THE WD-FAB: DEVELOPMENT AND VALIDATION TESTING

Leighton Chan MD, MPH
Chief, Rehabilitation Medicine Department
National Institutes of Health Clinical Research Center

2 March 2018

Acknowledgement

This research was funded through an Inter-Agency Agreement with the U.S. Social Security Administration and by the Intramural Research Program of the NIH, Clinical Research Center.

National Institutes of Health
Clinical Center

Boston University School of Public Health
Health & Disability Research Institute

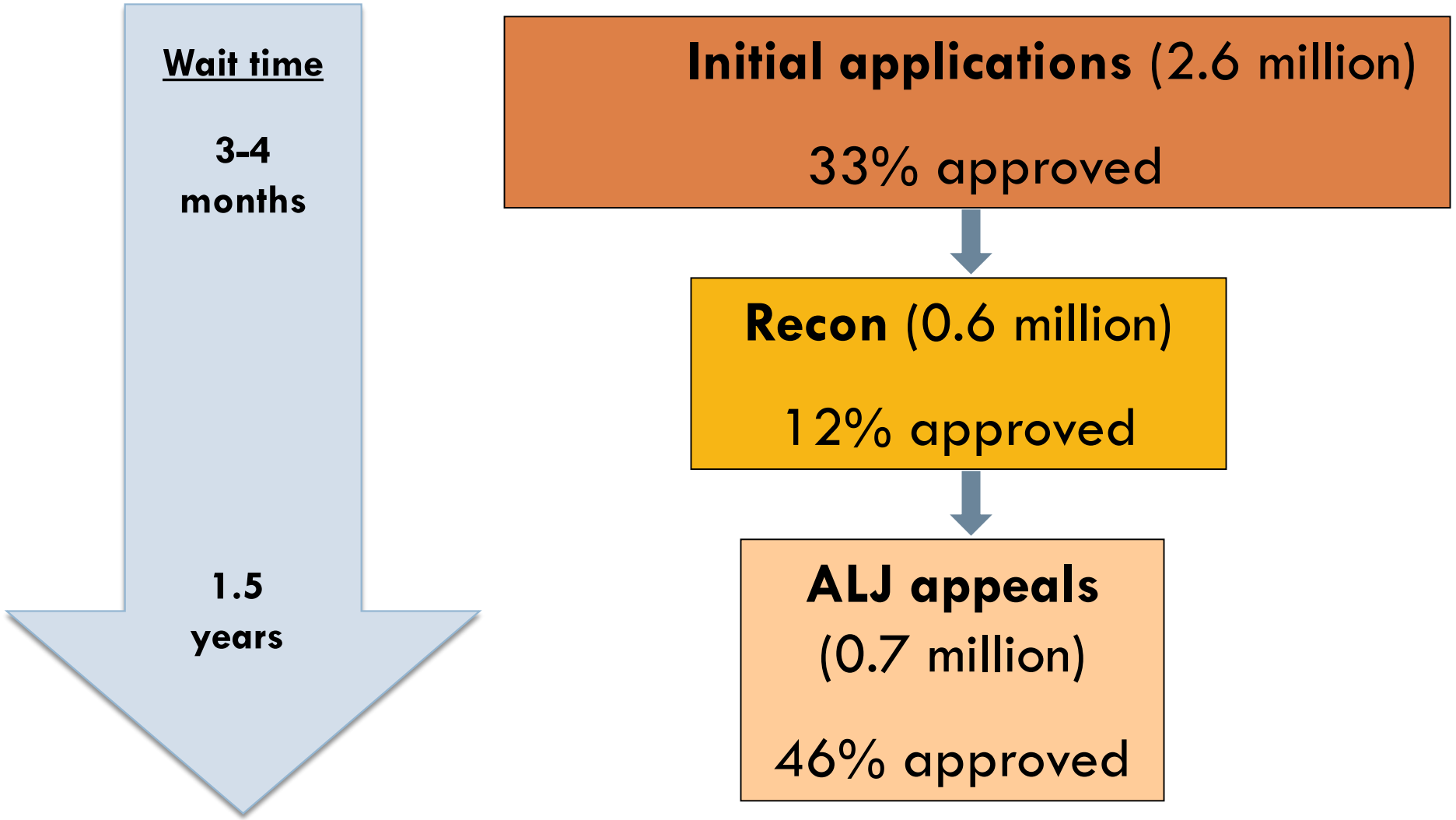
Why Build a New Disability Instrument?

- Comprehensive & feasible assessment for use on a large scale
- Computer-based: much more efficient than traditional measures
- User friendly: reduces irrelevant questions
- Can provide required accuracy & precision
- Allows for identification of aberrant response patterns to protect against intentional misreporting
- Instrument can be replenished & improved

U.S. Social Security Administration

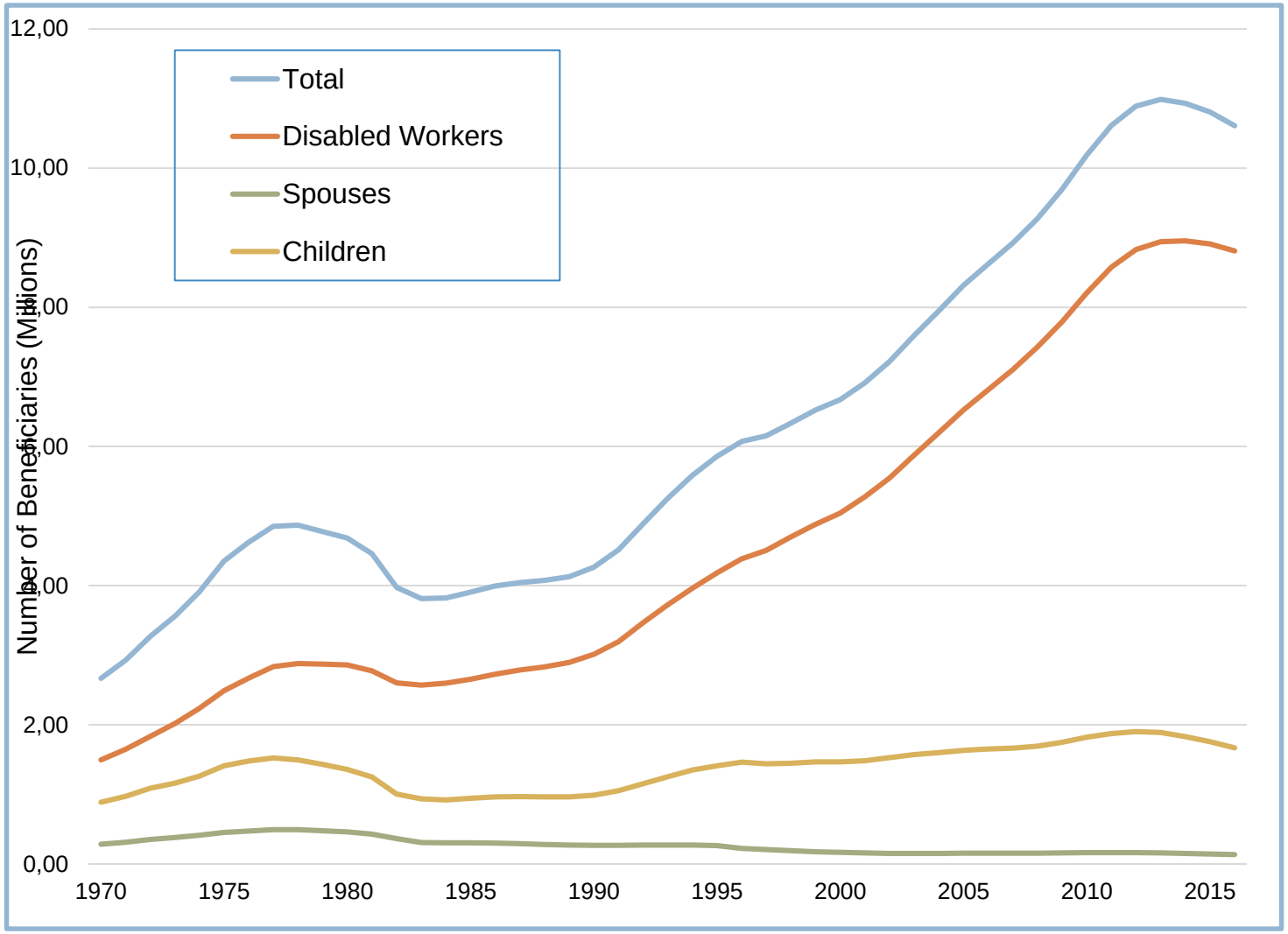
- Serves over 19 million adults and children
  - ~2-3 million new applications/year
- Benefits
  - 700-1700 USD per month (75% of income)
  - Health Insurance
- Annual costs ~ 187 billion USD
- Programmatic Definition of Disability:
  - Based on impairment and income (SGA)
  - 2017 SGA: 1170 USD/month
  - ‘All or nothing’ – no partial disability

Approval Rates



Escalating Demand

- Disability Programs pressured by rise in applications
- Projected to run out of funds in 2034*
  - *Disability and Retirement combined



SSA-NIH Collaboration

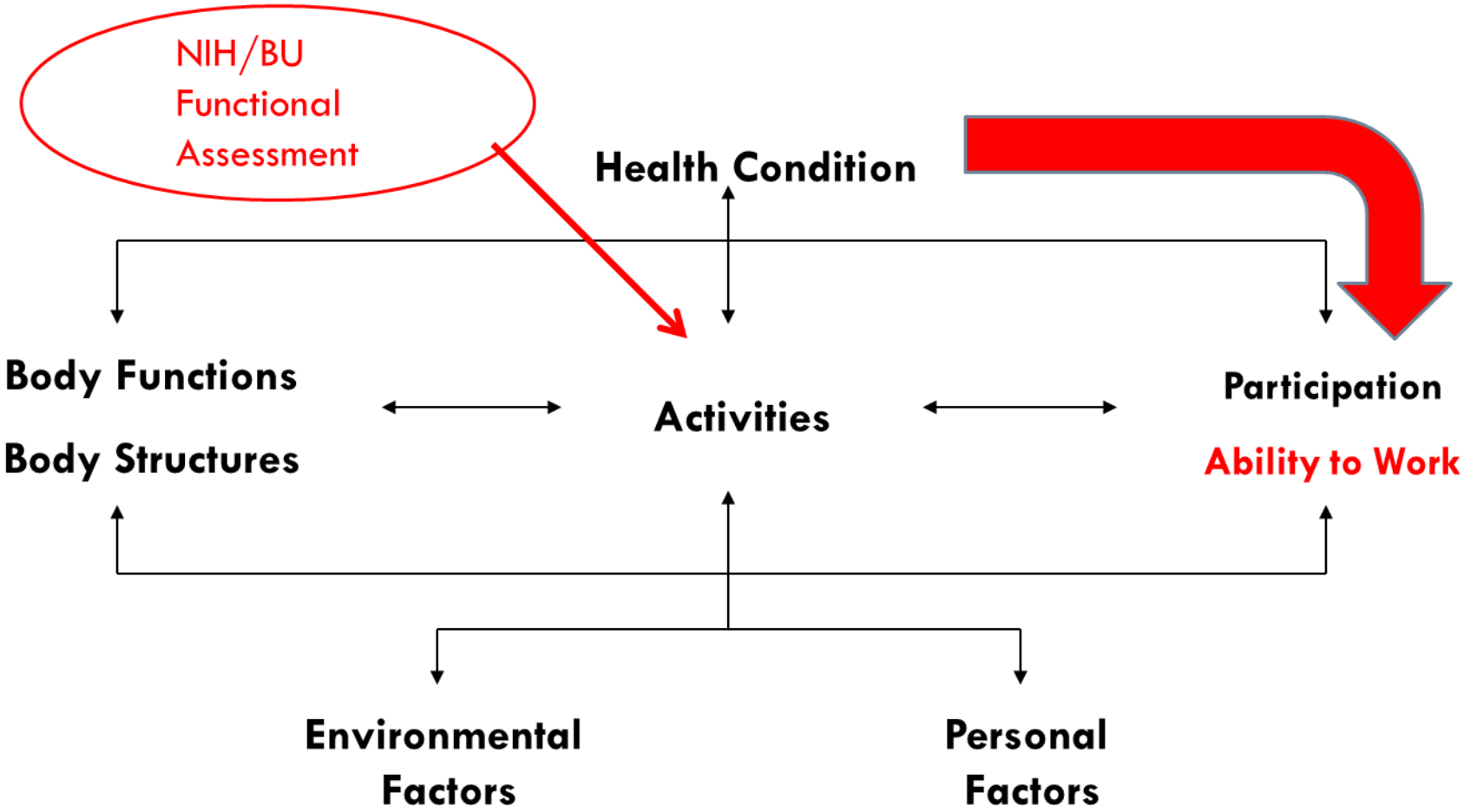
- SSA-NIH Interagency Agreement established in 2008
  - Initial objective:
    - Assess feasibility of developing functional tests to:
      - Improve standardization, completeness, & comprehensiveness of medical evidence on function
      - Collect functional data early in the decision making process
      - Improve program efficiency & effectiveness

Framing the Problem

- In the U.S., the historical focus in work disability determination has been on assessing physical and mental impairments
- Diagnosis and impairment alone are frequently poor predictors of work disability

- Contemporary disability models such as the ICF depict disability as the gap between an individual's functional capabilities and environmental demands

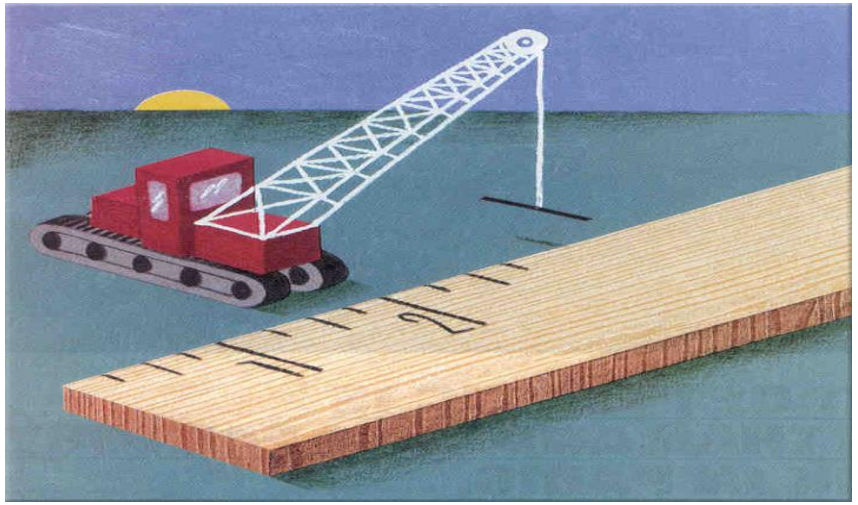
The WD-FAB and the ICF



ICF: World Health Organization, 2001

Methodology: Item Response Theory (IRT)

- Model the likelihood of a “correct” answer given the person’s ability level
- Questions are calibrated to a scale to cover the range of function in one dimension (e.g., mobility)
- Provides the platform for efficient administration using computer adaptive testing (CAT)



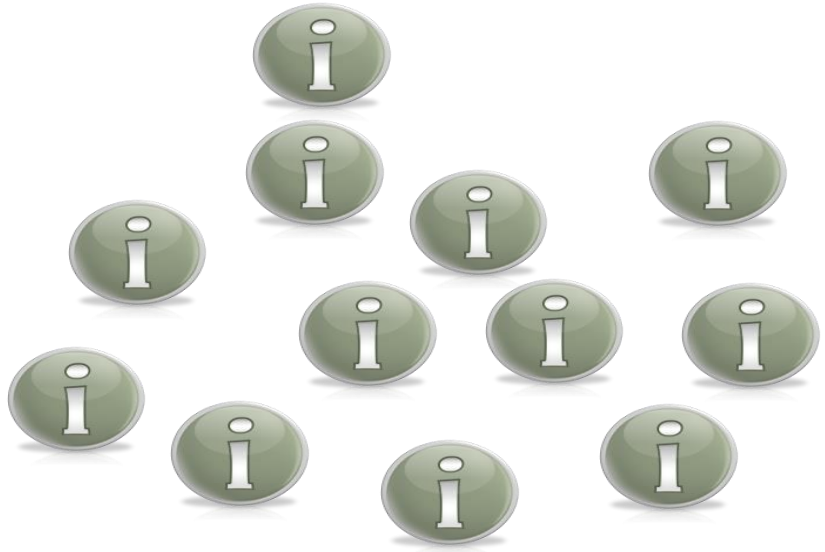
Methodology: Computer Adaptive Test (CAT)

CAT Instruments…

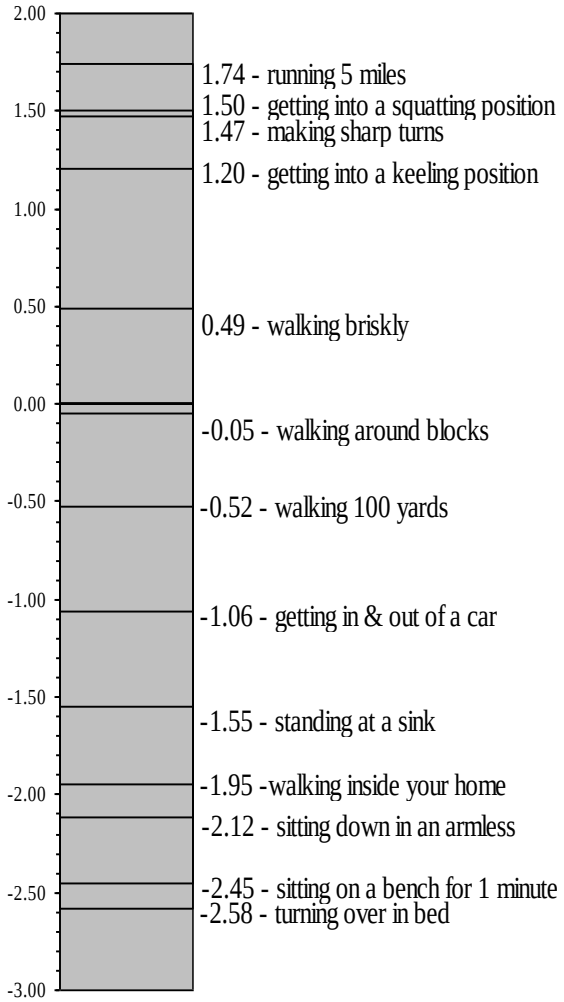
- Are highly efficient
- Administer small sample of questions from the IRT calibrated ‘item bank’
- Choose items based on how a person responds to previous items
- Stop when the person’s functional level is estimated to a pre-determined level of precision or a set number of items are answered

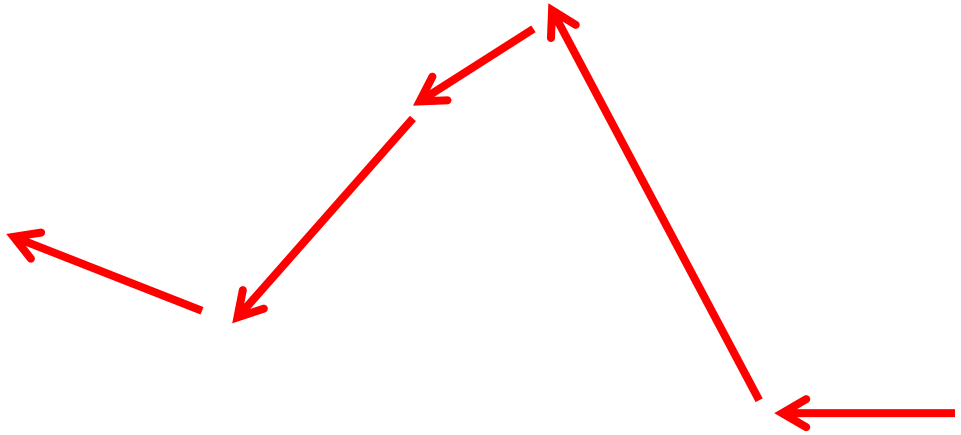
IRT and CAT methods create a tailored, individualized measure that best measures the ‘ability’ of that person

Physical Function

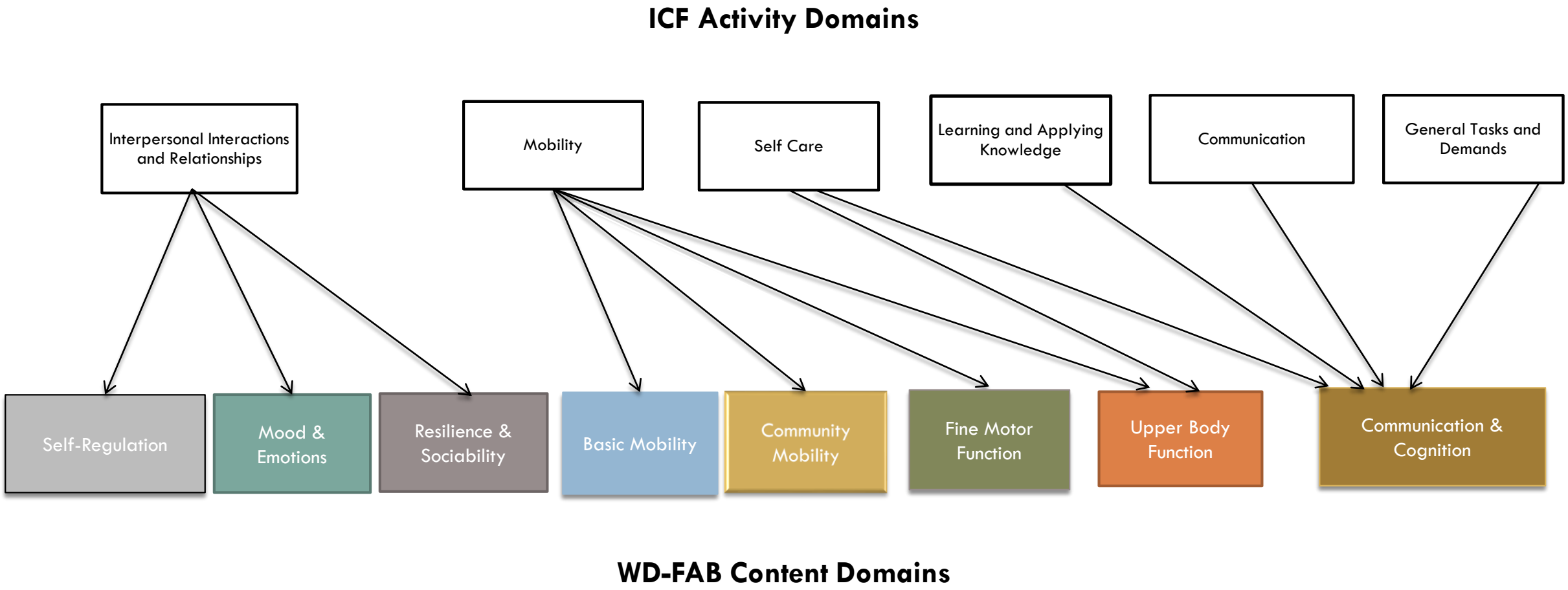


Example Physical Function Domain





ICF Content Coverage in WD-FAB

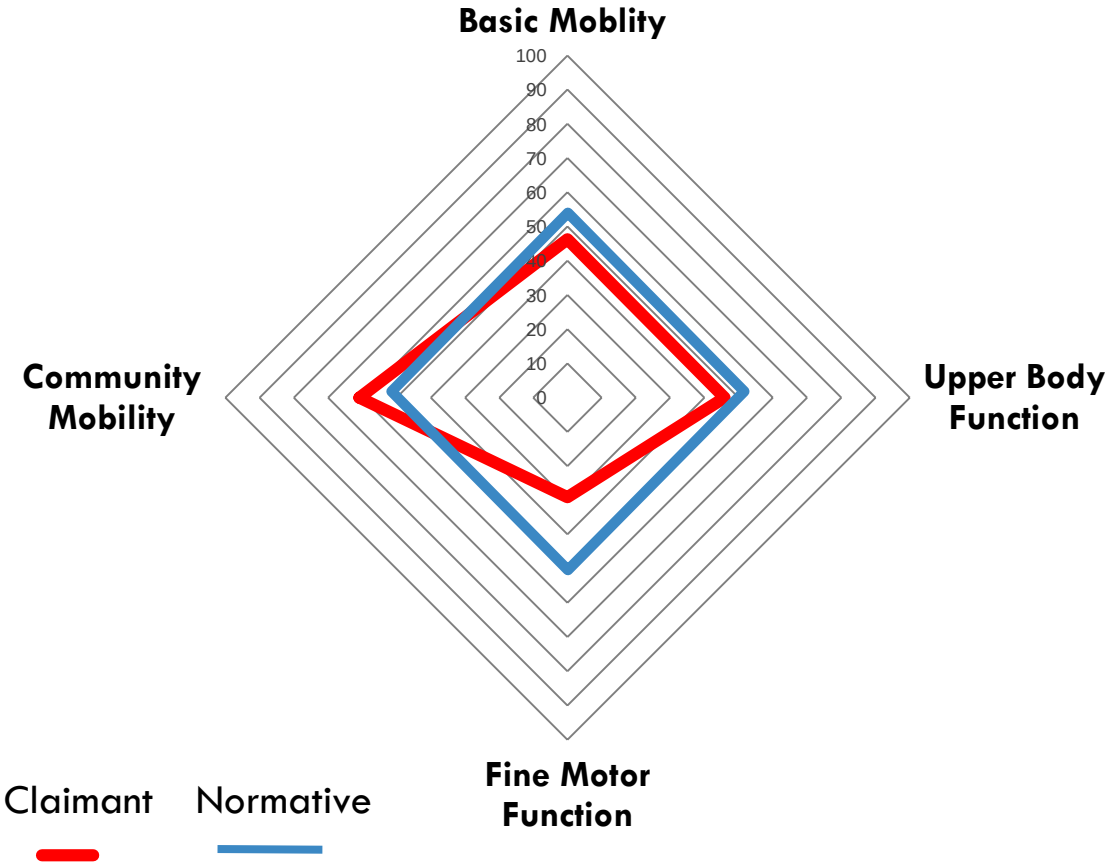


Sample WD-FAB Items

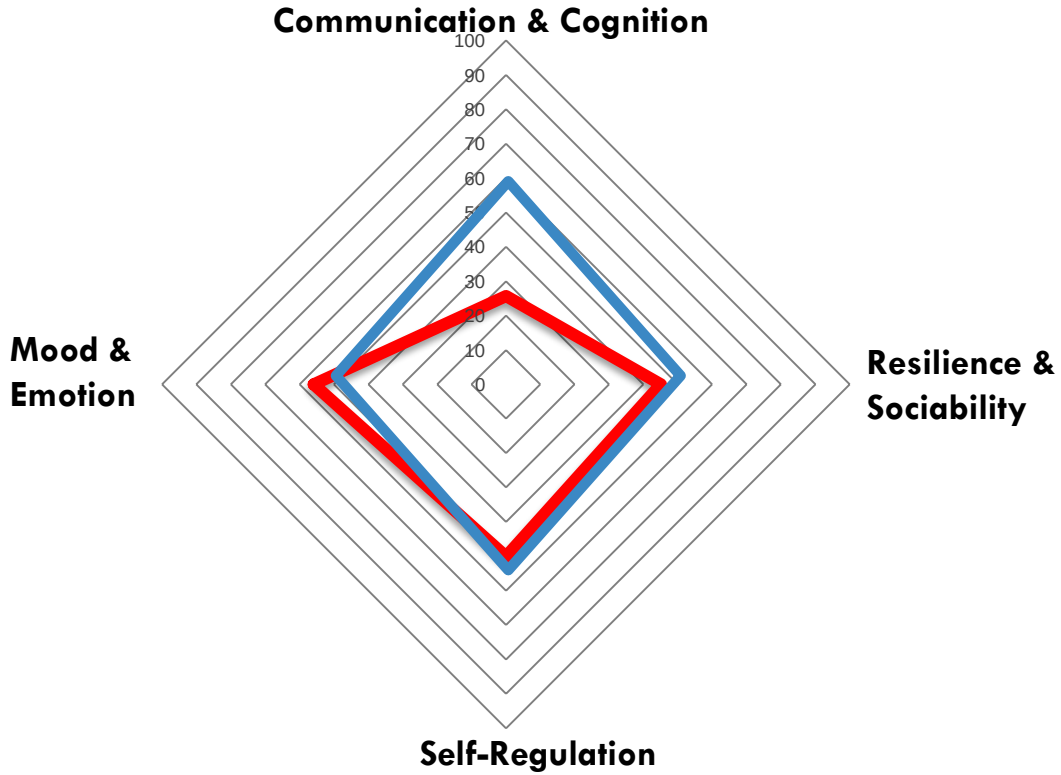
WD-FAB Scale	Item Content	Response Scale	Activity Measured
Basic Mobility	Are you able to bend to look under a car?	Ability (Yes, without difficulty; Unable to do)	Bending
Upper Body Function	Are you able to push open a heavy door?	Ability (Yes, without difficulty; Unable to do)	Pushing
Fine Motor Function	Are you able to remove a gas cap from a car?	Ability (Yes, without difficulty; Unable to do)	Turning or twisting the hands or arms
Community Mobility	Are you able to get on to a bus or train?	Ability (Yes, without difficulty; Unable to do)	Using public motorized transportation
Communication & Cognition	I have trouble putting my thoughts together.	Agreement (Strongly agree; Strongly disagree)	Thinking
Self-Regulation	I have difficulty following the rules.	Agreement (Strongly agree; Strongly disagree)	Interacting according to social rules
Self-Regulation	In the past 7 days I had trouble controlling my temper.	Frequency (Never; Always)	Regulation of emotion

WD-FAB Functional Profiles

Physical Function



Mental Health Function



Development of the WD-FAB

Scientifically Rigorous Development Process

- Used the *ICF* to conceptualize function
- Extensive literature review
- Focus groups with providers & individuals with disability
- Met with content experts
- Performed Cognitive Testing of all items to check clarity & comprehension
- Administered items to user groups

Field Calibration and Replenishment Studies

SSA Claimant Samples

- Phase 1 Sample (2011/12) 2,032 claimants
- Phase 2 Sample (2014/15) 3,720 claimants
- Phase 3 Sample (2016/17) 1,051 claimants

- Sample was from applicants in the past 2 months
- Stratified by urban/rural status across the 10 national SSA regions

General Population Sample

- Phase 1 Sample (2011/12) 1,999 adults
- Phase 2 Sample (2014/15) 2,025 adults
- Phase 3 Sample (2016/17) 1,000 adults

- Stratified by age, sex, race, ethnicity, US census region urban/rural classifications based on zip code, and education level
- Over sampled racial minorities for Differential Item Analysis (DIF)

Analysis of the WD-FAB

Three phase analytic approach:

1. Initial analyses completed in the claimant sample
2. Results replicated in the general population sample to allow score comparisons between claimants and US adults
3. Items merged from calibration studies

Analytic Steps:

- **Exploratory Factor Analyses (EFA):**
  - Establish underlying subdomain structure
- **Confirmatory Factor Analyses (CFA):**
  - Test uni-dimensionality of each subdomain

Differential Item Functioning (DIF)

Question: Do subjects at the same ability level respond differently to items?

1. Tested DIF by claimant age, gender, and race/ethnicity
2. Tested claimant sample vs. general population sample
3. Tested new WD-FAB items vs. existing items.

Validation of the WD-FAB

WD-FAB Development Studies

Tested the WD-FAB to see how it performs:

- Concurrent Tests
  - Comparison to legacy instruments
- Criterion Tests
  - Situational assessments
- Reliability
  - Test-retest
- Refinement & Replenishment
  - Item bank expansion
  - Ongoing replenishment
- User Simulation Study
  - 4 SSA Field Offices

Reliability and Validity of the WD-FAB

- Good test-retest reliability in adults with work-disability and general adult samples
- Low respondent burden
- Measurement accuracy:
  - Very high for Physical Function
  - More variability in Mental Health
- Convergent validity correlations with legacy measures were moderate to strong

WD-FAB Strengths

- Selects questions most relevant to the respondent
  - Low respondent burden
- Comprehensively assesses functional activity efficiently
  - Over 300 questions in Item Bank
  - Administered in ~15-20 minutes
- IRT/CAT instruments have been successfully translated into other languages

WD-FAB Strengths cont.

- User friendly
  - Administer in-person, over the phone, online, with paper/pencil via short forms as needed
- Track functional changes over time
- Item pools are not static
- Standardized and consistent assessment of function
- Instrument precision can be adjusted
- Thresholds for minimal detectable differences have been established

WD-FAB Limitation

- WD-FAB outcomes must be linked to workplace demand
  - WD-FAB measures at the activity level according to the ICF
  - Work disability must link activity (whole person functioning) to participation (work)
    - No known gold standard
    - A challenge confronted by all social security programs
    - Potential approach to this key issue:
      - Use WD-FAB to develop functional profiles by occupation

Potential Applications of the WD-FAB

- **Research**
  - Monitor function over time as an indicator of population health
  - Track influence of intervention strategies on functioning over time

- **Claimant support**
  - Who needs help? Identify functional profile thresholds and sociodemographic characteristics of program constituency.
  - Examine functional profiles relative to occupational demand to allow assessment of “fit”
  - Identify functional thresholds relative to inability to sustain work

Additional studies to further enhance WD-FAB

- Response pattern recognition
- Proxy respondent study
- Translation
  - Requires calibration to target population
- Replenishment
- Administration feasibility
- Optimize reporting for users

WD-FAB accomplishments

- Completion of WD-FAB prototype in 2016
  - Additional replenishment completed 2017
- Scientific validation – 14 publications in peer reviewed journals
- Now ready for testing/further development
- International collaborations?

WD-FAB Access and Translation

- WD-FAB hosted at Northwestern University Assessment Center
- Access available through Amazon Web Services
  - Global users will have local access
  - In Europe:
    - Current Availability Zones include Frankfurt, Ireland, and London
    - Plans to expand availability to France and Sweden
- Translation can also be done through the Assessment Center
  - Conceptual (semantic) translation following approved processes

WD-FAB Demonstration: https://ysurvey.alphce.com/

Q & A