



Published in final edited form as:

Arch Phys Med Rehabil. 2013 September ; 94(9): 1653–1660. doi:10.1016/j.apmr.2013.03.011.

Development of a Self-Report Physical Function Instrument for Disability Assessment: Item Pool Construction and Factor Analysis

Christine M. McDonough, PT, PhD^{1,2}, Alan M. Jette, PT, PhD¹, Pengsheng Ni, MD, MPH¹, Kara Bogusz, BA¹, Elizabeth E Marfeo, PhD, MPH¹, Diane E Brandt, PT, MA, PhD³, Leighton Chan, MD, MPH³, Mark Meterko, PhD¹, Stephen M. Haley, PT, PhD*, and Elizabeth K. Rasch, PT, PhD³

¹The Health & Disability Research Institute, Department of Health Policy & Management, School of Public Health, Boston University Medical Campus, Boston, MA

²The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine, Lebanon NH

³Clinical Research Center, Rehabilitation Medicine Department, National Institutes of Health, Bethesda, MD

Abstract

Objectives—To build a comprehensive item pool representing work-relevant physical functioning and to test the factor structure of the item pool. These developmental steps represent initial outcomes of a broader project to develop instruments for the assessment of function within the context of Social Security Administration (SSA) disability programs.

Design—Comprehensive literature review; gap analysis; item generation with expert panel input; stakeholder interviews; cognitive interviews; cross-sectional survey administration; and exploratory and confirmatory factor analyses to assess item pool structure.

© 2013 The American Congress of Rehabilitation Medicine. Published by Elsevier Inc. All rights reserved.

Corresponding Author: Christine M. McDonough, Health & Disability Research Institute, Boston University School of Public Health, 715 Albany St. T5W, Boston, MA 02118, Ph: (617) 638-1982, Fax: (617) 638-1999, cmm@bu.edu.

*Deceased.

PRESENTATION OF THIS MATERIAL:

“Development of a Computer Adaptive Test to Assess Physical Capabilities for Work Disability Determination” presented to the Disability Forum of the American Public Health Association, Washington, DC, November 1, 2011.

“Innovations in Self-Reported Function and Disability Assessment.” Presented at the Psychiatric Research Center, Geisel School of Medicine, Lebanon, NH. April, 2012

An additional presentation was made of this material to the 4th Congress of the International Association of Bodily Impairment (AIDC), Montreal, Quebec on September 12, 2012.

COMMERCIAL SUPPORT/CONFLICTS STATEMENT:

We certify that no party having a direct interest in the results of the research supporting this article has or will confer a benefit on us or on any organization with which we are associated AND, we certify that all financial and material support for this research and work are clearly identified in the title page of the manuscript (Christine M. McDonough, Alan M. Jette, Pengsheng Ni, Kara Bogusz, Elizabeth E Marfeo, Diane E Brandt, Leighton Chan, Mark Meterko, Stephen M. Haley, Elizabeth K. Rasch).

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.

Setting—In-person and semi-structured interviews; internet and telephone surveys.

Participants—A sample of 1,017 SSA claimants, and a normative sample of 999 adults from the US general population.

Interventions—Not Applicable.

Main Outcome Measure—Model fit statistics

Results—The final item pool consisted of 139 items. Within the claimant sample 58.7% were white; 31.8% were black; 46.6% were female; and the mean age was 49.7 years. Initial factor analyses revealed a 4-factor solution which included more items and allowed separate characterization of: 1) Changing and Maintaining Body Position, 2) Whole Body Mobility, 3) Upper Body Function and 4) Upper Extremity Fine Motor. The final 4-factor model included 91 items. Confirmatory factor analyses for the 4-factor models for the claimant and the normative samples demonstrated very good fit. Fit statistics for claimant and normative samples respectively were: Comparative Fit Index = 0.93 and 0.98; Tucker-Lewis Index = 0.92 and 0.98; Root Mean Square Error Approximation = 0.05 and 0.04.

Conclusions—The factor structure of the Physical Function item pool closely resembled the hypothesized content model. The four scales relevant to work activities offer promise for providing reliable information about claimant physical functioning relevant to work disability.

Keywords

Disability Evaluation; Disabled Persons; Insurance; Work Disability; Questionnaires; Factor Analysis; Statistical

Social Security Administration (SSA) disability programs are the largest federal source of assistance, providing support for more than 12 million people with disabilities.¹(http://www.ssa.gov/policy/docs/statcomps/di_asr/2011/di_asr11.pdf) These programs have experienced a substantial increase in applications over time, presenting critical challenges to the production of timely, accurate disability determinations, evidenced by prolonged wait times for a final decision.² Therefore, innovative approaches to support the efficiency of the disability determination process are of significant interest because they have the potential to impact program performance and decrease costs. The 2007 Institute of Medicine report “Improving the Social Security Disability Decision Process” made several recommendations relating to the development and testing of different approaches to assessing disability, including the assessment of function.³

Self-report assessment methodologies within health care and rehabilitation have been shown to provide valuable information about symptoms and function that cannot be provided by observer-based approaches.^{4,5} One such methodology, computer-adaptive testing (CAT), employs a computer algorithm to tailor questions to the specific ability level of the respondent.⁶ This approach supports a comprehensive and systematic assessment of functioning and has the potential to improve the precision and efficiency of SSA’s disability evaluation process. CAT measurement systems employ contemporary measurement methods including Item Response Theory (IRT).⁷ These methods have been applied for decades in

educational testing, and more recently employed in self-reported health outcomes measurement to improve the breadth of coverage and precision of outcome measurement.^{8,9}

The overall goal of the present project is to develop instruments for the assessment of functioning within the Social Security Administration (SSA) disability programs, integrating CAT methodology. A CAT system requires a pool of items that covers the range of ability within a construct of interest. The item pool is ordered, or calibrated, so that each item represents a position from low to high levels, in this case, of physical functioning. The specific aims of this study include building a comprehensive item pool for the physical function domain covering the full range of function, and assessing the underlying factor structure of the item pool. The number of sub-domains revealed by factor analyses indicates the number of item pools necessary to represent physical functioning within a CAT instrument system. The results of this work provide the foundation for next steps in the CAT development process, including IRT analyses to calibrate the items in each item pool and to build the algorithms to determine their sequence of administration for CAT applications. These steps and the resulting CAT measurement system, which we call the Social Security Administration Physical Function Instrument (SSA-PF) are described in a companion paper.¹⁰

Methods

All study procedures were approved by the Boston University Internal Review Board.

Content Expert and Stakeholder Input

Four content experts were identified based on their clinical and research expertise in the evaluation of physical function and/or work disability in the context of the SSA programs, and assisted with item pool development. A series of telephone conferences and an in-person working session were held to solicit input on item content, framing, and response options. Health care and social service providers who regularly work with claimants and beneficiaries of SSA disability programs were also selected to provide input via in-person and telephone interviews.

Physical Function Item Pool Development

The development of the content model, including identification of work-related constructs for the physical function domain is reported in a companion paper.¹¹ A comprehensive literature review was conducted to identify existing items relevant to physical function in the context of work, to evaluate item candidates, and to inform the establishment of an item pool representing the full range of function. The literature review targeted work literature, work capacity assessment and generic self-report measures of health and physical function. Relevant items from previously existing instruments such as PROMIS¹² and Neuro-QOL¹³ were included to allow for the future development of crosswalks with these legacy instruments. The content model, based on the World Health Organization International Classification of Functioning, Disability & Health (ICF)¹⁴ was used to organize items by category and to identify redundant items and content gaps.

Redundant items were removed and new items were written to address the identified gaps, incorporating content expert input and the published literature to inform work relevance.

Instrument Construction

The initial instrument developed for field testing consisted of a pool of items representing a wide variety of physical activities about which claimants were asked to indicate their “usual ability during a typical day,” without the help of another person or equipment or devices not mentioned in the question. For example, “Are you able to *push a full wheelbarrow?*” Response options included; “*Yes, without difficulty; yes, with a little difficulty; yes with some difficulty; yes, with a lot of difficulty; unable to do; and I don’t know.*” Cognitive interviews were conducted by administering the items to persons with disabling physical impairments and eliciting their interpretation of the questions and decision process for choosing a response. Results of cognitive interviews informed item deletion or rewriting. Following these changes the final item pool was ready for administration to participants in the field study.

Field Study to Examine Item Pool Structure

Participants and Sampling—A large cross-sectional field study of the items developed to assess physical function was conducted. Data were collected from 2 samples: a claimant sample, which consisted of adults that had submitted a claim for Social Security disability benefits within a 3-month period, and a normative sample, consisting of U.S. adults from an opt-in internet respondent pool.

Claimant and Normative Samples—The claimant sample was selected from an initial pool of 10,000 applicants who applied to the Social Security Disability Insurance or Supplemental Security Income programs within a 3-month time period. Data were collected by the survey research firm Westat, Inc. Stratified random selection from the initial pool was conducted across the 10 SSA geographic regions and by urban or rural designation. Mailings of study information and consent materials were sent to 7,800 claimants and follow-up telephone calls were made with the recruitment target of 1,000 participants. Westat staff screened potential participants via the telephone using the following eligibility criteria: 21 years of age or older; and ability to read and understand English and that the primary allegation included a physical condition.

To allow comparison of study results from claimant data to those from a general population sample, a normative sample of U.S. adults was selected from an opt-in internet respondent pool of greater than 1 million respondents. A proximity sample matching technique developed by YouGov Inc. was used to target a normative sample of 1000 participants matched to the distribution of U.S. adults on sex, racial/ethnic background, age, and education.¹⁵

Data Collection Procedures—The full 139-item pool and demographic questions were administered to each subject in the claimant and normative samples. The normative sample self-administered the items via the internet and administration to the claimant sample was either via the internet or by a trained interviewer at Westat over the telephone. Demographic

questions included self-reported age, race, ethnicity, sex, marital status, and education level. Geographic location was coded as urban or rural based upon address. Quality control procedures included periodic monitoring of all aspects of recruitment, survey administration and data collection.

Data Analytic Procedures—Descriptive statistics were calculated for each demographic question and item, including the percentage of missing responses, and the frequency distribution for each item response category. The items containing response options with zero frequency were documented. The factor structure of the item pool was examined using exploratory factor analyses (EFA), confirmatory factor analyses (CFA) and expert content review. Factor analyses were conducted on a polychoric correlation matrix due to the ordinal nature of the item response categories. Analyses were conducted using Mplus.¹⁶

Beginning with the claimant data, we conducted the EFA followed by Geomin rotation to examine correlations across the factors. EFA models were estimated using Unweighted Least Squares methods. We assessed the eigenvalues and cumulative percentage of variance explained by one to four factors and the factor loading patterns for each model. Items with loadings < 0.4 were considered weak and were excluded from further analysis. Items loading on two factors with a loading > 0.4 on the second factor were removed. However, we also considered item content coverage during this process.

Following EFA, we conducted CFA. To guide decision-making for the final factor structure, we considered 1) the relationship between factor loading patterns and the hypothesized structure in the content model; 2) item content relative to the factor content (face validity); 3) reasonable fit on CFA fit indices; 4) strong (> 0.4) item loading on one factor;¹⁷ 5) discrimination between the factors (inter-factor correlations < 0.8);¹⁸ and 6) at least 6 items per factor. The 6 item requirement per factor was informed by tradeoffs in the benefits of CAT methods for scales with very few items, and the need for increasingly high item performance with decreasing numbers of items per scale. The CFA models were performed using robust weighted least squares ('weighted least squares with robust standard errors, mean- and variance-adjusted') estimation, and the factors' means and variances were constrained to 0 and 1 respectively to identify the model for estimation of the factor correlations.¹⁶

Based on the final number of factors selected, separate "uni-dimensional" CFAs were conducted on each factor to test the local independence assumption for IRT analyses in subsequent work. For physical function, this assumption holds that responses to different items are independent for subjects who have the same level of function. Inter-item residual correlations were examined to test for violation of the local independence assumption with residual correlation $> .20$ considered evidence of dependence and the basis for exclusion of the item.¹² Using the final set of items remaining after uni-dimensional CFAs, we conducted replication CFA of the 4-factor model. Finally, we used CFA to examine whether the data from the normative sample was consistent with the factor structure established in claimant sample.

To assess the fit of each model examined, we used 2 incremental fit indices, the Tucker-Lewis Index (TLI)¹⁹ and Bentler's Comparative Fit Index (CFI).²⁰ The Chi-square statistic divided by degrees of freedom (χ^2/df) and the difference between the chi-square statistics and the degrees of freedom were used to calculate TLI and CFI respectively, characterizing the difference between the proposed model and independent or worst case model, which assumes no correlations between items. We considered TLI and CFI values greater than 0.90 to represent good fit, greater than 0.95 to represent extremely good model fit.²⁰⁻²² We also computed the root mean square error of approximation (RMSEA),²³ an estimate of fit that considers complexity of the models. The RMSEA represents the difference between the proposed model and the saturated model, or perfect fit, which would result in a RMSEA of 0.00. Values of < .06 indicate very good fit, < .08 indicate reasonable fit, and > .08 indicates poor fit.^{21,22,24}

Results

The flow of item pool development is presented in Figure 1, beginning with 381 initial items identified from the literature review, of which 119 originated from PROMIS or NeuroQoL systems. After cognitive interviewing, 10 items were re-written and 8 were removed. The final item pool used for the calibration field test is shown in Figure 2, consisting of 139 items, including 24 items specific to wheelchair and walking aid use.

The characteristics of the claimant and normative samples are presented in Table 1. Among claimants the mean age was 49.7 years; 58.7% were white; 31.8% were black; 46.6% were female; 115 used a wheelchair; 439 used a walking aid; 578 reported upper extremity impairment. In the normative sample the mean age was 49.7 years; 78% were white; 11% were black; and 48% were female.

Walking aid and wheelchair-specific items were excluded from the factor analyses because responses were not available for the entire sample for those items; results of IRT analyses for these items are reported in a companion paper.¹⁰ Results of EFA/CFA on the remaining 115 items indicated that both the 69 item 3-factor and 82-item 4-factor solutions provided strong fit statistics; however, the 4-factor model included 12 more items and allowed separate characterization of overall upper body function and fine motor abilities (Figure 2). The factors identified and their content are: Changing and Maintaining Body Position, which includes the ability to assume, maintain and transfer among various positions such as lying, kneeling, sitting, squatting and standing; Whole Body Mobility, which includes the ability to move around from one place to another including crawling, walking and running; Upper Body Function, which entails reaching, lifting, pulling, pushing, and carrying; and Upper Extremity Fine Motor, which includes manipulation of objects requiring dexterity. After adding 11 items based on content, the 4-factor model, comprised of 93 items, demonstrated very good fit (Table 2).

Eigen values associated with the 4-factor solution were [54.84, 11.87, 4.47, 3.47] and cumulative percentage of variance explained was 64.9%. The factor pattern matrix is provided in Appendix A. The results of replication CFAs for each factor are shown in Table 3. The amount of variance explained by the first factor for each was: 58% (Changing &

Maintain Body Position), 62% (Whole Body Mobility), 68% (Upper Body Function), and 69% (Upper Extremity Fine Motor). The ratio between the first and second eigenvalue was: 8.21 (Changing & Maintain Body Position), 8.02 (Whole Body Mobility), 12.72 (Upper Body Function), and 20.81 (Upper Extremity Fine Motor). High residual correlation (>0.20) with all other items were identified for 2 items in the Upper Extremity Fine Motor factor and the items were removed (“How long are you able to use a computer keyboard?” and “How long are you able to use a computer mouse?”). In 3 of the scales, locally dependent item pairs were identified. To optimize content coverage, we chose to keep them in the item pool, using “enemy item” programming in the CAT setting to prohibit administration of the second item in a locally dependent pair to the same person once the first one has been administered. Replication CFA results, shown in Table 2, indicated improved fit after the 2 items were removed. After factor analyses the item pool consisted of 115 items, including 89 new items developed for this study, 14 PROMIS/Neuro-QOL items and 12 from other sources; 24 of the total were walking aid or wheelchair-specific items.

The results of the CFA for the normative sample were consistent with those of the claimant sample (Table 4).

Discussion

This paper reports on the initial stages of the development of an instrument to measure physical functioning, the Social Security Administration Physical Function Instrument (SSA-PF). Factor analyses of the item pool using data from a sample of 1,017 SSA claimants and a normative sample of 999 U.S. adults revealed 4-factor structure. The practical implication of this finding is the construction of 1 scale for each of the 4 factors plus a wheelchair mobility scale. This will allow the user to obtain a Physical Function score profile characterizing a claimant’s functional level in Changing & Maintaining Body Position; Whole Body Mobility; Upper Body Function; Upper Extremity Fine Motor; and Wheelchair Mobility.¹⁰ The final SSA-PF item pools provide the foundation for enhancing efficiency and performance of the process to accumulate medical evidence to support determinations for SSA disability programs.

The factor structure of the Physical Function item pool closely resembled the hypothesized content model, which included 3 domains: Changing & Maintaining Body Position, Whole Body Mobility, and Carrying, Moving & Handling Objects.¹¹ Rather than one factor covering Carrying, Moving & Handling Objects, we found 2 separate factors covering this content. Other research has reported 1 or more factors underlying the physical function construct.^{25–28} PROMIS researchers, for example, reported the development of a single general physical function scale; however, this required the removal of items addressing upper extremity function and assistive devices use.²⁷ The claimant data revealed better fit for the 3 and 4-factor models than the 1 and 2 factor models, and the fit for the 4-factor model was supported in the data from the normative sample. These findings are similar to those of a recent project to develop a physical function instrument for persons with spinal cord injury, the Spinal Cord Injury –Functional Index (SCI-FI).²⁹ The factors found in the SCI-FI were similar to those revealed in this study, except that the SCI-FI included a factor for self-care where the Physical Function Item Pool included Upper Body Function.

The present findings are also similar to those of earlier work in the development of a measure for older adults, the Late-Life Function & Disability Instrument, in which a 3-factor model demonstrated acceptable fit while a 1-factor model did not.³⁰ There were 2 lower extremity/mobility factors and one upper extremity factor within the 3-factor model, and fewer poorly fitting items. In summary, the evolving body of research work investigating the factor structure of the physical function domain demonstrates increasing evidence to support 2–4 factors. This multi-factor structure provides the important advantage of offering claimant profiles that characterize ability across key subdomains of physical function relevant to work.

Key strengths of the item pool development process in this project include the incorporation of important advances in the conceptualization of disability such as the ICF,¹⁴ and knowledge gained from research and methodological advances in physical function and disability evaluation.^{31–33} In addition, content expert and stakeholder input informed the development of items relevant to work and framed within everyday environments, eliminating the need to ask claimants about workplace-specific tasks that they may not have performed for several months.

The sampling strategy in this study provided diverse geographic representation of the 10 SSA administrative regions as well as urban and rural location. In addition, the claimant sample included representation of persons using wheelchair and assistive devices, and those with upper extremity impairments. The diversity of claimant characteristics provides a sound basis for interpretation of the results of this study in the context of SSA disability programs. The inclusion of data from a normative sample further strengthens the study by providing validation of the item pool factor structure sample among persons with a broader range of physical functioning.

The majority of the items in the item pool were new items. Although there is overlap between the content in the Physical Function Item Pools and that of existing instruments found in the literature, our developmental work indicated a need to generate new items and rewrite existing items in order to make the instrument informative in the context of work disability. The chief limitation of existing CAT measures of physical functioning found were that they focused on self-care from the perspective of health care outcome assessment, with limited focus on work-relevant aspects of functioning, such as the duration, repetition, distance, and body position required to perform the activity.³³ While developing a new instrument offers substantial potential for enhancing physical function measurement related to disability, it necessitates future research to investigate the instrument's measurement properties.

Study Limitations

The results of the replication unidimensional CFA's (Table 3) were mixed regarding fit, with relatively high RMSEAs and chi-square/df values. However, the body of evidence produced by this study support the unidimensional 4-factor solution, including CFI/TLI results, 4-factor CFA results (Table 2); no residual correlation >0.2; the large amount of variance explained by the first factor for each scale (58–69%), and a ratio between the first and second Eigen value greater than 4 (8.0 – 20.8).³⁴ In addition, it has been demonstrated

that RMSEA is closely related to factor loading. In a high factor loading model (e.g. high reliability test), even with a small residual covariance, a large RMSEA may occur.^{35, 36} In the Physical Function scale, most of the items demonstrate high loading on a unidimensional scale; therefore, the higher RMSEA may be caused by the higher reliability of the scale and the small model error. The traditional cutoff for fit indices has been questioned, and literature suggests a greater appreciation for the incremental fit indices compared to RMSEA alone, in such situations.^{37,38}

Our CFA revealed items that demonstrated local dependence. One solution would have been to drop one of the items in each locally dependent pair. For reasons of content coverage, we elected to retain these items in the item pool. We plan to use “enemy item” programming within the CAT development phase, which will avoid administration of both locally dependent items to the same person. This approach allows the use of the maximum number of items in administrations across different claimants. Although the results from these analyses of the Physical Function item pool are encouraging, additional research is needed to assess the measurement properties of the instrument as developed.

Conclusions

The current work integrates major advances in the conceptualization of disability and in measurement methodology to provide a contemporary and comprehensive representation of physical functioning for use within SSA disability programs. The development of the SSA-PF instrument is a critical step toward the important goal of integrating standardized functional information into the disability determination process within Social Security Disability programs. The item pool development methods incorporated stakeholder and expert input, recent conceptual advances, existing items and knowledge regarding physical function, work, and disability. Comprehensive factor analysis of the items in claimant and normative samples revealed a multi-factor structure similar to the hypothesized content model, and informed subsequent IRT analyses and CAT development study.¹⁰ The SSA-PF covers a wide range of physical function content relevant to work activities, offering promise for providing reliable information relevant to disability determination decision-making.

Acknowledgments

ACKNOWLEDGMENT OF FINANCIAL SUPPORT: 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. Dr. McDonough was funded by a New Investigator Fellowship Training Award from the Foundation for Physical Therapy.

List of Abbreviations

CAT	computer-adaptive test, computer-adaptive testing
CFA	confirmatory factor analysis
CFI	comparative fit index
EFA	exploratory factor analysis

ICF	World Health Organization's International Classification of Functioning, Disability & Health
IRT	item response theory
PROMIS	Patient-Reported Outcomes Measurement Information System
RMSEA	root mean square error of approximation
SCI-FI	Spinal Cord Injury – Functional Index
SSA	Social Security Administration
SSA-PF	Social Security Administration Physical Function Instrument
TLI	Tucker-Lewis Index

References

1. Social Security Administration. Annual Statistical Supplement to the Social Security Bulletin, 2011. Office of Retirement and Disability Policy and Office of Research Evaluation and Statistics; 2011. p. 1-3.
2. Brandt DE, Houtenville AJ, Huynh MT, Chan L, Rasch EK. Connecting contemporary paradigms to the Social Security Administration's Disability Evaluation Process. *Journal of Disability Policy Studies*. Sep; 2011 22(2):116–128.
3. IOM (Institute of Medicine). Improving the Social Security disability decision process. Washington, D.C: 2007.
4. Cella D, Gershon R, Lai J, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Quality of Life Research*. 2007; 16 (Suppl 1): 133–141. [PubMed: 17401637]
5. US Department of Health and Human Services FDA. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. Silver Spring: US Department of Health and Human Services Food and Drug Administration; 2009.
6. Wainer, H. Computer Adaptive Testing: A Primer. Hillsdale, NJ: Lawrence Erlbaum Associates; 2000.
7. Lord, F. Applications of Item Response Theory to Practical Testing Problems. Hillsdale, NJ: Erlbaum Associates; 1990.
8. Hambleton, R.; Pitoniak, M.; Pashler, H., editors. Testing and measurement. Advances in item response theory and selected testing practices. NY, NY: John Wiley & Sons, Inc; 2002.
9. Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Research*. 2007; 16 (Suppl 1):133–141.
10. Ni PS, MC, 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 Rehabil*. 2013; XX(XX):XX–XX.
11. Marfeo EE, HS, Jette AM, Eisen SE, et al. A conceptual foundation for measures of physical function and behavioral health function for Social Security work disability evaluation. *Arch Phys Med Rehabil*. 2013; XX(XX):XX–XX.
12. Reeve BB, Hays RD, Bjorner JB. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care*. May; 2007 45(5 Supple 1):S22–31. [PubMed: 17443115]
13. 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. *Quality of Life Research*. Apr; 2012 21(3):475–486. [PubMed: 21874314]

14. World Health Organization. International Classification of Functioning, Disability and Health (ICF). Geneva: 2001.
15. Rivers, D. A white paper on the advantages of the sample matching methodology. Palo Alto, CA: Sample matching: Representative sampling from internet panels.
16. Mplus Statistical Analysis with Latent Variables User's Guide [computer program]. Los Angeles, CA: Muthén & Muthén; 2007.
17. Hair, JF.; Anderson, RE.; Tatham, RL.; Black, W. Multivariate Data Analysis. New Delhi: Prentice-Hall; 1998.
18. Kline, RB. Principles and Practices of Structural Equation Modeling. 2. New York: The Guilford Press; 2005.
19. Tucker L, Lewis C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*. 1973; 38:1–10.
20. Bentler P. Comparative fit indices in structural models. *Psycho Bull*. 1990; 190(107):238–246.
21. Hu, L.; Bentler, PM., editors. Evaluating Model Fit. Thousand Oaks, CA: Sage Publications; 1995. Hoyle, RH., editor. Structural Equation Modeling: Concepts, Issues and Applications.
22. Hu LT, Bentler P. Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55.
23. Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. Bollen, KA.; Long, JS., editors. Newbury Park, CA: Sage Publications; 1993. Testing Structural Equation Models
24. Steiger JH. Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research*. 1990; 25:173–180.
25. Jette AM, McDonough CM, Haley SM, et al. A computer-adaptive disability instrument for lower extremity osteoarthritis research demonstrated promising breadth, precision, and reliability. *Journal of Clinical Epidemiology*. Aug; 2009 62(8):807–815. [PubMed: 19216052]
26. Jette AM, McDonough CM, Ni P, et al. A functional difficulty and functional pain instrument for hip and knee osteoarthritis. *Arthritis Research & Therapy*. 2009; 11(4):R107. [PubMed: 19589168]
27. Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *Journal of Clinical Epidemiology*. Jan; 2008 61(1):17–33. [PubMed: 18083459]
28. Haley SM, Coster WJ, Andres PL, et al. Activity outcome measurement for postacute care. *Medical Care*. Jan; 2004 42(1 Suppl):I49–61. [PubMed: 14707755]
29. Jette AM, Tulskey DS, Ni P, et al. Development and initial evaluation of the Spinal Cord Injury-Functional Index. *Arch Phys Med Rehabil*. 2012; XX
30. Haley SM, Jette AM, Coster WJ, et al. Late Life Function and Disability Instrument: II. Development and evaluation of the function component. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences*. Apr; 2002 57(4):M217–222.
31. Gaudino EA, Matheson LN, Mael FA. Development of the Functional Assessment Taxonomy. *Journal of Occupational Rehabilitation*. 2001; 11(3):155–175. [PubMed: 11822194]
32. Matheson LN. Disability Methodology Redesign: Considerations for a new approach to disability determination. *J Occup Rehabil*. 2001; 11 (3):135–141. [PubMed: 11822192]
33. Occupational Information Development Advisory Panel (OIDAP). Content Model and Classification Recommendations for the Social Security Administration Occupational Information System. Social Security Administration; 2009.
34. Hattie J. Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological*. 1984; 20:1–14.
35. Browne MW, MacCallum RC, Kim C-T, Andersen BL, Glaser R. When fit indices and residuals are incompatible. *Psychological Methods*. 2002; 7:403–421. [PubMed: 12530701]
36. Saris WE, Satorra A, van der Veld WM. Testing structural equation models or detection of misspecifications? *Structural Equation Modeling*. 2009; 16:561–582.

37. Cook KF, Kallen MA, Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Qual Life Res.* May 18.2009 :447–460. [PubMed: 19294529]
38. Miles JNV, Shevlin M. A time and a place for incremental fit indices. *Personality and Individual Differences.* 2007; 42:869–874.

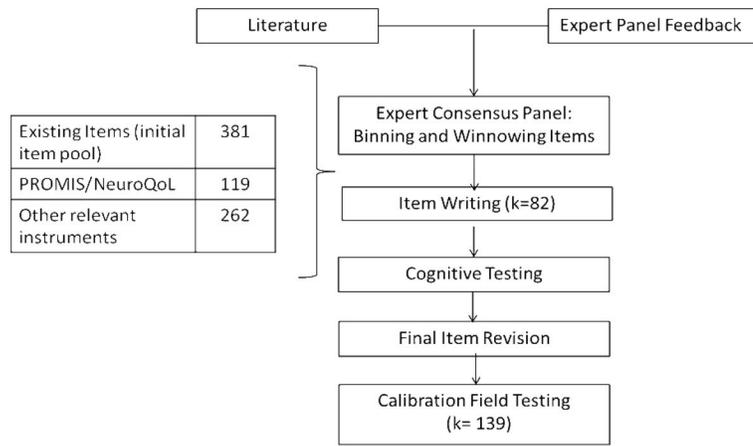


Figure 1.
Physical Function Item Pool Development Process

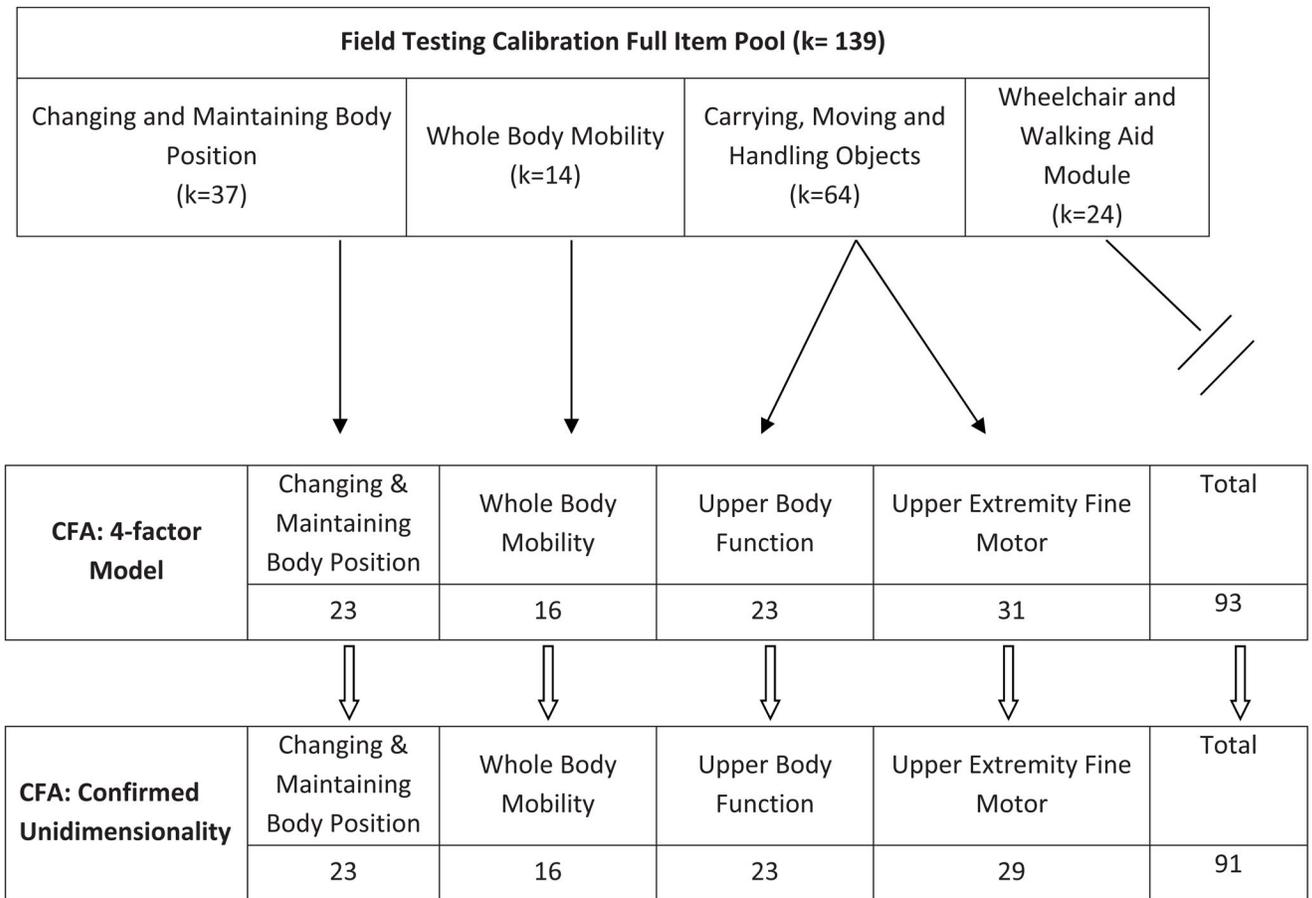


Figure 2.
Physical Function Final Model

Table 1
Background Characteristics of the SSA Claimant and Normative Study Samples

Values are expressed as Mean \pm SD or N (%)

Characteristic	SSA Claimants (N= 1017)		Normative Sample (N=999)	
	N	%	N	%
Age mean \pm SD	49.65 \pm 9.85	-	49.72 \pm 16.12	N=983
Under 40	167	16.42	260	26.45
40–55	485	47.69	289	29.40
55+	365	35.89	434	44.15
Sex				996
Female	474	46.61	480	48.19
Male	543	53.39	516	51.81
Geography				
Urban	682	67.06	-	
Rural	335	32.94	-	
Race				
White	597	58.7	782	78.28
Black/African American	323	31.76	110	11.01
Other	63	6.2	105	10.51
missing	34	3.34	2	0.20
Marital Status				
Never married	230	22.62	215	21.52
Married/partner	424	41.69	581	58.16
Divorced/separated	314	30.98	138	13.81
Widowed	45	4.42	54	5.41
missing	4	0.39	11	1.1
Education				
Less than high school	199	19.57	40	4
High School/GED	397	39.03	361	36.14
Greater than high school	419	41.2	591	59.16
missing	2	0.2	7	0.7
Primary Complaint				
Physical	982	96.56	-	
Physical & Mental	35	3.44	-	
missing	0	0	-	

Table 2

Confirmatory Factor Analyses for 4-factor Physical Function domain

	# of items	CFI	TLI	RMSEA
Initial CFA on 4-factor model	93	0.915	0.913	0.057
Replication CFA	91	0.925	0.923	0.054

Table 3
Goodness-of-fit Indices for Unidimensional CFAs for the Final Physical Function Item Pools

Scale	# of items	Chi-square/df	CFI	TLI	RMSEA	Removed Items*
Changing & Maintaining Body Position	23	6312.232/230	0.862	0.848	0.161	
	16	1357.128/103	0.952	0.944	0.109	PD201; PD082; PD447; PD450; PD448; PD436; PD453
Whole Body Mobility	16	1350.153/104	0.904	0.930	0.113	
	14	833.574/76	0.954	0.945	0.103	PD512; PD245_142
Upper Body Function	23	4021.379/230	0.943	0.937	0.127	
	19	1620.320/151	0.969	0.964	0.098	PD496; PD259; PD505; PD497
Upper Extremity Fine Motor	29	2503.832/376	0.973	0.971	0.075	

* Items that were locally dependent were removed for CFA. These items were kept in the scale, with programming to avoid administration to the same person.

Table 4

Confirmatory Factor Analyses Results for the Normative Sample (n= 999) and the Claimant Sample (n=1,017).

4-factor Model			
Sample	CFI	TLI	RMSEA
Normative Sample	0.977	0.977	0.042
Claimant Sample	0.925	0.923	0.054