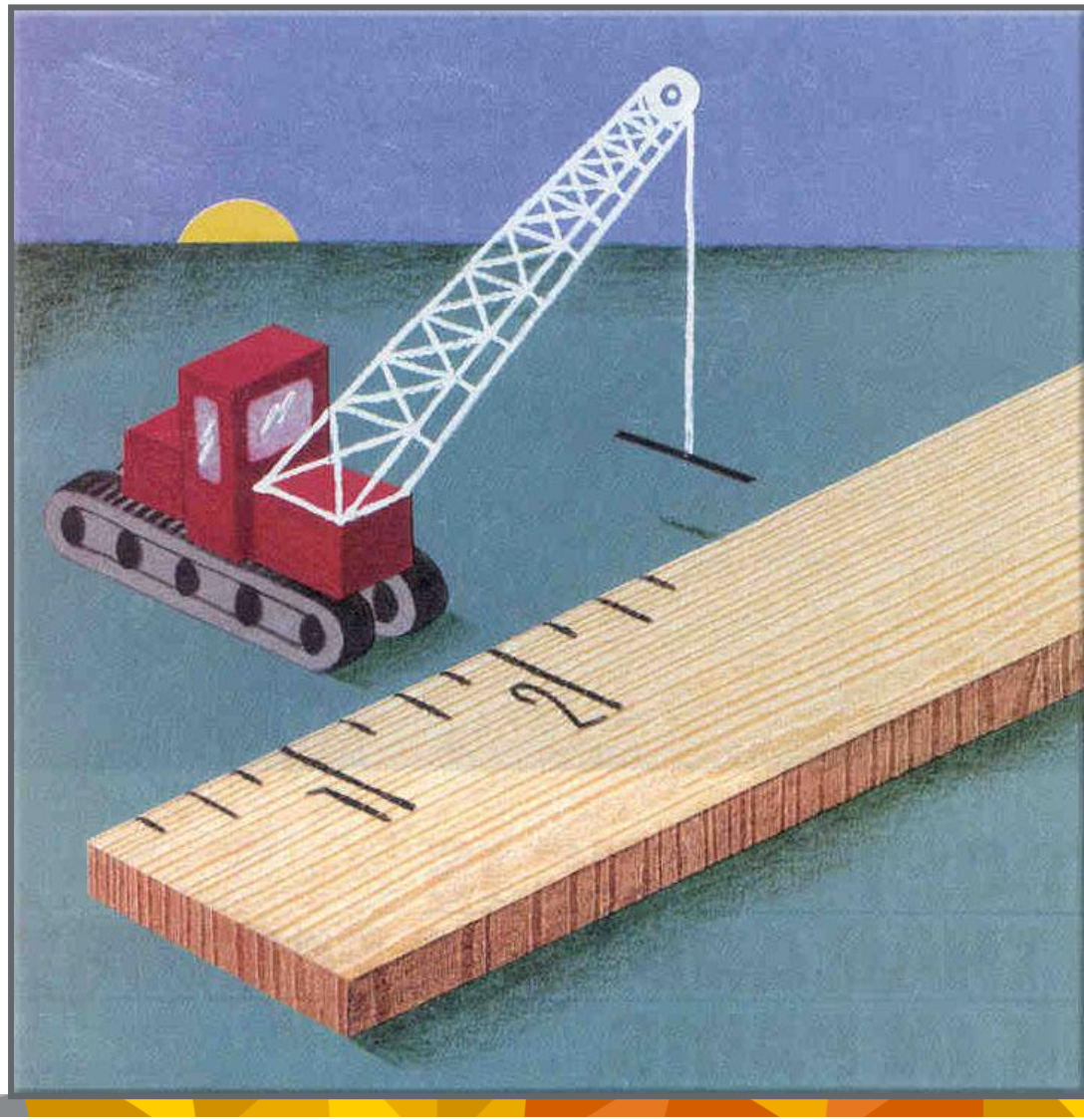
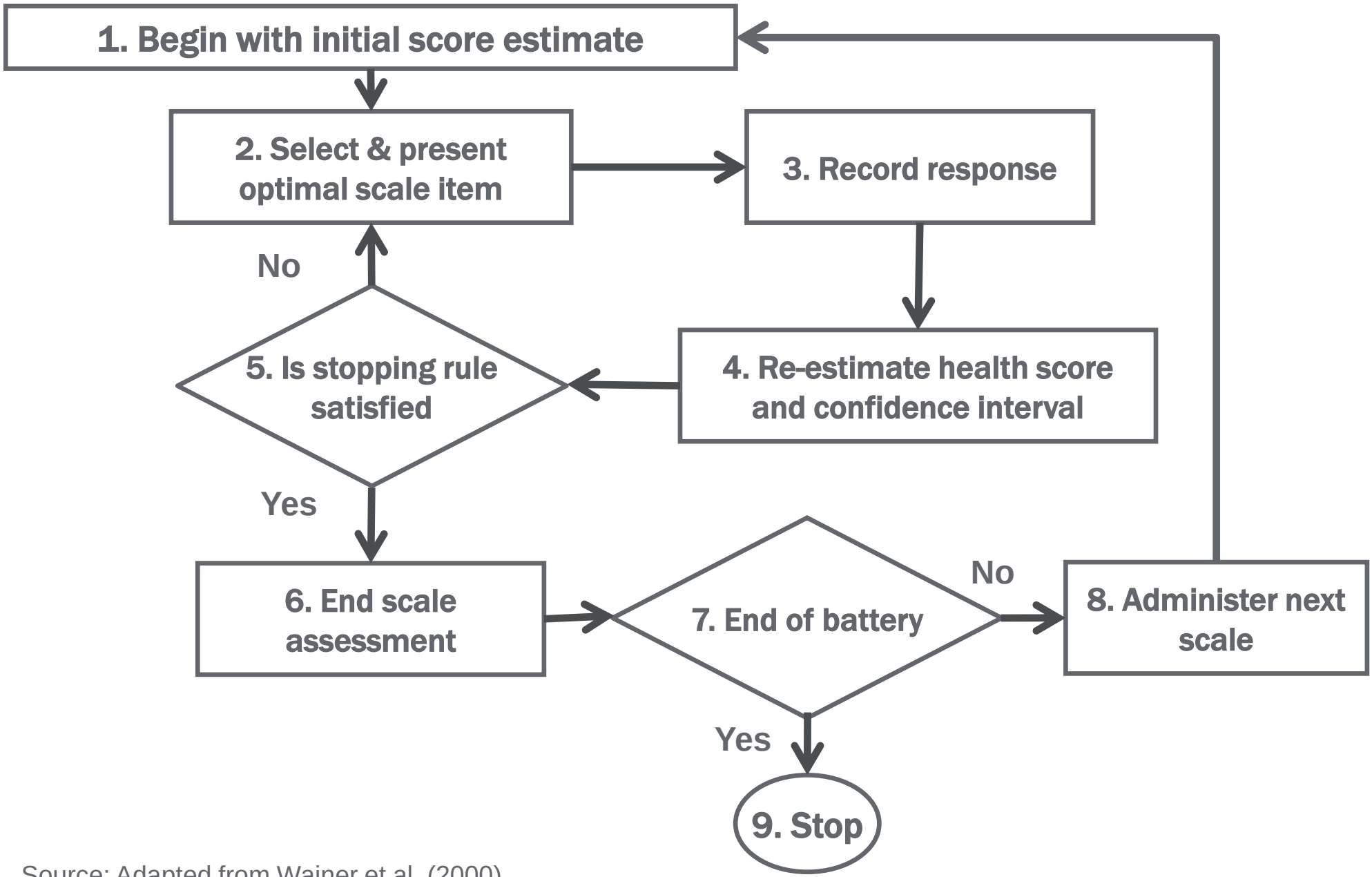
Tailored Instruments: Item Banking and Computerized Adaptive Assessment



Jakob Bue Bjorner, Optum PatientInsights / University of Copenhagen

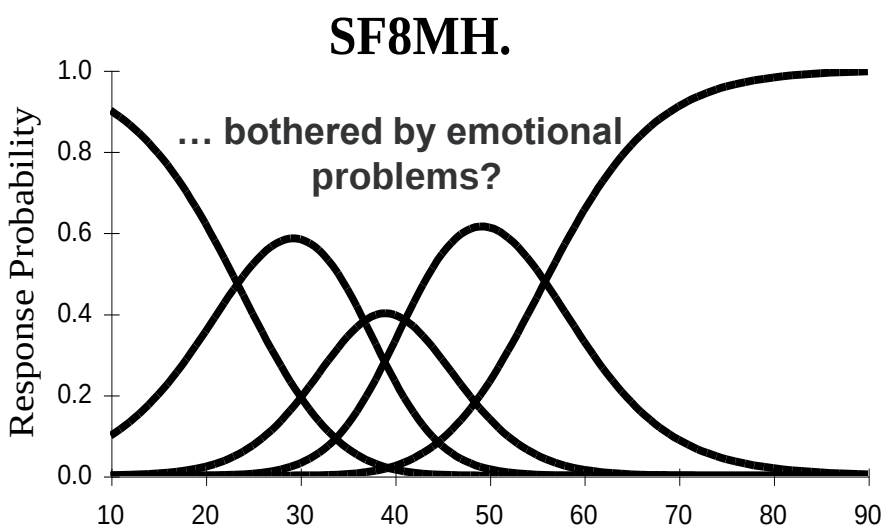
2nd Symposium on IRT in Health Measurement, University of Twente 17/9/2013

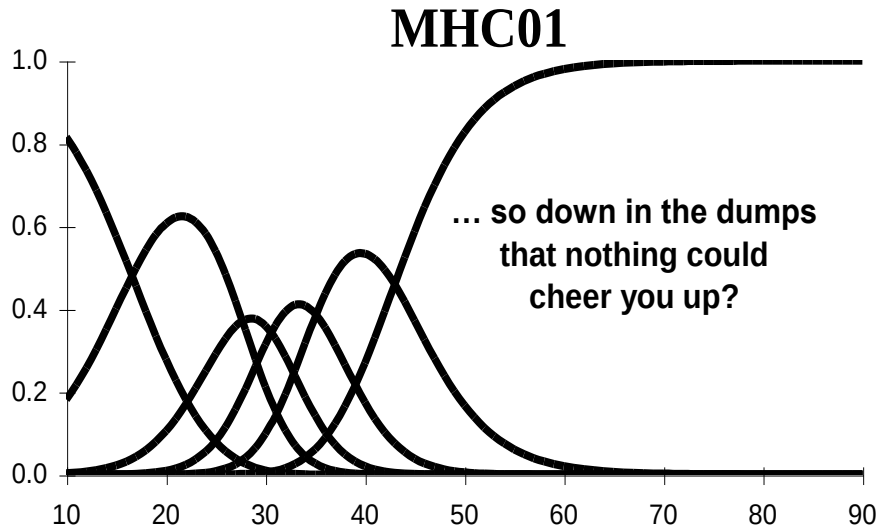
Logic of Computerized Adaptive Testing



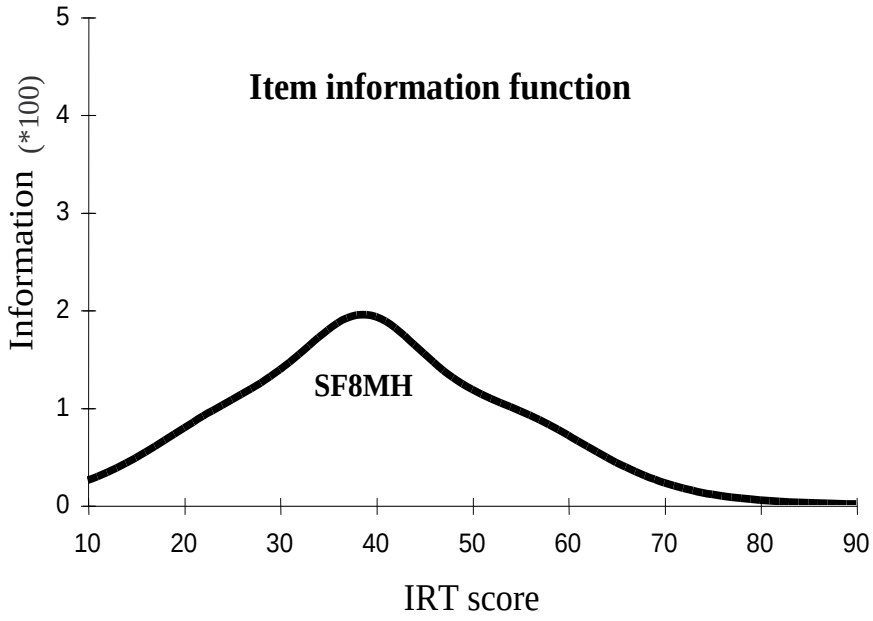
Source: Adapted from Wainer et al. (2000)

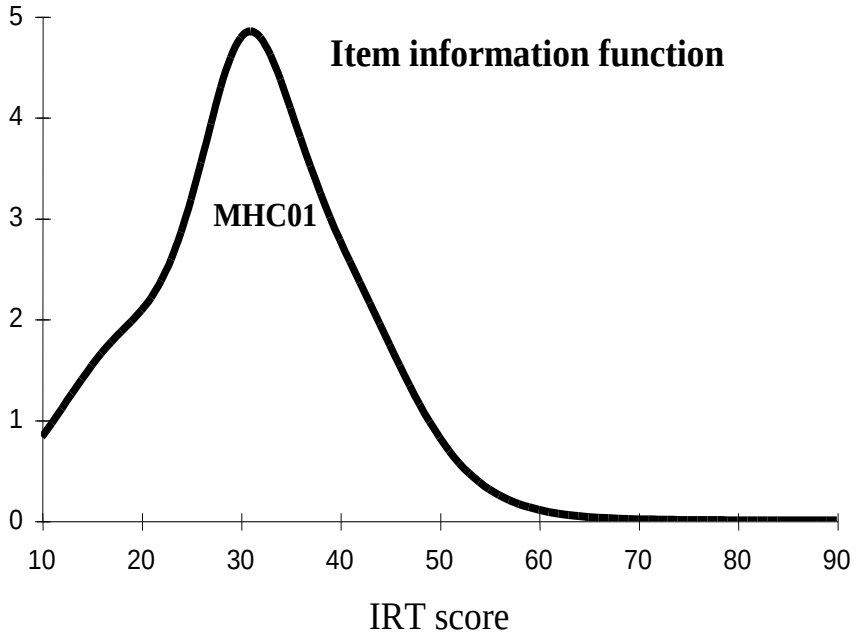
Item Information





$$I_i(\theta) = a_i^2\left[\sum_{k=0}^{m_i} k^2 P_{ik}(\theta) - \left(\sum_{k=0}^{m_i} kP_{ik}(\theta)\right)^2\right] = a_i^{\,2}\,\mathrm{var}(X_i).$$

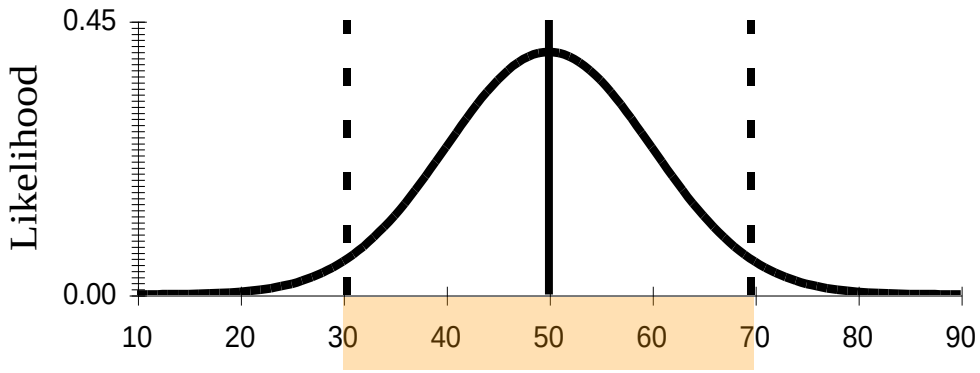




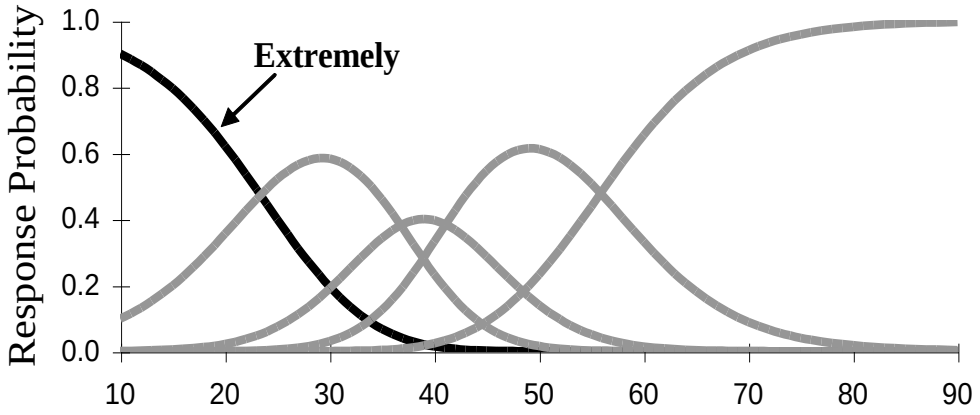
IRT Scoring in a CAT

(EAP Approach)

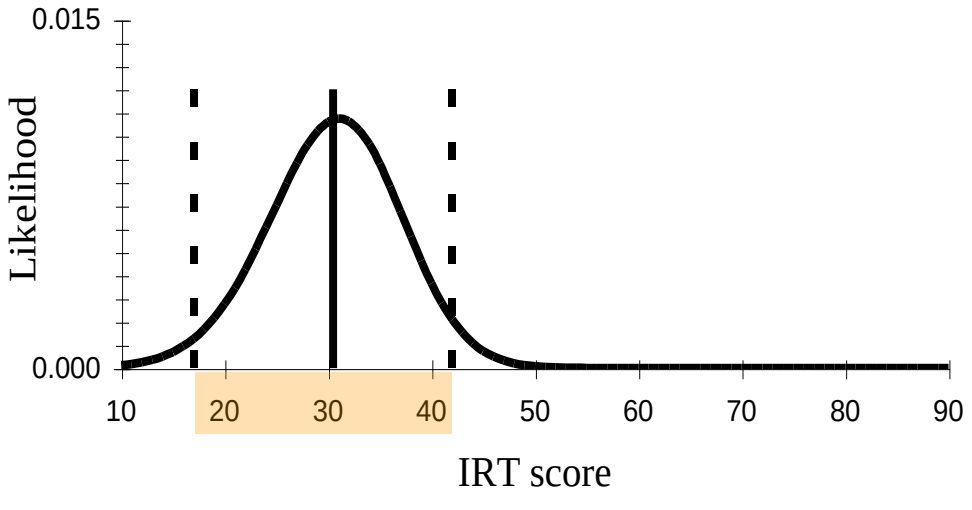
Prior distribution



SF8MH



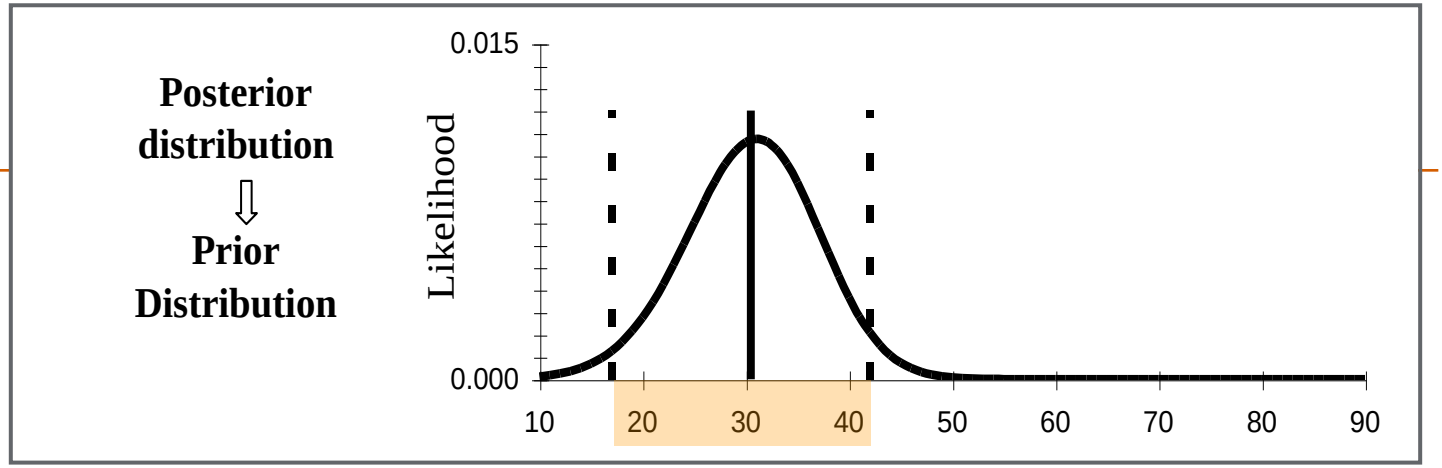
Posterior distribution

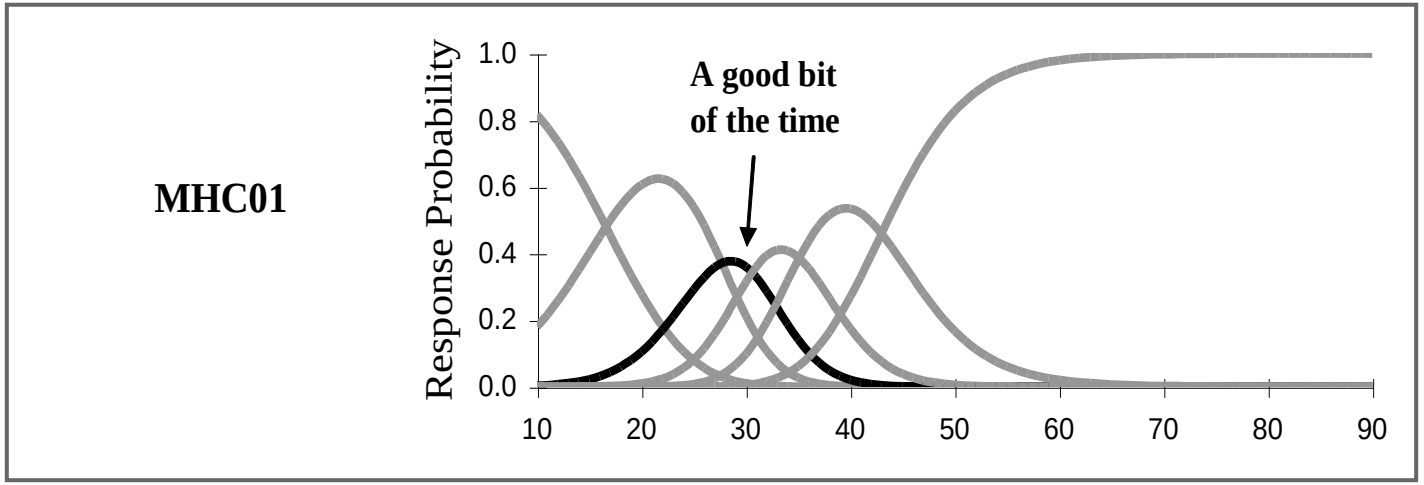


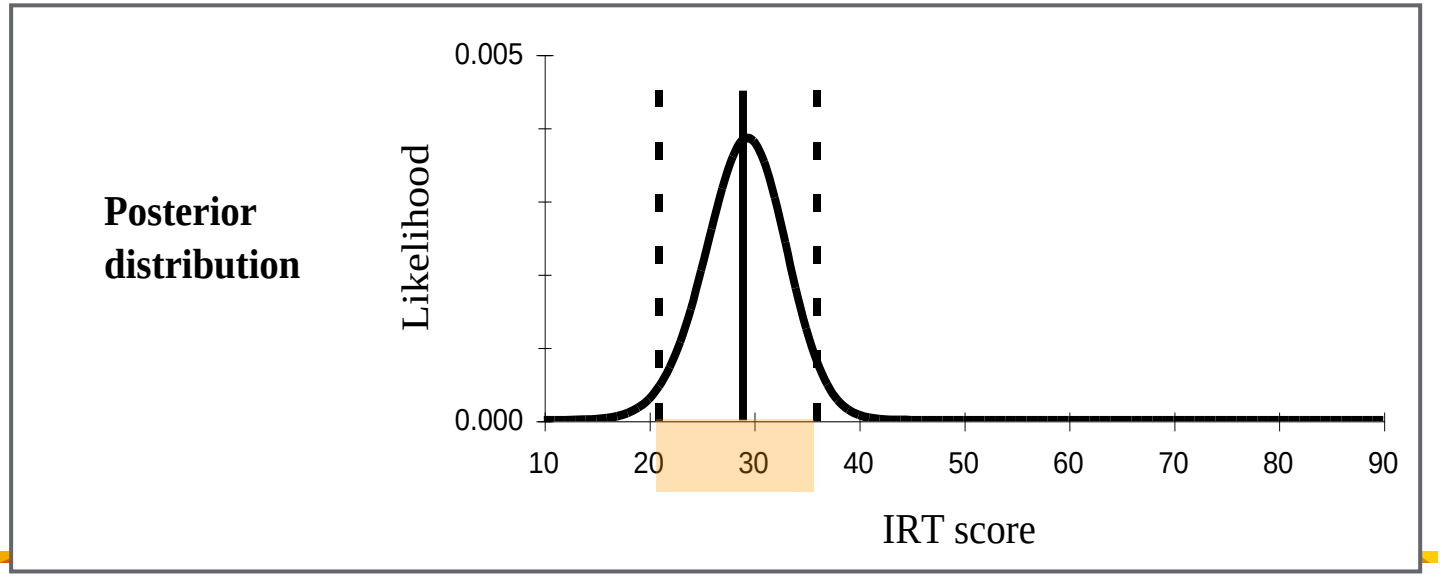
OPTUMInsight™

IRT Scoring in a CAT

(EAP Approach)







Steps to Build an IRT based CAT

1. Definition of the Construct
2. Data collection
3. Evaluation of Dimensionality
4. Evaluation of Item Properties
5. Test for Differential Item functioning
6. Evaluation of Measurement Precision
7. Choosing methods of administration
8. Deciding on CAT algorithms

US-National Institutes of Health (NIH)

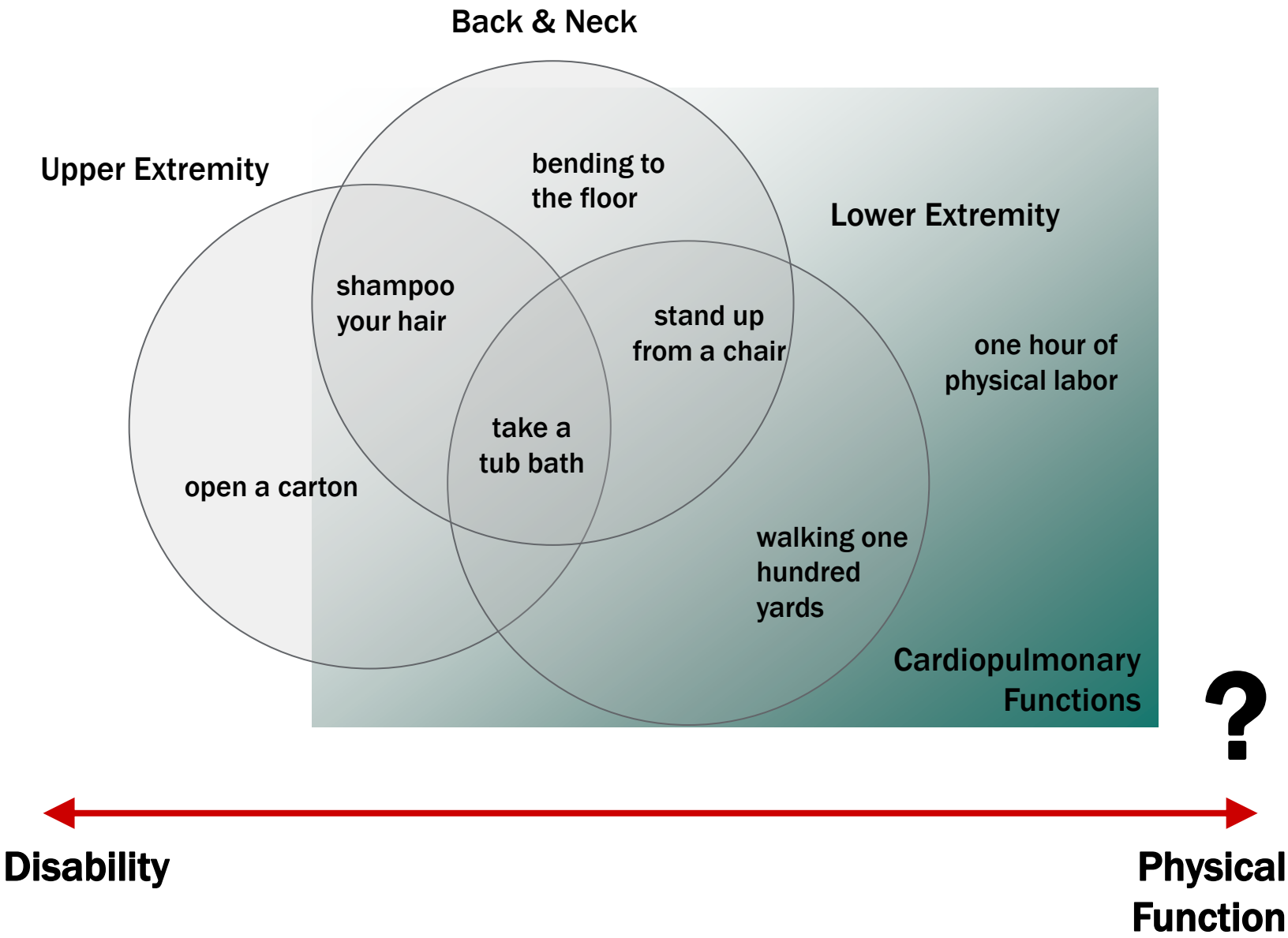
Generic Health Related Quality of Life - like SF-36

Key-Road-Initiative
"Re-engineering the Clinical Research Enterprises"
2004-2009+

n=21.133 participants

http://www.nihpromis.org/

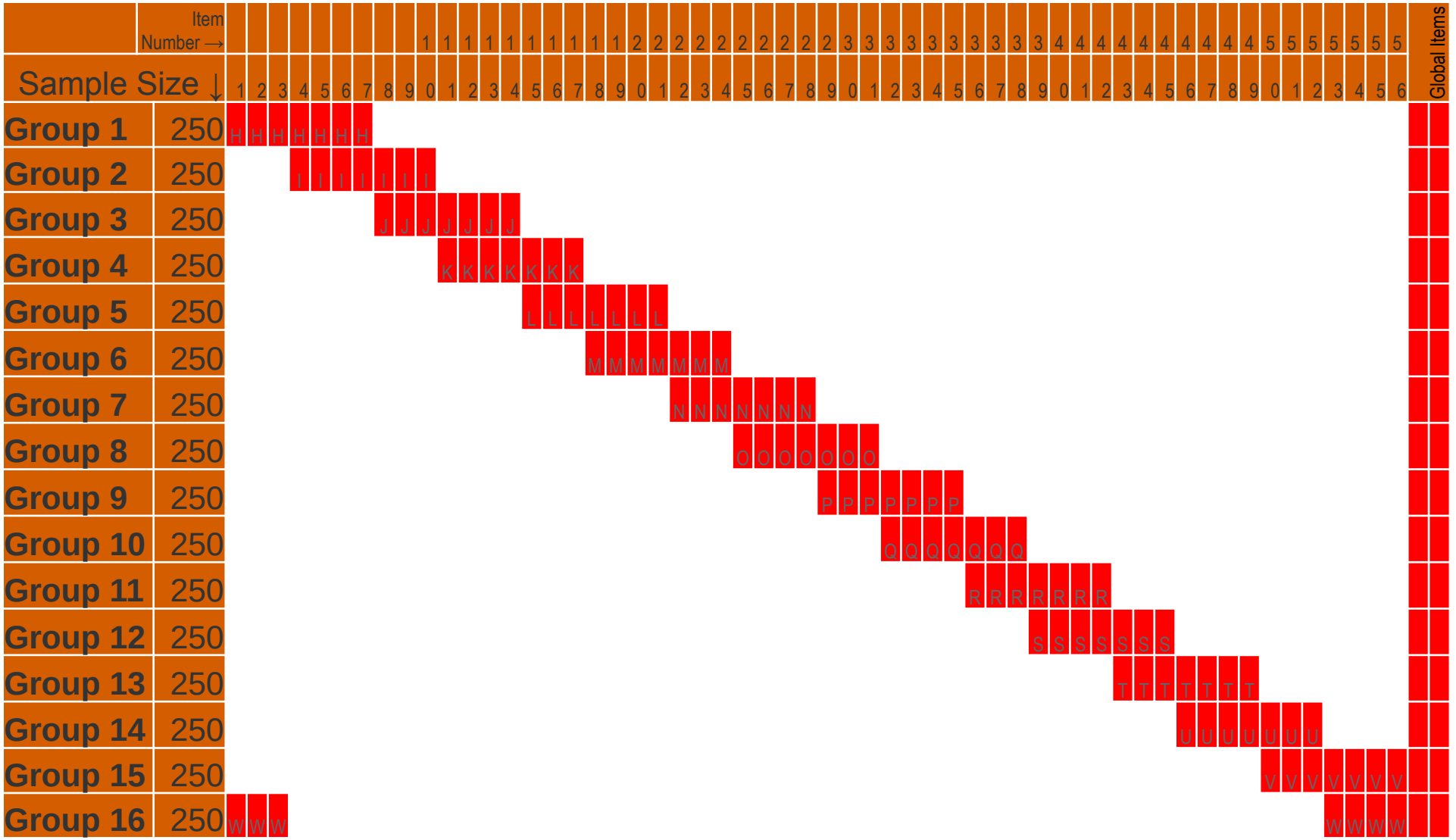
Physical Function



Sampling strategies – “total sample”

	Item 1	Item 2	Item 3	…					Item 112
Person 1									
Person 2									
Person 3									
…									
…									
Person 500									

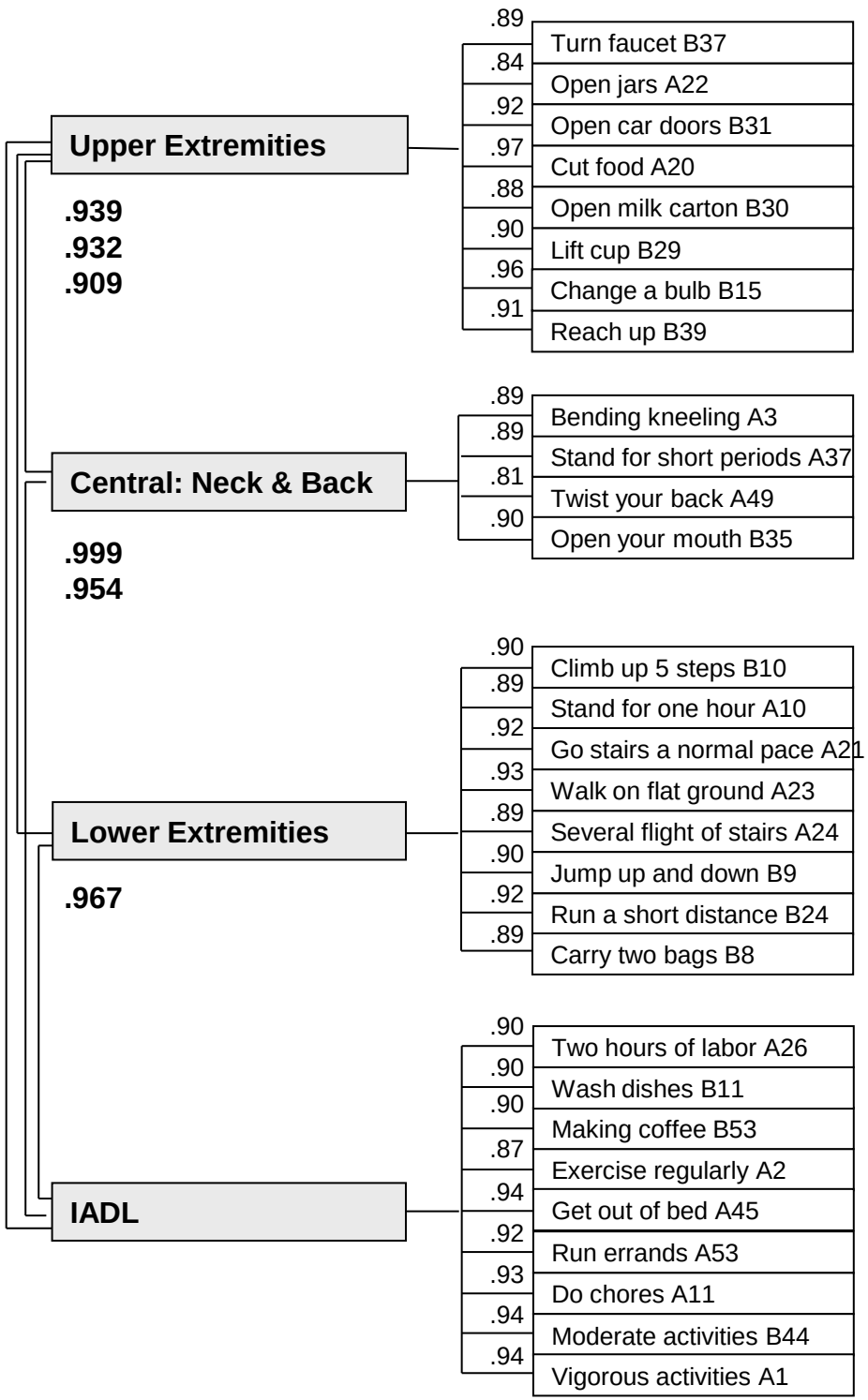
Sampling strategies – block sampling



PF Categories

CFA
n=502 General Population

RMSEA 0.083
CFI .959



examples
out of 112 items

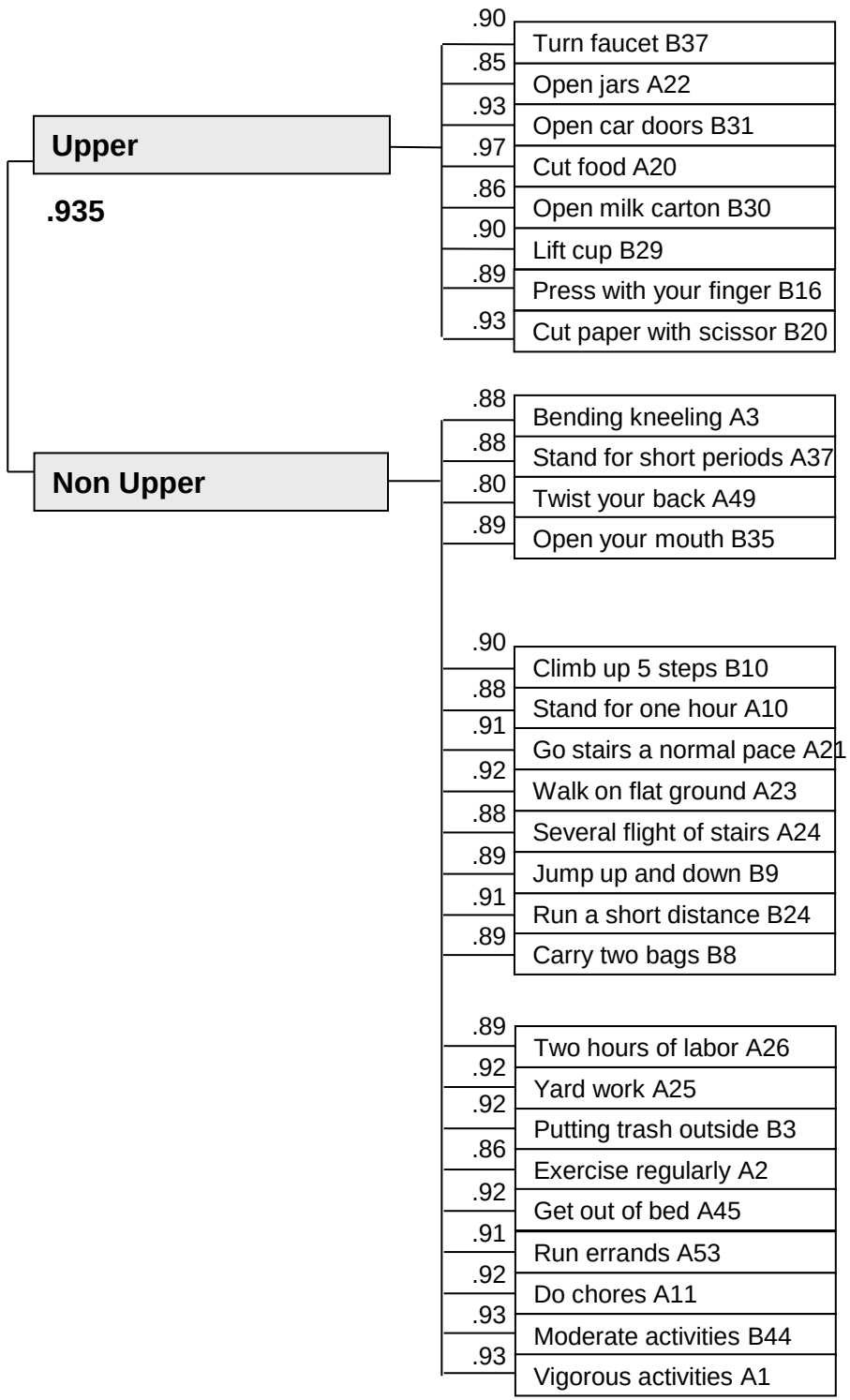
PF Categories

CFA

n=502 General Population

RMSEA 0.088

CFI .954



examples
out of 112 items

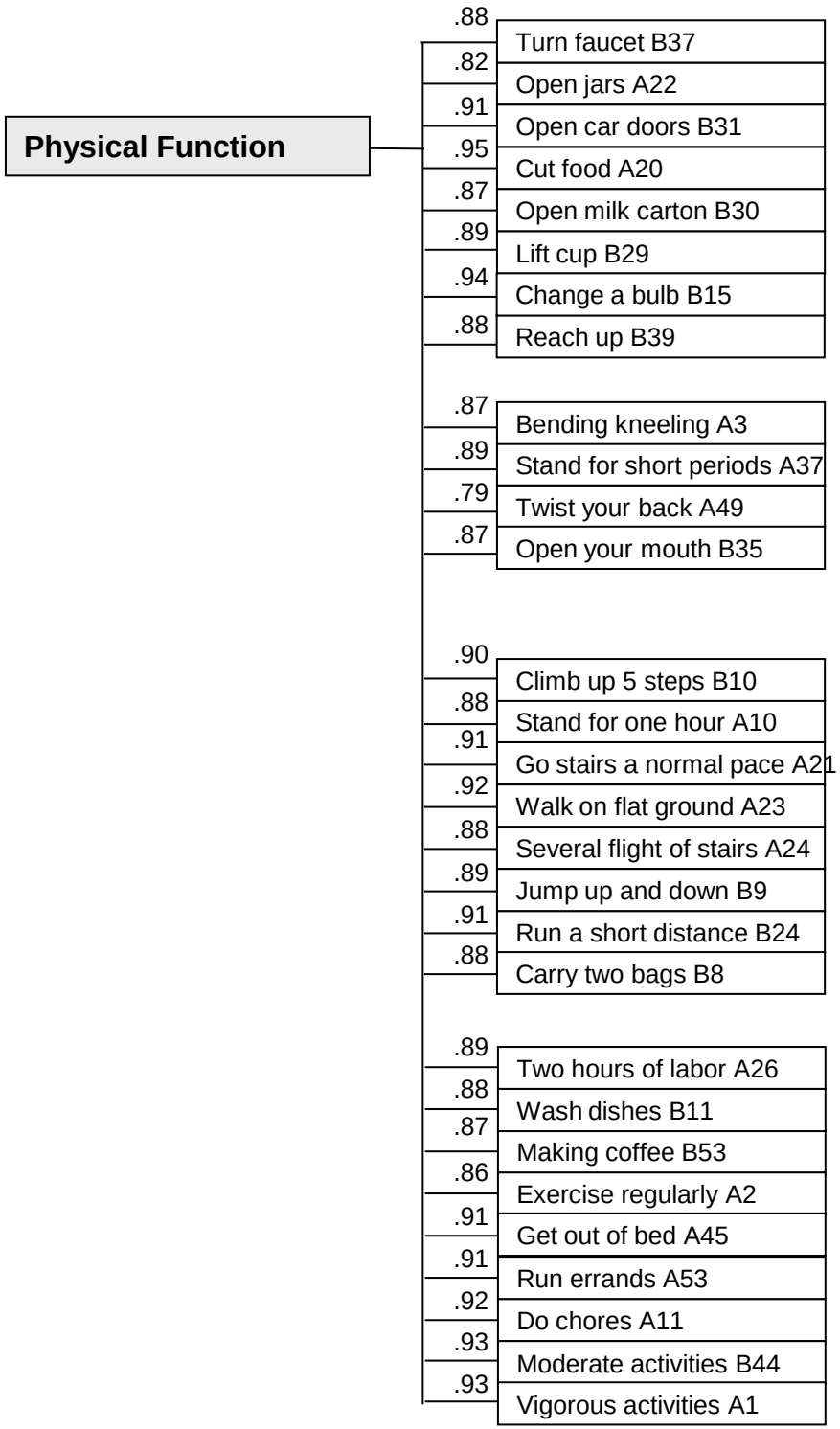
PF Categories

CFA

n=502 General Population

RMSEA 0.094

CFI .952



examples
out of 112 items

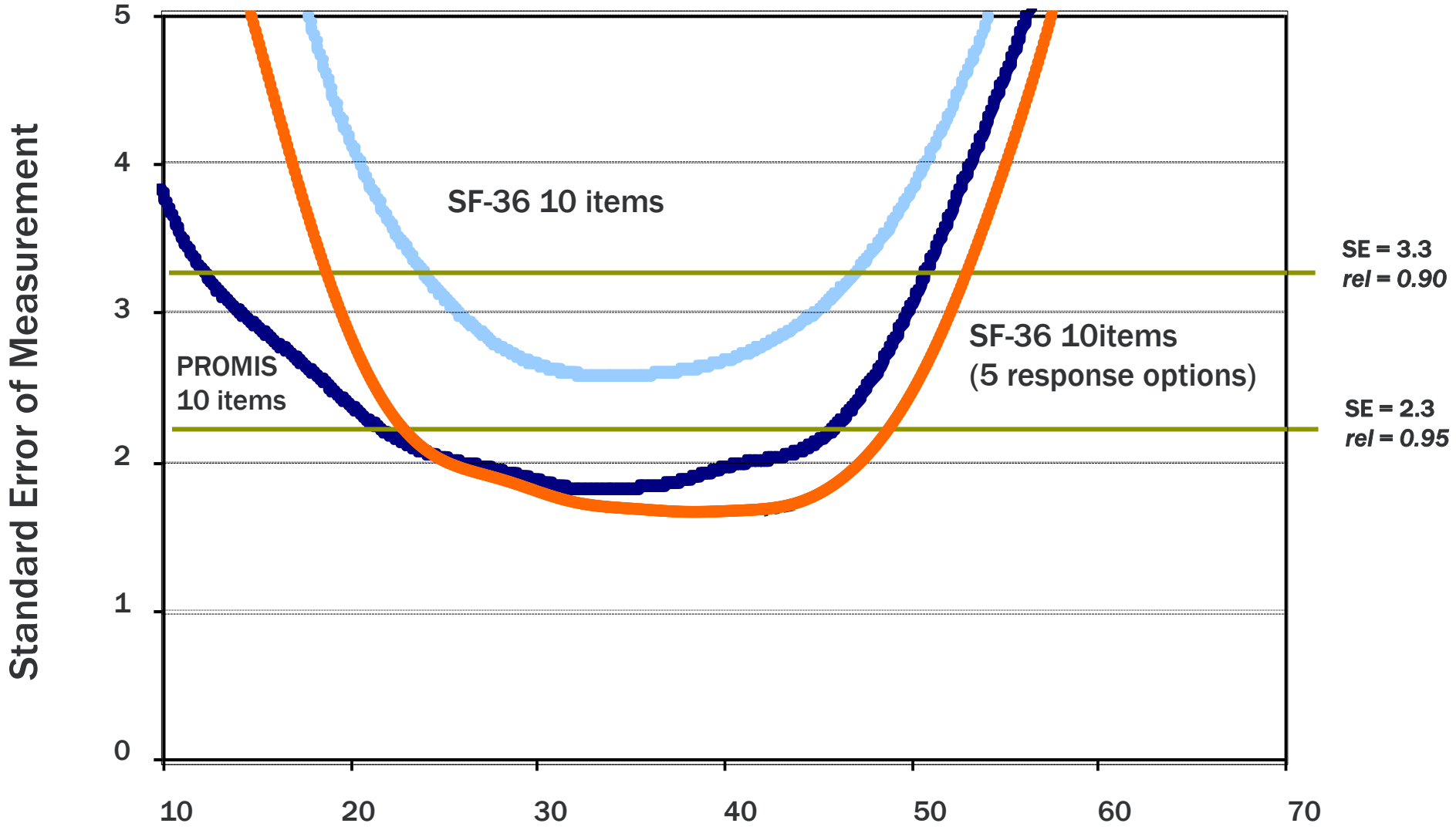
Evaluation of Item Properties

- Is the item a good measure of the domain in question?
  - Does the item fit the IRT model?
  - Does the item provide sufficient information (is the IRT discrimination parameter high)?
  - Is the item unique (i.e. not too similar to other items – no local dependence between pairs of items)?
- Are the item properties robust across population subgroups?
  - Lack of Differential Item Function (DIF)

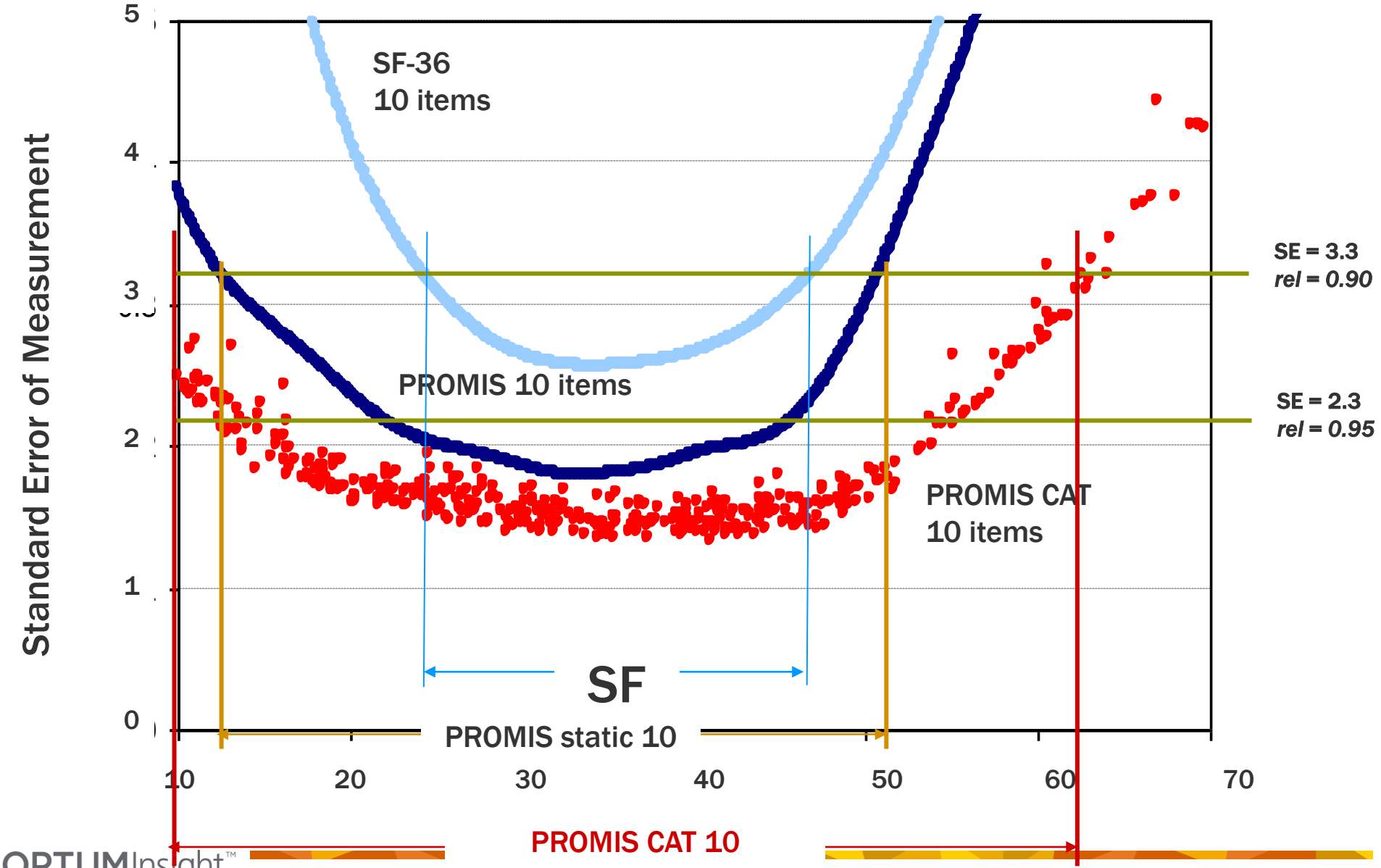
Description of Item Properties

	Item Label and Stem	a priori dom.	skew-ness	corr. to global item	CFA	DIF	slope	step1	step2	step3	step4	Fit G2	I_{max} at Θ	CAT util. %
C46	Are you able to transfer from a bed to a chair and back?	Lower	-3.63	0.75	0.89		3.61	15.2	20.7	26.7	34.4		1.31 14.6	0.4 24.0
B26	Are you able to shampoo your hair?	Up	-6.00	0.74	0.88		3.52	18.7	21.5	27.2	32.7		2.82 17.8	0.2 15.7
A56	Are you able to get in and out of a car?	Cent.	-2.19	0.75	0.88		3.24	13.6	22.5	30.0	38.9	0.01 0.77 0.03	2.33 20.4	0.0 5.6
A16	Are you able to dress yourself. including tying shoelaces and doing buttons?	Upper	-3.36	0.71	0.92	0.04 sex	3.31	18.1	24.4	30.7	37.5		2.7 20.7	0.0 0.0
A11	Are you able to do chores such as vacuuming or yard work?	IADL	-1.67	0.81	0.93		4.72	30.1	34.7	39.1	45.8		5.58 32.9	45.7 75.0
B22	Are you able to hold a plate full of food?	Upper	-4.86	0.76	0.91	0.03 edu	3.31	15.7	22.6	27.4	34.3		2.41 22.0	0.0 0.0
C12	Does your health now limit you in doing two hours of physical labor?	IADL	-0.71	0.77	0.91		4.49	35.8	40.9	46.3	50.9		5.2 46.1	59.8 31.7
A01	Does your health now limit you in doing vigorous activities. such as running, lifting heavy objects, participating in strenuous sports?	IADL	0.11	0.76	0.94	0.03 age 0.04 dis	2.99	38.2	45.0	51.7	56.5	0.02 0.00 0.01	2.68 52.2	39.3 0.6

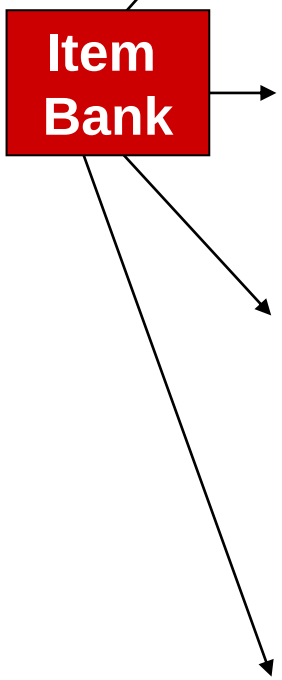
PROMIS PF-Static Form 10



Measurement Precision and Range



PROMIS Tools



Pick a PRO

Build a PRO

static

Predefined static PROMIS instruments
PROMIS-29, -43, -57 Profiles
(one size fits all)

User picks the items of the bank with most appropriate face validity
(e.g. literacy level, measurement range target population)

Use a CAT

Most precise and less burdensome individually tailored test

Standardization of Measurement

Transfer Scores from established tools *into IRT scores*

Transfer Scores from established instruments *into each other*

Summary

- CAT instruments can provide more precise measurement without increased respondent burden
- Item banks have been created by several research groups, including PROMIS, QualityMetric/Optum PatientInsights, the EORTC Quality of Life Study Group, the Charite Hospital
- CAT is also available through several groups
- Evaluations show feasibility of CAT in research and clinical practice
- Item banks enable a flexible measurement strategy using either:
  - Standard fixed forms
  - Adapted fixed forms
  - CAT

Remaining opportunities and challenges

- Developing systems for updating item parameters – using data from patients
- Improved statistical methods for analyzing data from item banks
- Optimize CAT algorithms for longitudinal studies

Thank you