Arch Phys Med Rehabil. Author manuscript; available in PMC 2014 September 01.

Published in final edited form as:
Arch Phys Med Rehabil. 2013 September ; 94(9): 1670–1678. doi:10.1016/j.apmr.2013.03.013.

Development of an Instrument to Measure Behavioral Health Function for Work Disability: Item Pool Construction and Factor Analysis

Elizabeth E. Marfeo, PhD, MPH[1], **Pengsheng Ni, MD, MPH**[1], **Stephen M. Haley, PT, PhD**[1,*], **Alan M. Jette, PT, PhD**[1], **Kara Bogusz, BA**[1], **Mark Meterko, PhD**[1], **Christine M. McDonough, PT, PhD**[1,3], **Leighton Chan, MD, MPH**[2], **Diane E. Brandt, PT, MS, PhD**[2], and **Elizabeth K. Rasch, PT, PhD**[2]

[1]Boston University School of Public Health; Health & Disability Research Institute 715 Albany St., T5W Boston, MA 02118-2526

[2]National Institutes of Health, Mark O. Hatfield Clinical Research Center; Rehabilitation Medicine Department 6100 Executive Boulevard, Suite 3C01, MSC 7515 Bethesda, MD 20892-7515

[3]The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine, Lebanon, NH 03766

Abstract

Objectives—To develop a broad set of claimant-reported items to assess behavioral health functioning relevant to the Social Security disability determination processes, and to evaluate the underlying structure of behavioral health functioning for use in development of a new functional assessment instrument.

Design—Cross-sectional.

Setting—Community.

Participants—Item pools of behavioral health functioning were developed, refined, and field-tested in a sample of persons applying for Social Security disability benefits (N=1015) who reported difficulties working due to mental or both mental and physical conditions.

Interventions—None.

Main Outcome Measure—Social Security Administration Behavioral Health (SSA-BH) measurement instrument

Corresponding Author and Reprints: Elizabeth E. Marfeo, PhD, MPH, OTR/L, Boston University, School of Public Health, Health & Disability Research Institute, 715 Albany St., T5W, Boston, MA 02118-2526, T 617-638-1990 F 617-638-1999 emarfeo@bu.edu.
*Deceased

Work was performed at Boston University School of Public Health, Health and Disability Research Institute

There are no conflicts of interest to disclose of the author and co-authors

Results—Confirmatory factor analysis (CFA) specified that a 4-factor model (self-efficacy, mood and emotions, behavioral control, and social interactions) had the optimal fit with the data and was also consistent with our hypothesized conceptual framework for characterizing behavioral health functioning. When the items within each of the four scales were tested in CFA, the fit statistics indicated adequate support for characterizing behavioral health as a unidimensional construct along these four distinct scales of function.

Conclusion—This work represents a significant advance both conceptually and psychometrically in assessment methodologies for work related behavioral health. The measurement of behavioral health functioning relevant to the context of work requires the assessment of multiple dimensions of behavioral health functioning. Specifically, we identified a 4-factor model solution that represented key domains of work related behavioral health functioning. These results guided the development and scale formation of a new SSA-BH instrument.

Keywords

Work disability; Behavioral health; Outcome assessment (health care); Psychometrics

---

In 2011, mental health impairments represented one of the largest categories of disabling conditions for which individuals receive Social Security Administration's Disability Insurance (SSDI) benefits.[1] Mental or behavioral health related work disability encompasses various factors beyond a singular disease including social, cultural, and environmental considerations.[2–5] Due to its multifactorial nature, behavioral health related work disability is one of the more challenging areas to assess. Additionally, the specific mechanisms whereby mental health impairments of varying types and degrees actually affect a person's ability to work are complex and poorly understood.[6, 7]

Given the complexity of behavioral health related work disability, one would expect many aspects of behavioral health to affect an individual's potential ability to work. Now, with the advancement in instrument development utilizing modern psychometric methods such as item response theory (IRT) and computer adaptive testing (CAT), new health status measures can be developed to efficiently assess multifactorial scales of health. By utilizing an assessment methodology that allows for complex factor solutions, the characterization of a person's underlying functional abilities may be represented with high degrees of breadth and precision.[8, 9]

Considering the steady annual increases in the number of disability applications along with other challenges to the disability determination process, there is a particular interest in applying new assessment methods within the context of the SSDI and SSI programs.[10] The Social Security Administration (SSA) currently defines disability with language that is heavily grounded on the narrowly defined medical model.[11–13] A broader conceptualization of behavioral health related disability would align more closely with contemporary notions of disability as articulated in the World Health Organization's International Classification of Functioning, Disability, and Health (ICF).[10, 14] This taxonomy highlights the interactive nature of disability and specifies components of whole person function that are key factors in characterizing a person's overall health and ability to function.[15]

Current behavioral health assessments typically target specific clinical populations and focus on characterizing symptom severity rather than measuring a broader concept of behavioral health function as would be relevant for a more heterogeneous population such as applicants for SSA disability benefits.[16, 17] Additionally, recent work in the area of patient reported outcomes measures have produced more general health assessment tools. The Patient Reported Outcomes Measurement Instrument System (PROMIS) and Quality of Life Outcomes in Neurological Disorders (Neuro-Qol) assessments offer the flexibility of general health assessment and include some components of mental health, but these instruments were not specifically designed to assess behavioral health in the context of work.[18–22]

In order to enhance the current SSA disability evaluation processes, reliable, valid, and feasible tools are needed to comprehensively assess work related behavioral health outcomes in the SSA claimant population. One method to achieve this goal is to utilize modern psychometric techniques of item response theory (IRT) and computer adaptive testing (CAT).[8] IRT combined with CAT implementation allows questions to be administered by responses to a given question guiding the selection of the next question.[23] This iterative process allows multiple dimensions of health to be assessed in a comprehensive, efficient, and precise way.[24]

The aims of this study were to test a set of newly developed items that characterize aspects of behavioral health using the SSA Behavioral Health measurement instrument (SSA-BH). Specifically, our goal was to evaluate the robustness of utilizing a multifactorial model for use in assessing work related behavioral health functioning in a SSA claimant population as compared to a 1- or 2- factor model of behavioral health functioning. The current work lays the groundwork for item response theory (IRT) analysis and the development of a new behavioral health measurement instrument implemented through computer adaptive testing (CAT), as described in a complementary article.[25]

METHODS

Behavioral Control Item Pool Development

The SSA-BH item pool examined in this study was grounded on a theoretical framework that characterizes behavioral health functioning across five hypothesized domains: Behavioral Control, Temperament & Personality, Adaptability, Basic Interactions, and Workplace Behaviors.[26] The initial content model served as the conceptual foundation for constructing items and was developed using a structured literature review, expert interviews, and stakeholder feedback. The item pool consisted of 165 questions that used a four-point agreement response scale ranging from “Strongly Agree” to “Strongly Disagree” with an opt-out option of “I don’t know.” In terms of content, questions asked about various behaviors, thoughts, or feelings that a person may experience in the context of work or their daily environment. See Table 1 for sample item content within each domain. A systematic and iterative process of literature review, cognitive testing, and extensive content expert feedback was utilized to ensure content validity of the item pools (See Figure 1). The resulting item pool included three types of items: (1) core items relevant to all claimants, (2) a set of legacy items from previously validated instruments (PROMIS and NeuroQol) which may allow for future crosswalks between the SSA-BH instrument and the legacy instruments, and (3) a work module for people who were more recently (within 6 months) out of work.

Calibration Study

Participants—The study included two cross-sectional samples—a SSA claimant sample and a comparative US adult sample. The claimant sample was composed of adults who filed a disability application through the Social Security disability programs within a 3-month time period in 2011. Claimants had to be 21 years of age or older, able to speak, read, and understand English, and filed the claim him/herself rather than through a proxy. Additionally, claimants must have provided the name and contact information for at least one medical provider familiar with health condition for which they were seeking SSA benefits. Claimants who filed for conditions including paranoia, psychosis, autism, intellectual disability, or Down's Syndrome were excluded. A random sample of claimants who met these criteria was selected, geographically stratified across all 10 SSA regions, including consideration of urban/rural distributions within each region.

In order to compute norm-based scoring of the SSA-BH instrument, a normative sample of 1000 US adults was included. A unique sampling procedure developed by YouGov was used to obtain the normative sample whereby a sample of US adults was drawn from a large opt-in internet panel.[27] Proximity sample-matching methods were used to obtain a target normative population approximating the US adult population matched on gender, racial/ethnic background, age, and education, and weighted equally. All variables and measures of behavioral health were developed using methods that were identical to those used for the claimant sample. A university institutional review board approved this study and all sample participants provided informed consent.

Data Collection Procedures—The full item pool was administered to the SSA claimant sample via phone or internet by Westat research personnel. All study personnel completed survey administration training prior to conducting data collection activities. Periodic monitoring of the data collection procedures was performed for data quality control purposes. In addition to the 165 items regarding behavioral health, the survey included self-reported demographic variables [age, gender, marital status, race, ethnicity, education, and urban/rural geographic location] and variables characterizing the nature of the disabling condition [mental condition only vs. both mental and physical health conditions]. Responses to a screener question were used to select administration of the workplace module (persons who had been out of work 6 months or less). Participants were asked to respond to all survey questions based on a typical day, with the exception of a subset of 57 items that specifically referenced "the past seven days." Data were collected from the normative sample on the same 165 items via an opt-in web survey format.

Data analysis—Descriptive statistics were calculated to assess missing data, examine frequency distributions for each item response category, and characterize demographics of the claimant and normative samples. "I don't know" responses were considered missing data and analyzed separately. Responses were re-coded as necessary based on the phrasing of the question so that all items with a higher numeric score were indicative of higher functioning.

To establish construct validity, the dimensionality of the data were examined using both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) techniques. An initial EFA was performed using a geomin rotation to examine eigenvalues and cumulative percentage of variance in behavioral health functioning explained by different numbers of factors. Factor loading patterns were examined to characterize the latent dimension represented by the items that loaded on each factor. For this purpose, individual items with factor loadings less than 0.40 were considered weak and removed from further analysis as were items with cross-loadings greater than 0.30 on two or more different factors. In addition to these statistical criteria, the item removal process involved consideration of each item's meaning and context prior to final removal.

A process of CFA was used for developing and testing several alternative model structures of behavioral health functioning. To develop the most robust yet parsimonious factor structure, a series of models was developed beginning from a simple 1-factor model and advancing to a more complex 5-factor model. Each model was tested using multiple goodness-of-fit indices. The Tucker-Lewis Index (TLI) and Bentler's comparative fit index (CFI) are the incremental relative fit indices, which assess the difference between a proposed model and the null model (model specifying that all measured variables are uncorrelated), making adjustments for the number of degrees of freedom in the model.[28, 29] The chi-square statistics for the null and target (i.e. hypothesized) models divided by their respective degrees of freedom ($\chi 2$/df), and the differences between the chi-square statistics and the degrees of freedom for both models were used to calculate TLI and CFI respectively. For both TLI and CFI, values greater than .90 indicate good model fit, with those above .95 indicating extremely good model fit.[29–31] In addition, the root mean square error of approximation (RMSEA) was calculated. The RMSEA is a parsimony corrected index of model fit and estimates the difference between the lack of fit in the proposed model and a perfect or saturated model, accounting for degrees of freedom.[32] For the RMSEA, values of .06 or lower indicate a very close fit; values greater than .08 indicate reasonable model fit; and,; values greater than or equal to 0.10 indicate poor model fit.[30, 31, 33]

Additionally, the CFA model testing included consideration of four criteria: (1) each factor should have a sufficient number of items to represent the measured construct, (2) the factor must demonstrate face validity, in that the factor loading patterns and item content matched hypothesized model constructs, (3) item loadings must be greater than 0.40 and load on only one factor, (4) reasonable discrimination between factors was indicated by correlations between factors less than 0.80. The next step in the CFA model testing was analysis of unidimensionality of each CFA supported factors specifying the same fit statistics criteria as described above for model evaluation.

In preparation for subsequent IRT analysis, local independence assumptions were tested in each of the factor item banks. This is to ensure that responses among items are independent for individuals with the same level of the hypothesized construct. If local independence is not met, then the measure will be violating the assumption of unidimensionality. To test for local item dependence, residual correlations between items were examined and items with residual inter-item correlation greater than 0.20 were removed.[20]

Content expert feedback guided empirical analysis regarding the important subdomains of behavioral health functioning in order to yield a final model that would be supported both conceptually and empirically. To confirm the resultant model structure and to provide an initial demonstration of comparative group validity of the models, the same series of analyses were performed in the normative sample as was performed in the claimant data analysis. All CFAs were conducted using MPlus software; descriptive statistics were performed using SAS software; and, data transformations were performed using SPSS software.[34–36]

RESULTS

Demographic data from the calibration study characterizing the SSA claimant and normative samples are presented in Table 2. The SSA claimant sample was approximately 56% female, predominantly (61%) white, 73% from urban areas, and average age of 44±11 years. With regard to the work-related disability that was motivating their application for benefits, 20% of the SSA claimant sample reported having a mental condition only (e.g. depression, anxiety, bipolar, PTSD) and 80% reported both mental and physical conditions (e.g. depression/anxiety with back pain, arthritis, fibromyalgia, etc.). The normative sample was approximately 52% male and 77% white; average age was 49 ±15 years. Missing data were sparse; therefore, no imputation of data was performed. “I don’t know” responses were endorsed infrequently in both samples (Normative sample: average % missing =2.21%, SD 2.56%, range 0%–14.9%; Claimant sample: average % missing = 1.42%, SD 1.34%, range 0%–7.68%).

Results from the initial exploratory analysis supported models ranging from 2 to 5- factors. Parallel analysis was performed as a secondary EFA technique. Results from the parallel analysis showed there could be up to 8 factors within the behavioral health domain, but factors above 5 have few to no items loading in these higher level factors. Eigenvalues and the cumulative percentage of variance explained from each model are: 2-factor model, 8.95, 38.7%; 3-factor model, 6.8, 43.9%; 4-factor model, 3.55, 46.6%; and 5-factor model, 3.22, 49.1% respectively. Table 3 shows that the EFA fit statistics improve with the number of factors, but the content and number of items in the higher order factors are limited.

Prior to CFA analysis individual item quality was considered: 24 items were deleted due to item-to-total correlations less than 0.3; 46 items were removed with cross-factor loading greater than 0.4 on two or more factors; and 12 work module items were removed due to missing data (87% of the sample did not answer this item set). Table 4 outlines the sequence of models tested by CFA ranging from the simple 1-factor solution to the most complex 5-factor solution. The models specified in the 1-, 2-, and 3- factor solutions yielded fit statistics that were unacceptable statistically. In addition, content expert input confirmed that the proposed models did not meet conceptual criteria consistent with the hypothesized framework.

The two model solutions that emerged as optimal both psychometrically and conceptually were the 4- and 5-factor model solutions. These models yielded adequate CFI, TLI, RMSEA statistics (4-Factor solution: CFI=0.90, TLI=0.897, RMSEA=0.055; 5-factor solution:

CFI=0.905, TLI=0.901, RMSEA=0.058). Two additional model analyses were conducted with the goal of maximizing the number of items in each scale while preserving the legacy items that were not included in the initial CFA results. These modified 4- and 5-factor CFA results yielded similarly acceptable goodness of fit statistics (4-Factor solution: CFI=0.882, TLI=0.879, RMSEA=0.056; 5-Factor solution: CFI=0.885, TLI=0.882, RMSEA=0.057).

Table 5 presents results from the final round of CFAs conducted to confirm the unidimensionality of the modified 4-factor solution, which we favored on grounds of parsimony and conceptual clarity. The unidimensionality check suggested the removal of up to 2 items in each of the three factors to meet psychometric criteria. With removal of these few items, all of the items demonstrate good fit within each factor. Further, the CFI and TLI values meet or exceeded the 0.90 threshold for all factors, with the social interactions factor demonstrating exceptionally good fit statistics (CFI=0.974; TLI=0.956). The final model solution, comprised of 78 items representing four unidimensional factors of behavioral health functioning is presented in Figure 2.

To further verify the final structure of the model, all CFA analyses were replicated in a normative sample. The four-dimension model structure was supported in the normative sample and similarly demonstrated adequate goodness of fit indices as presented in Table 6. The combination of results from the claimant and normative samples provide robust justification for proceeding to develop calibrated item banks to represent (1) mood and emotions, (2) social interactions, (3) behavioral control, and (4) self-efficacy factors of work related behavioral health function.

DISCUSSION

Empirical evidence, conceptual elements, and utility of the instrument were all important components of the model refinement process. We consulted an expert panel to impart conceptual meaning and clinical relevance to the items specified in each solution. In addition, the goal of maintaining the ability to crosswalk with legacy instruments was a priority. With feedback from the experts, we reached consensus on the modified 4-factor model representing the solution that aligned best with the conceptual framework while demonstrating parsimony and acceptable goodness of fit statistics.

The resultant model represents a measure that characterizes four domains of behavioral health function relevant to the context of work. These four factors are presented as separate item banks and demonstrate strong unidimensionality as four distinct scales of work related behavioral health. Robustness of the 4-factor solution was verified in the normative sample. Results from the normative replication represent an important step in initial demonstration of the SSA-BH instrument being able to characterize distinct factors of behavioral health functioning important in the work context within various adult populations. The overall model solution fit statistics suggests that retaining these factors as separate scales would be an appropriate and useful characterization of behavioral health, allowing functional profiles to be developed along four dimensions of work related behavioral health.

Development of the SSA-BH instrument system represents a significant opportunity to enhance current disability evaluation processes. The content of this instrument expands current disability assessment to include aspects of whole person behavioral health that are relevant in the context of work. A unique feature of this assessment development process was the extensive involvement of key stakeholders including claimants, content experts, and representatives from SSA to ensure that the content of the items was both relevant and comprehensive. To date, the SSA-BH instrument marks a significant contribution to the assessment of behavioral health functioning in that it comprehensively measures four distinct, unidimensional dimensions: mood and emotions, social interactions, self-efficacy, and behavioral control.

Our goal was to test a set of newly developed items that reflect a whole person perspective of behavioral health functioning relevant to the context of work. Further, by using both items that were developed to be tailored to the target population and legacy items from PROMIS and NeuroQol measures, scores derived from the SSA-BH have the potential to be linked to those previously validated assessments.[37] The benefit of this potential score linking method may facilitate comparisons between different behavioral health function in general adult populations and SSA claimant samples.

Study limitations

In conceptualizing behavioral health function relevant to work, we specifically utilized a broad conceptualization of behavioral health function rather targeting a narrow construct. This is unique compared to the approach existing measures often utilize, which often focus on measuring a single component of behavioral health such as depression, anxiety, or anger in isolation. Because of this broad conceptualization, the overall variance explained by initial hypothesized model reported in the EFA results is lower compared to measures of more narrowly defined constructs. The resultant 4-factor model solution is based on a balancing both conceptual and empirical criteria for new measurement development and requires future validation through further field-testing.[26] Although the factor structure was validated in the normative sample, future work should examine the model robustness in other populations and its ability to discriminate between known groups along these four dimensions of behavioral health. An initial study of discriminate validity between the normative and claimant samples was performed and is discussed in a complementary article[25]. Comparison of behavioral health functioning as assessed using the 4-factor model solution with existing SSA disability determination processes is essential to establish the utility and validity of the newly developed instruments.

Further, the value of the CFI and TLI fit statistics were right on the criterion value of 0.90, optimally they would have exceeded that threshold. Lastly, due to various threshold criteria (item-to-total correlation, cross-loadings, and missing data) the final item set yields a smaller number of items than desired. Of particular note is the Social Interactions factor. The resultant item count retained in this factor was six items, and the fit was marginal with RMSEA = 0.101. This limitation offers an opportunity for future work to improve the SSA-BH measure through item replenishment processes. Item replenishment will allow for improving this particular factor by increasing the number of items retained in the factor.

This should strengthen the model both conceptually, by increasing the content coverage of the social interactions factor and improve the overall model fit statistics. One benefit of utilizing the IRT and CAT based assessment is the flexibility in modifying assessments to improve measures both in terms of item content as well as psychometric properties.[38]

CONCLUSIONS

The development of the SSA-BH instrument represents a groundbreaking effort in moving SSDI assessment processes forward both conceptually and psychometrically. The items utilized in the SSA-BH instrument offer an opportunity to assess functioning from a “whole person” perspective as compared to traditional assessment methods based upon the medical model of disability. This conceptual shift provides a unique opportunity to assess behavioral health functioning across a variety of key domains that may be important in predicting a person’s ability to work. Study methods including extensive stakeholder, expert feedback and advanced psychometric methodologies were utilized in an effort to ensure the utility of the SSA-BH instrument both conceptually and empirically. These results support a 4-factor model of behavioral health functioning relevant for work including domains characterizing aspects of self-efficacy, mood and emotions, behavioral control, and social interactions. These analyses and results provide the groundwork for finalizing the SSA-BH instrument using IRT methods to calibrate the items representing these four behavioral health dimensions.

Acknowledgments

Funding for this project was provided through SSA-NIH Interagency Agreements under NIH Contract # HHSN269200900004C, NIH Contract # HHSN269201000011C, and NIH Contract # HHSN269201100009I and through the NIH intramural research program.

Abbreviations

CAT	Computer Adaptive Testing
CFA	Confirmatory Factor Analysis
CFI	Bentler’s Comparative Fit Index
EFA	Exploratory Factor Analysis
ICF	International Classification of Functioning, Disability and Health
IRT	Item Response Theory
Neuro-QoL	Quality of Life Outcomes in Neurological Disorders
PROMIS	Patient Reported Outcomes Measurement Instrument System
RMSEA	Root Mean Square Error of Approximation
SSA	Social Security Administration
SSA-BH	Social Security Administration; Behavioral Health measurement instrument
SSDI	Social Security Disability Insurance

TLI	Tucker-Lewis Index

References

1. Social Security Administration. Annual Statistical Report on the Social Security Disability Insurance Program, 2011. Washington, DC: Social Security Administration; 2012.
2. Cieza A, Bickenbach J, Chatterji S. The ICF as a conceptual platform to specify and discuss health and health-related concepts. Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)). 2008; 70:e47–56. [PubMed: 18932116]
3. Escorpizo R, Reneman MF, Ekholm J, et al. A Conceptual Definition of Vocational Rehabilitation Based on the ICF: Building a Shared Global Model. J Occup Rehabil. 2011
4. Giannangelo K, Bowman S, Dougherty M, Fenton S. ICF: representing the patient beyond a medical classification of diagnoses. Perspectives in health information management/AHIMA, American Health Information Management Association. 2005; 2:7.
5. Anthony WA, Rogers ES, Cohen M, Davies RR. Relationships between psychiatric symptomatology, work skills, and future vocational performance. Psychiatr Serv. 1995; 46:353–358. [PubMed: 7788456]
6. Tsang H, Lam P. Predictors of Employment Outcome for People with Psychatric Disabilities: A Review of the Literature Since the Mid ‘80s. Journal of Rehabilitation. 2000; 66:19–31.
7. MacDonald-Wilson K, Rogers ES, Anthony WA. Unique issues in assessing work function among individuals with psychiatric disabilities. J Occup Rehabil. 2001; 11:217–232. [PubMed: 11822197]
8. Reeve, B. [Accessed 9/6/2012.] Applications of Item Response Theory (IRT) Modeling for Building and Evaluating Questionnaires Measuring Patient-Reported Outcomes. 2004. Available from: http://outcomes.cancer.gov/conference/irt/reeve.pdf
9. Tulsky DS, Jette AM, Kisala PA, et al. Spinal cord injury-functional index: item banks to measure physical functioning in individuals with spinal cord injury. Arch Phys Med Rehabil. 2012; 93:1722–1732. [PubMed: 22609299]
10. Brandt DE, Houtenville AJ, Huynh M, Chan L, Rasch EK. Connecting Contemporary Paradigms to Social Security Administration’s Disability Evaluation Process. Journal of Disability Policy Studies. 2011; 20:1–13.
11. Social Security Administration. Social Security Administration: Titles II and XVI: Residual Functional Capacity for Mental Impairments. May 30.2001b 2001 20 CFR 402.1545(c); SSR 85–16; 20 CFR, Appendix 1, Part A, Section 12.00, et seq.
12. Social Security Administration. Administrative review process for adjudicating initial disability claims. Final rule. Fed Regist. 2006; 71:16423–16462. [PubMed: 16578949]
13. Social Security Administration. Annual Statistical Report on the Social Security Disability Insurance Program, 2010. 2011. SSA Publication No. 13–11827. Available from: http://www.ssa.gov/policy/docs/statcomps/di_asr/2010/di_asr10.pdf
14. Sanderson K, Nicholson J, Graves N, Tilse E, Oldenburg B. Mental health in the workplace: using the ICF to model the prospective associations between symptoms, activities, participation and environmental factors. Disability and rehabilitation. 2008; 30:1289–97. [PubMed: 18608382]
15. Bufka LF, Stewart D, Stark S. The International Classification of Functioning, Disability and Health (ICF) is emerging as the principal framework for the description of health and health related status. Canadian journal of occupational therapy. 2008; 75:134.
16. Reed GM, Leonardi M, Ayuso-Mateos JL, et al. Implementing the ICF in a psychiatric rehabilitation setting for people with serious mental illness in the Lombardy region of Italy. Disability and rehabilitation. 2009; 31 (Suppl 1):S170–3. [PubMed: 19852707]
17. Sanderson K, Andrews G. Prevalence and severity of mental health-related disability and relationship to diagnosis. Psychiatr Serv. 2002; 53:80–86. [PubMed: 11773654]
18. Cella D, Lai JS, Nowinski CJ, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. Neurology. 2012; 78:1860–1867. [PubMed: 22573626]

19. 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; 63:1179–1194. [PubMed: 20685078]
20. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for Patient-Reported Outcomes Measurement Information System (PROMIS). Medical care. 2007; 45:S22–S31. [PubMed: 17443115]
21. Cella, D. Quality of Life Outcomes in Neurological Disorders (NeuroQOL). Northwestern University Medical Center: NINDS/NIH; 2006.
22. Cella D, Young S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap Cooperative Group during its first two years. Medical Care. 2007; 45(5 Suppl):S3–S11. [PubMed: 17443116]
23. Fries JF, Bruce B, Bjorner J, Rose M. More relevant, precise, and efficient items for assessment of physical function and disability: moving beyond the classic instruments. Ann Rheum Dis. 2006; 65(Suppl 3):iii16–21. [PubMed: 17038464]
24. Haley SM, Ni P, Hambleton RK, Slavin MD, Jette AM. Computer adaptive testing improved accuracy and precision of scores over random item selection in a physical functioning item bank. J Clin Epidemiol. 2006; 59:1174–1182. [PubMed: 17027428]
25. Marfeo EE, Pengsheng Ni, Haley SM, et al. Scale Refinement and Initial Evaluation of a Behavioral Health Function Measurement Tool for Work Disability Evaluation. Archives of Physical Medicine and Rehabilitation. XXXX; XX: XX–XXX.
26. Marfeo EE, Haley SM, Jette AM, et al. A Conceptual Foundation for Measures of Physical Function and Behavioral Health Function for Social Security Work Disability Evaluation. Archives of Physical Medicine and Rehabilitation. XXXX; XX: XX–XXX.
27. Rivers, D. A white paper on the advantages of the sample matching methodology. YouGovPolymetrix; Sample Matching: Representative Sampling from Internet Panels.
28. Tucker L, Lewis C. A reliability coefficient for maximum likelihood factor analysis. Psychometrika. 1973; 38:1–10.
29. Bentler P. Comparative fit indices in structural models. Psychological Bulletin. 1990; 107:238–246. [PubMed: 2320703]
30. Hu L, Benter PM. Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling. 1999; 6
31. Hu, L.; Bentler, PM. Evaluating model fit. In: Hoyle, RH., editor. Structural Equation Modeling: Concepts, Issues and Applicaitons. Thousand Oaks, CA: Sage Publications; 1995. p. 76-79.
32. Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. Testing Structural Equation Models. Newbury Park, CA: Sage Publications; 1993. p. 136-172.
33. Steiger JH. Structural model evaluation and modification: an interval estimation approach. Multivariate Behavioral Research. 1990; 25:173–180.
34. Muthen, B.; Muthen, L. Mplus Statistical Analysis with Latent Variables User's Guide. Los Angeles, CA: Muthen & Muthen; 2007.
35. SPSS. SPSS Base 15.0 user's guide. Chicago: 2006.
36. SAS Institute. SAS 9.1.3. Cary, NC: 2004.
37. Noonan VK, Cook KF, Bamer AM, Choi SW, Kim J, Amtmann D. Measuring fatigue in persons with multiple sclerosis: creating a crosswalk between the Modified Fatigue Impact Scale and the PROMIS Fatigue Short Form. Qual Life Res. 2012; 21:1123–1133. [PubMed: 22048931]
38. Haley SM, Ni P, Jette AM, et al. Replenishing a computerized adaptive test of patient-reported daily activity functioning. Qual Life Res. 2009; 18:461–471. [PubMed: 19288222]

NIH-PA Author Manuscript
NIH-PA Author Manuscript

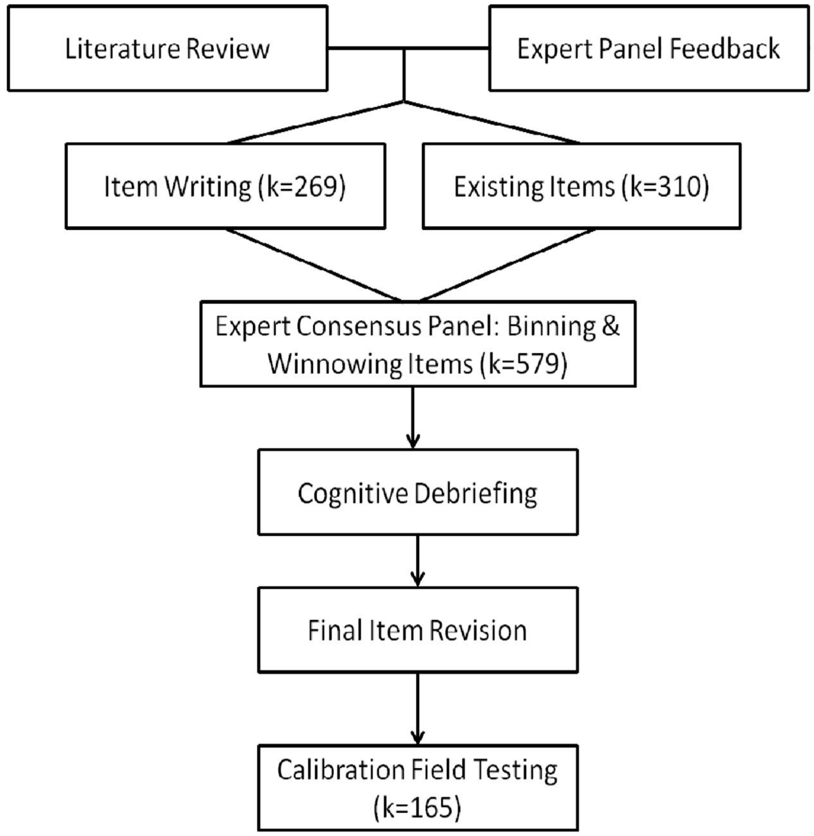
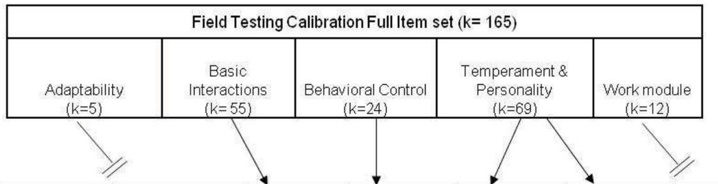


Figure 1.
Behavioral Health Function Item Pool Development Process



		Social Interactions	Behavioral Control	Mood and Emotions	Self-Efficacy	Total
CFA: Confirmed Model	#Agreement items	1	11	6	25	43
	#Frequency items (Promis/NeuroQol)	7	24	9	0	40
		8	35	15	25	83

		Social Interactions	Behavioral Control	Mood and Emotions	Self-Efficacy	Total
CFA: Confirmed Unidimensionality	#Agreement items	1	6	11	23	41
	#Frequency items (Promis/NeuroQol)	5	9	23	0	37
		6	15	34	23	78

Figure 2.

Initial Behavioral Health Function Framework and Final Model

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Sample Item Content

Domain	Item Content
Social Interactions	Please specify your level of agreement: I can keep up with my social commitments. Hint: Social commitments meaning plans you've made with others
Mood & Emotions	Please specify your level of agreement: I am so tired when I wake up, it's hard to get going.
Self-Efficacy	Please specify your level of agreement: I am comfortable making eye contact with others.
Behavioral Control	Please specify your level of agreement: I sometimes get physical when I'm angry.

Table 2

Background Characteristics of the Sample

Variable	Study Claimants (N=1015) Mean ± SD or n (%)	Normative Sample (N=1000) Mean ± SD or n (%)	
Age*	43.76 ± 11.09	49.07± 15.48	T(1791)=−8.8, $p<0.0001$
Under 40	341 (33.63)	264 (26.64)	
40–55	499 (49.21)	314 (31.69)	
55+	174 (17.16)	413 (41.68)	$X^2(2)=149.96$, $p<0.0001$
Sex*			
Female	571 (56.26)	484 (48.50)	
Male	444 (43.74)	514 (51.50)	$X^2(1)=12.15$, $p=0.0005$
Geography			
Urban	744 (73.30)	-	
Rural	271 (26.70)	-	
Race			
White	617 (60.79)	773 (77.30)	
Black/African American	266 (26.21)	105 (10.50)	
Other	111 (10.94)	104 (10.40)	$X^2(2)=87.53$, $p<0.0001$
missing	21 (2.07)	18 (1.80)	
Marital Status			
Never married	301 (29.66)	206 (20.60)	
Married/partner	352 (34.68)	627 (62.70)	
Divorced/separated	326 (32.12)	111 (11.10)	
Widowed	31 (3.05)	48 (4.80)	$X^2(3)=204.34$, $p<0.0001$
missing	5 (0.49)	8 (0.80)	
Education			
Less than high school	238 (23.52)	44 (4.40)	
High School/GED	361 (35.67)	361 (36.10)	
Greater than high school	413 (40.81)	591 (59.10)	$X^2(2)=164.9$, $p<0.0001$
missing	3	4 (0.40)	
Primary Complaint			
Mental	208 (20.49)	-	
Both Mental & Physical	807 (73.51)	-	
missing	0	-	

Age* Claimant sample (N=1014), Normative sample (N=991)

Sex** Claimant sample (N=998)

Table 3

Exploratory Factor Analysis Fit Statistics

	CFI	TLI	RMSEA
1-factor model	0.767	0.762	0.057
2-factor model	0.857	0.851	0.045
3-factor model	0.921	0.917	0.034
4-factor model	0.934	0.929	0.031
5-factor model	0.947	0.942	0.028
6-factor model	0.954	0.949	0.027

Table 4

Model Specifications for Confirmatory Factor Analysis

Factor	Domains
1	Behavioral Health
2	Mood and Emotions; Confidence/Conscientiousness
3	Mood and Emotions; Self-Efficacy; Behavioral Control
4	Mood and Emotions; Self-Efficacy; Behavioral Control; Social Interactions
5	Mood and Emotions; Self-Efficacy; Behavioral Control; Social Interactions; Trust

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 5

Goodness-of-Fit Indices for Unidimensional CFAs for the Behavioral Health Function Factors

Scale	# of items	Chi-square/df	CFI	TLI	RMSEA	# of Removed Items
Self-Efficacy	25	2508.733/275	0.889	0.879	0.089	-
	23	1863.338/230	0.910	0.90	0.084	2
Mood and Emotions	35	5829.077/560	0.891	0.884	0.096	-
	34	4895.457/527	0.908	0.902	0.09	1
Behavioral Control	15	930.213/90	0.943	0.934	0.096	
Social Interactions	8	943.114/20	0.909	0.873	0.213	-
	6	102.788/9	0.974	0.956	0.101	2

Table 6

Comparison of CFA results for Normative Sample (N=1,000) with the SSA Claimant Sample (N=1015)

4-Factor Model (78 items)			
Sample	**CFI**	**TLI**	**RMSEA**
Claimant	0.882	0.879	0.056
Normative	0.872	0.868	0.056

A1. CFA Factor Loading Table—Behavioral Health Function