What can IRT do?

Ron D. Hays, Ph.D.

UCLA Department of Medicine

http://gim.med.ucla.edu/FacultyPages/Hays/

March 1, 2016 12 noon – 1pm

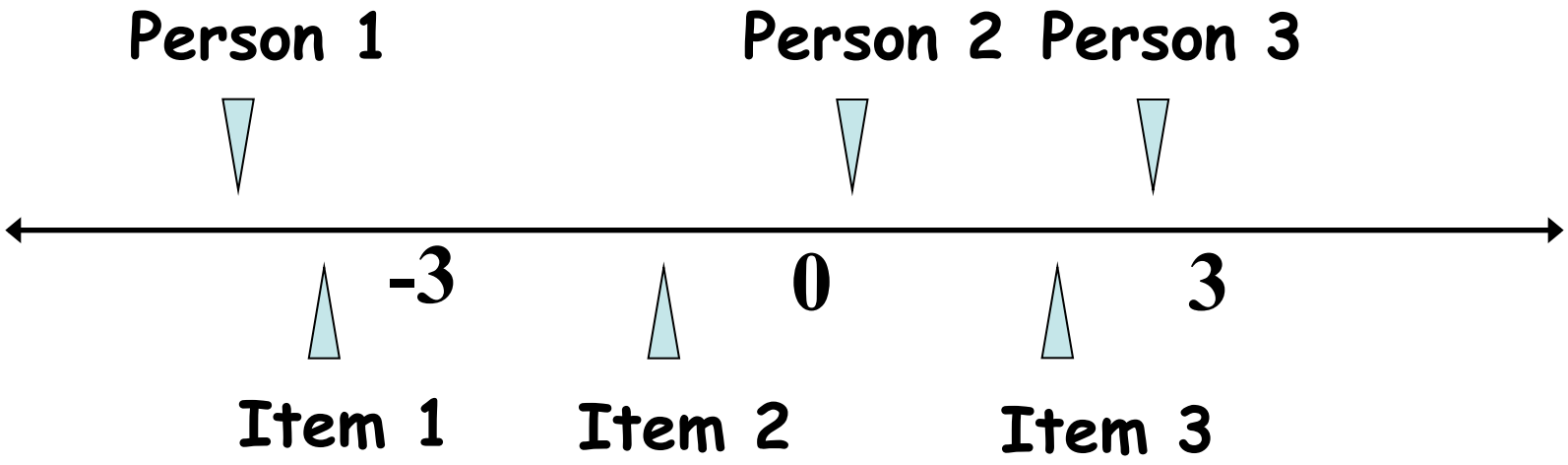
Center for Healthful Behavior Change

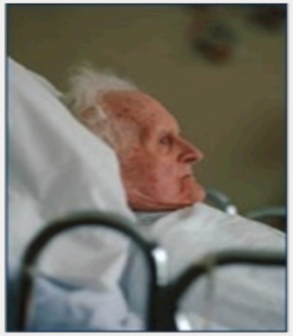
Translation Research Building 619/NYU School of Medicine

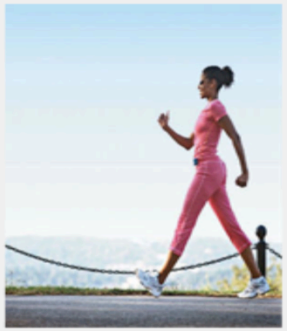
Department of Population Health

227 East 30th St. 6th Floor, NY, NY 10016

People and Items on Same *z-score* metric






-3 0 3

Physical Functioning Item Bank

Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 | Item 7 | Item 8 | Item 9 | Item *n*

Are you able to get in and out of bed?
Are you able to stand without losing your balance for 1 minute?
Are you able to walk from one room to another?
Are you able to walk a block on flat ground?
Are you able to run or jog for two miles?
Are you able to run five miles?

Item Response Theory (IRT)

IRT graded response model estimates relationship between a person's response Y_i to the question (i) and his or her level on the latent construct (θ):

$$\Pr(Y_i \geq k) = \frac{1}{1 + \exp(-a_i\theta + b_{ik})}$$

$\mathbf{b_{ik}}$ estimates how difficult it is to have a score of k or more on item (i).

$\mathbf{a_i}$ estimates item discrimination.

IRT is mainstreaming

- BIGSTEPS and WINSTEPS
- PARSCALE and MULTILOG

- IRTPRO and FLEXMIRT

- SAS and STATA

Computer Adaptive Testing (CAT)

2004

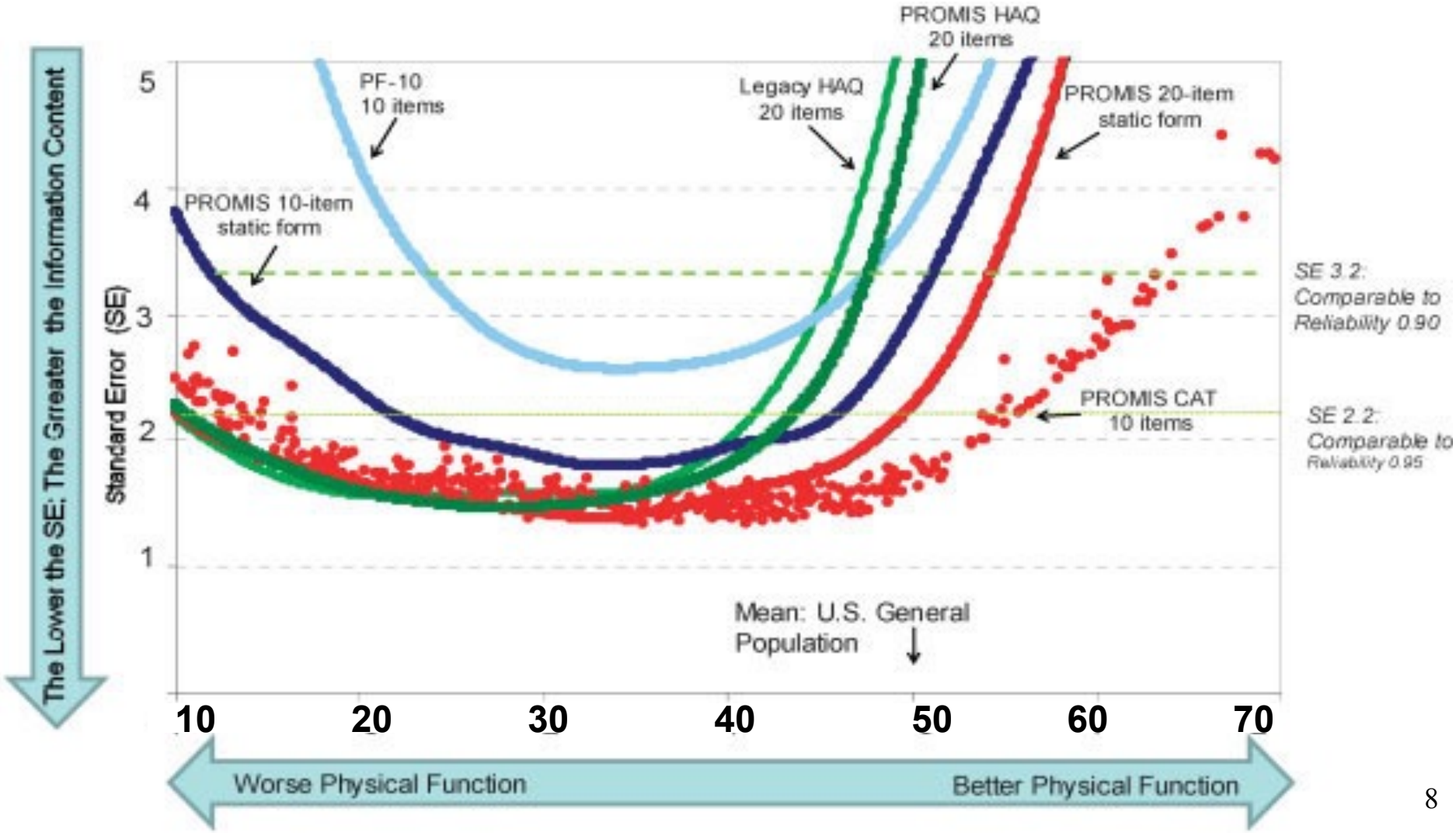
www.nihpromis.org

Reliability Target for Use of Measures with Individuals

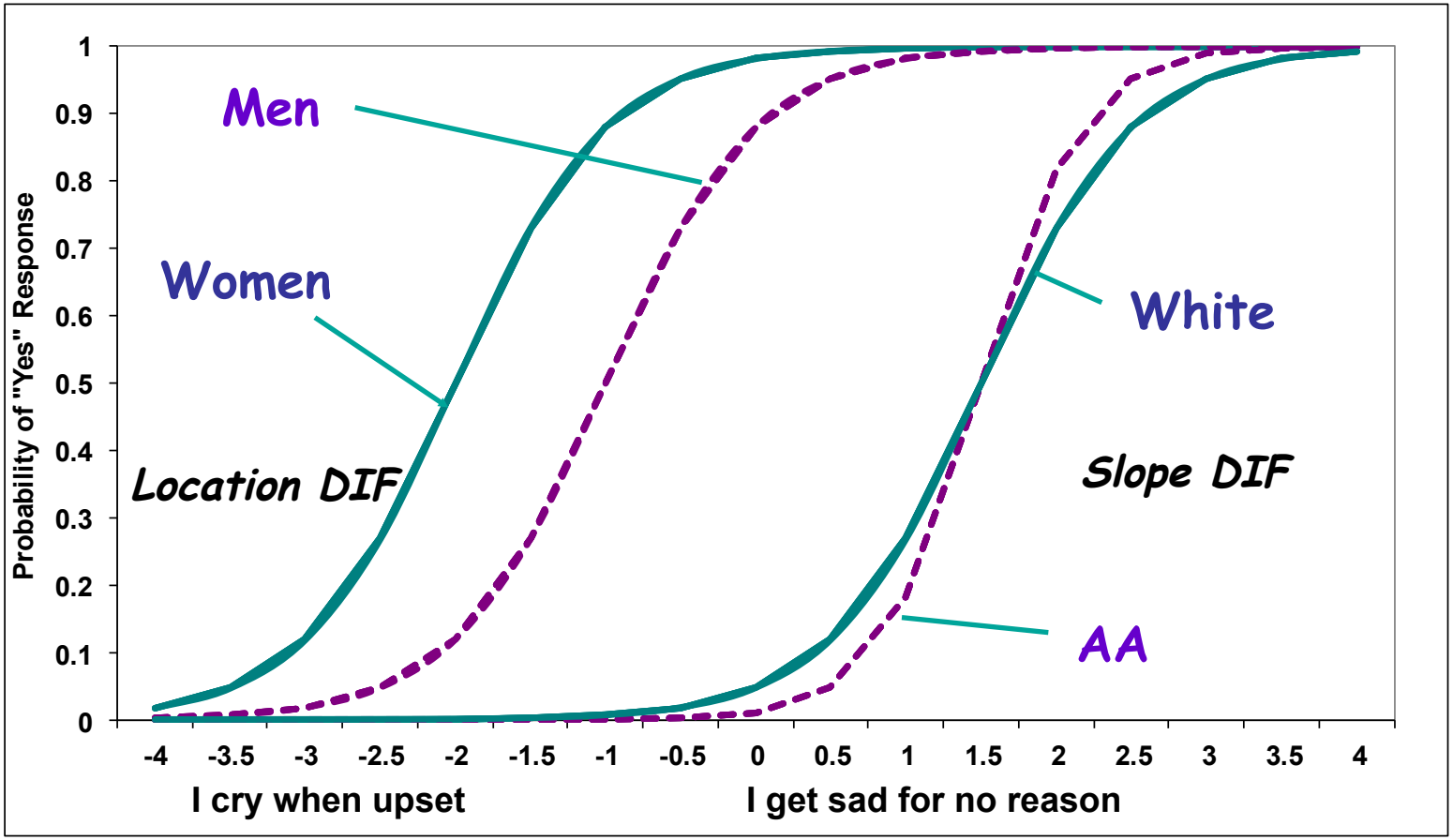
- z-score (mean = 0, SD = 1)
- Reliability ranges from 0-1
  - 0.90 or above is goal
  - SE = SD $(1 - \text{reliability})^{1/2}$
  - Reliability = $1 - SE^2$
  - Reliability = 0.90 when SE = 0.32
- 95% CI = true score +/- 1.96 x SE

  (CI = -0.63 → 0.63 z-score when reliability = 0.90)

PROMIS Physical Functioning vs. "Legacy" Measures



DIF (2-parameter model)



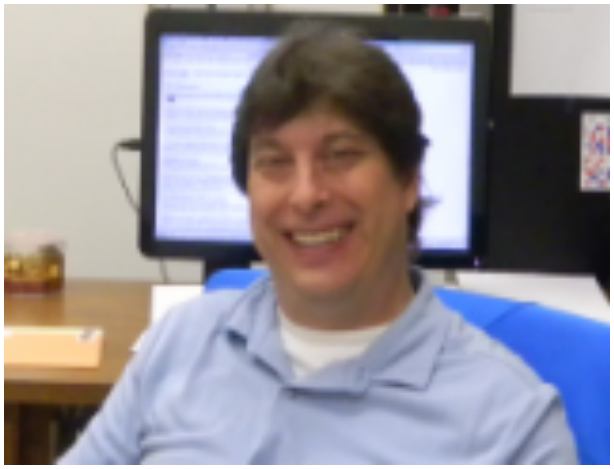
Higher Score = More Depressive Symptoms

Person Fit

- Large negative Z_L values indicate misfit.

- One person in PROMIS project had $Z_L = -3.13$
- This person reported that they could do 13 physical functioning activities (including running 5 miles) without any difficulty, but
  - This person reported a little difficulty being out of bed for most of the day.

My exposure to IRT (1990's)



- Hays, R. D., & Reise, S. P. (1998, November). *Item response theory*. Invited Workshop, International Society for Quality of Life Research, Baltimore, MD
- Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item Response Theory and Health Outcomes Measurement in the 21st Century. Medical Care, 38 (Suppl.), II-28-II-42.

Physical Functioning

- Ability to conduct a variety of activities ranging from self-care to running
  - 6 physical functioning items included in 2010 Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Medicare Survey

Hays, R. D., Mallett, J. S., Gaillot, S., & Elliott, M. N. (2015). Performance of the Medicare Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Physical Functioning Items. Medical Care, 54, 205-209

Because of a health or physical problem are you unable to do or have any difficulty doing the following activities?

- Walking?
- Getting in or out of chairs?
- Bathing?
- Dressing?
- Using the toilet?
- Eating?

  - I am unable to do this activity (0)
  - Yes, I have difficulty (1)
  - No, I do not have difficulty (2)

Medicare beneficiary sample (n = 366,701)

- 58% female
- 57% high school education or less
- 14% 18-64; 48% 65-74, 29% 75-84, 9% 85+

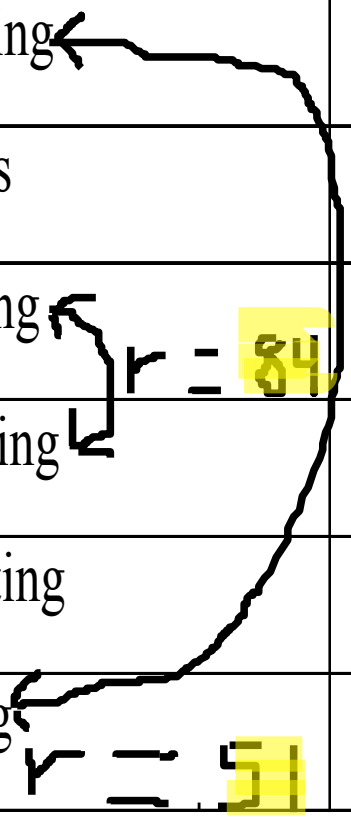
% of Medicare beneficiaries (n = 366,701) selecting each response option

Item **(Some difficulty)**	Unable to do	Have difficulty	No difficulty
Walking **(1/3)**	4	27	69
Chairs **(1/5)**	3	19	78
Bathing **(1/7)**	4	11	85
Dressing **(1/9)**	3	9	88
Toileting **(1/10)**	3	6	91
Eating **(1/16)**	3	3	94

Possible 6-item scale range: 0-12 (2% floor, 65% ceiling)

% of Medicare beneficiaries (n = 366,701) selecting each response option

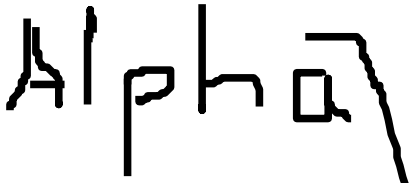
Item	Unable to do	Have difficulty	No difficulty
Walking	4	27	69
Chairs	3	19	78
Bathing	4	11	85
Dressing	3	9	88
Toileting	3	6	91
Eating	3	3	94



Item-Scale Correlations

Item	Item-Scale Correlations
Walking (0, 1, 2)	0.71
Chairs (0, 1, 2)	0.80
Bathing (0, 1, 2)	0.83
Dressing (0, 1, 2)	0.86
Toileting (0, 1, 2)	0.84
Eating (0, 1, 2)	0.75

Possible 6-item scale range: 0-12 (2% floor, 65% ceiling)



Reliability Formulas

Model	Reliability	Intraclass Correlation
Two-way random	$\frac{N(MS_{BMS} - MS_{EMS})}{NMS_{BMS} + MS_{JMS} - MS_{EMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS} + k(MS_{JMS} - MS_{EMS})/N}$
Two-way mixed	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS}}$
One-way	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS} + (k-1)MS_{WMS}}$

BMS = Between Ratee Mean Square **N = n of ratees**

WMS = Within Mean Square **k = n of items or raters**

JMS = Item or Rater Mean Square

EMS = Ratee x Item (Rater) Mean Square

Internal Consistency Reliability (Coefficient Alpha)

- Coefficient alpha = 0.92

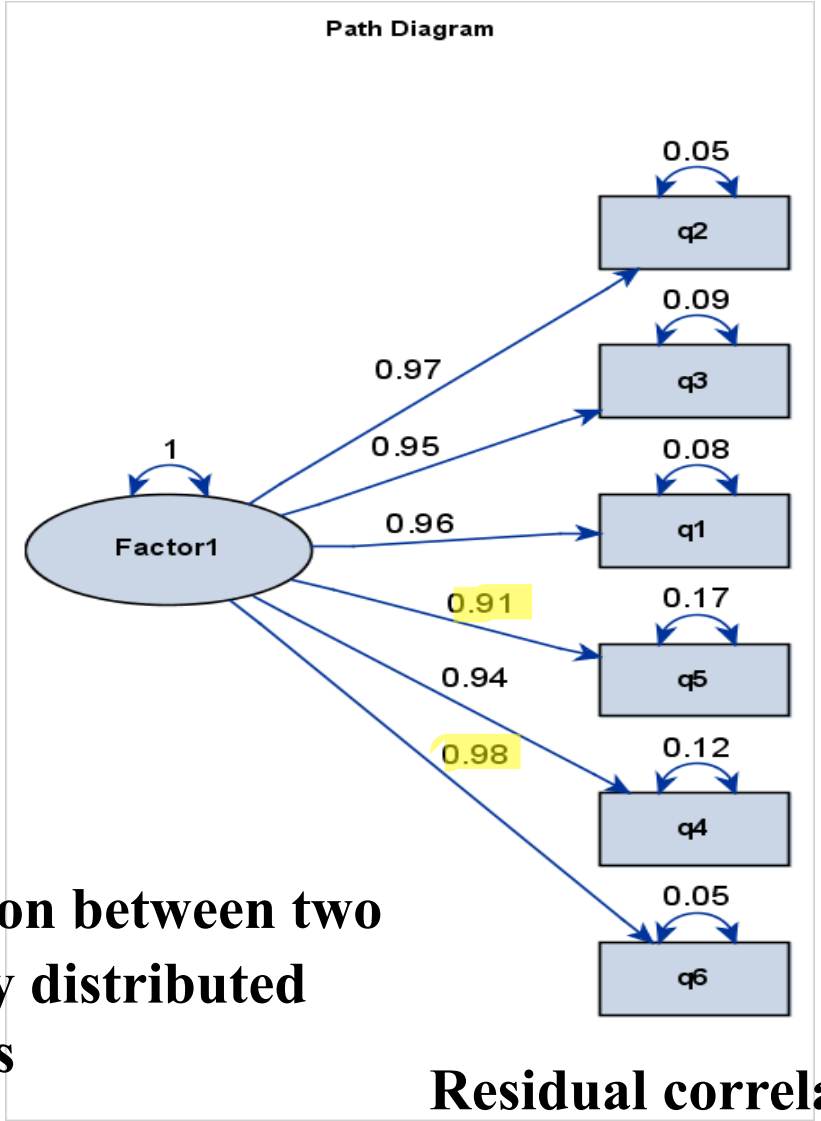
  $(MS_{bms} - MS_{ems})/MS_{bms}$

- Ordinal alpha = 0.98

  -http://support.sas.com/resources/papers/proceedings14/2042-2014.pdf

  -http://gim.med.ucla.edu/FacultyPages/Hays/utils/

Confirmatory Factor Analysis (Polychoric* Correlations)



Dressing

Eating

Bathing

Walking

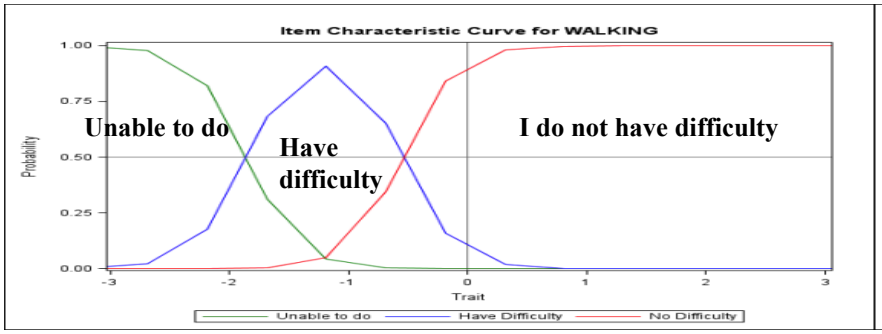
Chairs

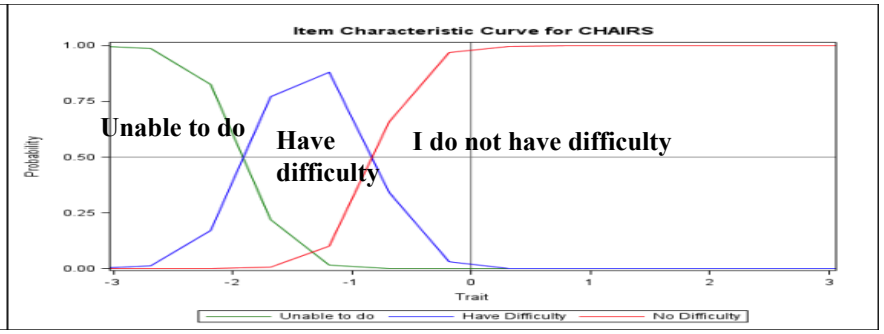
Toileting

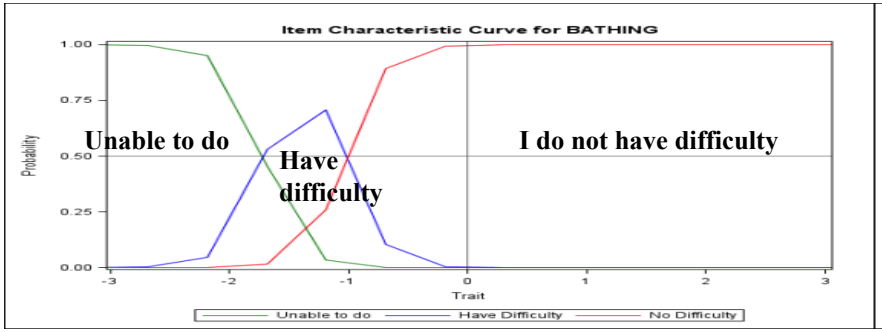
***Estimated correlation between two underlying normally distributed continuous variables**

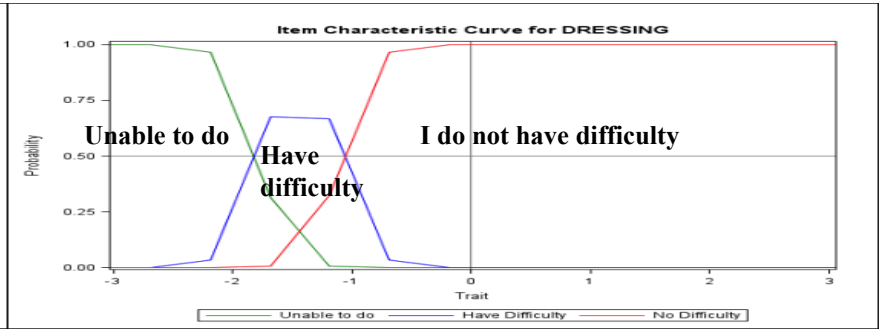
Residual correlations <= 0.04

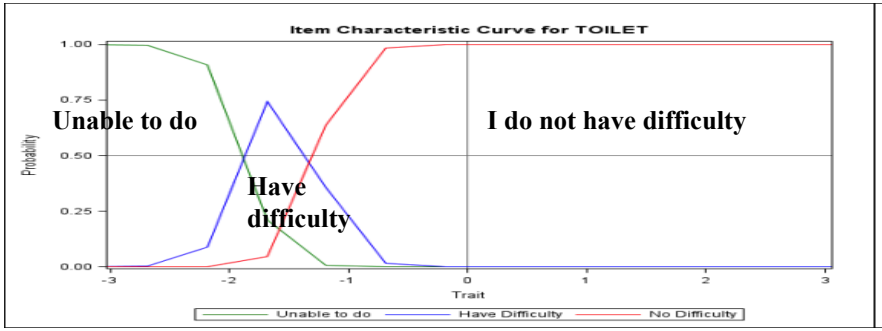
Item Characteristic Curves











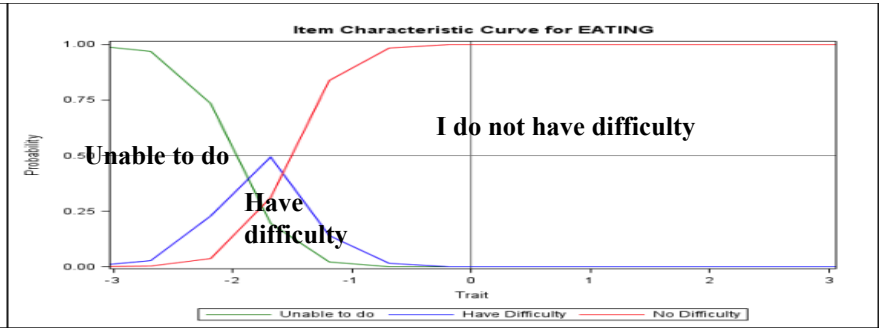
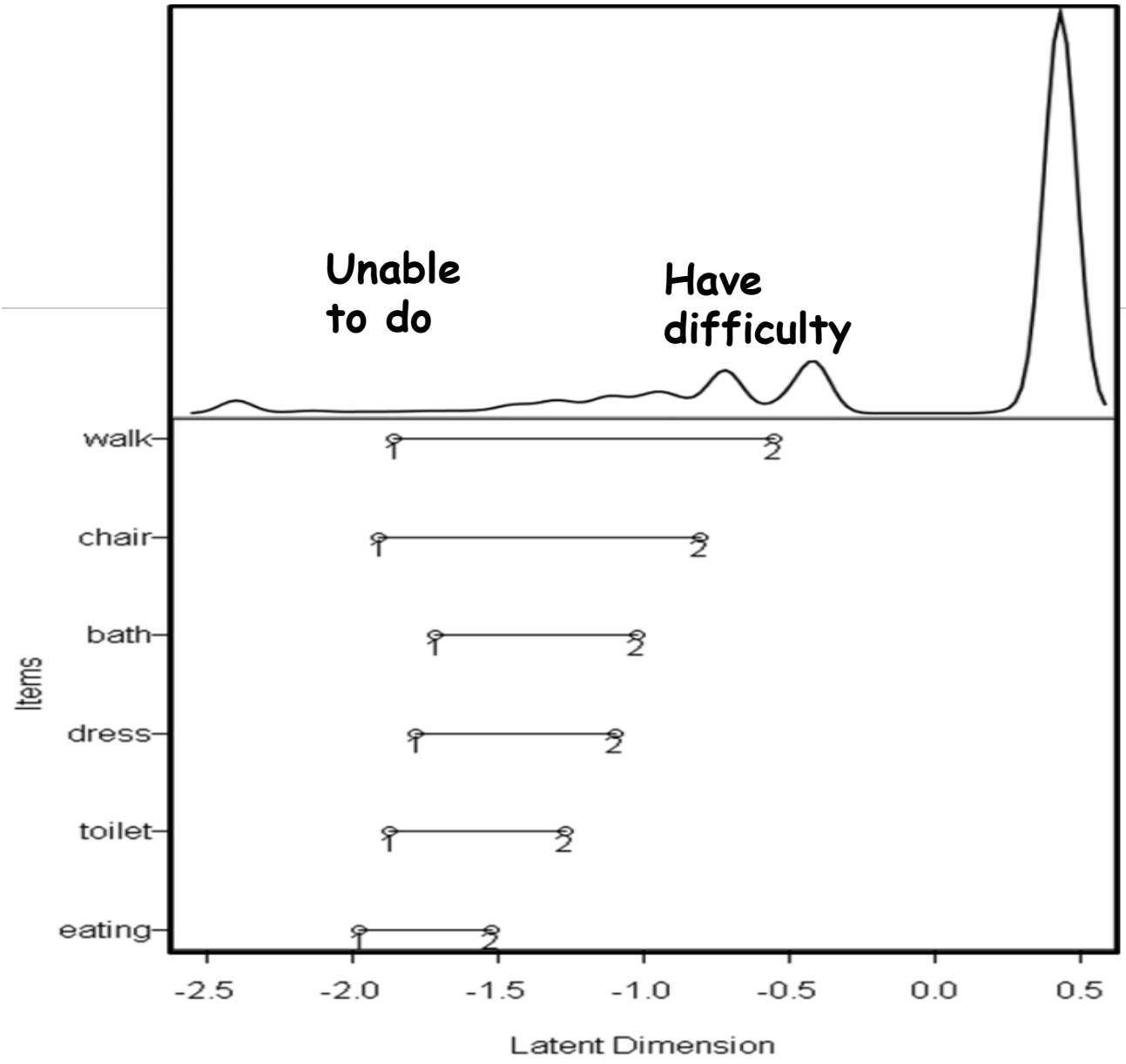
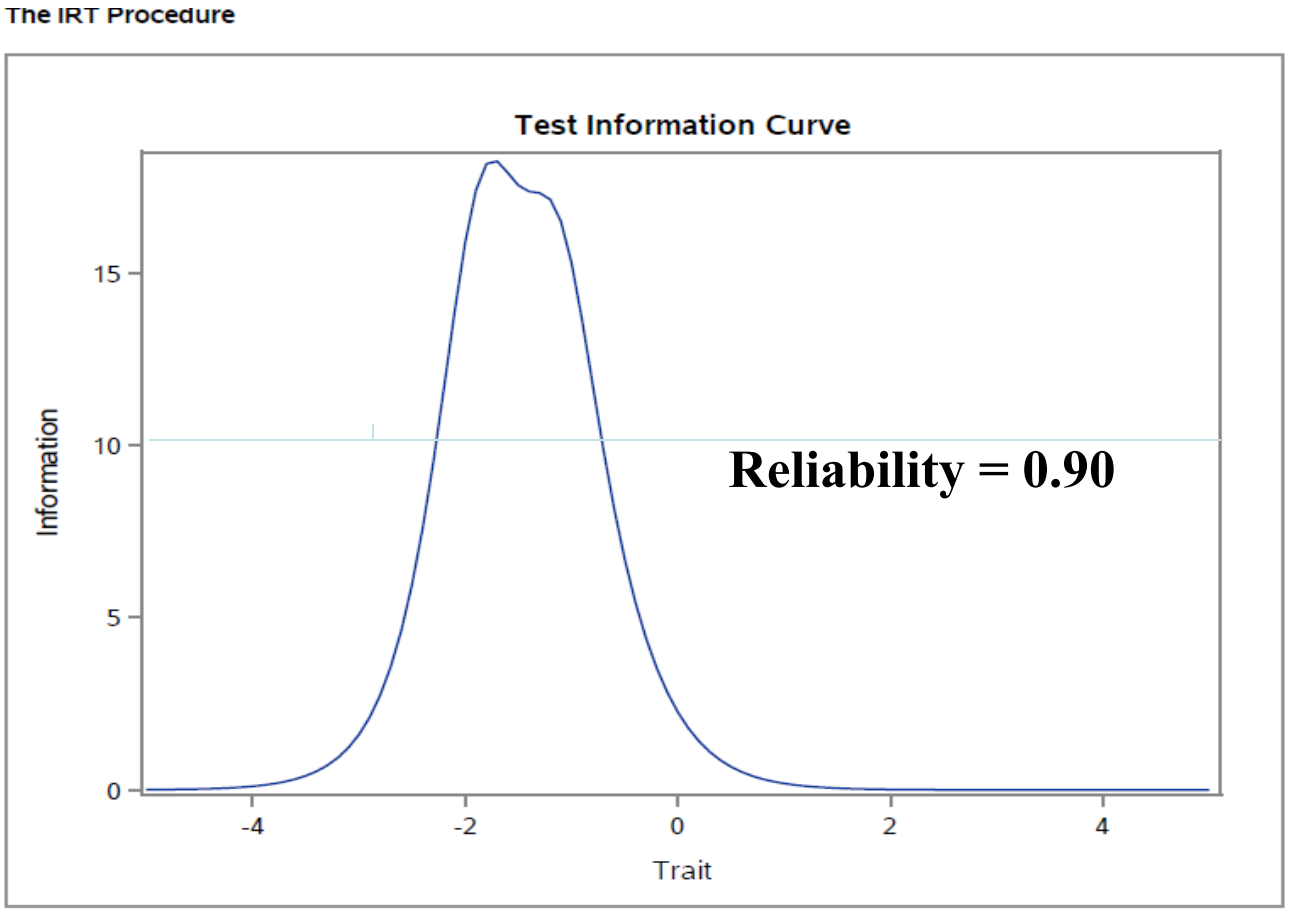


Figure 2. Person-Item Map



Reliability = (Info – 1) / Info



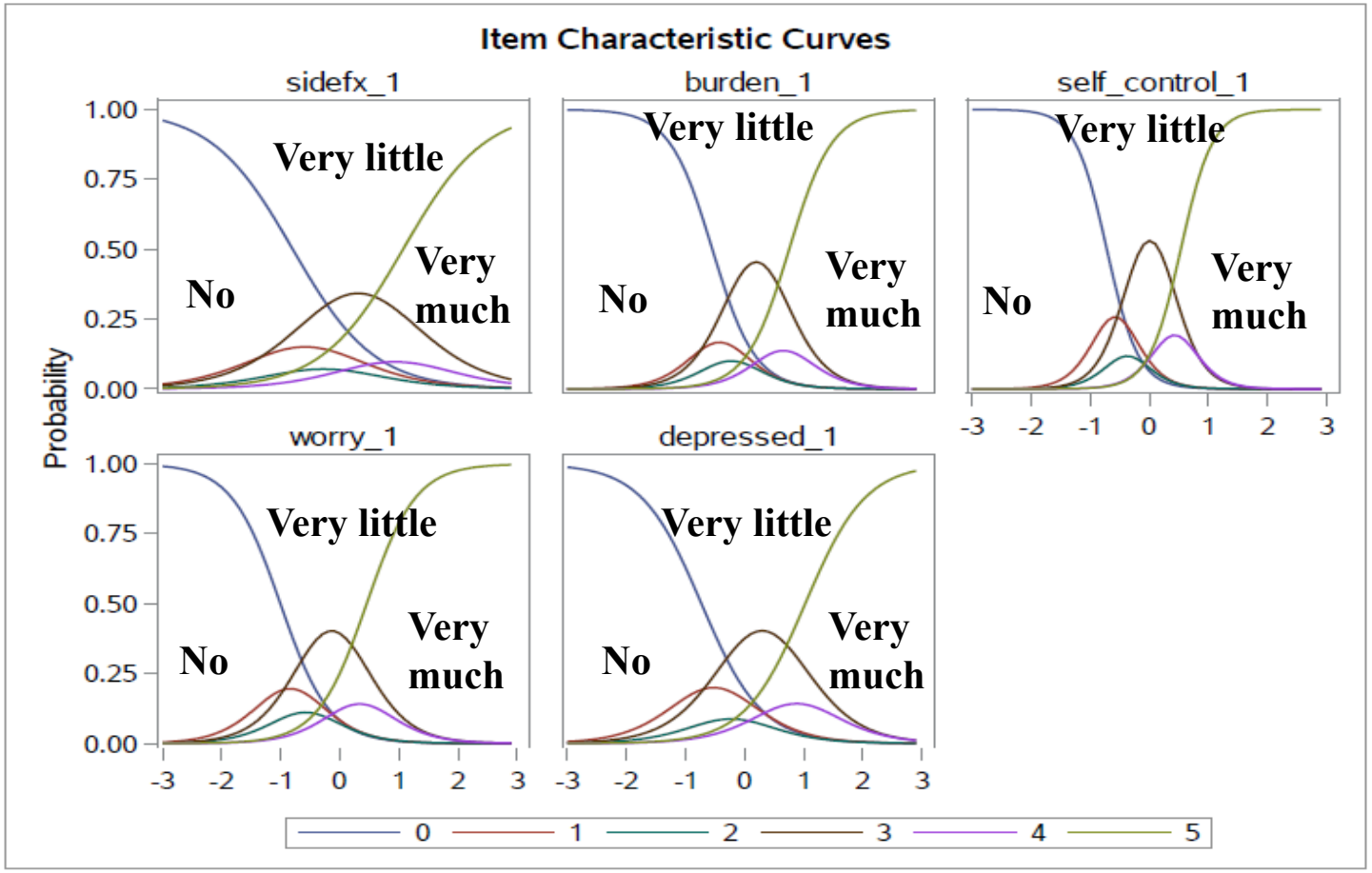
MINNESOTA LIVING WITH HEART FAILURE® QUESTIONNAIRE

The following questions ask how much your heart failure (heart condition) affected your life during the past month (4 weeks). After each question, circle the 0, 1, 2, 3, 4 or 5 to show how much your life was affected. If a question does not apply to you, circle the 0 after that question.

Did your heart failure prevent you from living as you wanted during the past month (4 weeks) by -	**No**			**Very Little**		**Very Much**
1. causing swelling in your ankles or legs?	0	1	2	3	4	5
2. making you sit or lie down to rest during the day?	0	1	2	3	4	5
3. making your walking about or climbing stairs difficult?	0	1	2	3	4	5
4. making your working around the house or yard difficult?	0	1	2	3	4	5
5. making your going places away from home difficult?	0	1	2	3	4	5
6. making your sleeping well at night difficult?	0	1	2	3	4	5
7. making your relating to or doing things with your friends or family difficult?	0	1	2	3	4	5
8. making your working to earn a living difficult?	0	1	2	3	4	5
9. making your recreational pastimes, sports or hobbies difficult?	0	1	2	3	4	5
10. making your sexual activities difficult?	0	1	2	3	4	5

Item Characteristic Curve for Emotional Health Scale

The IRT Procedure



IRT Distortions

- "*Longer tests are more reliable than shorter tests* vs. *Shorter tests can be more reliable than longer tests*."
  - The new rules of measurement, Psychological Assessment, 1996, Susan E. Embretson
- "Parameter values are identical in separate subgroups or across different measurement conditions."
  - It is the often misunderstood feature of parameter invariance that is frequently cited in introductory or advanced texts" (Rupp & Zumbo, 2006).

Ben Wright or Been Wrong?

- "Modern day psychometric analyses such as Rasch analysis convert ordinal data to an interval scale so that response scores meet the criteria for measurement"
- "Application of the Rasch model to the data set estimates a measure that can be considered valid."
- The "Rasch model is the only valid approach to measurement"
  - Bergan, 2013, *Rasch versus Birnbaum: New arguments in an old debate* (p. 3)

Questions?