

TBI-QOL: Development and Calibration of Item Banks to Measure Patient Reported Outcomes Following Traumatic Brain Injury

David S. Tulsky, PhD; Pamela A. Kisala, MA; David Victorson, PhD; Noelle Carlozzi, PhD; Tamara Bushnik, PhD; Mark Sherer, PhD; Seung W. Choi, PhD; Allen W. Heinemann, PhD; Nancy Chiaravalloti, PhD; Angelle M. Sander, PhD; Jeffrey Englander, MD; Robin Hanks, PhD; Stephanie Kolakowsky-Hayner, PhD; Elliot Roth, MD; Richard Gershon, PhD; Mitchell Rosenthal, PhD; David Cella, PhD

Objective: To use a patient-centered approach or participatory action research design combined with advanced psychometrics to develop a comprehensive patient-reported outcomes (PRO) measurement system specifically for individuals with traumatic brain injury (TBI). This TBI Quality-of-Life (TBI-QOL) measurement system expands the work of other large PRO measurement initiatives, that is, the Patient-Reported Outcomes Measurement Information System and the Neurology Quality-of-Life measurement initiative. **Setting:** Five TBI Model Systems centers across the United States. **Participants:** Adults with TBI. **Design:** Classical and modern test development methodologies were used. Qualitative input was obtained from individuals with TBI, TBI clinicians, and caregivers of individuals with TBI through multiple methods, including focus groups, individual interviews, patient consultation, and cognitive debriefing interviews. Item pools were field tested in a large multisite sample ($n = 675$) and calibrated using item response theory methods. **Main Outcomes Measures:** Twenty-two TBI-QOL item banks/scales. **Results:** The TBI-QOL consists of 20 independent calibrated item banks and 2 uncalibrated scales that measure physical, emotional, cognitive, and social aspects of health-related quality of life. **Conclusions:** The TBI-QOL measurement system has potential as a common data element in TBI research and to enhance collection of health-related quality-of-life and PRO data in rehabilitation research and clinical settings. **Key words:** common data elements, clinical research, health-related quality of life, item response theory, patient-reported outcomes, quality of life, traumatic brain injury

Author Affiliations: Department of Physical Therapy, University of Delaware College of Health Sciences, Newark (Dr Tulsky and Ms Kisala); Kessler Foundation Research Center, West Orange, New Jersey (Drs Tulsky, Chiaravalloti, and Rosenthal); New York University Langone Medical Center, New York (Dr Bushnik); Northwestern University Feinberg School of Medicine, Chicago, Illinois (Drs Victorson, Heinemann, Roth, Gershon, and Cella); University of Michigan Medical School, Ann Arbor (Dr Carlozzi); TIRR Memorial Hermann, Houston, Texas (Drs Sherer and Sander); McGraw-Hill Education CTB, Monterey, California (Dr Choi); Rehabilitation Institute of Chicago, Chicago, Illinois (Drs Heinemann and Roth); Rutgers NJ University Medical Center, Newark, New Jersey (Dr Chiaravalloti); Baylor College of Medicine, Houston, Texas (Dr Sander); Santa Clara Valley Medical Center, Santa Clara, California (Drs Englander and Kolakowsky-Hayner); and Wayne State University School of Medicine, Detroit, Michigan (Dr Hanks).

All authors have contributed significantly to the design, analysis, and writing of the manuscript. The contents represent original work and have not been published elsewhere. No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the authors or upon any organization with which the authors are associated.

All TBI-QOL items, parameters, and data are © 2013 David Tulsky and Kessler Foundation. All rights reserved. All items are freely available to the

public via the Assessment Center platform (www.assessmentcenter.net) and there are currently no plans for Dr Tulsky or Kessler Foundation to profit from the use of the TBI-QOL instrument.

This study was supported by grant nos. H133G070138, H133A070037, H133A070043, H133A080045, H133A080044, and H133A70038 from the National Institute on Disability and Rehabilitation Research.

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.headtraumarehab.com).

The authors declare no conflicts of interest.

Corresponding Author: David S. Tulsky, PhD, Department of Physical Therapy, University of Delaware, College of Health Sciences, STAR Health Sciences Complex, 540 S College Ave, Newark, DE 19713 (dtulsky@udel.edu).

DOI: 10.1097/HTR.0000000000000131

TRAUMATIC BRAIN INJURY (TBI) can instantly and dramatically change a person's life, leaving lifelong challenges. Many individuals who sustain a TBI experience lifelong impairments that profoundly impact health-related quality of life (HRQOL) across a wide range of functioning, including physical, emotional, and social health.¹ Some of the most common post-TBI deficits are in the areas of cognition (e.g., attention, arousal, memory, executive functioning, emotional and behavioral control, and inhibition),^{2,3} mobility (e.g., balance and motor coordination problems), and physical symptoms (e.g., fatigue, headache, sleep disturbances, sensory impairments, and weakness).⁴ Many individuals with TBI experience difficulties carrying out self-care activities such as eating, dressing, and personal hygiene.⁵ Behaviorally, individuals with TBI may experience a lack of inhibition resulting in issues such as impulsivity and irritability. Sexual function and sexual desire may also be affected.⁶ Recovery may be complicated by mood disorders such as depression and anxiety,⁷ and effects can include a lack of awareness of deficits. While individuals' subjective experience of their deficits is relevant for both clinicians and researchers, there is no widely accepted instrument with which to obtain a comprehensive HRQOL assessment of individuals with TBI.

Health-related quality of life is a subjectively evaluated construct that refers to the impact of a disease or disability or its treatment on one's physical, mental, and social well-being.⁸ The measurement of HRQOL is well established in many health settings; yet, there is a paucity of HRQOL measures developed for TBI populations. Outcomes measures used in TBI research have often included generic instruments such as the SF-36, Nottingham Health Profile, EuroQOL, and the WHO-QOL, although they have not been well validated in TBI samples,⁹ and have focused more on measuring a limited domain such as community participation¹⁰⁻¹⁵ rather than capturing a complex, multidimensional construct encompassing physical, emotional, and social functioning. Such limitations in available TBI outcomes measures challenge the construct validity of the outcomes measurement process through construct underrepresentation, "as the assessment is too narrow and fails to include important aspects of the construct."^{16(p742)} Furthermore, such generic measures fail to address TBI-specific issues while also including irrelevant and uninformative items. For instance, the Satisfaction with Life Scale asks respondents whether they would go back and change anything in their life, given the opportunity. Such a question can seem offensive to an individual who has sustained substantial deficits due to a single, traumatic event. Such irrelevant questions distort the meaning of the scores obtained on the given assessment and therefore seriously challenge the construct valid-

ity of the outcomes measurement process.¹⁷ For these reasons, TBI researchers need a better method for measuring multifaceted patient-reported health outcomes, or HRQOL.

While there are no generally accepted HRQOL measures focusing on the specific needs and concerns of the TBI population, over the last 10 years, there have been significant advances in HRQOL measurement more broadly. Most notably, the Patient-Reported Outcomes Measurement Information System (PROMIS) was a key initiative within the National Institutes of Health (NIH) Roadmap process (now called the NIH Common Fund).¹⁸ As part of a cooperative agreement, several institutes, agencies, and centers within the US Department of Health and Human Services funded a network of scientists whose goal was to create a cutting-edge tool using advanced psychometric theory and computer technology that could be used as a common data element across NIH's research portfolio.^{18,19}

A key feature of the PROMIS is the use of item banks^{20,21} across a variety of domains that can be calibrated using item response theory (IRT) and administered via computerized adaptive testing (CAT) or short forms (SFs). An item bank contains many carefully selected items that represent a unidimensional concept or area of functioning, with individual items spread out across the continuum of the concept being measured.²² The item banking process is a laborious one in which a comprehensive set of questions about functioning, symptoms, or clinical problems is identified and developed as potential items, employing qualitative research with key stakeholders (e.g., individuals with the disease or condition in question, clinicians, and/or caregivers)²³ to generate and test ideas. Calibration involves the use of advanced quantitative psychometric methodology to determine the difficulty (level of underlying trait being measured) and discrimination (ability to distinguish between individuals possessing different levels of the underlying trait) of each item.^{20,24,25} The integration with CAT technology provides an efficient method of administration in that only the most informative items for the individual being tested are administered.²⁶ Unnecessary or uninformative items are thereby avoided and test administration is tailored to each individual. Examinee burden is reduced because only a small subset of the items in each bank is administered. Moreover, the measurement precision of CAT very closely approximates the precision of the full item bank containing 4 to 6 times the number of items.^{27,28} A detailed description of the item banking process, including IRT methodology, and its advantages in rehabilitation outcomes measurement can be found in the works of Velozo et al.²⁹ and Tulskey et al.³⁰

The initial release of PROMIS included 11 Physical (e.g., Physical Function, Fatigue), Emotional (e.g.,

Depression, Anger), and Social Health (e.g., Satisfaction with Social Roles) item banks that were calibrated with data from a large sample of 21,133 individuals drawn from the general population.³¹ In addition, a 10-item PROMIS Global Health Scale was included with items cutting across all domains as a measure of general health status.³² New (version 2.0) item banks have subsequently been developed, calibrated, and released measuring additional areas (e.g., Applied Cognition, Psychosocial Illness Impact, Social Isolation, and Sexual Functioning). PROMIS may also be used with children, as several pediatric and pediatric-proxy item banks (e.g., Anxiety, Mobility, Peer Relationships) have been developed. All PROMIS item banks are available as CATs or SFs, and all PROMIS instruments are publicly available on the Assessment Center^{SM33–35} platform.

The Neurology Quality of Life Measurement Initiative (Neuro-QOL),³⁶ funded by the National Institute of Neurological Disorders and Stroke, is a clinically relevant, psychometrically robust approach to measuring HRQOL in individuals with neurologic disorders that is linked directly to PROMIS through common items. It was designed to precisely measure issues that are relevant to individuals with major neurologic disorders. Following extensive qualitative research about key neurologic conditions,³⁶ the scale was developed targeting individuals with stroke (adult only), multiple sclerosis (adult only), Parkinson disease (adult only), epilepsy (adult and pediatric), amyotrophic lateral sclerosis (adult only), and muscular dystrophy (pediatric only). Traumatic brain injury was rated as very important (ranking as the eighth most important condition), but individuals with TBI were not part of the Neuro-QOL development. Furthermore, most of the Neuro-QOL calibration testing was conducted with general population samples.³⁷ Like PROMIS, Neuro-QOL was designed to be a leading outcomes tool in clinical trials research and includes item banks assessing Physical, Emotional, Cognitive, and Social Health. Neuro-QOL item banks may also be administered through CATs or SFs.³⁶

Both the PROMIS and Neuro-QOL mark significant advances in HRQOL measurement development, implementing new standards in advanced psychometrics and computer-based technology to enhance assessment. However, neither system includes content that is TBI-specific, nor have they been tailored to individuals with TBI. Despite the significant contributions that PROMIS and Neuro-QOL have made to the field of patient-reported outcomes (PRO) and HRQOL measurement, a clinically relevant, psychometrically sound measure of HRQOL for use among individuals with TBI is still needed. To address this need, the Traumatic Brain Injury Quality-of-Life (TBI-QOL) project was funded by the National Institute on Disability and

Rehabilitation Research as both a Field Initiated Research grant and a collaborative Traumatic Brain Injury Model Systems (TBIMS) project involving 5 TBIMS centers. The TBI-QOL project's primary goal was to develop a meaningful, relevant, and psychometrically sound measure that provides comprehensive assessment of HRQOL for persons with TBI. Entirely new item banks were developed (e.g., Headache Pain) that address TBI-specific issues. The TBI-QOL has been designed to integrate with PROMIS and Neuro-QOL: when measuring generic constructs that are measured by PROMIS and Neuro-QOL, TBI-QOL scores are directly comparable with the relevant Neuro-QOL and PROMIS item banks and serve to optimize these measurement systems for TBI populations. This article provides an overview of the TBI-QOL development and calibration process.

METHODS

The TBI-QOL was developed in 2 distinct phases. The initial phase focused on development of the domain structure, evaluation of extant items, and creation and refinement of new item pools.³⁸ The second phase focused on large-scale field testing, item calibration, finalization of item banks, and creation of CATs and SFs. Both phases followed PROMIS Measurement Standards³⁹ throughout the development process. This process included (1) determination of domain coverage through multiple methods (including focus groups with individuals with TBI, expert/clinician input, and literature review); (2) identification of extant items; (3) item pool construction; (4) qualitative item review including expert item review and cognitive debriefing interviews; (5) literacy-level review and translatability review³⁶; (6) large-scale field testing of item pools in representative samples; (7) advanced psychometric analyses to calibrate and finalize item banks; and (8) development of CAT and SFs to provide options for brief administration of item banks.

Phase 1: Domain selection and item pool development

Identification of domains to be included in the TBI-QOL was achieved through qualitative research with patients and clinicians, literature review, and expert input. Emphasis was placed upon the qualitative analyses previously reported by Carlozzi et al.³⁸ Thirteen focus groups were conducted involving 33 individuals with TBI, 17 family member/caregivers, and 15 TBI clinicians. Participants in all groups discussed quality of life in individuals with TBI and were asked to consider ways in which a TBI affected overall well-being and quality of life. Rigorous qualitative analysis of focus group transcripts, as described in Kisala and Tulskey,²³ identified

general as well as more targeted domains of HRQOL. In accordance with the goal of aligning the TBI-QOL to the fullest extent possible with the PROMIS and Neuro-QOL systems, the literature reviews, domain definitions, and even item content borrowed heavily from the extant projects, although specific literature reviews were conducted for new, TBI-specific item banks not included in the existing measures. The TBI-QOL collaborators, all experts in TBI and/or measurement science, met on several occasions to review the domain structure of the TBI-QOL and to develop and refine item content.

Item bank development

Once the most important HRQOL domains and subdomains were identified, a variety of sources were used to develop a large pool of relevant items (item pool) for each topic area. First, items were drawn verbatim from the PROMIS and/or Neuro-QOL systems in content areas that overlapped with other health populations. For example, the TBI-QOL item pools for Fatigue, Depression, Anxiety, Anger, and Pain Interference were made up primarily of PROMIS items, whereas items in the Ability to Participate in Social Roles and Activities, Satisfaction with Social Roles and Activities, Positive Affect and Well-being, Stigma, Mobility, and Upper Extremity pools were drawn primarily from Neuro-QOL. Next, the TBI-QOL focus group transcripts were used to identify and address any conceptual gaps in the existing item banks and to draft new items in areas where content was more unique to TBI. For topics and subdomains specific to TBI (or to individuals who had experienced a sudden life-changing injury), new items were developed to capture the constructs. The focus group transcripts were reviewed in an iterative fashion,²³ often using comments and feedback directly from people with TBI, to develop the initial candidate items. Literature reviews were also performed to develop construct definitions and guide further item development. Content experts reviewed the item pools and helped develop additional items in clinically relevant areas that were not discussed frequently during focus groups. Finally, the item pools were reviewed informally by subgroups of coinvestigators who edited and revised each item to reduce ambiguous, colloquial, and inconsistent wording.

The preliminary item pool for each domain was then subjected to a rigorous qualitative item review process through more formal expert item review by the entire team, cognitive debriefing interviews of individuals with TBI, a literacy-level review, and an item translatability review by translation scientists with experience in adapting and translating tests. The expert item review phase consisted of all study team members, each of whom is

an expert in TBI and/or measurement theory, reviewing each draft item pool. Within each pool, each item was reviewed for relevance, construct representativeness, wording, and conciseness. Unacceptable items were either reworded or deleted at this time. To the extent possible, team members organized the content along a “difficulty” continuum (i.e., level of underlying trait), with items relevant to individuals with lower functioning in a given domain at one end of the continuum and items relevant to higher-functioning individuals at the opposite end. The experts also helped identify new items to close conceptual gaps within each subdomain including—most importantly—gaps near the ceiling and floor of the continuum.

Next, a series of cognitive debriefing interviews was conducted with individuals with TBI ($n \geq 5$ per item). Each participant was given a subset of items to complete. Following the response to each item, participants would discuss their understanding of the item with the trained interviewer, describing their interpretation of the item and commenting on its relevance and wording. Finally, participants were asked whether they felt any items were “missing” from each topic area. This “cognitive interviewing” process helped ensure that items selected for testing would be understood as intended by respondents with TBI.

Because the TBI-QOL may be completed as a self-administered instrument, steps were taken to ensure the comprehensibility of each item to all participants capable of reading English, regardless of their education level. The Lexile Framework⁴⁰ was used to ensure that none of the final TBI-QOL items exceeded a fifth-grade reading level.

The final step, prior to field testing, was to ensure that TBI-QOL items would be appropriate for translation to Spanish at a later time. We conducted a translatability review⁴⁰ in which Spanish-speaking translation science experts reviewed each item to identify potential concerns, such as items containing wording or concepts that would be difficult to translate into Spanish. The translation science team had 2 reviewers from different Spanish-speaking countries conduct both forward- and back-translations while reviewing all items and noting any concerns that could potentially affect translatability potential. A third reviewer reviewed these comments and items, reconciling and aggregating the feedback between the 2 independent reviews. All translatability review feedback was provided in detail to the principal investigator and item development teams, who modified the item pools to ensure that all vocabulary and grammar would be amenable to future translation into Spanish.

The final item pools consisted of 922 items covering a wide range of functioning. The TBI-QOL preliminary subdomains and items consisted of 28 item

pools (see Supplemental Digital Content Table A, available at: <http://links.lww.com/JHTR/A137>) across the 4 larger domains of Physical, Emotional, Cognitive, and Social Health.

Phase 2: Calibration field testing and development of item banks

The a priori goal of this phase was to calibrate the item pools using a 2-parameter IRT model, more specifically the Graded Response Model⁴¹ (GRM). Despite the somewhat larger sample size requirement, the GRM was chosen because of its use in the calibration of PROMIS and Neuro-QOL items and availability of programming new CATs in Assessment CenterSM.⁴² In addition to item difficulty that is estimated by the Rasch or 1-parameter model, the GRM also estimates the ability of each item to discriminate between individuals at different levels of the underlying trait. Estimation of the GRM requires a participant sample that is both heterogeneous with regard to functioning (i.e., representative of the population) and fairly large (e.g., $n \geq 500$); a sample of this size is considered adequate to produce stable parameter estimates.⁴³ For this study, we planned to collect a sample of at least 500 individuals with medically documented complicated-mild⁴⁴ (i.e., emergency room Glasgow Coma Scale score of 13-15 accompanied by positive neuroimaging findings), moderate, or severe TBI from multiple sites around the United States.

Participants

A sample of 675 individuals with TBI was recruited from 5 TBI Model System centers: Kessler Foundation/Kessler Institute for Rehabilitation; Rehabilitation Institute of Chicago; Rehabilitation Institute of Michigan/Wayne State University; TIRR Memorial Hermann; and Santa Clara Valley Medical Center. The study protocol was reviewed and approved by each site's Institutional Review Board. Persons with documented complicated-mild, moderate, or severe TBI who were 16 years or older at the time of injury, at least 18 years old at the time of data collection, and able to read and speak English fluently were eligible to participate. Each participant's diagnosis and severity of injury were confirmed through medical record review. The heterogeneous sample was stratified by severity of injury (i.e., complicated-mild,⁴⁴ moderate, and severe). Furthermore, although this study used a community-dwelling sample of individuals who tended to be several years postinjury at the time of study participation, care was taken to include more recently injured individuals; namely, those who had been injured in the past 6 to 18 months and in the past 18 months to 3 years.

Data collection procedures

Prior to field testing, a customized software package was developed and beta tested to facilitate data collection at multiple collaborative sites via a Web-based interface. Each site was set up with LAN, wireless, and/or cellular Internet access to facilitate data collection from any location. All site coordinators and data collectors were trained to collect TBI-QOL data in an interview format using the Web-based interface/software program. Training was performed via in-person and/or Web-based training sessions. A detailed Manual of Procedures was prepared and distributed to all sites, and all items were presented to participants in interview format, either in person or over the phone, by a trained data collector.

Because of the large number of items in the calibration version of the TBI-QOL ($k = 922$), data collection was divided into 4 separate interview sessions. All items within an individual item pool (e.g., Depression) were presented during the same session. To facilitate the responses, participants were shown a series of response cards with response options to guide them through the interview. All data were entered directly into the Web-based data collection system and automatically stored on a secure server in real time. Response data were reviewed biweekly for completeness and quality, and the sampling matrix was reviewed for accrual and stratification (i.e., severity and time since injury) targets. For the duration of data collection, representatives from all sites participated in biweekly conference calls to discuss progress and goals, specifically with regard to meeting sampling stratification goals.

RESULTS

Participant demographic characteristics

Of the total sample of 675 individuals, 590 completed session 1 (containing items related to physical function and symptoms), 521 completed session 2 (emotional well-being items), 569 completed session 3 (cognition items, anger, and emotional and behavioral dyscontrol), and 556 completed session 4 (social participation and stigma items). While each participant was encouraged to complete each of the sessions, this was not always feasible, hence the differing sample sizes for each session. Detailed demographic data for study participants by interview session are shown in Table 1.

Among the 675 total participants, mean age was 41.6 (SD = 14.7) years and 72% of participants were male. Fifty-eight percent of the sample self-identified as white, 26% as African American, 5% as more than 1 race, 2% Asian, and 10% other. In addition, 15% of participants were of Hispanic or Latino origin or descent. In terms of injury severity, 27% had complicated-mild injuries, 18% had moderate TBI, and 55% had severe TBI. Average

TABLE 1 Calibration sample demographics

Variable	Total sample (N = 675)	Session 1 (n = 590)	Session 2 (n = 521)	Session 3 (n = 569)	Session 4 (n = 556)
Age, mean (SD), y	41.6 (14.7)	41.9 (14.8)	40.8 (14.7)	41.0 (14.6)	41.0 (14.7)
Gender, n (%)					
Male	484 (71.7)	419 (71)	367 (70)	384 (69)	373 (68)
Female	191 (28.3)	171 (29)	154 (30)	177 (31)	176 (32)
Ethnicity, %					
Hispanic	15	14	17	16	16
Non-Hispanic	85	86	83	84	84
Race, %					
White	58	57	64	63	62
African American	26	25	18	21	22
Asian	2	2	2	3	3
American Indian/Alaska Native or Native Hawaiian/Pacific Islander	2	2	1	2	2
> 1 race	5	5	5	4	4
Other	8	8	8	7	7
Education, %					
High school or less	40	40	37	39	40
Some college	35	36	37	35	36
Bachelor's degree or more	24	24	26	26	24
Total household income, %					
<\$20 000	35	35	22	33	33
\$20 000-\$74 999	29	29	31	30	30
\$75 000	19	19	21	20	20
Unknown	18	17	17	18	18
Glasgow Coma Scale score	8.6	8.7	8.7	8.5	8.5
Injury severity, %					
Complicated mild	27	31	29	28	27
Moderate	18	14	14	13	14
Severe	55	54	55	58	58
Time since injury, %					
<18 mo	26	26	28	28	28
18 mo-3 y	19	19	21	20	20
>3 y	54	55	51	52	52
Cause of injury, %					
Motor vehicle accident	48.6	48.7	50.7	50.1	48.4
Gunshot wound or other violence	18.7	18.1	14.0	16.3	16.0
Fall	14.5	15.1	15.9	15.9	15.6
Pedestrian hit by car	4.4	5.1	4.8	4.9	4.9
Motorcycle accident	4.1	4.8	5.2	5.1	5.0
Bicycle accident	1.6	1.7	1.9	2.0	1.8
Sports	0.9	1.0	1.2	1.1	0.9
ATV accident	0.6	0.7	0.8	0.7	0.3
Airplane accident	0.3	0.3	0.4	0.2	0.2
Other	3.7	4.2	4.2	3.8	3.6

Abbreviation: ATV, all-terrain vehicle.

time since injury was 5.8 (SD = 6.8) years. Twenty-six percent of participants had been injured for less than 18 months at the time of study participation, 19% between 18 months and 3 years, and 54% for more than 3 years.

Psychometric analysis

Of the 28 initial draft item pools, 22 were analyzed.* Twenty calibrated item banks, 2 brief, fixed-length scales, and 6 uncalibrated item pools were developed (see Supplemental Digital Content Table B, available at: <http://links.lww.com/JHTR/A138>, for a list of final calibrated item banks and associated definitions).

Preliminary analyses included tests for internal consistency (Cronbach α), corrected item-total correlations, examinations for excessive missing data and sparse cells (i.e., response categories with <10 responses for a given item), and violations of unidimensionality. Final values for all item bank statistics including α and item-total correlations are shown in Table 2.

Since unidimensionality is a required assumption for IRT analysis and CAT programming, dimensionality of each bank was assessed using confirmatory factor analyses (CFAs). Specifically, we tested all items in the bank for their fit to a unidimensional model. Several goodness-of-fit indices served as criteria for acceptable unidimensionality, including the Bentler Comparative Fit Index (CFI)⁴¹ the Tucker-Lewis Index (TLI),⁴² (where values of ≥ 0.90 indicate acceptable fit to the model and values of ≥ 0.95 indicate good fit⁴³), and the root mean square error of approximation (RMSEA)⁴⁴ (where values range from 0 to 1, with 0 indicating perfect model fit, values below 0.08 acceptable fit, and values below 0.06 are considered good fit). Unidimensional models were tested in separate CFAs for each of the 20 item pools. When poor fit was indicated by the CFI, TLI, and/or RMSEA value(s), low item-total correlations, or local item dependence,⁴⁵ items were removed in an iterative fashion and unidimensional models were retested after the poorly functioning items were removed. The analyses were iteratively rerun until the fit statistics for the CFAs for each bank supported a unidimensional model, all items exhibited satisfactory factor loadings (i.e., ≥ 0.30 ; ideally ≥ 0.40), and local item dependence was minimized.

IRT parameters and IRT-based model fit were subsequently estimated using the GRM. Ranges for slope and threshold for each bank are given in Table 2. Within each item bank, each item was evaluated for misfit ($S-X^2$ index)⁴⁶ and differential item functioning.⁴⁷ If any

additional items were removed from the item pool at this time, the IRT and differential item functioning analyses were rerun following their removal. Final item counts and fit statistics are given in Table 2. Next, item banks containing a substantial number of items from PROMIS (e.g., Anger, Anxiety, Depression, Pain Interference) or Neuro-QOL (e.g., Positive Affect and Well-being, Ability to Participate in Social Roles & Activities, Satisfaction with Social Roles & Activities) were transformed onto the PROMIS or Neuro-QOL metric (as appropriate) using IRT linking techniques. This linking procedure used common (verbatim) items as “anchors,” using Stocking-Lord⁴⁸ equating techniques to identify slope and intercept transformation parameters and performing a linear transformation of each item calibration so that TBI-QOL transformations were placed on the respective PROMIS or Neuro-QOL metric. This procedure ensures the dual advantage of having a TBI-specific sample inform the CAT item selection algorithm to ensure the administration of the most informative item at each level of the underlying trait while still allowing direct comparison with the general population⁴⁹ and/or other studies via the respective PROMIS or Neuro-QOL metric. For item content that did not have a comparable PROMIS or Neuro-QOL item bank (see Table 3; e.g., Headache Pain, Resilience, Grief/Loss, Self-esteem, Communication/Comprehension, Independence), the calibrations based upon the TBI sample were used to develop the CAT.

Once all analyses were completed and all parameters transformed (where applicable), our final step was to develop CATs and static SFs for each item bank. The CATs were developed using final item calibration parameters obtained from the last iteration of the IRT analyses. The final IRT parameters were programmed into the Assessment CenterSM Web site⁴² to enable CAT administration. To develop the SFs, we met with co-investigators to review the item information functions (produced by the IRT analyses) and determine the most discriminating items and examined the difficulty (e.g., locations on the measurement continuum) to ensure that we had selected items across the entire continuum of each underlying trait. We also balanced these empirical indices with clinical judgment of each item’s relative importance. The SFs each contain between 8 and 10 items (see Table 2) and are also available through the Assessment CenterSM Web site.

Scoring metric for TBI-QOL item banks

All TBI-QOL scores have been transformed to a T metric, with a mean of 50 (SD = 10). For all banks that have been equated and placed on the PROMIS or Neuro-QOL metric, the population used to calibrate the extant item bank (often a general population sample) serves as the reference group (see Table 3). In these cases,

*The rationale for not analyzing and/or calibrating items varied by item pool. A variety of factors, including limited sample size, experimental nature of certain items, and multidimensionality of item pools, contributed to these decisions.

TABLE 2 TBI-QOL item bank statistics

Domain	Bank	Scoring direction (higher scores)	No. items tested	No. final items	No. items in SF	n	CFI	TLI	RMSEA	Cronbach α	Item-total correlations	Slope range	Threshold range	
Physical health	Mobility	Better function	37	32	9	590	0.977	0.973	0.075	0.977	0.62-0.84	2.07-3.94	-3.33 to 0.62	
	Upper Extremity	Better function	44	33	9	590	0.969	0.973	0.054	0.969	0.56-0.79	1.99-3.91	-3.66 to -0.69	
	Fatigue	Severe symptom	95	73	10	590	0.992	0.979	0.052	0.992	0.58-0.87	1.40-3.79	-1.37 to 3.69	
	Pain Interference	Severe symptom	40	40	10	590	0.988	0.980	0.076	0.988	0.68-0.89	2.07-5.44	0.11-2.56	
	Headache Pain	Severe symptom	13	13	10	590	0.975	0.997	0.075	0.975	0.71-0.91	3.40-7.90	-0.23 to 1.86	
	Positive Affect and Well-being	Better function	32	32	9	591	0.972	0.961	0.070	0.972	0.56-0.80	1.73-3.90	-2.14 to 1.55	
Emotional health	Depression	Severe symptom	32	28	8	591	0.967	0.981	0.061	0.967	0.50-0.82	1.57-4.06	-0.50 to 3.89	
	Anxiety	Severe symptom	35	28	10	591	0.959	0.967	0.066	0.959	0.46-0.78	0.99-3.13	-0.94 to 3.22	
	Stigma	Severe symptom	30	28	10	556	0.963	0.952	0.078	0.963	0.49-0.78	1.48-3.04	-2.86 to 0.82	
	Resilience	Better function	30	27	10	591	0.952	0.950	0.070	0.952	0.49-0.77	1.15-2.57	-3.47 to 2.00	
	Grief/Loss	Severe symptom	18	17	9	591	0.952	0.981	0.087	0.952	0.59-0.81	1.64-3.26	-0.93 to 2.67	
	Self-esteem	Better function	24	13	10	591	0.933	0.986	0.071	0.933	0.56-0.80	1.32-3.69	-2.67 to 0.58	
	Anger	Severe symptom	38	35	10	569	0.970	0.969	0.062	0.970	0.45-0.82	0.99-2.89	-1.74 to 4.78	
	Emotional and Behavioral	Severe symptom	36	26	10	569	0.958	0.952	0.079	0.958	0.51-0.76	1.27-2.41	-0.94 to 3.39	
	Cognition	Dyscontrol	Better function	65	37	10	569	0.972	0.954	0.070	0.972	0.55-0.79	2.25-4.15	-3.61 to -0.49
		Executive Function	Better function	56	39	10	569	0.974	0.971	0.058	0.974	0.55-0.83	1.64-4.53	-4.61 to 0.01
Cognition—General Concerns		Better function	48	31	9	569	0.962	0.949	0.078	0.962	0.54-0.75	1.36-2.85	-3.87 to 1.06	
Social participation	Communication/Comprehension	Better function	54	45	10	556	0.976	0.954	0.067	0.976	0.55-0.79	2.19-5.85	-2.02 to 0.24	
	Ability to Participate in Social Roles and Activities	Better function	46	41	10	556	0.973	0.954	0.069	0.973	0.58-0.71	3.10-6.46	-1.92 to 0.21	
	Satisfaction with Social Roles and Activities	Better function	14	13	8	556	0.914	0.969	0.092	0.914	0.50-0.77	1.20-3.55	-3.21 to 0.83	
	Independence	Better function	14	13	8	556	0.914	0.969	0.092	0.914	0.50-0.77	1.20-3.55	-3.21 to 0.83	

Abbreviations: CFI, comparative fit index; RMSEA, root mean square error of approximation; SF, short form; TBI-QOL, Traumatic Brain Injury Quality of Life; TLI, Tucker-Lewis Index.

TABLE 3 Linkages with PROMIS and Neuro-QOL^a

Subdomain/Bank	No. TBI-QOL items	Linked to	No. PROMIS items	No. Neuro-QOL items	Reference population
Mobility	32	Neuro-QOL	9	19	General + Neuro
Upper Extremity	33	Neuro-QOL	17	16	General + Neuro
Fatigue	73	PROMIS	67	6	General
Pain Interference	40	PROMIS	40	0	General
Headache Pain	13	...	0	0	TBI
Positive Affect and Well-being	32	Neuro-QOL	0	23	General
Depression	28	PROMIS	18	22	General
Anxiety	28	PROMIS	16	20	General
Stigma	28	Neuro-QOL ^b	0	23	Neuro
Resilience	27	...	0	0	TBI
Grief/Loss	17	...	0	0	TBI
Self-Evaluation	13	...	0	0	TBI
Anger	35	PROMIS	29	0	General
Emotional and Behavioral Dyscontrol	26	Neuro-QOL	2	16	Neuro
Executive Function	37	Neuro-QOL	0	11	General + Neuro
Cognition—General Concerns	39	Neuro-QOL	0	12	General + Neuro
Communication/Comprehension	31	...	0	5	TBI
Ability to Participate in Social Roles and Activities	45	Neuro-QOL	0	43	General
Satisfaction with Social Roles and Activities	41	Neuro-QOL	19	39	General
Independence	13	...	0	0	TBI

Abbreviation: Neuro, mixed neurologic population consisting of individuals with stroke, epilepsy, Parkinson disease, multiple sclerosis, and amyotrophic lateral sclerosis; Neuro-QOL, Neurology Quality-of-Life Measurement Initiative; PROMIS, Patient-Reported Outcomes Measurement Information System; QOL, quality of life; TBI, traumatic brain injury; TBI-QOL, Traumatic Brain Injury Quality of Life.

^aBold text indicates the number of items that are statistically linked. Neuro-QOL and PROMIS items are not mutually exclusive—many items are in both PROMIS and Neuro-QOL.

^bThe term “injury” has replaced “illness” in all Neuro-QOL Stigma items.

the person’s score on the TBI-QOL reflects his or her standing in the general population.[†] For all banks that are “new” to TBI-QOL (e.g., Resilience, Grief/Loss), an individual’s score represents his or her standing among individuals with medically confirmed TBI. For example, an individual with a score of 60 on Pain Interference and a score of 60 on Grief/Loss is experiencing 1 SD more pain interference than average for the general population and 1 SD more grief than average for individuals with TBI.

[†]For the Neuro-QOL scale, the majority of banks have been developed using general population samples. However, some item banks were targeted for neurologically impaired populations (e.g., Stigma) and were developed using a mixed neurologic population.

DISCUSSION

The primary purpose of the TBI-QOL was to develop a state-of-the-art PRO measurement system for individuals with TBI. The TBI-QOL was created with extensive input from individuals with TBI, clinicians who specialize in TBI populations, and measurement experts. Creation of this measure used advanced measurement development technology following the PROMIS development standards.³⁷ The TBI-QOL is the first comprehensive measurement scale covering a wide array of domains that has been developed specifically for use among individuals with TBI. It includes newly developed items or, in some cases, item banks designed to reflect constructs that affect individuals with TBI that are not measured by generic scales. Moreover, in contrast to measures developed using classical test

theory methods, the TBI-QOL benefits from the use of IRT-based methods. Use of IRT increases measurement precision even while reducing respondent burden (i.e., with shorter forms), facilitates evaluation of change over time through the use of standardized scores, and informs development of CATs. A secondary goal of this project was to enhance and extend the PROMIS and Neuro-QOL measurement systems, which were designed for the general population, for use with persons with TBI. On the surface, these 2 goals would appear to be at odds; however, the TBI-QOL appears to have achieved both of these goals. The TBI-QOL consists of a core set of PROMIS and Neuro-QOL item banks that have been optimized for TBI so that only the most relevant and discriminating items for individuals with TBI are administered. In addition, the TBI-QOL scales that overlap with PROMIS or Neuro-QOL have been re-anchored so that scores on the more generic and universal constructs (such as Depression) are directly interpretable with the PROMIS or Neuro-QOL systems, enabling cross-condition or disease comparisons to be made. Researchers and clinicians using the TBI-QOL banks can compare individuals with TBI with individuals with other neurologic disorders or even with individuals in other health populations or within the general population. The TBI-QOL offers rehabilitation researchers and clinicians an opportunity to use the leading measurement systems while being assured that the content and items are relevant and have utility when testing individuals with TBI. Furthermore, the availability of all of these related measures (i.e., PROMIS, Neuro-QOL, TBI-QOL) on the Assessment CenterSM platform gives researchers the flexibility to easily select the most appropriate measure for use with a given population and within a given study. Our analyses indicated that each of the TBI-QOL item banks measures a unidimensional construct and that items within each bank demonstrate a high degree of internal consistency.

Ultimately, the TBI-QOL can improve TBI rehabilitation by instantaneously informing treating clinicians of an individual's standing on a wide variety of subjectively important HRQOL domains. Given the brevity of each assessment, TBI-QOL measures may be administered repeatedly throughout rehabilitation, and clinicians can

access scores in real time. Future directions include automation of this process so that clinicians would receive automated messages when patients score significantly worse than their baseline or when they exceed a predetermined cutoff for severity.

Study limitations

While the TBI-QOL calibration sample met stratification criteria and is demographically diverse and medically heterogeneous, the sample excluded individuals who had more severe cognitive impairment that would preclude them from completing self-report items. Each site tried to include individuals with as wide a range of functioning as possible and as such collected all data in interview format using trained and certified examiners. However, the subgroup of individuals with TBI who had the most profound cognitive deficits (e.g., unable to speak and/or respond to questions about HRQOL) was not involved in the development or calibration processes. In addition, the TBI-QOL studies have yet to systematically evaluate convergent and divergent validity with other measures. Future work is also needed to examine the TBI-QOL banks' responsiveness to individuals' change over time, as such work was beyond the scope of the current project. Finally, the potential for a mode of administration effect (interview vs. self-report) should be evaluated.

Nevertheless, the TBI-QOL (version 1.0) consists of 20 calibrated item banks across 20 subdomains of 4 larger HRQOL domains (Physical, Emotional, Cognitive, and Social Functioning). These banks are conceptually grounded, having incorporated feedback of individuals with TBI from the ground up. All item banks have been calibrated using IRT and are available at no cost on the Assessment CenterSM study administration platform. The end user will have a great deal of flexibility in item administration, with options including presentation of the full item bank, CAT, or 8- to 10-item (depending on bank) SF. In addition, SFs may be downloaded and administered by "paper and pencil" when computerized administration is unavailable. The TBI-QOL therefore offers brief, yet precise, relevant measurement of 22 different areas of HRQOL developed specifically for individuals with TBI.

REFERENCES

1. Dijkers MP. Quality of life after traumatic brain injury: a review of research approaches and findings. *Arch Phys Med Rehabil*. 2004;85(4)(suppl 2):S21-S35. doi:10.1016/j.apmr.2003.08.119.
2. Rosenthal M, Ricker J. Traumatic brain injury. In: Frank R, Rosenthal M, Caplan B, eds. *Handbook of Rehabilitation Psychology*. 2nd ed. Washington, DC: American Psychological Association; 2007: 49-74.
3. Draper K, Ponsford J. Cognitive functioning ten years following traumatic brain injury and rehabilitation. *Neuropsychology*. 2008;22(5):618-625. doi:10.1037/0894-4105.22.5.618.
4. Whyte J, Hart T, Laborde A, Rosenthal M. Rehabilitation issues in traumatic brain injury. In: DeLisa J, Gans B, eds. *Physical Medicine and Rehabilitation: Principles and Practice*. 4th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2005:1677-1713.

5. Bottari C, Swaine B, Dutil E. Interpreting activity of daily living errors for treatment and discharge planning: the perception of occupational therapists. *J Head Trauma Rehabil.* 2007;22(1):26–30.
6. Sherer M, Madison C. Moderate and severe traumatic brain injury. In: Larabee GJ, ed. *Forensic Neuropsychology: A Scientific Approach.* New York, NY: Oxford University Press; 2005:237–270.
7. Ponsford J, Draper K, Schonberger M. Functional outcome 10 years after traumatic brain injury: its relationship with demographic, injury severity, and cognitive and emotional status. *J Int Neuropsychol Soc.* 2008;14(2):233–242. doi:http://dx.doi.org/10.1017/S1355617708080272.
8. Cella DF, Bonomi AE. Measuring quality of life: 1995 update. *Oncology (Williston Park).* 1995;9(11)(suppl):47–60.
9. Bullinger M; The TBI Consensus Group. Quality of life in patients with traumatic brain injury—basic issues, assessment and recommendations. Results of a consensus meeting. *Restor Neurol Neurosci.* 2002;20(3/4):111–124.
10. Whiteneck GG, Dijkers MP, Heinemann AW, et al. Development of the participation assessment with recombined tools-objective for use after traumatic brain injury. *Arch Phys Med Rehabil.* 2011;92(4):542–551. doi:10.1016/j.apmr.2010.08.002.
11. Willer B, Ottenbacher KJ, Coad ML. The Community Integration Questionnaire—a comparative-examination. *Am J Phys Med Rehabil.* 1994;73(2):103–111.
12. Whiteneck GG, Charlifue SW, Gerhart KA, Overholser JD, Richardson GN. Quantifying handicap: a new measure of long-term rehabilitation outcomes. *Arch Phys Med Rehabil.* 1992;73(6):519–526.
13. Johnston MV, Goverover Y, Dijkers M. Community activities and individuals' satisfaction with them: quality of life in the first year after traumatic brain injury. *Arch Phys Med Rehabil.* 2005;86(4):735–745. doi:10.1016/j.apmr.2004.10.031.
14. Malec JF, Kragness M, Evans RW, Finlay KL, Kent A, Lezak MD. Further psychometric evaluation and revision of the Mayo-Portland Adaptability Inventory in a national sample. *J Head Trauma Rehabil.* 2003;18(6):479–492.
15. Willer B, Rosenthal M, Kreutzer JS, Gordon WA, Rempel R. Assessment of community integration following rehabilitation for traumatic brain injury. *J Head Trauma Rehabil.* 1993;8(2):75–87.
16. Messick S. Validity of psychological assessment—validation of inferences from persons responses and performances as scientific inquiry into score meaning. *Am Psychol.* 1995;50(9):741–749. doi:10.1037/0003-066X.50.9.741.
17. Messick S. Test validity and the ethics of assessment. *Am Psychol.* 1980;35(11):1012–1027. doi:10.1037/0003-066X.35.11.1012.
18. Quatrano LA, Cruz TH. Future of outcomes measurement: impact on research in medical rehabilitation and neurologic populations. *Arch Phys Med Rehabil.* 2011;92(10)(suppl):S7–S11. doi:10.1016/j.apmr.2010.08.032.
19. Reeve BB, Burke LB, Chiang YP, et al. Enhancing measurement in health outcomes research supported by agencies within the US Department of Health and Human Services. *Qual Life Res.* 2007;16(suppl 1):175–186. doi:10.1007/s11136-007-9190-8.
20. Bjorner JB, Chang CH, Thissen D, Reeve BB. Developing tailored instruments: item banking and computerized adaptive assessment. *Qual Life Res.* 2007;16(suppl 1):95–108. doi:10.1007/s11136-007-9168-6.
21. Bode RK, Lai JS, Cella D, Heinemann AW. Issues in the development of an item bank. *Arch Phys Med Rehabil.* 2003;84(4):S52–S60. doi:10.1053/apmr.2003.50247.
22. 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):133–141. doi:10.1007/s11136-007-9204-6.
23. Kisala PA, Tulsy DS. Opportunities for CAT applications in medical rehabilitation: development of targeted item banks. *J Appl Meas.* 2010;11(3):315–330.
24. Thissen D, Reeve BB, Bjorner JB, Chang CH. Methodological issues for building item banks and computerized adaptive scales. *Qual Life Res.* 2007;16(suppl 1):109–119. doi:10.1007/s11136-007-9169-5.
25. Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. *Clin Exp Rheumatol.* 2005;23(5)(suppl 39):S53–S57.
26. Reeve BB, Hays RD, Chang CH, Perfetto EM. Applying item response theory to enhance health outcomes assessment. *Qual Life Res.* 2007;16:1–3. doi:10.1007/s11136-007-9220-6.
27. Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short-forms and computerized adaptive testing. *J Rheumatol.* 2009;3(9):2061–2066. doi:10.3899/jrheum.090358.
28. Lai JS, Cella D, Choi S, 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)(suppl):S20–S27. doi:10.1016/j.apmr.2010.08.033.
29. Velozo CA, Seel RT, Magasi S, Heinemann AW, Romero S. Improving measurement methods in rehabilitation: core concepts and recommendations for scale development. *Arch Phys Med Rehabil.* 2012;93(8)(suppl):S154–S163. doi:10.1016/j.apmr.2012.06.001.
30. Tulsy DS, Carlozzi NE, Cella D. Advances in outcomes measurement in rehabilitation medicine: current initiatives from the National Institutes of Health and the National Institute on Disability and Rehabilitation Research. *Arch Phys Med Rehabil.* 2011;92(10)(suppl 1):S1–S6. doi:10.1016/j.apmr.2011.07.202.
31. Cella D, Riley W, Stone A, 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):1179–1194. doi:10.1016/j.jclinepi.2010.04.011.
32. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the Patient-Reported Outcomes Measurement Information System (PROMIS) global items. *Qual Life Res.* 2009;18(7):873–880. doi:10.1007/s11136-009-9496-9.
33. DeWitt EM, Stucky BD, Thissen D, et al. Construction of the eight-item Patient-Reported Outcomes Measurement Information System pediatric physical function scales: built using item response theory. *J Clin Epidemiol.* 2011;64(7):794–804. doi:10.1016/j.jclinepi.2010.10.012.
34. Varni JW, Stucky BD, Thissen D, et al. PROMIS Pediatric Pain Interference Scale: an item response theory analysis of the pediatric pain item bank. *J Pain.* 2010;11(11):1109–1119. doi:10.1016/j.jpain.2010.02.005.
35. Yeatts KB, Stucky B, Thissen D, et al. Construction of the Pediatric Asthma Impact Scale (PAIS) for the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Asthma.* 2010;47(3):295–302. doi:10.3109/02770900903426997.
36. Cella D, Nowinski C, Peterman A, et al. The neurology quality of life measurement initiative. *Arch Phys Med Rehabil.* 2011;92(10)(suppl):S28–S36. doi:10.1016/j.apmr.2011.01.025.
37. Gershon RC, Lai JS, Bode R, 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):475–486. doi:10.1007/s11136-011-9958-8.
38. Carlozzi NE, Tulsy DS, Kisala PA. Traumatic brain injury patient-reported outcome measure: identification of health-related

- quality of life issues relevant to individuals with traumatic brain injury. *Arch Phys Med Rehabil*. 2011;92(10)(suppl):S52–S60. doi:10.1016/j.apmr.2010.12.046.
39. PROMIS. PROMIS instrument development and psychometric evaluation scientific standards. http://www.nihpromis.org/Documents/PROMIS_Standards_050212.pdf. Published 2012. Accessed December 27, 2012.
40. Eremenco SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval Health Prof*. 2005;28(2):212–232. doi:10.1177/0163278705275342.
41. Samejima F, van der Liden W, Hambleton R. The graded response model. In: *Handbook of Modern Item Response Theory*. New York, NY: Springer; 1996:85–100.
42. Gershon R, Rothrock NE, Hanrahan RT, Jansky LJ, Harniss M, Riley W. The development of a clinical outcomes survey research application: Assessment Center. *Qual Life Res*. 2010;19(5):677–685. doi:10.1007/s11136-010-9634-4.
43. Reise SP, Yu J. Parameter recovery in the graded response model using MULTILOG. *J Educ Meas*. 1990;27(2):133–144. doi:10.1111/j.1745-3984.1990.tb00738.x.
44. Williams DH, Levin HS, Eisenberg HM. Mild head injury classification. *Neurosurgery*. 1990;27(3):422–428.
45. Steinberg L, Thissen D. Use of item response theory and the testlet concept in the measurement of psychopathology. *Psychol Methods*. 1996;1(1):81–97. doi:10.1037/1082-989X.1.1.81.
46. Orlando M, Thissen D. Further investigation of the performance of $S-X^2$: An item fit index for use with dichotomous item response theory models. *Appl Psychol Meas*. 2003;27:289–298. doi:10.1177/0146621603027004004.
47. Choi SW, Gibbons LE, Crane PK. lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *J Stat Softw*. 2011;39(8):1–30.
48. Stocking ML, FM L. Developing a common metric in item response theory. *Appl Psychol Meas*. 1983;7(2):201–210. doi:10.1177/014662168300700208.
49. Liu H, Cella D, Gershon R, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. *J Clin Epidemiol*. 2010;63(11):1169–1178. doi:10.1016/j.jclinepi.2009.11.021.