Published in final edited form as:
J Pain Symptom Manage. 2014 October ; 48(4): 639–648. doi:10.1016/j.jpainsymman.2013.12.236.

Linking Fatigue Measures on a Common Reporting Metric

Jin-Shei Lai, PhD, **David Cella, PhD**, **Betina Yanez, PhD**, and **Arthur Stone, PhD**
Departments of Medical Social Sciences (J.-S.L., D.C., B.Y.) and Pediatrics (J.-S.L.), Feinberg School of Medicine, Northwestern University, Chicago, Illinois; and Department of Psychiatry and Behavioral Science, Stony Brook University, Stony Brook, New York, USA

Abstract

Context—Fatigue is one of the most common and debilitating symptoms experienced by patients living with chronic conditions and is also commonly experienced in the general U.S. population. Linking fatigue scores from some of the most widely used measure of fatigue to the same metric will facilitate interpretation of fatigue outcomes.

Objectives—The goal of this study is to report the methods used to develop linking (crosswalk) tables to enable the direct comparison of PROMIS fatigue with fatigue scores on the FACIT-Fatigue Scale, the Medical Outcomes Study Short Form-36 4-item Vitality Scale, and the Quality of Life in Neurological Disorders Fatigue Scale.

Methods—Participants came from two data sets (n= 1,120 and n= 803). Two item response theory based linking methods, the Stocking-Lord calibration and fixed parameter calibration, were used to establish linking between measures. IRT calibrations were derived using the graded response model.

Results—Both the Stocking-Lord calibration and fixed parameter calibration linking methods produced comparable results. Final crosswalk tables are reported for the fixed parameter calibration.

Conclusion—Findings can facilitate comparison of scores across some of the most widely used fatigue measures and assist in comparing patient-reported fatigue outcomes in clinical trials, comparative effectiveness research, and clinical practice.

Keywords

Fatigue; item response theory; patient-reported outcomes; comparative effectiveness measurement

Corresponding Author: Jin-Shei Lai, PhD, Department of Medical Social Sciences and Pediatrics, Feinberg School of Medicine, Northwestern University, 633 St. Clair, 19th Floor, Chicago, IL 60611, TEL: 312-503-3370, FAX: 312-503-9800, js-lai@northwestern.edu.

Introduction

Fatigue is one of the most common and debilitating symptoms experienced by patients living with chronic conditions and is also commonly experienced in the general U.S. population. Defined as a subjective sensation of overwhelming and debilitating weakness, lack of energy, or tiredness, fatigue affects millions of people with various chronic conditions such as systemic lupus erythematosus, (1) rheumatoid arthritis,(2) cancer, (3) and Parkinson's disease.(4) Given its high prevalence, and the impact it has on quality of life, it is important to assess fatigue using standard, reproducible instrumentation in order to inform clinical practice and future research.

Because patient-reported outcomes such as fatigue can supplement more traditional clinical endpoints, measurement tools for fatigue such as the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue),(5) European Organisation for Research and Treatment of Cancer Fatigue Subscale,(6) Quality of Life in Neurological Disorders (Neuro-QOL),(7) the Medical Outcomes Study Short Form (SF-36) Vitality Scale,(8) Cancer-related Fatigue Item Bank,(9) the Multidimensional Fatigue Inventory,(10) the Piper Fatigue Scale, (11) and the Fatigue Symptom Inventory,(12) have become increasingly popular in research and practice. However, a growing problem with this vast selection of fatigue instruments is proliferation without a common metric to report fatigue severity, which results in difficulty in comparing results between studies. Each instrument has its own unique scoring routine, disconnected from each other, making comparison of scores quite difficult, if not impossible. Yet, the questions asked by these measures bear striking resemblance to one another.

To address the growing concern of score comparisons, the National Institutes of Health (NIH) initiated the Patient Reported Outcomes Measurement Information System (PROMIS®)(13) to develop a fatigue item bank measuring the experience of fatigue and its impact upon daily functioning (14). This bank of 95 items included items from the FACIT-Fatigue and SF-36 Vitality Scale in its development and was tested on the US general population, including a range of people with chronic health conditions. Applications from a well-established item bank include computerized adaptive tests (CAT) or short-forms measures that can provide brief, yet very precise, measurements that are comparable to those obtained from administering the full-length item banks (15, 16). The similarity of the PROMISE fatigue item bank to established measures and its ability to be used in CATs makes it a desirable new measure.

Despite the development of the PROMIS Fatigue Item Bank, some researchers and clinicians will continue to use previously validated and widely used instruments. Relevant to the purpose of the current work, the PROMIS Fatigue Item Bank provides a standardized fatigue T score metric against which other highly similar fatigue instruments can be compared and linked. A common (shared) fatigue metric can be achieved by aligning various fatigue assessment tools on the same metric, for example, the metric defined by the PROMIS.(17-19) Once instruments are linked on a common metric, a simple linking (“crosswalk”) table can then associate scores from one measure to corresponding scores on another, facilitating direct comparisons on a common scale. This will permit more refined

use of fatigue scores across any fatigue measure as long as they are linked to the PROMIS metric.(20, 21) This common metric enables interpretation of fatigue outcomes across multiple research contexts such as comparative effectiveness research, which seeks to inform healthcare decisions by comparing different strategies and interventions to prevent, treat, and monitor health conditions.

In this paper, we demonstrated the development of such linking (crosswalk) tables using three fatigue measurement tools as examples: the FACIT-Fatigue Scale(5), the Medical Outcomes Study Short Form-36 4-item Vitality Scale (8), and the Quality of Life in Neurological Disorders (Neuro-QOL) Fatigue Scale.(7) PROMIS Fatigue Item Bank was served as the reference to allows for a common metric among these three sets of analyses.

Methods

Measures

The Patient-Reported Outcomes Measurement Information System Fatigue Item Bank (PROMIS FIB) consists of 95 items (13 items of which are from the FACIT-Fatigue) with a 7-day time frame and 5-point scales (either frequency or intensity responses depending on item content).(14) The item bank was developed using comprehensive mixed methods (i.e., qualitative and quantitative).(13) (22, 23) consisted of items measuring fatigue experience and impact of fatigue upon daily living. Higher scores represent higher levels of fatigue.

The Medical Outcomes Study Vitality Scale is a subset of the 36-item SF-36,(8) which consisted of four items using a 5-point frequency rating scale ranging from “all of the time” to “none of the time”. Higher scores indicate more vitality. The Vitality scale of the SF-36 has demonstrated validity across a range of studies(8).

The FACIT-Fatigue is a 13-item questionnaire that assesses self-reported tiredness, weakness, and difficulty conducting usual activities due to fatigue.(5) A 5-point intensity rating scale (from “not at all” to “very much”) is used. Higher scores reflect less fatigue. Respondents are instructed to answer questions with respect to their experiences and functioning over the previous 7 days. The FACIT-Fatigue is a psychometrically sound instrument and has been widely used to measure fatigue among people with various chronic illnesses(24) and the US general population.(25) Of note, the 13-items of the FACIT-Fatigue are included in the 95-item PROMIS FIB. As a result, the FACIT-F and PROMIS FIB have already been calibrated on the same scale. Therefore, building a linking (crosswalk) table for these two instruments only required creation of a short-form scoring table.(26, 27)

The Quality of Life in Neurological Disorders Fatigue Scale (Neuro-QOL Fatigue) (7) consists of 19 items with a 7-day time frame and a 5-point intensity scale (from “not at all” to “very much”). Similar to the PROMIS FIB, Neuro-QOL Fatigue was developed using mixed methods, was validated on people with neurological conditions.(7, 28) Higher scores represent higher levels of fatigue.

The four instruments used in this study measured the same underlying construct; fatigue experienced by individuals and its effect on daily living upon fatigue. Moreover, the

wording across measures was similar. For example, “How often were you too tired to do your household chores?” in PROMIS versus “I was too tired to do my household chores “ in Neuro-QOL, and “Did you feel tired?” in SF-36/Vitality versus “How often did you feel tired?” in PROMIS. These similarities allowed for linking exercises among these instruments. PROMIS, FACIT-F and Neuro-QOL Fatigue used a 7-day time frame while the SF-36/Vitality used a 4-week time frame. Results of previous research (29) did not indicate a significant difference between self-reported fatigue using 4-week and 7-day time frames. In light of this data we considered it acceptable to link PROMIS to SF-36/Vitality.

Sample

Sample 1—The SF-36 Vitality Scale and FACIT Fatigue to PROMIS Fatigue linking data were collected from the PROMIS wave 1 full-length testing as reported by Lai et al (see Table 1 for demographic information). The average age of the 803 participants was 51.8 (SD=17.8; range: 18-89); 55% of the sample was female and 45% was male; 11% was of Hispanic origin; 81% was white, followed by 10% African American and 9% multiple races. In terms of education, 20% high school or lower, 44% some college, 18% college, and 18% advanced degree. Thirty-four percent of participants reported having a diagnosis of hypertension, 22% arthritis or rheumatism, 20% depression, 15% anxiety, and 14% asthma, 14% OA or degenerative arthritis, and 14% migraines or severe headache.

Sample 2—The Neuro-QOL Fatigue to PROMIS Fatigue linking data were collected by an Internet survey company, Op4G (www.op4g.com), that maintains a panel of respondents from the US general population. To ensure adequate demographic diversity and accurate representation of the U.S. population based on the 2010 census, we imposed minimum requirements for age, gender, race, ethnicity, and education (Table 1). In addition to providing socio-demographic and clinical information, and responding to questions on other health domains, participants responded to items on 14 items from the PROMIS fatigue item bank and the 19-item Neuro-QOL Fatigue. In order not to increase the response burden, unlike with sample 1 where all 95 PROMIS fatigue items were administered, 14 items were administered in this testing, which were selected based on the content (no overlapping with the Neuro-QOL) and psychometric properties (e.g., better information function and coverage of the fatigue continuum). Since all 95 PROMIS fatigue items were calibrated onto the same measurement continuum using IRT, using all or portion of the item would produce similar linking results.(30)

Analysis

Confirmatory factor analysis was used to assess the unidimensionality of the combined scales (i.e., PROMIS and Neuro-QOL Fatigue; PROMIS and Vitality) before linking.(31) The combined scales were considered unidimensional if the comparative fit index (CFI) was >0.9 and the root mean square estimate of approximation (RMSEA) was <0.1.(32) MPlus 6.0(31) was used for factor analysis. Multiple linking strategies are available, including both traditional procedures (e.g., equipercentile) and IRT (e.g., fixed-parameters and Stocking-Lord linking). Since the PROMIS Fatigue bank was developed using IRT, we reported IRT-based linking methods in this manuscript. Two IRT based linking methods were used in this study: 1) the Stocking-Lord separate calibration method that produces additive and

multiplicative constants to transform item parameters; and 2) fixed parameter calibration that places non-PROMIS items on the same metric as PROMIS items. The graded response model (GRM),(33) implemented in MULTILOG(34) computer software, was used to derive IRT calibrations. For each item the GRM estimates a slope or discrimination parameter (*a*), which indicates the degree of association between the item responses and the underlying construct and four threshold parameters (b_k) (for five category items) that reflect the degree of fatigue where the most probable response occurs in a given category or higher.

The Stocking-Lord method, as implemented in Plink(35) (a package for R), was used to link IRT estimated parameters from different scales using several steps.(17, 19, 36) First, it freely calibrated the PROMIS and target measures (e.g., Neuro-QOL Fatigue, SF-36 Vitality Scale) concurrently (without fixing the PROMIS parameters). Secondly, the older, established PROMIS parameters were used as an “anchor” to estimate multiplicative and additive constants needed to transform the newly calibrated PROMIS parameters on to the metric of the established PROMIS parameters. These constants were then used to linearly transform all the non-PROMIS items parameters to the PROMIS metric.

The second approach, fixed parameter calibration, fixed the PROMIS Fatigue item parameters and calibrated only SF-36 Vitality Scale and Neuro-QOL Fatigue items using GRM model as described earlier. As a result, the non-PROMIS item parameters could be placed on the same metric as the PROMIS items enabling the creation of the crosswalk tables.

Finally, we created crosswalk tables to convert the SF-36 Vitality Scale and Neuro-QOL raw scores to the PROMIS FIB using the PROMIS scoring system as described in Lai et al (2011) and available in Assessment Center (http://www.assessmentcenter.net/). PROMIS's T score distributions are standardized such that a score of 50 represents the average (mean) for the US general population, and the standard deviation around that mean is 10 points. A high PROMIS score represents more fatigue.

Results

Linking SF-36 Vitality Scale to PROMIS Fatigue

Factor analysis confirmed the assumption of unidimensionality (CFI=0.968, RMSEA=0.062). Item-total correlations ranged from 0.510 to 0.883 and Cronbach's Alpha=0.995. The correlation between the SF-36 Vitality Scale and the PROMIS FIB was 0.89. The correlations between the combined score (i.e., Vitality Scale + PROMIS FIB) and the measures were 1.0 and 0.90 for PROMIS FIB and SF-36 Vitality Scale, respectively.

In the Stocking-Lord method, two constants were obtained (an additive constant of 0.996 and a multiplicative constant of 0.591) and were used to transform the SF-36 Vitality items onto the PROMIS FIB metric. The transformed slope parameter estimates for the SF-36 Vitality Scale ranged from 2.32 to 3.31 with a mean of 2.91. When using the fixed parameter calibration method, the slope parameter estimates ranged from 2.36 to 3.35 with a mean of 2.95. The correlations of the parameters (slope and threshold parameters) from these two methods are shown on Table 2, which ranged from 0.94 to 0.99. We then

compared IRT scaled scores of participants obtained from both methods. The person scaled scores from these two methods were almost identical (r=1, p<0.001). The T score discrepancies (Stocking-Lord minus fixed-parameter) ranged from -0.30 to 1.10 with a mean of 0.06 (SD=0.01) and only one participant had a discrepancy greater than 1 T score unit (i.e., 0.1 SD). Figure 1a depicts the T score discrepancies (Y-axis) against the scores from fixed-parameter calibration method (X-axis), in which most large discrepancies occurred on participants with negligible fatigue (i.e., lower T scores). Finally, we produced the conversion table using the fixed-parameter method (Table 3).

Linking Neuro-QOL to PROMIS FIB

Factor analysis confirmed the assumption of unidimensionality (CFI=0.973 and RMSEA=0.105). Item-total correlations ranged from 0.74 to 0.888 and Alpha=0.987. The correlation between Neuro-QOL and PROMIS FIB was 0.88. The correlations between the combined score and the measures were 0.98 and 0.99 for PROMIS FIB and Neuro-QOL, respectively. Given the high correlations between PROMIS FIB and Neuro-QOL and high Cronbach's alpha, we concluded the combined 33 items (i.e., 19 Neuro-QOL and 14 PROMIS fatigue items) were sufficiently unidimensional to proceed despite the elevated RMSEA.

Using the Stocking-Lord methods, two constants (an additive constant of 1.193 and a multiplicative constant of 0.395) were obtained and used to transform Neuro-QOL onto the PROMIS FIB metric. The transformed slope parameter estimates for the Neuro-QOL Fatigue ranged from 2.00 to 3.76 with a mean of 3.07. In the fixed-parameter method, the slope parameter estimates for Neuro-QOL ranged from 2.04 to 3.84 with a mean of 3.12. The correlations of the parameters from these two methods are shown on Table 2, which ranged from 0.99 to 1.00. We then compared IRT scaled scores from both methods. The person scaled scores from these two methods were almost identical (r=1, p<0.001). T score discrepancies (Stocking-Lord minus fixed-parameter) ranged from -0.87 to 1.24 with a mean of 0.01 (SD=0.30) and only one participant had a discrepancy greater than 1 T score unit (i.e., 0.1 SD). Figure 1b depicts the T score discrepancies (Y-axis) against the scores from fixed-parameter calibration method (X-axis). Finally, we produced the conversion table using the fixed-parameter method (Table 4).

Linking FACIT-F to PROMIS FIB

Given that the FACIT-Fatigue items have been included in the PROMIS FIB, the FACIT-Fatigue was regarded as a short-form of the PROMIS FIB. Therefore, a cross-walk table (Table 5) was created using existing calibrations as described in Lai et al(14) and scoring without additional analysis.

Discussion

Fatigue is a common complaint among people with or without chronic conditions. The current study used IRT to provide information on the methods for linking the FACIT-Fatigue, SF-36 Vitality Scale, and Neuro-QOL Fatigue to the PROMIS FIB. Results from both fix-parameter and Stocking-Lord methods were similar. We decided to produce

conversion tables using the fixed-parameters method with PROMIS FIB as the reference group. This approach allows for comparative effectiveness research as scores from these four scales were all on the same metric (i.e., PROMIS T score metric; mean = 50 and SD=10).

The increasing integration of patient-reported outcomes in clinical trials and practice has fueled the need to establish a common metric among patient-reported outcomes. It is difficult to discern whether differences in patient-reported fatigue outcomes across clinical trials are due to scale content, differences in psychometric properties, or differences in the actual treatment effect.(37) Item response theory based linking methods provide a basis for comparing effect sizes across measures in studies. (20, 21) The crosswalk tables provided in this manuscript have immense practical value. Researchers and clinicians can easily convert raw scores from the SF-36 Vitality Scale, the FACIT-Fatigue Scale, or the Neuro-QOL Fatigue Scale into PROMIS FIB T score metric to facilitate comparative effectiveness research, including meta analytic studies, and provide a common language when reporting fatigue in performance improvement or other practice-based research or evaluation.

Our linking results between PROMIS FIB and FACIT-F were similar to those reported in our previous study, in which we linked the PROMIS FIB with the FACIT-F scale using the sample mean (IRT scaled score =0.12) and the standard deviation (IRT scaled score=0.94).(26) These methods contrast to the current project in which we used the population mean (IRT scaled score =0) and standard deviation (IRT scaled score =1) to convert IRT scaled scores into T score metric. Despite the different approaches, the results from the current linking result produced estimates of PROMIS T Scores that differed by less than 2.5 points (i.e., 0.25 SD).(26) The larger discrepancies appeared to the most and least severe ends (especially the least fatigue end), where larger estimation errors were also found. Given that differences less than 0.25 SD are considered to be small, our findings indicate that the two linking approaches produced similar results.

The findings in this study have several limitations. Because estimated T score errors were higher at two ends (most severe and least fatigue), thus the conversions between T scores and raw scores are not as precise as T scores in the middle range. For example, the range of a converted PROMIS T score for the extremely low Vitality score of 4 could be 9.6 points (i.e., 0.96 SD) higher or lower. Therefore, while the current tables can be used to generate PROMIS scores from the other fatigue measure, the precision of these converted scores, especially those at extreme ends, should be used with caution. Future research should also be conducted to test the validity and applicability of these results in clinical samples, especially for those with most severe fatigue, and to expand IRT linking to other measures of fatigue. Further similar research with other fatigue instruments, such as the FSI, MFSI, PFS and others, can potentially bring these instruments into the same common reporting metric.

This paper has several notable strengths. To begin with, this is the first study to use IRT methods to link the PROMIS FIB to three of the most widely used measures of fatigue. Second, we used two IRT-based linking methods to minimize differences between observed and linked scores. Third, fixed parameter calibrations were not determined by the current

sample, but were anchored on the PROMIS calibrations that were derived from the larger standardization sample(38) and centered on the 2000 US Census.(23)

In summary, the increasing use of patient-reported outcomes to supplement more traditional clinical endpoints in clinical research and practice has necessitated the ability to compare scores across popular patient-reported measures of fatigue. This study provides information on two IRT-based methods used to link the, PROMIS FIB to the SF-36 Vitality Scale, Neuro-QOL Fatigue, and FACIT-Fatigue. Three crosswalk tables were generated to enable the direct comparison of some of the most widely used fatigue measures for groups and will facilitate the comparison of patient-reported outcomes in clinical trials, comparative effectiveness research, and clinical practice.

Acknowledgments

Disclosures: This manuscript was supported by an NIH/National Cancer Institute grant PROSETTA STONE (1RC4CA157236-01, PI: David Cella).

References

1. Iaboni A, Ibanez D, Gladman DD, Urowitz MB, Moldofsky H. Fatigue in systemic lupus erythematosus: contributions of disordered sleep, sleepiness, and depression. J Rheumatol. 2006; 33(12):2453–7. [PubMed: 17143980]
2. Repping-Wuts H, Fransen J, Achterberg Tv, Bleijenberg G, Riel Pv. Persistent severe fatigue in patients with rheumatoid arthritis. J Clin Nurs. 2007; 16(11c):377–83. [PubMed: 17931330]
3. Mendoza TR, Wang XS, Lu C, Palos GR, Liao Z, Mobley GM, et al. Measuring the Symptom Burden of Lung Cancer: The Validity and Utility of the Lung Cancer Module of the M. D. Anderson Symptom Inventory. The oncologist. 2011 Feb 1; 16(2):217–27. 2011. [PubMed: 21285393]
4. Morris NS, MacLean CD, Littenberg B. Literacy and health outcomes: a cross-sectional study in 1002 adults with diabetes. BMC family practice. 2006; 7:49. English. [PubMed: 16907968]
5. Yellen SB, Cella DF, Webster K, Blendowski C, Kaplan E. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. J Pain Symptom Manage. 1997; 13(2):63–74. [PubMed: 9095563]
6. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993; 85(5):365–76. [PubMed: 8433390]
7. Gershon R, Lai J, Bode R, Choi S, Moy C, Bleck T, et al. Neuro-QOL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. Qual Life Res. 2012; 21(3):475–86. Epub August 27, 2011. [PubMed: 21874314]
8. Ware JE Jr, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36). I. Conceptual Framework and Item Selection. Med Care. 1992; 30(6):473–83. [PubMed: 1593914]
9. Lai JS, Cella D, Dineen K, Von Roenn J, Gershon R. An item bank was created to improve the measurement of cancer-related fatigue. J Clin Epidemiol. 2005; 58(2):190–7. [PubMed: 15680754]
10. Smets EMA, Garssen B, Bonke B, DeHaes JCJM. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res. 1995; 39:315–25. [PubMed: 7636775]
11. Badger TA, Segrin C, Figueredo AJ, Harrington J, Sheppard K, Passalacqua S, et al. Who benefits from a psychosocial counselling versus educational intervention to improve psychological quality of life in prostate cancer survivors? Psychol Health. 2013; 28(3):336–54. Epub 2012/10/11. eng. [PubMed: 23045995]

12. Hann DM, Jacobsen PB, Azzarello LM, Martin SC, Curran SL, Fields KK, et al. Measurement of fatigue in cancer patients: Development and validation of the Fatigue Symptom Inventory. Qual Life Res. 1998; 05(4):7. 301–10.
13. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol. 2010; 63(11):1179–94. Eng. [PubMed: 20685078]
14. Lai JS, Cella D, Choi SW, Junghaenel DU, Christodolou C, Gershon R, et al. How Item Banks and Their Application Can Influence Measurement Practice in Rehabilitation Medicine: A PROMIS Fatigue Item Bank Example. Arch Phys Med Rehabil. 2011; 92(10 Supplement):S20–S7. [PubMed: 21958919]
15. Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. Qual Life Res. 2007; 16(Suppl 1):133–41. [PubMed: 17401637]
16. Bjorner, JB.; Kosinski, M.; Ware, JE. Computerized adaptive testing and item banking. In: Fayers, PM.; Hays, RD., editors. Assessing quality of life in clinical trials: methods and practice. Oxford; New York: Oxford University Press; 2005. p. 95-112.
17. Chen WH, Revicki DA, Lai JS, Cook KF, Amtmann D. Linking pain items from two studies onto a common scale using item response theory. J Pain Symptom Manage. 2009; 38(4):615–28. [PubMed: 19577422]
18. McHorney CA, Cohen AS. Equating health status measures with Item Response Theory: Illustrations with functional status items. Med Care. 2000; 38(9 Suppl):1143–59.
19. Kolen, MJ.; Brennan, RL. Test equating, scaling, and linking : methods and practices. New York: Springer; 2004.
20. Cavanaugh K, Huizinga MM, Wallston KA, Gebretsadik T, Shintani A, Davis D, et al. Association of Numeracy and Diabetes Control. Ann Intern Med. 2008; 148(10):737–46. [PubMed: 18490687]
21. Reeve BB, Chang CH, Perfetto E. Applying item response theory to enhance health outcomes assessment. Qual Life Res. 2007 Aug 25; 16(0):1–3. [PubMed: 17033892]
22. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric Evaluation and Calibration of Health-Related Quality of Life Item Banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care. 2007; 45(5 Suppl 1):S22–S31. [PubMed: 17443115]
23. Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, et al. Representativeness of the PROMIS Internet Panel. J Clin Epidemiol. 2010; 63(11):1169–78. [PubMed: 20688473]
24. Cella D, Yount S, Sorensen M, Chartash E, Sengupta N, Grober J. Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. J Rheumatol. 2005; 32:811–9. [PubMed: 15868614]
25. Cella D, Lai JS, Chang CH, Peterman A, Slavin M. Fatigue in cancer patients compared with fatigue in the general United States population. Cancer. 2002; 94(2):528–38. [PubMed: 11900238]
26. Smith E, Lai JS, Cella D. Building a measure of fatigue: the functional assessment of chronic illness therapy fatigue scale. PM R. 2010; 2(5):359–63. [PubMed: 20656617]
27. Lai JS, Stucky B, Thissen D, Varni J, DeWitt E, Irwin D, et al. Development and psychometric properties of the PROMIS® pediatric fatigue item banks. Qual Life Res. 2013 (Epub ahead of print):1-11. Epub 2013/02/01. English.
28. Cella D, Lai JS, Nowinski C, Victorson D, Peterman A, Miller D, et al. Neuro-QOL: Brief Measures of Health-related Quality of Life for Clinical Research in Neurology. Neurology. 2012; 78:1860–7. [PubMed: 22573626]
29. Lai JS, Cook K, Stone A, Beaumont JL, Cella D. Classical test theory and item response theory/Rasch model to assess differences between patient-reported fatigue using seven-day and four-week recall periods. J Clin Epidemiol. 2009; 62(9):991–7. [PubMed: 19216054]
30. Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. Qual Life Res. 2007; 16(Suppl 1):133–41. [PubMed: 17401637]
31. Muthen, LK.; Muthen, BO. Mplus User's Guide. Los Angeles, CA: Muthen & Muthen; 2006.

32. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychol Methods. 1996; 1(2):130–49.
33. Samejima, F. The graded response model. In: van der Linden, WJ.; Hambleton, R., editors. Handbook of modern item response theory. New York: Springer-Verlag; 1997. p. 85-100.
34. Thissen, D. MULTILOG Windows 7.0 ed. Lincolnwood, IL: Scientific Software International, Inc.; 2003.
35. Weeks JP. Plink: An R package for linking mixed-format tests using IRT-based methods. J Stat Softw. 2010; 35(12):1–33. English. [PubMed: 21603108]
36. Kim SH, Cohen AS. A Comparison of Linking and Concurrent Calibration Under the Graded Response Model. Appl Psychol Meas. 2002; 26(1):25–41. English.
37. Vadiraja HS, Rao MR, Nagarathna R, Nagendra HR, Rekha M, Vanitha N, et al. Effects of yoga program on quality of life and affect in early breast cancer patients undergoing adjuvant radiotherapy: a randomized controlled trial. Complement Ther Med. 2009 Oct-Dec;17(5-6):274–80. Epub 2009/11/28. eng. [PubMed: 19942107]
38. Schnur JB, David D, Kangas M, Green S, Bovbjerg DH, Montgomery GH. A randomized trial of a cognitive-behavioral therapy and hypnosis intervention on positive and negative affect during breast cancer radiotherapy. J Clin Psychol. 2009 Apr; 65(4):443–55. Epub 2009/02/20. eng. [PubMed: 19226611]

Figure 1a

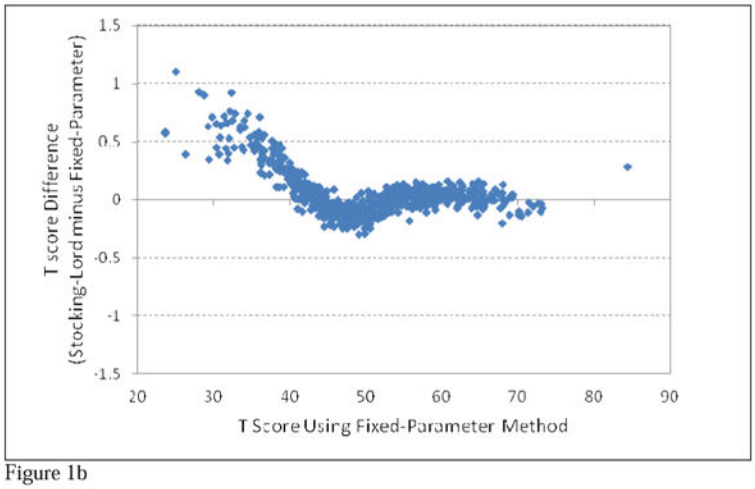

Figure 1b

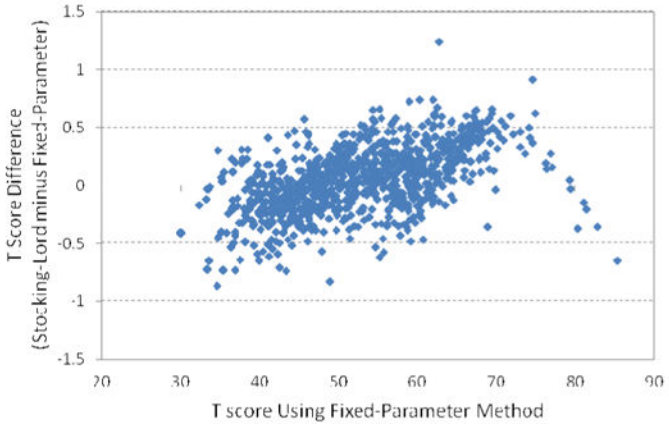

Figure 1.

Comparisons of T scores from Fixed-Parameter and Stocking-Lord Linking Methods. All but one subject in the SF-36 Vitality Scale and one in Neuro-QOL showed discrepancies between these two methods were within 1 T score range (0.1 standard deviation).

1a. Linking between PROMIS FIB and Vitality

1b. Linking between PROMIS FIB and NeuroQOL Fatigue

Table 1
Demographic characteristics of the samples

	Samples	
Characteristic	**Sample 1 (N=803)**	**Sample 2 (N= 1,120)**
Gender		
Male	43.9	47.4
Ethnicity		
Hispanic	15.2	14.7
Race		
White	80.1	72.0
Black / African American	9.1	11.3
Asian	2.8	5.2
Multiracial	NA	2.4
Other	10.1	9.2
Education		
Less than high school, high school diploma, GED, or some college/vocational training	80	42
College degree or advanced degree	20	58
Mean age (SD)	54.1 (15.2)	46.4 (17.5)

Table 2

Correlations of slope (a) and threshold (cb1, cb2, cb3 cb4) parameter estimations from fixed parameter and Stocking-Lord linking methods.

	a_fixed	***cb1_fixed***	***cb2_fixed***	***cb3_fixed***	***cb4_fixed***
Vitality					
a_Stocking-Lord	0.94				
cb1_ Stocking-Lord		0.99			
cb2_ Stocking-Lord			0.99		
cb3_ Stocking-Lord				0.98	
cb4_ Stocking-Lord					0.97
Neuro-QOL Fatigue					
a_ Stocking-Lord	0.99				
cb1_ Stocking-Lord		1.00			
cb2_ Stocking-Lord			1.00		
cb3_ Stocking-Lord				1.00	
cb4_ Stocking-Lord					1.00

[1] fixed: parameter estimation from “fixed parameter estimation” where PROMIS parameters were fixed.

[2] Stocking-Lord: parameter estimations from “Stock-Lord” transformation linking method.

[3] a: slope parameter; cb1-cb4: threshold parameters of the 5-point rating scale where 4 thresholds are available.

J Pain Symptom Manage. Author manuscript; available in PMC 2015 October 01.

Table 3

Raw SF-36 Vitality Scale Score to T Score Conversion Table (IRT Fixed Parameter Calibration Linking)

Vitality Expected Total Raw Score	PROMIS T score	T score SE
4	29.0	4.9
5	34.2	4.0
6	38.2	3.8
7	41.6	3.6
8	44.8	3.5
9	47.5	3.4
10	49.9	3.4
11	52.1	3.3
12	54.1	3.3
13	56.0	3.3
14	57.9	3.3
15	59.8	3.3
16	61.8	3.3
17	64.0	3.3
18	66.3	3.4
19	69.3	3.6
20	73.9	4.5

NOTE: T score SE is the standard error of the estimated PROMIS T score in the adjacent column.

Table 4

Raw Neuro-QOL Score to T Score Conversion Table (IRT Fixed Parameter Calibration Linking)

Neuro-QOL Expected Total Raw Score	PROMIS T score	T score SE
19	30.7	4.8
20	35.3	3.3
21	37.7	2.8
22	39.4	2.5
23	40.8	2.3
24	42.0	2.1
25	43.0	2.0
26	43.9	1.9
27	44.7	1.8
28	45.5	1.8
29	46.2	1.7
30	46.8	1.7
31	47.4	1.6
32	48.0	1.6
33	48.6	1.6
34	49.1	1.5
35	49.6	1.5
36	50.2	1.5
37	50.7	1.5
38	51.1	1.5
39	51.6	1.5
40	52.1	1.5
41	52.5	1.5
42	53.0	1.4
43	53.4	1.4
44	53.9	1.4
45	54.3	1.4
46	54.7	1.4
47	55.2	1.4
48	55.6	1.4
49	56.0	1.4
50	56.5	1.4
51	56.9	1.4
52	57.3	1.4
53	57.8	1.4
54	58.2	1.4

Neuro-QOL Expected Total Raw Score	PROMIS T score	T score SE
55	58.6	1.4
56	59.1	1.4
57	59.5	1.4
58	59.9	1.4
59	60.4	1.4
60	60.8	1.5
61	61.2	1.5
62	61.7	1.5
63	62.1	1.5
64	62.6	1.5
65	63.0	1.5
66	63.5	1.5
67	63.9	1.5
68	64.4	1.5
69	64.9	1.5
70	65.3	1.5
71	65.8	1.5
72	66.3	1.5
73	66.8	1.5
74	67.2	1.5
75	67.7	1.5
76	68.2	1.5
77	68.7	1.5
78	69.2	1.5
79	69.7	1.5
80	70.2	1.5
81	70.8	1.5
82	71.3	1.5
83	71.8	1.5
84	72.4	1.5
85	73.0	1.5
86	73.5	1.6
87	74.2	1.6
88	74.8	1.7
89	75.5	1.7
90	76.3	1.8
91	77.2	1.9
92	78.2	2.1
93	79.5	2.4

Neuro-QOL Expected Total Raw Score	PROMIS T score	T score SE
94	81.1	2.7
95	83.3	3.0

NOTE: T score SE is the standard error of the estimated PROMIS T score in the adjacent column.

Table 5

Raw FACIT-Fatigue Score to T Score Conversion Table (IRT Fixed Parameter Calibration Linking)

FACIT-F Expected Total Raw Score*	PROMIS T score	T score SE
52	30.3	4.8
51	35.0	3.5
50	38.0	3.0
49	40.3	2.8
48	42.1	2.6
47	43.7	2.5
46	45.0	2.3
45	46.3	2.2
44	47.3	2.1
43	48.3	2.0
42	49.3	2.0
41	50.1	1.9
40	51.0	1.9
39	51.7	1.9
38	52.5	1.9
37	53.2	1.9
36	53.9	1.8
35	54.6	1.8
34	55.3	1.8
33	55.9	1.8
32	56.6	1.8
31	57.2	1.8
30	57.8	1.8
29	58.4	1.8
28	59.0	1.8
27	59.6	1.8
26	60.2	1.8
25	60.8	1.8
24	61.4	1.8
23	62.0	1.8
22	62.6	1.8
21	63.2	1.8
20	63.8	1.8
19	64.4	1.8
18	65.0	1.8
17	65.6	1.8

FACIT-F Expected Total Raw Score*	PROMIS T score	T score SE
16	66.2	1.9
15	66.9	1.9
14	67.5	1.9
13	68.2	1.9
12	68.9	2.0
11	69.6	2.0
10	70.4	2.0
9	71.2	2.1
8	72.0	2.2
7	72.9	2.3
6	73.9	2.4
5	75.0	2.5
4	76.2	2.7
3	77.5	2.9
2	79.1	3.1
1	81.2	3.3
0	83.5	3.4

[a] T score SE is the standard error of the estimated PROMIS T score in the adjacent column.

[b] FACIT-Fatigue was analyzed using the same direction as the PROMIS FIB (i.e., higher scores mean more fatigue). However, to be consistent with the FACIT measurement System manual, in this table, higher FACIT-F raw scores represent better quality of life (i.e., lower fatigue) and lower scores represent worse quality of life (i.e., more fatigue).