as "response pattern scoring." Response pattern scoring tools within Assessment Center can be used even if data was collected on paper or in another software package. Because response pattern scoring is more accurate than the use of raw score/scale score look up tables, it is preferred. However, if you aren't able to use response pattern scoring, you can use the instructions below which rely on raw score/scale score look-up tables.

For the parent proxy form, each question 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 parent proxy 8-item form, the lowest possible raw score is 8; the highest possible raw score is 40 (see all short form scoring tables in Appendix 1). All questions must be answered in order to produce the most valid summed score on a short form.

A score can be approximated if a participant skips a question. If items are missing, first check how many items were answered. For the SF-8, confirm that 4 or 50% of items were answered. The SF-8 can be scored if 4 items were answered. The SF-4 form can only be scored with complete data. After confirming that enough responses were provided, sum the response scores from the items that were answered (not including any screening question). Multiply this sum by the total number of items in the short form. Finally, divide by the number of items that were answered. For example, if a respondent answered 5 of 8 questions and answered all items with the second lowest response option (2), you would sum all responses (10), multiply by the number of items in the short form (8) and divide by the number of items that were

answered (5). Here (10x8)/5=16. If the result is a fraction, round up to the nearest whole number. This is a pro-rated raw score.

Again, the formula is:
(Raw sum x number of items on the short form) Number of items that
were actually answered

Locate the applicable score conversion table in Appendix 1 and use this table to translate the total raw score or pro-rated 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. It is important to note that Assessment Center will convert a participant's pattern of responses to a standardized T-score after they have finished a CAT. The standardized T-score is reported as the final score for each participant.

For the parent proxy PROMIS Sleep-related Impairment SF-8, a raw score of 10 converts to a T-score of 47.0 with a standard error (SE) of 3.9 (see scoring table for the SF-8 short form in appendix). Thus, the 95% confidence interval around the observed score ranges from 39.4 to 54.6 (T-score $\pm$ (1.96*SE) or 47.0 $\pm$ (1.96*3.9).
For pro-rated scores, this calculation assumes that responses are missing at random. This isn't always true. Therefore, use caution when interpreting the final pro-rated T-score.
CAT: A minimum number of items (4 for Pediatric and Parent Proxy CATs) must be answered in order to receive a score for Sleep-related Impairment CAT. The first item is selected because it provides the most information about Sleep-related Impairment for the U.S. general population. The response to this item will guide the system's choice of the next item for the participant. The participant's response to this 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, or the participant has answered the maximum number of questions (12), whichever occurs first.

For most PROMIS instruments including Parent Proxy Sleep-related Impairment, 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. This instrument was calibrated on data combined from two samples: 1) a large sample that answered all items and was comprised of participants from school, clinic and internet panel settings and 2) an additional large sample that answered only the 8-item short form and was comprised of participants from a nationally representative internet panel. 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.

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-related
Impairment:

Figure 18

- **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 = $1/SE^2$).
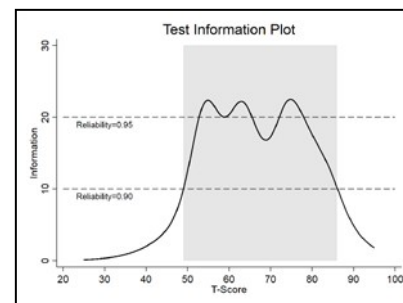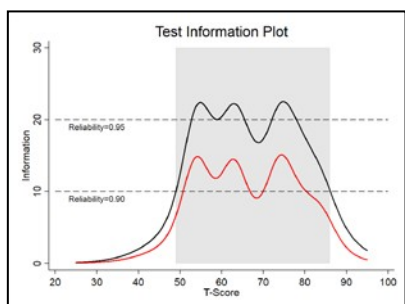

Figure 19

- **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 $^+$ (1.96*SE) = 52 $^+$ 3.9 = 48.1 to 55.9).

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 (Parent Proxy 8-item short form), the two dotted horizontal lines each represent a degree of internal consistency reliability (ie, .90 or .95) typically regarded as sufficient for an accurate individual score. The shaded blueregion 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 the 4-item Sleep-related Impairment short form.

Figure 3 (Parent Proxy 4- and 8-item short forms) also tells us where on the scale the form is most informative based upon the T-score: the 8-item form (black) is more informative than the 4-item form (red).

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

More information is available online via Assessment Center (assessmentcenter.net).

PREVIEW OF SAMPLE ITEM

Scaling Model Used For Calibration	Graded Response Model (GRM)
Total Number of Items	28

Sample	N	Alpha Reliability
PROMIS Wave 1 Full Bank ▼	782	0.98

Score Distributions								
	Mean	SD	P5	P10	P25	P50	P75	P90	P95
Raw	49.16	22.58	28.00	29.00	31.00	41.00	60.00	83.00	96.95
Scale	49.31	9.63	33.49	37.66	42.42	48.62	55.25	62.01	66.00

										Min	Max
Scale Score	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0	0.0	0.0
SE		13.90	4.00	1.10	.30	.10	.10	.10	.20	.50	
Reliability		.00	.00	.00	.92	.98	.99	.99	.97	.75	

Figure 20

114

Figure 5 shows a Parent Proxy Sleep-related Impairment item from the full item bank as it would appear to a study participant during data collection in Assessment Center. Several formats for presenting the items are available for computer-based administration through Assessment Center (see FAQ section).

Figure 6 is an excerpt from the paper version of the Parent Proxy eight-item short form. This is the paper version format used for all Sleep-related Impairment instruments. It is important to note, CAT is not available for paper administration.

In the past 7 days,

My child had problems during the day because of poor sleep.

- Never
- Almost Never
- Sometimes
- Almost Always
- Always

Figure 21

In the past 7 days…		Never	Almost Never	Sometimes	Almost Always	Always
w001p	My child was sleepy during the daytime.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
w002p	My child had trouble staying awake during the day.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Figure 6

FREQUENTLY ASKED QUESTIONS (FAQ)

Q: *I am interested in learning more. Where can I do that?*
All instruments are available on the PROMIS website through Assessment Center, which houses all PROMIS instruments for each domain. Assessment Center is a free online data collection tool. It enables researchers to create study-specific websites for capturing participant data securely. Studies can include measures within the Assessment Center library, as well as custom instruments created or entered by the researcher. PROMIS instruments (short forms, CATs, profiles) are a central feature of the instrument library within Assessment Center. Any PROMIS measure can be included in an online study or downloaded for administration on paper. Detailed statistical information and development history about PROMIS items and instruments are available for review at nihpromis.org or assessmentcenter.net. To learn more, contact help@assessmentcenter.net.

Q: *Do I need to register with PROMIS to use these instruments?*
PROMIS instruments require endorsement of Terms and Conditions. You can click on the Request PDF button on www.assessmentcenter.net. You can also get access to instruments within Assessment Center, REDCap, and other data collection platforms (e.g., Epic).

Q: *Are these instruments available in other languages?*
Many of these instruments are currently available in Spanish in Assessment Center. The PROMIS group is also working to translate this form into other languages. Information on available translations is updated periodically at http://nihpromis.org/measures/translations.
Q: *Can I make my own short form?*

Yes, custom Sleep-related Impairment short forms can be made by selecting any items from the item bank. Instructions for creating a custom short form in Assessment Center can be found in the Assessment Center User Manual https://www.assessmentcenter.net/UserManuals.aspx.

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?*

This question is related to an area of active research in the PROMIS network, namely the determination of the "minimally important difference" or "MID" for a PROMIS instrument. A manuscript in the *Journal of Clinical Epidemiology* outlines the process for MIDs for adult PROMIS measures and estimates the MIDs for six PROMIS-Cancer scales: Yost, K. J., Eton, D. T., Garcia, S. F., & Cella, D. (2011). Minimally important differences were estimated for six PROMIS-Cancer scales in advanced-stage cancer patients. *Journal of Clinical Epidemiology, 64(5)*, 507-16. As described in that manuscript, the MID is a tool to enhance the interpretability of patient-reported outcomes and is often defined as the "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management" (Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials* 1989; 10(4):407-415).

APPENDIX-SCORING TABLES

SF8 Sum Score to T-Score Conversion Table

Raw Score	T Score	T Score SE
8	37.9	6.1
9	43.8	4.5
10	47	3.9
11	49.6	3.3
12	51.7	2.8
13	53.5	2.5
14	55	2.3
15	56.4	2.3
16	57.7	2.3
17	58.9	2.3
18	60.2	2.3
19	61.5	2.3
20	62.9	2.3
21	64.3	2.3
22	65.6	2.3
23	67	2.4
24	68.4	2.4
25	69.8	2.4
26	71.2	2.4
27	72.6	2.4
28	73.9	2.3
29	75.3	2.3
30	76.6	2.3
31	77.9	2.4
32	79.2	2.5
33	80.6	2.6
34	82.1	2.7
35	83.7	2.7
36	85.3	2.5
37	86.6	2.2

SF4 Sum Score to T-Score Conversion Table

Raw Score	T Score	T Score SE
4	40	6.4
5	46.7	4.4
6	50.7	3.7
7	54	3
8	56.5	2.8
9	58.7	2.8
10	61.1	2.8
11	63.7	2.9
12	66.4	2.9
13	68.9	3.1
14	71.3	3.1
15	73.8	3
16	76.1	2.9
17	78.4	3
18	80.9	3.2
19	84.3	3