Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Arch Phys Med Rehabil. Author manuscript; available in PMC 2016 June 01.

Published in final edited form as:

Arch Phys Med Rehabil. 2015 June ; 96(6): 1028–1035. doi:10.1016/j.apmr.2014.11.025.

The Work Disability Functional Assessment Battery (WD-FAB): Feasibility and Psychometric Properties

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

[1]Boston University School of Public Health, Health & Disability Research Institute, Boston, MA

[2]VA Boston Healthcare System, Center for Healthcare Organization and Implementation Research, Boston, MA

[3]National Institutes of Health, Rehabilitation Medicine Department, Mark O. Hatfield Clinical Research Center, Bethesda, MD

Abstract

Objectives—To assess the feasibility and psychometric properties of eight scales covering two domains of the newly developed Work Disability Functional Assessment Battery (WD-FAB): physical function (PF) and behavioral health (BH) function.

Design—Cross-sectional.

Setting—Community.

Participants—Adults unable to work due to a physical (n=497) or mental (n=476) disability.

Interventions—None.

Main Outcome Measures—Each disability group responded to a survey consisting of the relevant WD-FAB scales and existing measures of established validity. The WD-FAB scales were evaluated with regard to data quality (score distribution; percent “I don’t know” responses), efficiency of administration (number of items required to achieve reliability criterion; time required to complete the scale) by computerized adaptive testing (CAT), and measurement accuracy as tested by person fit. Construct validity was assessed by examining both convergent and discriminant correlations between the WD-FAB scales and scores on same-domain and cross-domain established measures.

Corresponding Author: Mark Meterko, PhD, Health and Disability Research Institute, Boston University School of Public Health, 715 Albany Street T5W, Boston, MA 02118, Meterkom@bu.edu, Phone: 617-638-1994, Fax: 617-638-1999.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosure of Conflicts of Interest

No commercial party having a direct financial interest in the results of the research described in this article has or will confer a benefit on the authors or on any organization with which the authors are associated.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Results—Data quality was good and CAT efficiency was high across both WD-FAB domains. Measurement accuracy was very good for the PF scales; BH scales demonstrated more variability. Construct validity correlations, both convergent and divergent, between all WD-FAB scales and established measures were in the expected direction and range of magnitude.

Conclusions—The data quality, CAT efficacy, person fit and construct validity of the WD-FAB scales were well supported and suggest that the WD-FAB could be used to assess physical and behavioral health function related to work disability. Variation in scale performance suggests the need for future work on item replenishment and refinement, particularly regarding the Self-Efficacy scale.

Keywords

Validation Studies; Disability Evaluation; US Social Security Administration; Outcomes Assessment; Psychometrics

The concept of work disability is evolving as represented by the World Health Organization's International Classification of Functioning, Disability and Health.[1] Rather than defining work disability based on a person's medical diagnosis and impairments, current models and occupational rehabilitation practice consider the relationships between a person's functioning and the demands of the work and working environment.[2] This conceptual development has wide-ranging implications for rehabilitation, policy and research. A fundamental issue relevant to all three is the question of how to identify and measure work disability within this contemporary framework. The United States' Social Security Administration (SSA), which provides support to people with disabilities through insurance coverage and income replacement programs, faces challenges in the assessment of work disability within this context.[3,4]

Currently the SSA defines work disability as the inability to take part in "substantial gainful activity due to any medically determinable physical or mental impairment that can be expected to result in death or to last for a continuous period of not less than 12 months."[5] Determination of disability is accomplished through a 5-step process using information including symptoms, medical evidence, and to some extent function. Limitations to this process have been acknowledged, and have served as the subject of recent reform efforts.[3] In 2007, an Institute of Medicine panel made recommendations for new approaches to measuring disability that would incorporate functional assessment.[6] However, the measurement of person functioning presents several challenges, including the lack of a gold standard assessment. While observer-based approaches to assessing function provide useful information, it can be difficult and expensive to administer them to large numbers of persons as in the case of the SSA disability determination process. Self-report measures using questionnaire formats have demonstrated reliability and validity, but can be hampered by significant respondent burden and poor precision.

More recently, measurement methods used in educational testing have been applied to the measurement of health and functioning. These methods, including item response theory (IRT) and computerized -adaptive testing (CAT), involve the development of large banks of questions representing the full range of functioning in an area of interest. Using field study

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

data, items are calibrated so that each is assigned a location on a unidimensional scale. Advantages of these approaches include: 1) a score representing the person's ability level on the latent trait and associated precision level is estimated after the response to each question; 2) a CAT program tailors selection of questions to test-takers based on their prior responses, and thus these instruments can measure a broad construct with precision while limiting respondent burden.

We developed instruments to measure physical and behavioral functioning for work disability determination using IRT and CAT methods, resulting in the Work Disability Physical Function (PF) and Behavioral Health Function (BH) scales, collectively referred to as the Work Disability Functional Assessment Battery (WD-FAB).[2,7–10] Four multi-item physical function scales were identified: Changing and Maintaining Body Position, Upper Body Function, Upper Extremity Fine Motor, and Whole Body Mobility. Four multi-item scales were also identified in the behavioral health domain: Self-Efficacy, Mood and Emotions, Behavioral Control, and Social Interactions. Results regarding internal consistency reliability, precision of measurement across the range of possible scores, and the pattern of correlations among the scales generated during the development process provided preliminary evidence supportive of the scales' having desirable psychometric properties.[8, 10]

However, this previous work was conducted in single samples of adults seeking SSA disability benefits in each domain, and only involved the examination of relationships among the WD-FAB scales themselves. The objective of the present study, therefore, was to further evaluate the feasibility of administration and psychometric properties of the WD-FAB CAT scales. This was done in new samples and included assessments on several parameters of measurement performance: data quality, including score distribution and amount of missing data; administrative efficiency; accuracy; and construct validity.

Methods

Participants/Sample

The study sample consisted of adults unable to work due to a permanent disability, and was obtained through a survey research organization that maintains an opt-in internet survey panel of approximately 1.5 million US adults. For the preset study, the subset of that panel who had indicated in their enrollment background information that they were unable to work due to a permanent disability were invited to participate. No independent assessment of their disability status was conducted. Consenting participants indicated whether their disability w as primarily physical in nature, mental, or both. Those who selected "both" were randomly assigned to receive either the physical or behavioral measurement batteries to achieve the target of 500 subjects in each disability category. The study was approved by a university institutional review board and all subjects provided informed consent.

Established Measures

All subjects received the Veterans Rand 36-item Health Survey (VR-36), a modification of the RAND SF-36 which measures eight domains of both physical and mental health.[11] Two standard composite indices were computed and used in the present study: the Physical

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Component Summary (PCS) and the Mental Component Summary (MCS). Scores on both the PCS and MCS range from 0–100 with higher scores indicating better functioning.

The physical disability group also received the PROMIS Physical Function 10-item Short Form (PROMIS), which measures current capability for mobility, walking, hand and arm use, and activities of daily living where a higher score always indicates better functioning.[12] The PROMIS measure yields a single summary score, which is then converted to a T-score (mean of 50, standard deviation of 10).

The behavioral disability group also received the Behavior and Symptom Identification Scale (BASIS-24), which consists of 24 items that ask respondents to describe their experience regarding various cognitive (e.g., concentrating) and emotional (e.g., depression) issues during the past week using Likert-type five-point (0 to 4) response scales.[13] Higher scores represent greater symptom severity (e.g., more frequent depression). For this study we administered five of the six standard BASIS scales: Depression/Functioning, Relationships, Self-harm, Emotional Lability, and Psychosis.

New Measures: The WD-FAB

The physical disability group received the WD-FAB PF scales assessing Upper Body Function, Upper Extremity Fine Motor Function, Whole Body Mobility, and Changing and Maintaining Body Position. All items used a 5-point difficulty response scale ranging from “unable to do” to “no difficulty,” with higher scores indicating better functioning. Each raw PF scale score was standardized on a national normative sample and then converted to a T-score (mean of 50, standard deviation of 10).[7]

The behavioral disability group received the WD-FAB BH scales assessing Self-Efficacy, Mood and Emotions, Behavioral Control and Social Interactions. The BH items include two types of item structures: 37 items using a 4-point frequency response categories with options ranging from “Never” to “Always,” and 41 items using 5 agreement response categories ranging from “Strongly Agree” to “Strongly Disagree,” including an “I don’t know” option. All items are scored such that higher scores indicate higher functioning. Each raw BH scale score was standardized on a national normative sample and then converted to a T-score (mean of 50, standard deviation of 10).[7]

Procedure

Respondents reported their sex, marital status, race, ethnicity, age, when they became unable to work, and if they were receiving disability support benefits. They were then administered the established measures followed by the new WD-FAB scales. The CAT algorithm specified that, for each scale, a minimum of five items would be administered up to a maximum of 8 for the PF scales and 10 for the BH scales, to achieve score reliability >=0.85.

Data Analysis

To assess *data quality*, we examined the distribution of scores by computing the mean, standard deviation (SD), coefficient of variation and score range for each of the WD-FAB

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

scales as well as the percentage of respondents with the lowest (floor) and highest (ceiling) possible scale values. The number of "I don't know" responses selected per respondent for each scale was also examined as an indicator of possible item problems.

To assess the *efficiency of CAT administration* we computed the mean, SD and range of the number of items selected for CAT administration for each WD-FAB scale. We also examined the percent of respondents who used the minimum and maximum number of items for each scale, and the amount of time required to complete the scale.

Measurement accuracy was assessed by computing the standardized log-likelihood statistic (l_z) for polytomous items proposed by Drasgow and colleagues, which assesses the difference between the observed response pattern and the pattern that would be expected given the respondent's estimated level of function.[14, 15] Specifically, we used a l_z value less than −1.645 as the criterion for significant misfit at alpha=0.05 (one-tailed). The fit mean, SD, and range were computed for each WD-FAB scale, as well as the percent of respondents with fit statistic values < −1.645, indicating unacceptable fit. Adequate person fit for a scale is achieved if <=5% of respondents demonstrate significant misfit.

Evidence of *construct validity* was obtained by examining the pattern of convergent and discriminant correlations computed among the established and new WD-FAB scales. Regarding convergent validity, we hypothesized that the WD-FAB measures of physical and behavioral health function would be significantly and positively correlated with the established measures of their respective domains. We examined discriminant validity by comparing the strength of same-domain to cross-domain correlations, expecting the latter to be notably weaker and ideally near zero.

Results

Demographic characteristics for the samples are reported in Table 1. The average PF respondent was a 56 year old non-Hispanic married white male with some college education who had been unable to work due to his disability for 3.8 years. The average BH respondent was a 51 year old non-Hispanic white female, also with some college education, who had been unable to work due to her disability for 7.4 years.

PF Scale Data Quality

Table 2 reports means on the PF scales standardized scores, which ranged from 33.6 (Whole Body Mobility) to 43.4 (Upper Extremity Fine Motor). The average score range across the four PF scales was 46.9 points. Coefficients of variation were very similar on three of the four scales (20.6 – 22.2), with relatively less score dispersion in Upper Extremity Fine Motor functioning (17.6). Missing data ranged from less than 1 – 4.3% per respondent. The percent of respondents with the minimum possible score (floor) was 2% or less across all the PF scales. Less than 1 percent of respondents had the maximum possible score (ceiling) on any PF scale with the exception of Upper Extremity Fine Motor (12.4%).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

CAT Administration Efficiency

The number of items administered by the CAT for each of the PF scales ranged from a minimum of 5 items to a maximum of 8, with the vast majority of respondents achieving the score reliability criterion (>=0.85) after the minimum administration of only 5 items (Table 3). The exception was Upper Extremity Fine Motor, where 24.3 percent of respondents were administered the maximum of 8 items. The amount of time required to complete each PF scale averaged less than two minutes per scale. Overall, the CAT administration of the full set of four PF scales represented low respondent burden, requiring an average of only 22 items and just over 6.5 minutes to complete.

Person Fit

The standardized log-likelihood fit statistics for the four PF scales show the percentage of misfit subjects well below the 5% threshold criterion (Table 4).

Construct Validity

All correlations between the two established measures of physical functioning and the four new PF scales were significant ($p<.05$) in the predicted direction and were, with one exception, moderate to strong, ranging from 0.23 to 0.70 (median 0.55) (Table 5). Discriminant validity correlations between the PF scales and the MCS were also statistically significant but weak, ranging from 0.12 to 0.24 (median 0.18). An exception to this pattern was observed regarding the new Upper Extremity Fine Motor scale, which was as strongly related to the divergent MCS as it was to the convergent PCS. However, the relationship between upper extremity scale and the PROMIS PF 10 item short form (0.54) was in the predicted direction and more than twice as strong as the correlations with the two VR-36 component summary scores.

BH Scale Data quality

Means on the standardized BH scales ranged from 34.8 (Self-Efficacy) to 42.9 (Behavioral Control) (Table 2). Considerable range was observed on each of the four BH scales. Coefficients of variation (CV) were very similar on three of the four scales (24.3 – 27.4), with relatively greater score dispersion evident regarding Self-Efficacy (39.5). The amount of missing data ranged from less than 1% per respondent (Mood and Emotion) to 2.2% per respondent (Social Interactions). The exception again was Self-Efficacy with an average of 10.4% of item presentations eliciting an “I don’t know” response per respondent. Across the BH scales the percent of respondents with floor values was 2.5% or less, and the percent at the ceiling was less than 2%.

CAT Administration Efficiency

The number of items administered by the CAT ranged from a minimum of 5 items to a maximum of 6 items on the Social Interactions scale and 10 items on the other three scales (Table 3). On two of the four scales (Behavioral Control and Mood and Emotions) the modal number of items needed to achieve the CAT stopping criterion was the minimum number administered. On the other two scales (Self-Efficacy; Social Interaction) the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

maximum number of items (10) was administered to the majority of respondents. The four scales were completed in an average of 7.1 minutes (SD=6.6 minutes).

Person Fit

Variation in measurement accuracy was observed across the four BH scales (Table 4). The percentage of misfit subjects was below the 5% criterion for adequate measurement person fit for both Mood and Emotions and Social Interactions, but exceeded this criterion for Behavioral Control (6.9% misfit) and Self-Efficacy (13.7% misfit).

Construct Validity

All of the 24 convergent validity correlations computed between the established measures and the four BH scales were significant ($p<.05$) in the predicted direction and were generally moderate to strong in magnitude, ranging from 0.24 to 0.74 (median 0.40) (Table 6). Discriminant correlations between the BH scales and the VR-36 PCS were much weaker, ranging from 0.06 to 0.31 (median 0.14).

Discussion

This study extends previous work in the development and evaluation of a new set of measures, the WD-FAB, that assess work-related physical and behavioral health functioning. We found substantial support for the psychometric quality and efficiency of the CAT administration with regard to both PF and BH scales, and for the construct validity of the four PF scales. Support for the four proposed BH scales was also observed, but was not as consistent across all four scales.

With regard to the PF scales, psyhchometric quality was very good. There was minimal missing data, substantial score variation was observed with virtually no clustering at the floor and ceiling values, and respondent burden was low. These findings suggest that the scales are discriminating different levels of ability across the range of function, and that there is potential sensitivity to group differences. A minor exception was observed regarding Upper Extremity Fine Motor scale, where just over 12% of respondents scored at the ceiling value. However, we did not see this as a concern given that the ceiling proportion was small in an absolute sense and given the overall strength of the scale's performance. Measurement accuracy was excellent, with a maximum of 2.4% (Changing and Maintaining Body Position) of respondents exceeding the misfit threshold.

Results provided strong evidence of construct validity for the PF scales: the median convergent correlation was 0.55 as compared to the median discriminant correlation of 0.18. One partial exception concerned the Upper Extremity Fine Motor scale, which was as strongly related to the VR-36 mental health component summary score (r=0.23) as it was to the physical health component summary score (r=0.24). However the relationship between the Upper Extremity Fine Motor scale and the other established physical function measure (the PROMIS SF-10) was twice as strong (0.54).

With regard to the BH scales, psychometric quality was also very good. The proportion of missing data was low, and substantial variation was observed across the four BH scales.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Respondents were neither clustered at the bottom or the top of the raw score distribution, suggesting that the BH scales are discriminating different levels of ability across the range of function, and that there is potential sensitivity to group differences and change. CAT administration of the full set of BH scales represented very low respondent burden. Two scales – Mood and Emotions, and Social Interactions – met the accuracy criteria of <5% of cases with significant misfit, whereas Behavioral Control was slightly above at 6.8% misfit, and Self-Efficacy was noticeably above at 13.7%. These findings suggest that the existing items in the latter two scales leave gaps in the representation of some levels of their respective underlying latent characteristics. Item replenishment work is underway to improve measurement precision, particularly of the Self-Efficacy scale, by using IRT methods to calibrate potential new or substitute items in the context of the existing questions.

Results for the BH scales were generally supportive of construct validity. The median of convergent correlations of 0.40 would be regarded as between “medium” (0.30) and “large” (0.50) according to widely-accepted criterion.[16] The median of discriminant correlations was 0.14, indicative of appropriately weak relationships. An exception to this pattern was the correlation of 0.31 between Social Interactions and the VR-36 *physical* component summary score. Although we expected this discriminant correlation to be weak, it was roughly equal to or greater than 7 of 24 convergent correlations between the new BH scales and the established measures of mental health function. However, this finding is consistent with the literature suggesting an interaction between physical health and behavioral functioning such that managing the physical aspects of a health condition may be associated with less interest in or ability to participate in social activities.[17–21]

Although developed to be an element of the SSA evaluation process for work disability benefits, we would suggest that the WD-FAB has broader clinical relevance. Many clinicians practice in the area of work and work rehabilitation and need to assess patients’ readiness to return to work. Further, although the dimensions of physical and emotional functioning assessed by the WD-FAB are work-relevant, the areas assessed are not specific to work and thus could be applied in other clinical contexts such as vocational rehabilitation or work disability prevention programs.

Study Limitations

The present study represents the logical next step in the psychometric assessment of a set of new measures of work-relevant physical and behavioral function developed on the basis of a sound theoretical framework (2), the application of modern item response measurement theory (7–10), and administration by computer-adaptive testing. The overall study design was strong, including substantial sample and counterbalanced administration of new and established criterion measures.

Some limitations should be noted, however. The etiology of participants’ disability was ascertained via self-report and could not be independently verified by examination or documentation; however, a growing body of evidence supports the validity and reliability of self-reported information about medical conditions.[22–25] More broadly, the reference population of interest for the application of the WD-FAB is all work-disabled individuals,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and we did not have a way to compare the study sample to that population. Consequently it is not clear how representative our sample is of that population. We would also note that present sample of survey panel participants who have voluntarily enrolled to participate in numerous internet surveys may be different in unmeasured ways from the clinical or disability claimant populations in which the WD-FAB might be applied. Further, in the BH domain, a higher degree of misfit as well as greater utilization of the "I don't know" response option was observed on the Self-Efficacy scale. These findings support the need for future work on this scale. A major advantage of the IRT methodology is the opportunity to update and refine the item bank for any given scale as needed by administering new candidate items in conjunction with the current item pool.[26] This procedure efficiently provides for the evolution of an instrument so that it reflects the most recent substantive and psychometric advances.

Conclusions

The present study provides a robust array of evidence supporting the validity of the WD-FAB PF and BH function scales. Overall, data quality, CAT efficacy, person fit and construct validity were well supported and suggest that that the WD-FAB could be used to assess physical and behavioral health function related to work disability. Future work will focus on item replenishment and refinement to increase the overall performance of these scales.

Acknowledgement

Financial Support

This study was supported by Social Security Administration-National Institutes of Health Interagency Agreements under the National Institutes of Health (contract nos. HHSN269200900004C, HHSN269201000011C, HHSN269201100009I), and by the National Institutes of Health Intramural Research Program.

List of Abbreviations

BASIS	Behavior and Symptom Identification Scale
BH	behavioral health
CAT	computer adaptive testing
IRT	item response theory
MCS	mental component summary
PCS	physical component summary
PF	physical function
PROMIS	Patient-Reported Outcomes Measurement Instrument System
SSA	Social Security Administration
VR-36	Veterans RAND SF-36 Health Survey
WD-FAB	Work Disability Functional Assessment Battery

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

References

1. World Health Organization. Geneva: 2001. International Classification of Functioning, Disability and Health (ICF).
2. Marfeo EE, Jette AM, Eisen SE, Bogusz K, McDonough CM, Ni PS, Rasch EK, Brandt DE, Meterko M, Haley SM, Chan L. A conceptual foundation for measures of physical function and behavioral health function for Social Security work disability evaluation. Arch Phys Med Rehab. 2013; 94:1645–1652.
3. [Accessed November 3rd, 2013] Social Security Advisory Board Round Table. A disability system for the 21st century. Available at: http://www.ssab.gov/documents/disability-system-21st.pdf
4. Brandt DE, Houtenville AJ, Huynh MT, Chan L, Rasch EK. Connecting contemporary paradigms to the Social Security Administration's Disability Evaluation Process. J Disability Policy Studies. 2011; 22(2):116–128.
5. Social Security Administration. [Accessed Nov 3, 2013] Annual Statistical Report on the Social Security Disability Insurance Program. 2011. SSA publication No. 13-11826. Available at: http://www.ssa.gov/policy/docs/statcomps/di_asr/2011/di_asr11.pdf
6. IOM (Institute of Medicine). Improving the Social Security disability decision process. Washington, D.C.: 2007.
7. Marfeo EE, Ni PS, Jette AM, Bogusz K, McDonough CM, Meterko M, Haley SM, Chan L, Brandt DE, Rasch EK. Development of an instrument to measure behavioral health function for work disability: Item pool construction and factor analysis. Arch Phys Med Rehab. 2013; 94:1670–1678.
8. Marfeo EE, Ni PS, Bogusz K, McDonough CM, Meterko M, Haley SM, Chan L, Brandt DE, Rasch EK, Jette AM. Scale refinement and initial evaluation of a behavioral health function measurement tool for work disability evaluation. Arch Phys Med Rehab. 2013; 94:1679–1686.
9. McDonough CM, Jette AM, Ni P, Bogusz K, Marfeo EE, Brandt DE, Chan L, Meterko M, Haley SM, Rasch EK. Development of a self-report physical function instrument for work disability assessment: Item pool construction and factor analysis. Arch Phys Med Rehab. 2013; 94:1653–1660.
10. Ni PS, McDonough CM, Jette AM, Bogusz K, Marfeo EE, Rasch EK, Brandt DE, Meterko M, Haley SM, Chan L. Development of a computer-adaptive physical function instrument for Social Security Administration disability determination. Arch Phys Med Rehab. 2013; 94:1661–1669.
11. Kazis LE, Miller DR, Clark J, Skinner K, Lee A, Rogers W, Spiro A 3rd, Payne S, Fincke G, Selim A, Linzer M. Health-related quality of life in patients served by the Department of Veterans Affairs: results from the Veterans Health Study. Archives of Internal Medicine. 1998; 158(6):626–632. [PubMed: 9521227]
12. Bruce B, Fries JF, Ambrosini D, Lingala B, Gandek B, Rose M, Ware JE Jr. Better assessment of physical function: Item improvement is neglected but essential. Arthritis Research and Therapy. 2009; 11(6):R191. (PMCID: PMC3003539). [PubMed: 20015354]
13. Eisen SV, Gerena M, Ranganathan G, Esch D, Idiculla T. Reliability and validity of the BASIS-24© mental health survey for whites, African-Americans, and Latinos. The journal of behavioral health services & research. 2006; 33(3):304–323. (2006). [PubMed: 16752108]
14. Drasgow F, Levine MV, Williams EA. Appropriateness measurement with polytomous item response models and standardized indices. Brit J Mathematical Statistical Psych. 1985; 38:67–86.
15. Drasgow F, Levine MV, McLaughlin ME. Appropriateness measurement for some multidimensional test batteries. Applied Psychological Measurement. 1991; 15:171–191.
16. Cohen, J. Statistical power analysis for the behavioral sciences (Revised edition). New York: Academic Press; 1977.
17. Frick U, Irving H, Rehm J. Social relationships as a major determinant in the valuation of health states. Qual Life Res. 2012; 21:209–213. [PubMed: 21633877]
18. Goud AL, de Lange J, Scholtes VA, Bulstra SK, Ham SJ. Pain, physical and social functioning, and quality of life in individuals with multiple hereditary exostoses in the Netherlands: A national cohort study. J Bone Joint Surg Am. 2012; 94:1013–1020. [PubMed: 22637207]
19. Hannah SD. Psychosocial issues after a traumatic hand injury: Facilitating adjustment. J Hand Therapy. 2011; 24:95–102.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

20. Montin L, Suominen T, Haaranen E, Katajisto J, Lepisto J, Leino-Kilpi H. The changes in health-related quality of life and related factors during the process of total hip arthroplasty. Int J Nurs Pract. 2011; 17:19–26. [PubMed: 21251150]

21. Poulsen T, Siersma VD, Lund R, Christensen U, Vass M, Avlund K. Impact of social capital on 8-year mortality among older people in 34 Danish municipalities. J Aging Health. 2012; 24:1203–1222. [PubMed: 22869899]

22. Jensen MP, Abresch RT, Carter GT. The reliability and validity of a self-report version of the FIM instrument in persons with neuromuscular disease and chronic pain. Arch Phys Med Rehabil. 2005; 86:116–122. [PubMed: 15641001]

23. Mason JH, Anderson JJ, Meenan RF, Haralson KM, Lewis-Stevens D, Kaine JL. The rapid assessment of disease activity in rheumatology (radar) questionnaire. Validity and sensitivity to change of a patient self-report measure of joint count and clinical status. Arthritis Rheum. 1992; 35:156–162. [PubMed: 1734905]

24. Latham NK, Mehta V, Nguyen AM, et al. Performance-based or self-report measures of physical function: which should be used in clinical trials of hip fracture patients? Arch Phys Med Rehabil. 2008; 89:2146–2155. [PubMed: 18996244]

25. Reuben DB, Siu AL, Kimpau S. The predictive validity of self-report and performance-based measures of function and health. J Gerontol. 1992; 47:M106–M110. [PubMed: 1624692]

26. 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]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Respondent Demographics

Demographic Characteristics	Physical Function Sample (n=497)		Behavioral Health Sample (n=476)	
A. Continuous	**Mean**	**SD**	**Mean**	**SD**
Age (Years)	56.02	8.52	51.2	9.81
Work Disability Duration (Years)	3.83	2.43	7.43	5.21
B. Categorical	**Count**	**Percent**	**Count**	**Percent**
Sex				
Male	260	52.31	160	33.61
Race				
White	415	83.5	404	84.87
Black/African American	32	6.44	28	5.88
Other/more than one	42	8.45	38	7.98
Missing	8	1.61	6	1.26
Hispanic				
Yes	21	4.23	17	3.57
No	467	93.96	455	95.59
Refused	9	1.81	4	0.84
Education				
High school diploma or less	132	26.56	130	27.31
Associate degree or vocational training	88	17.70	89	18.69
Some college	155	31.19	130	27.31
College graduate or more	122	24.55	127	26.68
Relationship status				
Never married	69	13.88	87	18.28
Married or living with partner	255	51.30	214	44.95
Divorced, Separated or Widowed	172	34.60	173	36.34
Refused	1	0.20	2	0.42

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Basic Descriptive Statistics for Established and New WD-FAB Measures (PF and BH)

A. Physical Function	Sample Size	Mean (SD)	CV[a]	Min -Max	Range	Mean Pct. "I don't know"
Established						
VR-36 MCS	497	43.26 (12.25)	28.3	15.8 – 69.35	53.6	NA[b]
VR-36 PCS	497	24.79 (8.45)	34.1	5.14 – 58.23	53.1	
PROMIS PF 10 item short form	497	34.19 (5.81)	17.0	14.1 – 55.3	41.2	
WD-FAB: PF						
Changing & maintaining body position	493	34.99 (7.49)	21.4	8.18 – 60.44	52.26	0.99
Upper body function	491	34.69 (7.69)	22.2	13.88 – 59.61	45.73	4.27
Upper extremity fine motor	494	43.42 (7.66)	17.6	21.26 – 58.43	37.17	1.15
Whole body mobility	471	33.56 (6.91)	20.6	7.07 – 59.52	52.45	3.72
B. Behavioral Health Function	Sample Size	Mean (SD)	CV[a]	Min -Max	Range	Mean Pct. "I don't know"
Established						
VR-36 MCS	476	29.6 (10.53)	35.57	6.29 – 63.86	57.57	NA[b]
VR-36 PCS	476	33.95 (11.24)	33.11	6.66 – 67.36	60.70	
BASIS24: Depression	476	2.28 (0.82)	35.96	0 – 4	4	
BASIS24: Relationships	476	2.04 (0.93)	45.59	0 – 4	4	
BASIS24: Self-Harm	476	0.66 (0.91)	137.88	0 – 4	4	
BASIS24: Emotional Lability	476	2.08 (0.98)	47.12	0 – 4	4	
BASIS24: Psychosis	476	0.92 (0.94)	102.17	0 – 4	4	
WD-FAB: BH						
Self-Efficacy	476	34.75 (13.74)	39.54	−14.39 – 90.7	105.09	10.41
Mood and Emotion	474	35.15 (9.63)	27.40	1.56 – 92.09	90.53	0.08

Arch Phys Med Rehabil. Author manuscript; available in PMC 2016 June 01.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Behavioral Control	468	42.89 (10.48)	24.43	9.22 – 85.22	76.0	1.06
Social Interactions	473	36.76 (8.94)	24.32	11.23 – 77.07	65.84	2.22

[a] Coefficient of variation, computed as (SD/mean)*100.

[b] Not applicable; answers to all established instrument items were required.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3

Efficiency of CAT Administration

A. Physical Function	Number of Items Used	Percent of Respondents Requiring:		Time to Completion (Minutes)	95% of Respondents Completed Within:
	Mean(SD)	Minimum Items (5)	Maximum Items (8)	Mean (SD)	Minutes
Changing & maintaining body position	5.41 (0.92)	80.3	8.7	6.55 (6.77)	13.27
Upper body function	5.23 (0.76)	90.9	6.1		
Upper extremity fine motor	5.86 (1.29)	66.4	24.3		
Whole body mobility	5.63 (1.09)	70.1	14.0		
B. Behavioral Health Function	Number of Items Used	Percent of Respondents Requiring:		Time to Completion (Minutes)	95% of Respondents Completed within:
	Mean (SD)	Minimum Items (5)	Maximum Items (10[a])	Mean (SD)	Minutes
Self-Efficacy	9.45 (0.92)	1.05	64.5	7.13 (6.57)	15.00
Mood and Emotion	5.34 (1.07)	87.13	3.59		
Behavioral Control	6.76 (1.97)	41.03	21.15		
Social Interactions	5.82 (0.38)	17.97	82.03		

[a]Maximum of 10 items for all scales except Social Interactions, for which the maximum number of items used was 6, the number in the item pool.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4

Person Fit (Standardized Log Likelihood)

A. Physical Function					
PF Dimension	N	Mean(SD)	Minimum	Maximum	Percent Misfit Subjects[a]
Changing & maintaining body position	493	0.44 (0.85)	−4.25	1.76	2.43%
Upper body function	491	0.49 (0.69)	−3.21	1.84	0.81%
Upper extremity fine motor	494	0.44 (0.71)	−5.07	1.80	1.42%
Whole body mobility	471	0.35 (0.79)	−4.92	1.91	2.12%
B. Behavioral Health Function					
BH Dimension	N	Mean(SD)	Minimum	Maximum	Percent Misfit Subjects[a]
Self-Efficacy	476	−0.09 (1.53)	−8.63	1.91	13.66
Mood and Emotion	474	0.43 (0.81)	−3.77	1.87	2.11
Behavioral Control	468	0.01 (1.04)	−4.90	2.06	6.84
Social Interactions	473	0.34 (0.94)	−4.96	2.09	4.23

[a] Fit statistic < −1.645.

Arch Phys Med Rehabil. Author manuscript; available in PMC 2016 June 01.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 5

Physical Function: Correlations Between Established and WD-FAB PF Measures

	MCS	PCS	PPF 10	CMBP	UBF	UEFM
VR-36 MCS	1.00					
VR-36 PCS	−0.25					
PROMIS PF 10 short form (PPF10)	0.24	0.59				
Changing & maintaining body position (CMBP)	*0.12*	**0.42**	**0.65**			
Upper body function (UBF)	*0.21*	**0.43**	**0.69**	0.63		
Upper extremity fine motor (UEFM)	*0.24*	**0.23**	**0.54**	0.49	0.53	
Whole body mobility (WBM)	*0.15*	**0.55**	**0.70**	0.60	0.67	0.39

Sample sizes varied from 417 to 497 due to differences in amount of missing data across scales.
Bold indicates convergent validity correlations; *italic* indicates discriminant validity correlations.
All correlations are significant at p<.05.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 6

Behavioral Health Function: Correlations Between Established and WD-FAB BH Measures

	VR-36		BASIS-24					BH		
	MCS	**PCS**	**DEP**	**RELATE**	**HARM**	**EMOT**	**PSYCH**	**SELF-E**	**MOOD**	**BC**
VR-36 MCS	1.00									
VR-36 PCS	−0.15									
BASIS24: Depression (DEP)	−0.71	−0.23								
BASIS24: Relationships (RELATE)	−0.38	−0.09	0.41							
BASIS24: Self-Harm (HARM)	−0.34	−0.13	0.41	0.24						
BASIS24: Emotional Lability (EMOT)	−0.45	−0.08	0.52	0.37	0.35					
BASIS24: Psychosis (PSYCH)	−0.27	−0.05	0.37	0.26	0.47	0.48				
Self-Efficacy (SELF-E)	**0.46**	*0.06*	**−0.46**	**−0.41**	**−0.30**	**−0.35**	**−0.32**			
Mood and Emotion (MOOD)	**0.67**	*0.21*	**−0.74**	**−0.42**	**−0.47**	**−0.54**	**−0.39**	0.49		
Behavioral Control (BC)	**0.32**	*0.07*	**−0.36**	**−0.28**	**−0.35**	**−0.59**	**−0.43**	0.43	0.43	
Social Interactions (SOCIAL)	**0.56**	*0.31*	**−0.63**	**−0.39**	**−0.28**	**−0.31**	**−0.24**	0.48	0.62	0.27

Sample sizes varied from 466 to 476 due to differences in amount of missing data across scales.
Bold indicates convergent validity correlations; *italic* indicates discriminant validity correlations.
All correlations are significant at p<.05 except for correlations (5) of less than 0.10.