FATIGUE
MEASURE DIFFERENCES

A brief guide to differences between the PROMIS® Fatigue instruments:

ADULT	PEDIATRIC	PARENT PROXY
PROMIS Item Bank v1.0 – Fatigue PROMIS Item Bank v1.0 – Fatigue (recommended) PROMIS Item Bank v1.0 – Fatigue (screen-to-CAT) PROMIS Short Form v1.0 – Fatigue 4a PROMIS Short Form v1.0 – Fatigue 6a PROMIS Short Form v1.0 – Fatigue 7a PROMIS Short Form v1.0 – Fatigue 7b Daily PROMIS Short Form v1.0 – Fatigue 8a PROMIS Short Form v1.0 – Fatigue 13a (FACIT-Fatigue) PROMIS – Ca Bank v1.0 Fatigue PROMIS Short Form v1.0 – Fatigue – Multiple Sclerosis 8a	PROMIS Pediatric Bank v2.0 – Fatigue PROMIS Pediatric Bank v2.0 – Fatigue (recommended) PROMIS Pediatric Bank v2.0 – Fatigue (screen-to-CAT) PROMIS Pediatric Short Form v2.0 – Fatigue 10a PROMIS Pediatric Bank v1.0 – Fatigue* PROMIS Pediatric Short Form v1.0 – Fatigue 10a*	PROMIS Parent Proxy Bank v2.0 – Fatigue PROMIS Parent Proxy Bank v2.0 – Fatigue (recommended) PROMIS Parent Proxy Bank v2.0 – Fatigue (screen-to-CAT) PROMIS Parent Proxy Short Form v2.0 – Fatigue 10a PROMIS Parent Proxy Bank v1.0 – Fatigue* PROMIS Parent Proxy Short Form v1.0 – Fatigue 10a*

*Retired Measure

ABOUT FATIGUE

The PROMIS Fatigue item banks assess a range of self-reported symptoms, from mild subjective feelings of tiredness to an overwhelming, debilitating, and sustained sense of exhaustion that likely decreases one's ability to execute daily activities and function normally in family or social roles. Fatigue is divided into the experience of fatigue (frequency, duration, and intensity) and the impact of fatigue on physical, mental, and social activities. The fatigue short forms are universal rather than disease-specific. All assess fatigue over the past seven days.

Fatigue 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 fatigue: short forms and a computer adaptive tests (CATs). When administering a short form, instruct participants to answer all of the items (i.e., questions or statements) presented. With a CAT, participant responses guide the system's choice of subsequent items from the full item bank (95 items in total for adults). Although items differ across respondents taking a 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 a CAT. This guide provides information on all fatigue short form and CAT instruments.

CAT: A minimum number of items (e.g., 4) must be answered in order to receive a score for Fatigue 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 (e.g., on the T-score metric 3.0), or the participant has answered the maximum number of questions (e.g., 12), whichever occurs first. For some CATs, specifically "recommended" and "screen-to-CAT" there are additional stopping rules. These include stopping when the standard error isn't improving much or if a respondent is asymptomatic. For details on the exact stopping rules for Fatigue CATs, see below.

CAT versus Short Form: Whether one uses a short form or a 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 fatigue represented by all items in the item bank. When choosing between a CAT and 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

Figure 1 illustrates the correlations (strength of relationship) of the full bank with a 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 a CAT to choose more informative questions offers more precision.
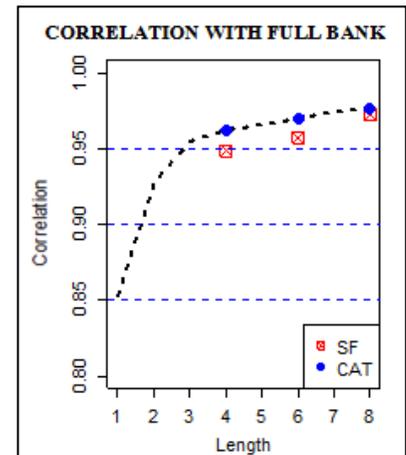
VERSION DIFFERENCES

Some PROMIS domains have multiple versions of instruments (i.e. v1.0, v1.1, v2.0). Generally, **it is recommended that you use the most recent version available which can be identified as the instrument with the highest version number.** In most cases, an instrument that has a decimal increase (v1.0 to v1.1) 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.

Adult
Standard, Recommended, and Screen-to-CAT Stopping Rules: The standard, recommended, and screen-to-CAT Adult Fatigue computer adaptive tests are based on the exact same item banks, but utilize different stopping rules. The PROMIS Bank v1.0 – Fatigue measure is administered by default as computer adaptive tests using the following standard stopping rules:

- Minimum number of items administered = 4
- Stop when one of these occurs:
    - 12 items are administered OR
    - Standard error is below 0.3 on the theta metric (3.0 on the T-score metric)

The PROMIS Bank v1.0 – Fatigue (recommended) measure uses the following stopping rules:

- Minimum number of items administered = 4
- Stop when one of these occurs:
    - 8 items are administered OR
    - Standard error is below 0.3 on the theta metric (3.0 on the T-score metric) OR
    - Standard error changes by less than 0.01 on the theta metric (0.1 on the T-score metric)

The PROMIS Bank v1.0 – Fatigue (screen-to-CAT) measure uses the following stopping rules:

- If the response to the first item is the "healthiest" response then stop.
- If the response to the first item is NOT the "healthiest" response, proceed with the "recommended" CAT stopping rules.

Pediatric and Parent Proxy

For fatigue, v2.0 pediatric and parent proxy measures replaced v1.0. The v2.0 measures 1) changed from using response scores of 0-4 to use 1-5 (item IDs amended with an "r") and 2) added new items (item IDs start with 7000). The calibrations between v1.0 and v2.0 are identical as is the item content on short forms.

Standard, Recommended, and Screen-to-CAT Stopping Rules: The standard, recommended, and screen-to-CAT Pediatric and Parent Proxy Fatigue computer adaptive tests are based on the exact same item banks, but utilize different stopping rules. The PROMIS Pediatric Bank v2.0 – Fatigue and PROMIS Parent Proxy Bank v2.0 – Fatigue measures are administered by default as computer adaptive tests using the following standard stopping rules:

- Minimum number of items administered = 5
- Stop when one of these occurs:
  - 12 items are administered OR
  - Standard error is below 0.4 on the theta metric (4.0 on the T-score metric)

The PROMIS Pediatric Bank v2.0 – Fatigue (recommended) and PROMIS Parent Proxy Bank v2.0 – Fatigue (recommended) measures use the following stopping rules:

- Minimum number of items administered = 5
- Stop when one of these occurs:
  - 12 items are administered OR
  - Standard error is below 0.4 on the theta metric (4.0 on the T-score metric) OR
  - Standard error changes by less than 0.01 on the theta metric (0.1 on the T-score metric)

The PROMIS Pediatric Bank v2.0 – Fatigue (screen-to-CAT) and PROMIS Parent Proxy Bank v2.0 – Fatigue (screen-to-CAT) measures use the following stopping rules:

- If the responses to the first two items are both the "healthiest" responses then stop.
- If the responses to the first two items are NOT the "healthiest" responses, proceed with the "recommended" CAT stopping rules.

SHORT FORM DIFFERENCES

Adult Profile Short Forms

There are multiple fatigue short forms for adults. The 4a, 6a, and 8a short forms are part of PROMIS Profile measures. 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 (anxiety, pain interference, depression, sleep disturbance,

ability to participate in social roles and activities (v2.0) and physical function (6b and 8b NOT 6a and 8a)) as part of a PROMIS Profile (see PROMIS-29, 43 or 57 Profile v2.0), though they can also be administered individually.

Additional Adult Short Forms
The original adult short form (7a) 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 the short form to give input on the relevance of each item. Each domain group (anxiety, depression, fatigue, pain interference, physical function) 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.

The PROMIS 7b Daily short form includes 5 items from the 7a short form plus 2 other items from the PROMIS Fatigue v1.0 item bank. However, it asks respondents to evaluate his/her fatigue since waking up rather than over the past 7 days.

The PROMIS 13a (FACIT-Fatigue) short form includes 13 items that match the items in the FACIT Fatigue Scale. The item content and responses are identical to the fatigue items in the full bank.

The PROMIS Fatigue–Multiple Sclerosis 8a is an 8-item short form derived from the PROMIS Fatigue v1.0 Item Bank and designed to assess the experience and impacts of fatigue in adults with multiple sclerosis (MS). Candidate items from the Fatigue Item Bank were identified with input on relevance from people with MS and clinical experts. Final item selection was informed by stakeholder feedback, item discrimination and performance, and maximum coverage of the full Fatigue Item Bank content. Consistent with scoring guidelines for other PROMIS measures, the Fatigue–MS 8a short form is scored on a T-score metric, which is centered on the U.S. general population (mean = 50; SD = 10). The PROMIS Fatigue-Multiple Sclerosis 8a short form is appropriate for use with respondents diagnosed with MS although the items do not reference MS.

Pediatric and Parent Proxy Short Forms
There is 1 pediatric and 1 parent proxy short form. 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).

PROMIS ADULT CANCER MEASURES
PROMIS-Cancer (PROMIS-Ca) measures (Physical Function, Fatigue, Pain Interference, Depression and Anxiety) were developed under the PROMIS Cancer Supplement (CaPS) grant from NCI. The measures are highly similar to PROMIS measures. Some banks include unique items. In rare instances, a shared item uses different item-level calibrations in each bank.

- PROMIS-Ca Bank v1.1 - Physical Function contains 45 items, 33 of which are also in PROMIS Bank v2.0 - Physical Function.
- PROMIS-Ca Bank v1.0 - Fatigue contains 54 items, all of which are from PROMIS Bank v1.0 - Fatigue.
- PROMIS-Ca Bank v1.0 - Anxiety contains 22 items; 20 items from PROMIS Bank v1.0 - Anxiety, and 2 items unique to CaPS in which cancer specific calibrations were used: EDANX09 & EDANX39.

- PROMIS-Ca Bank v1.0 - Depression item bank contains 30 items; 23 items are from PROMIS Bank v1.0 - Depression and 7 items unique to CaPS in which cancer specific calibrations were used: EDANG09, EDANG29, EDDEP02, EDDEP12, EDDEP16, EDDEP38 & EDDEP55.
- PROMIS-Ca Bank v1.1 - Pain Interference contains 35 items; 32 items from PROMIS Bank v1.1 - Pain Interference v1.1 and 3 items unique to CaPS in which cancer specific calibrations were used: PAININ4, PAININ15 & PAININ30.

PROMIS-Cancer (PROMIS-Ca) measures were developed by having content experts review the adult PROMIS item banks for anxiety, depression, fatigue, pain interference, and physical function. Items were selected through expert consensus and informed by focus groups and cognitive interviews with cancer patients. Multidisciplinary clinical input was obtained to ensure content coverage and the relevance of PROMIS items to patients' cancer and/or cancer treatment experiences. Items' psychometric properties were reviewed when applicable. Next, calibration testing was conducted with cancer patients with different diagnoses and treatments. Data were analyzed to identify if items performed differently in people with cancer than people with other chronic conditions or in the general population. In most cases, PROMIS calibrations ("PROMIS Wave 1") were retained. In rare cases where differential item functioning was identified, calibrations for that item were revised for when that item is used in the PROMIS-Ca item bank. For items that exist only in a PROMIS-Ca item bank, new calibrations were created by using a fixed parameter linking strategy. This set of calibrations is named "Cancer" in the HealthMeasures Scoring Service.

A fixed parameter linking approach was taken because of the additional analyses that were conducted to evaluate the differences between the PROMIS item bank and the PROMIS-Ca item bank. The measures produce slightly different scores. This difference was determined to be so small that comparing scores from a PROMIS measure and PROMIS-Ca measure is acceptable. Because the PROMIS measures have demonstrated validity across diverse patient populations, are linked with other PRO measures (i.e., PROsetta Stone), and have continued to be improved through item bank expansion (e.g., PROMIS Physical Function item bank v2.0), it is recommended to use the general population PROMIS calibrations when assessing individuals with cancer.

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.

WHICH CALIBRATION SAMPLE SHOULD I USE?

Some PROMIS Parent Proxy instruments (Anxiety, Depressive Symptoms, Fatigue, Mobility, Pain Interference, Peer Relationships) have two calibration samples – "Parent Proxy" and "Parent Proxy Without Local Dependence." The former (Parent Proxy) includes calibrations for all items. This is the default calibration sample. If you aren't sure which calibration sample to use, utilize this one. The Parent Proxy Without Local Dependence does not include calibrations for some items. The items without calibrations are enemy items. That is, a dyad or triad of items was identified in which there are psychometric reasons to only administer one of those items to a given respondent.

SCORES

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.

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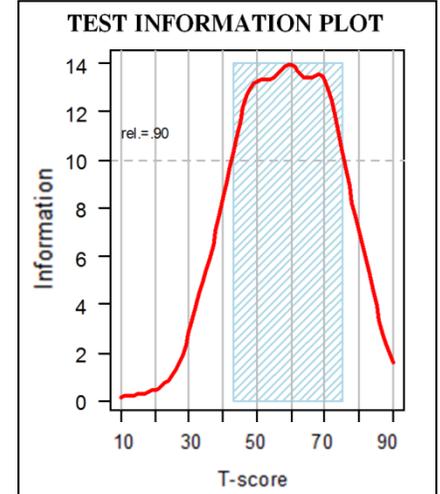

TEST INFORMATION PLOT

Figure 2

STATISTICAL CHARACTERISTICS

There are four key features of the score for fatigue:

- **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).
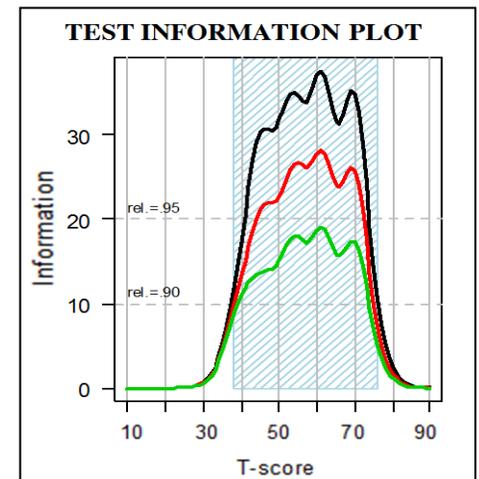

TEST INFORMATION PLOT

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 7a short form), the dotted horizontal line represents a degree of internal consistency reliability (i.e., .90) 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 seven-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 fatigue form with fewer items.

Scaling Model Used For Calibration	Graded Response Model (GRM)
Total Number of Items	124

Sample	N	Alpha Reliability
Promis ▾	14931	0.99

Score Distributions									
	Mean	SD	P5	P10	P25	P50	P75	P90	P95
Raw	220.69	91.34	98.00	109.00	142.00	206.00	284.00	356.00	392.00
Scale	51.23	9.41	36.46	39.02	44.31	50.82	57.83	63.97	67.33

										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	1.70	.70	.20	.10	.10	.10	.10	.10	.30		
Reliability	.00	.52	.94	.99	.99	.99	.99	.99	.90		

Figure 4

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. See additional test information figures for pediatric instruments in Appendix 1.

Figure 4 is a sample of the statistical information available for the adult fatigue CAT.

More information is available on www.HealthMeasures.net.

PREVIEW OF SAMPLE ITEM

Figure 5 is an excerpt from the paper version of the adult seven-item short form. This is the paper version format used for all fatigue instruments. It is important to note that the CAT is not available for paper administration.

	In the past 7 days…	Never	Rarely	Sometimes	Often	Always
FATEXP20	How often did you feel tired?...............	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
FATEXP5	How often did you experience extreme exhaustion?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Figure 5

FREQUENTLY ASKED QUESTIONS (FAQs)

Q: I am interested in learning more. Where can I do that?
Review the HealthMeasures website at www.healthmeasures.net.

Q: Are these instruments available in other languages?
Yes! Look at the HealthMeasures website (http://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/available-translations/117-available-translations) 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).
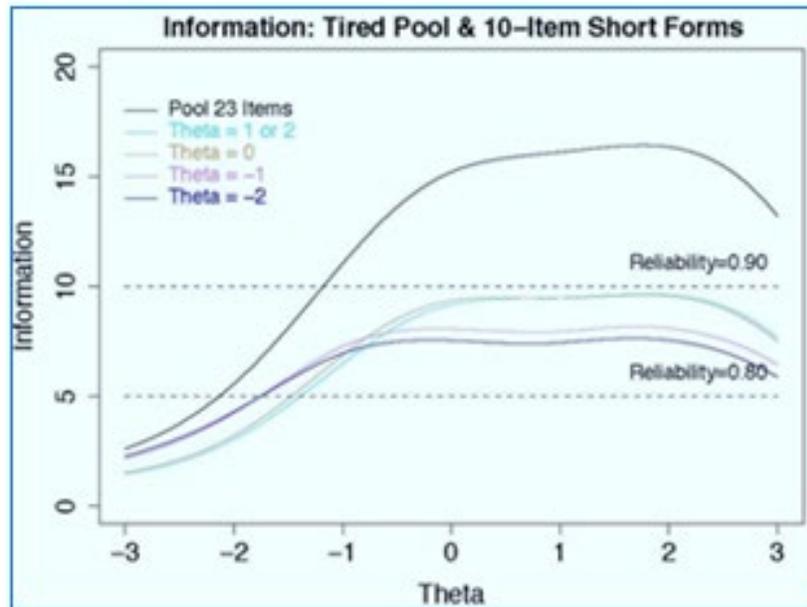
APPENDIX 1 – ADDITIONAL FIGURE

Figure 8 – Pediatric Test Information Tired