

Attachment 2: PROMIS Fatigue Scoring Manual, submitted to FDA with Initial Briefing Package on August 11, 2017

FATIGUE

A brief guide to the PROMIS® Fatigue instruments:

ADULT	ADULT CANCER	PEDIATRIC	PARENT PROXY
PROMIS Item Bank v1.0 – Fatigue 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 8a	PROMIS – Ca Bank v1.0 Fatigue	PROMIS Pediatric Bank v2.0 – Fatigue PROMIS Pediatric Bank v1.0 – Fatigue* PROMIS Pediatric Short Form v2.0 – Fatigue 10a PROMIS Pediatric Short Form v1.0 – Fatigue 10a*	PROMIS Parent Proxy Bank v2.0 – Fatigue PROMIS Parent Proxy Bank v1.0 – Fatigue* PROMIS Parent Proxy Short Form v2.0 – Fatigue 10a 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 computerized adaptive test (CAT). 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.

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 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 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.

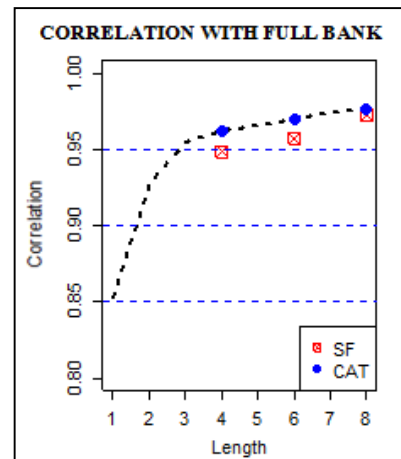


Figure 1

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.

DIFFERENCES IN CURRENTLY AVAILABLE SHORT FORMS

Adult Profile Short Forms

You will notice that there are 4 fatigue short forms for adults. Items in the 4a, 6a, and 8a short forms were selected based on rankings using two psychometric criteria: 1) maximum interval information; and 2) CAT simulations. Item rankings were similar for both criteria. For the maximum interval criterion, each item information function was integrated (without weighting) for the interval from the mean to 2 SDs worse than the mean. For the CAT simulations, responses to all items in each bank were generated using a random sample of 1,000 simulees drawn separately for each bank (centered on 0.5 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 4, 6, and 8 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.

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 short forms to give input on the relevance of each item. Each domain group worked independently and the original short forms are 6-10 items long depending on the domain. Psychometric properties and clinical input were both used and likely varied in importance across domains.



Pediatric and Parent Proxy Short Forms

There 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).

SELECTING THE ADULT CANCER INSTRUMENT

In selecting whether to use the adult cancer instrument (PROMIS-Ca) for this domain, it is important to consider the patient population being studied. All PROMIS-Ca instruments were developed for use with any cancer patient. This was done by having content experts review the adult PROMIS item bank to determine if there was a need to develop additional items or remove items because they conveyed a different meaning in cancer. Next, calibration testing with cancer patients with different diagnoses and treatments was conducted and data were analyzed to determine the final set of items and calibrations. The PROMIS-Ca Fatigue item bank contains a total of 54 items, all of which are also in the PROMIS Fatigue item bank. When scoring PROMIS-Ca items, we recommend using the default cancer calibrations. There are no short form instruments created specifically for an adult cancer population in Assessment CenterSM.

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?

The PROMIS parent proxy instruments have two calibration samples – parent proxy and parent proxy without Local Dependence. The former 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. For example, item Pf4fatigue12 and Pf4fatigue8 are enemy items. A participant should only see one of these items in a CAT.



SCORING THE INSTRUMENT

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

Each question usually has five response options ranging in value from one to five. To find the total raw score for a short form with all questions answered, sum the values of the response to each question. For example, for the adult 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 a valid score using the scoring tables.** If a participant has skipped a question, use the HealthMeasures Scoring Service (https://www.assessmentcenter.net/ac_scoring-service) to generate a final score.

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

For the adult PROMIS Fatigue 7a Short Form, a raw score of 10 converts to a T-score of 39.6 with a standard error (SE) of 4 (see scoring table for the 7a short form in Appendix 1). Thus, the 95% confidence interval around the observed score ranges from 47.4 to 31.8.7 (T-score \pm (1.96*SE) or 39.6 \pm (1.96*4)).

CAT: A minimum number of items (4 for adult and adult cancer CATs and 5 for peds and parent proxy CATs) must be answered in order to receive a score for the 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. The CAT will continue until either the standard error drops below a specified level (on the T-score metric 3.0 for adult and adult cancer CATs and 4.0 for peds and parent proxy CATs), or the participant has answered the maximum number of questions (12), whichever occurs first.

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

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.

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).

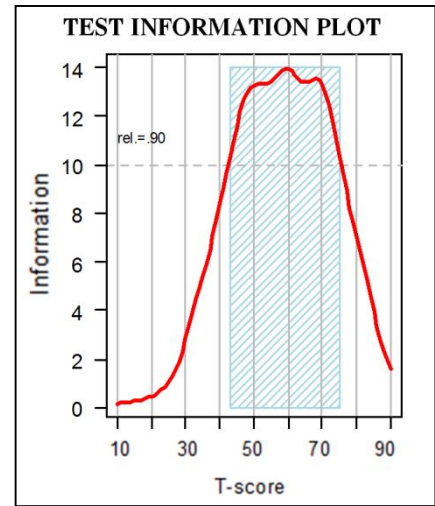


Figure 2

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 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 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.

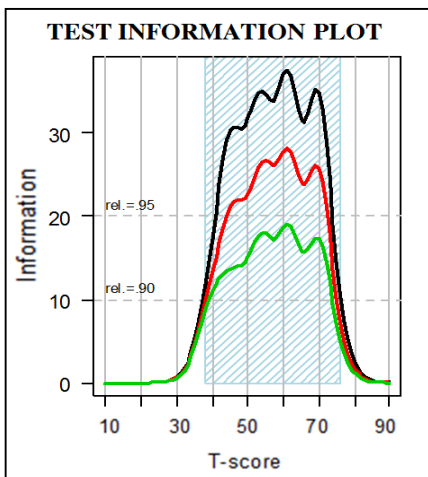


Figure 3

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 and parent proxy instruments in Appendix 1.

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

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

	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0	Min	Max
Scale Score										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

More information is available on www.HealthMeasures.net.

PREVIEW OF SAMPLE ITEM

Figure 5 shows an adult fatigue 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.

Figure 6 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.

Figure 5

In the past 7 days...		Never	Rarely	Sometimes	Often	Always
FATEXP20	How often did you feel tired?.....	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
FATEXP5	How often did you experience extreme exhaustion?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

Figure 6

DATA REPORTS

Upon completion of an adult CAT for physical function, depression, anxiety, pain interference, fatigue, sleep disturbance, and satisfaction with participation in social roles v1.0, a data report is available in Assessment Center. Figure 7 demonstrates some of the information available on the data reports.

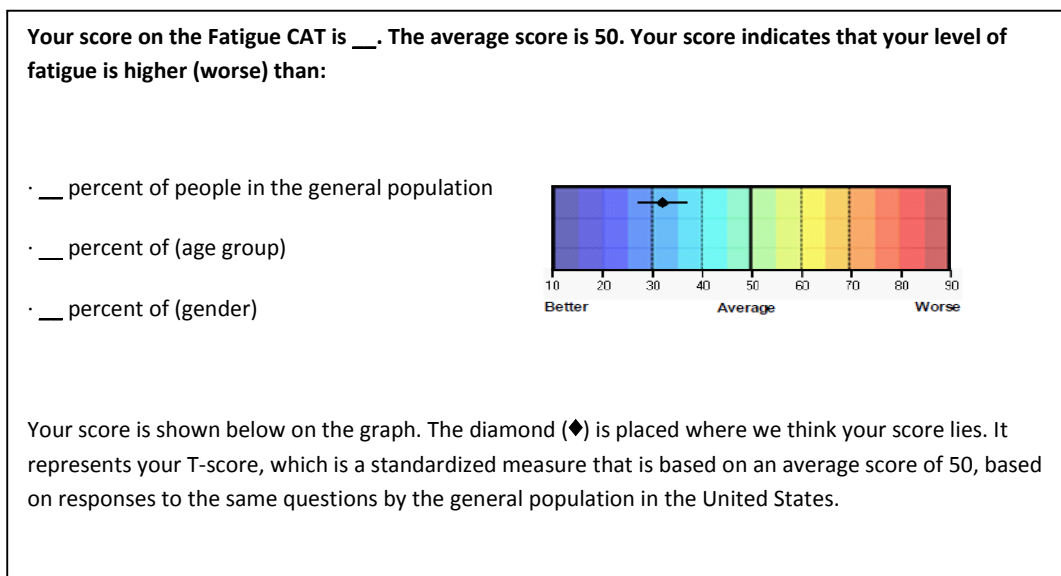


Figure 7



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

No.

Q: Are these instruments available in other languages?

Yes! Look at the HealthMeasures website (<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_scoring-service).

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

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

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

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

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



APPENDIX 1 - SCORING TABLES

It is recommended that you use the most recent version available which can be identified as the instrument with the highest version number

Fatigue 4a - Adult v1.0		
<i>Short Form Conversion Table</i>		
Raw Score	T-score	SE*
4	33.7	4.9
5	39.7	3.1
6	43.1	2.7
7	46.0	2.6
8	48.6	2.5
9	51.0	2.5
10	53.1	2.4
11	55.1	2.4
12	57.0	2.3
13	58.8	2.3
14	60.7	2.3
15	62.7	2.4
16	64.6	2.4
17	66.7	2.4
18	69.0	2.5
19	71.6	2.7
20	75.8	3.9

*SE = Standard Error on T-score

Fatigue 6a - Adult v1.0		
<i>Short Form Conversion Table</i>		
Raw Score	T-score	SE*
6	33.4	4.9
7	39.1	2.9
8	42.0	2.4
9	44.2	2.2
10	46.1	2.1
11	47.8	2.1
12	49.4	2.1
13	50.9	2.0
14	52.4	2.0
15	53.7	2.0
16	55.1	2.0
17	56.3	1.9
18	57.5	1.9
19	58.8	1.9
20	60.0	1.9
21	61.2	1.9
22	62.4	1.9
23	63.7	2.0
24	65.0	2.0
25	66.4	2.0
26	67.8	2.0
27	69.3	2.0
28	71.0	2.1
29	73.0	2.5
30	76.8	3.8

*SE = Standard Error on T-score

Fatigue 7a - Adult v1.0		
<i>Short Form Conversion Table</i>		
Raw Score	T-score	SE*
7	29.4	5.3
8	33.4	4.8
9	36.9	4.3
10	39.6	4.0
11	41.9	3.8
12	43.9	3.5
13	45.8	3.3
14	47.6	3.2
15	49.2	3.1
16	50.8	3.0
17	52.2	3.0
18	53.7	3.0
19	55.1	3.0
20	56.4	2.9
21	57.8	2.9
22	59.2	2.9
23	60.6	2.9
24	62.0	2.9
25	63.4	2.9
26	64.8	2.9
27	66.3	2.9
28	67.8	2.9
29	69.4	2.9
30	71.1	3.0
31	72.9	3.0
32	74.8	3.1
33	77.1	3.3
34	79.8	3.6
35	83.2	4.1

*SE = Standard Error on T-score



Fatigue 8a - Adult v1.0		
<i>Short Form Conversion Table</i>		
Raw Score	T-score	SE*
8	33.1	4.8
9	38.5	2.7
10	41.0	2.2
11	42.8	2.0
12	44.3	1.9
13	45.6	1.8
14	46.9	1.8
15	48.1	1.8
16	49.2	1.8
17	50.4	1.8
18	51.5	1.7
19	52.5	1.7
20	53.6	1.7
21	54.6	1.7
22	55.6	1.7
23	56.6	1.7
24	57.5	1.7
25	58.5	1.7
26	59.4	1.7
27	60.4	1.7
28	61.3	1.7
29	62.3	1.7
30	63.3	1.7
31	64.3	1.7
32	65.3	1.7
33	66.4	1.7
34	67.5	1.7
35	68.6	1.7
36	69.8	1.8
37	71.0	1.8
38	72.4	2.0
39	74.2	2.4
40	77.8	3.7

*SE = Standard Error on T-score !



Fatigue 10a - Pediatric v2.0		
<i>Short Form Conversion Table</i>		
Raw Score	T-Score	SE*
10	30.3	5.5
11	34.3	4.7
12	36.9	4.4
13	39.0	4.1
14	40.9	3.9
15	42.5	3.8
16	44.0	3.7
17	45.4	3.6
18	46.7	3.5
19	47.9	3.5
20	49.1	3.4
21	50.2	3.4
22	51.3	3.4
23	52.4	3.4
24	53.5	3.4
25	54.5	3.4
26	55.6	3.4
27	56.6	3.4
28	57.6	3.4
29	58.6	3.3
30	59.6	3.3
31	60.6	3.3
32	61.6	3.3
33	62.6	3.3
34	63.6	3.3
35	64.6	3.3
36	65.6	3.3
37	66.7	3.3
38	67.7	3.3
39	68.7	3.3
40	69.8	3.3
41	70.9	3.3
42	72.0	3.4
43	73.2	3.4
44	74.4	3.4
45	75.7	3.5
46	77.0	3.6
47	78.5	3.6
48	80.2	3.7
49	82.0	3.7
50	84.0	3.5

*SE = Standard Error on T-score



Fatigue 10a - Parent Proxy v2.0		
Short Form Conversion Table		
Raw Score	T-Score	SE*
10	34.0	5.0
11	39.0	4.0
12	42.0	3.0
13	44.0	3.0
14	45.0	3.0
15	47.0	3.0
16	48.0	2.0
17	49.0	2.0
18	50.0	2.0
19	51.0	2.0
20	52.0	2.0
21	53.0	2.0
22	54.0	2.0
23	55.0	2.0
24	56.0	2.0
25	57.0	2.0
26	58.0	2.0
27	59.0	2.0
28	60.0	2.0
29	61.0	2.0
30	62.0	2.0
31	63.0	2.0
32	64.0	2.0
33	65.0	2.0
34	66.0	2.0
35	67.0	2.0
36	68.0	2.0
37	69.0	2.0
38	70.0	2.0
39	71.0	2.0
40	72.0	2.0
41	72.0	2.0
42	73.0	2.0
43	74.0	2.0
44	75.0	2.0
45	76.0	2.0
46	77.0	2.0
47	79.0	3.0
48	80.0	3.0
49	82.0	3.0
50	85.0	4.0

*SE = Standard Error on T-score !

All scoring tables are based on default Parent Proxy calibrations. !



**APPENDIX 2 - SCORING
 TABLES FOR RETIRED
 MEASURES**

Fatigue 10a - Pediatric v1.0		
<i>Short Form Conversion Table</i>		
Raw Score	T-Score	SE*
0	30.3	5.5
1	34.3	4.7
2	36.9	4.4
3	39.0	4.1
4	40.9	3.9
5	42.5	3.8
6	44.0	3.7
7	45.4	3.6
8	46.7	3.5
9	47.9	3.5
10	49.1	3.4
11	50.2	3.4
12	51.3	3.4
13	52.4	3.4
14	53.5	3.4
15	54.5	3.4
16	55.6	3.4
17	56.6	3.4
18	57.6	3.4
19	58.6	3.3
20	59.6	3.3
21	60.6	3.3
22	61.6	3.3
23	62.6	3.3
24	63.6	3.3
25	64.6	3.3
26	65.6	3.3
27	66.7	3.3
28	67.7	3.3
29	68.7	3.3
30	69.8	3.3
31	70.9	3.3
32	72.0	3.4
33	73.2	3.4
34	74.4	3.4
35	75.7	3.5
36	77.0	3.6
37	78.5	3.6
38	80.2	3.7
39	82.0	3.7
40	84.0	3.5

*SE = Standard Error on T-score

Fatigue 10a - Parent Proxy v1.0		
<i>Short Form Conversion Table</i>		
Raw Score	T-Score	SE*
0	34.0	5.0
1	39.0	4.0
2	42.0	3.0
3	44.0	3.0
4	45.0	3.0
5	47.0	3.0
6	48.0	2.0
7	49.0	2.0
8	50.0	2.0
9	51.0	2.0
10	52.0	2.0
11	53.0	2.0
12	54.0	2.0
13	55.0	2.0
14	56.0	2.0
15	57.0	2.0
16	58.0	2.0
17	59.0	2.0
18	60.0	2.0
19	61.0	2.0
20	62.0	2.0
21	63.0	2.0
22	64.0	2.0
23	65.0	2.0
24	66.0	2.0
25	67.0	2.0
26	68.0	2.0
27	69.0	2.0
28	70.0	2.0
29	71.0	2.0
30	72.0	2.0
31	72.0	2.0
32	73.0	2.0
33	74.0	2.0
34	75.0	2.0
35	76.0	2.0
36	77.0	2.0
37	79.0	3.0
38	80.0	3.0
39	82.0	3.0
40	85.0	4.0

*SE = Standard Error on T-score

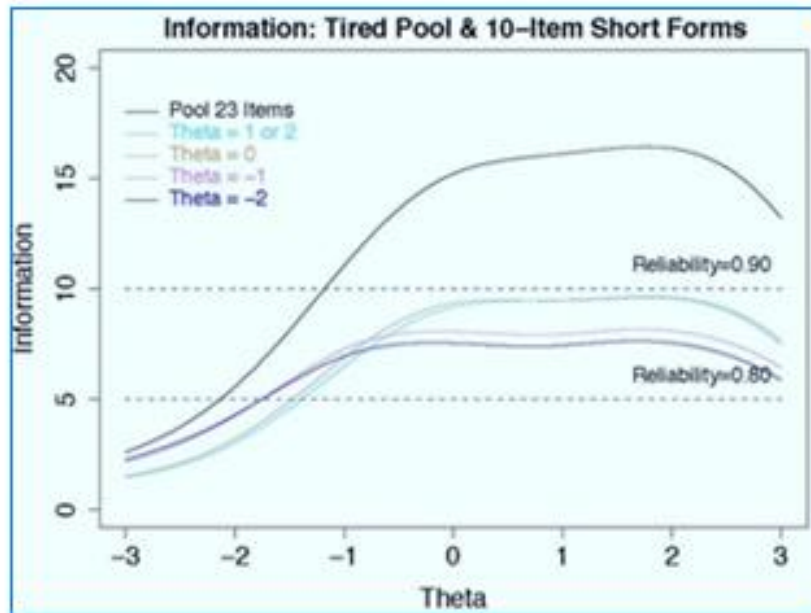


Figure 8 – Pediatric Test Information Tired