Primer on Evaluating Reliability and Validity of Multi-Item Scales

Questionnaire Design and Testing Workshop

October 25, 2013, 3:30-5:00pm

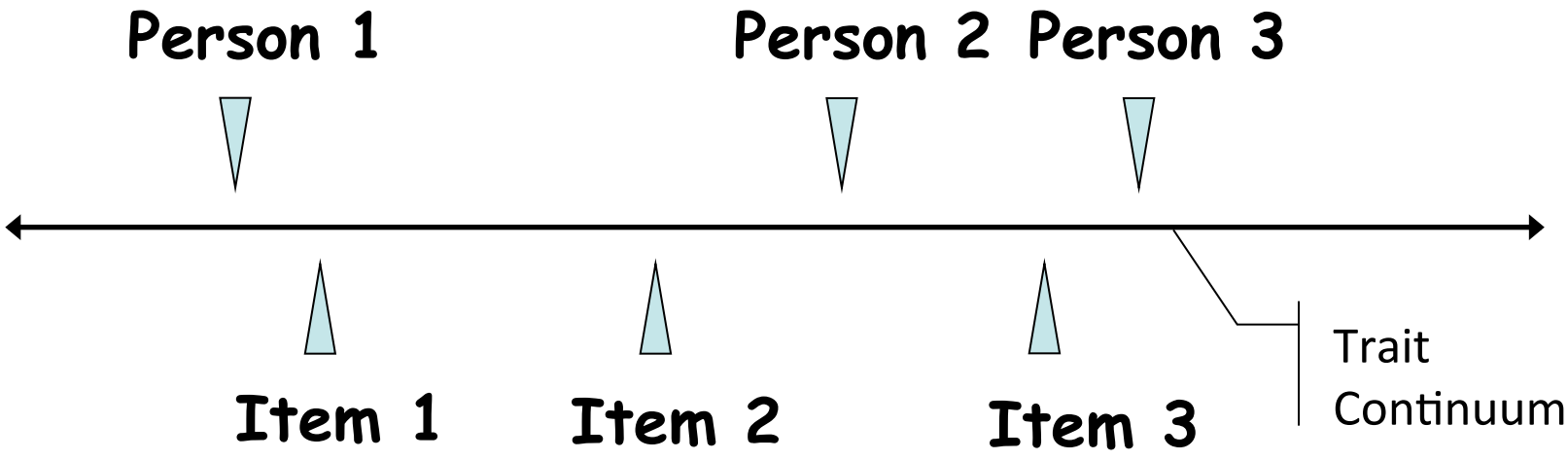
10940 Wilshire Blvd. Suite 710

Los Angeles, CA

www.nihpromis.org

- Patient Reported Outcomes Measurement Information System (PROMIS®)
- Funded by the National Institutes of Health
- One domain captured is “anger”
  - Mood (irritability, frustration)
  - Negative social cognitions (interpersonal sensitivity, envy, disagreeableness)
  - Needing to control anger

Item Responses and Trait Levels



www.nihpromis.org

Standard Error of Measurement (SEM)

$SEM = SD (1- reliability)^{1/2}$

95% CI = true score +/- 1.96 x SEM

- If z-score = 0 and reliability = 0.90
- CI: -.62 to +.62 (width is 1.24 z-score units)

Reliability (0-1)

- 0.70 or above for group comparisons
- 0.90 or above for individual assessment

z-scores (mean = 0 and SD = 1):

- Reliability = $1 - SE^2$
- So reliability = **0.90** when SE = 0.32

T-scores (mean = 50 and SD = 10):

$$T = 50 + (z * 10)$$

- Reliability = $1 - (SE/10)^2$
- So reliability = **0.90** when SE = 3.2

In the past 7 days …

I was grouchy [1st question]

- Never [39]
- Rarely [48]
- Sometimes [56]
- Often [64]
- Always [72]

Theta = 56.1 SE = 5.7 (rel. = 0.68)

In the past 7 days …

I felt like I was ready to explode

[2nd question]

- Never
- Rarely
- Sometimes
- Often
- Always

Theta = 51.9 SE = 4.8 (rel. = 0.77)

In the past 7 days …

I felt angry [3rd question]

- Never
- Rarely
- Sometimes
- Often
- Always

Theta = 50.5 SE = 3.9 (rel. = 0.85)

In the past 7 days …

I felt angrier than I thought I should
[4th question]

- Never
- Rarely
- Sometimes
- Often
- Always

Theta = 48.8 SE = 3.6 (rel. = 0.87)

In the past 7 days …

I felt annoyed [5th question]

- Never
- Rarely
- Sometimes
- Often
- Always

Theta = 50.1 SE = 3.2 (rel. = 0.90)

In the past 7 days …

I made myself angry about something just by thinking about it. [6th question]

- Never
- Rarely
- Sometimes
- Often
- Always

Theta = 50.2 SE = 2.8 (rel = 0.92)

Theta, SEM, and 95% CI

- 56 and 6 (reliability = .68) W = 22
- 52 and 5 (reliability = .77) W = 19
- 50 and 4 (reliability = .85) W = 15
- 49 and 4 (reliability = .87) W = 14
- 50 and 3 (reliability = .90) W = 12
- 50 and <3 (reliability = .92) W = 11

The following items are activities you might do during a typical day. Does your health limit you in these activities?

1. Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports
2. Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf
3. Lifting or carrying groceries
4. Climbing several flights of stairs
5. Climbing 1 flight of stairs
6. Bending, kneeling, or stooping
7. Walking more than a mile.
8. Walking several blocks.
9. Walking 1 block
10. Bathing or dressing yourself.

No, not limited at all

Yes, limited a little

Yes, limited a lot

11. In the past 4 weeks, to what extent did health problems limit you in your everyday physical activities (e.g., walking and climbing stairs)

- Not at all
- Slightly
- Moderately
- Quite a bit
- Extremely

12. How satisfied are you with your physical ability to do what you want to do?

- Completely satisfied
- Very satisfied
- Somewhat satisfied
- Somewhat dissatisfied
- Very dissatisfied
- Completely dissatisfied

13. When you travel around your community, does someone have to assist you because of your health?

- Yes, all of the time
- Yes, most of the time
- Yes, some of the time
- Yes, a little of the time
- No, none of the time

14. Are you in bed or in a chair most or all of the day because of your health?

- Yes, every day
- Yes, most days
- Yes, some days
- Yes, occasionally
- No, never

15. Are you able to use public transportation?

- No, because of my health
- No, for some other reason
- Yes, able to use public transportation

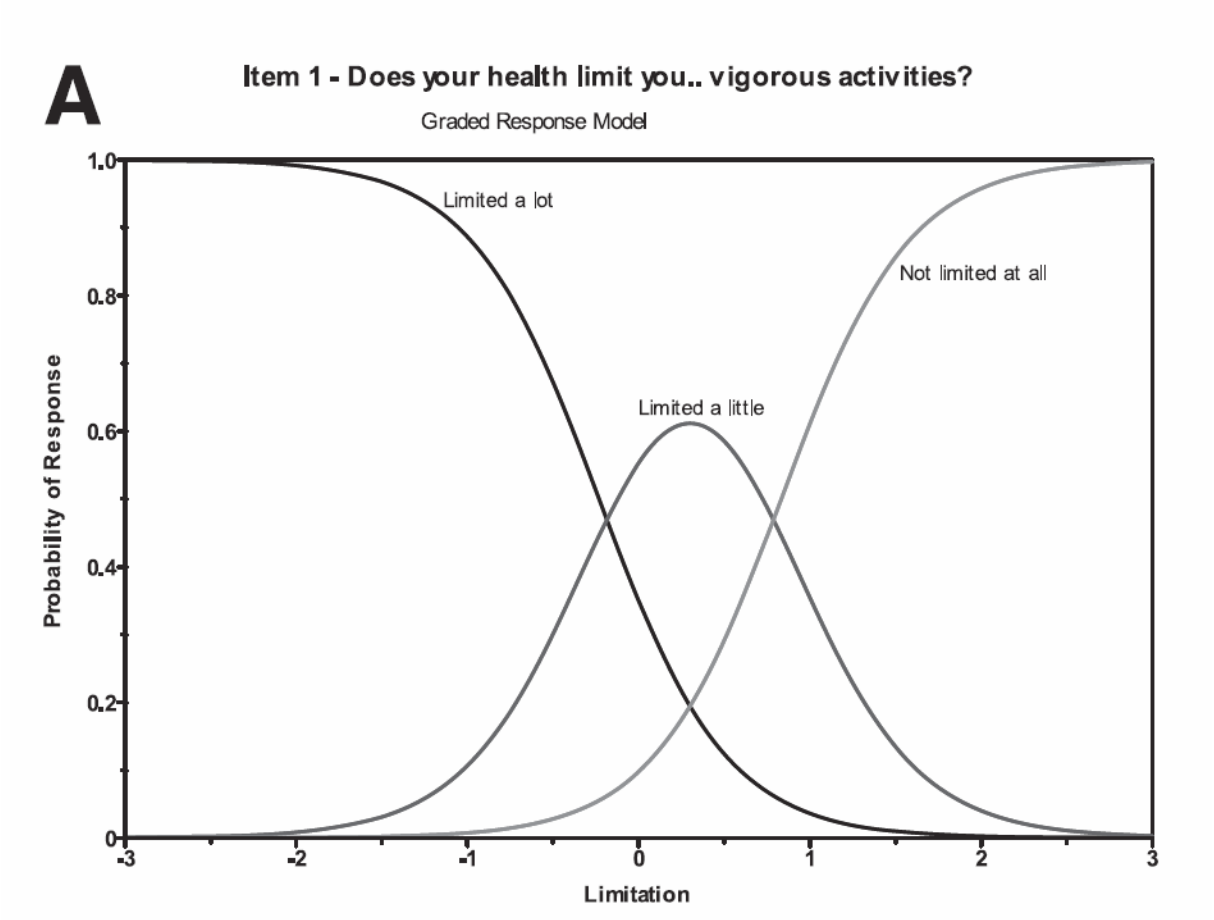
TABLE 1. Loadings from One-Factor Confirmatory Categorical Model (n = 3000)

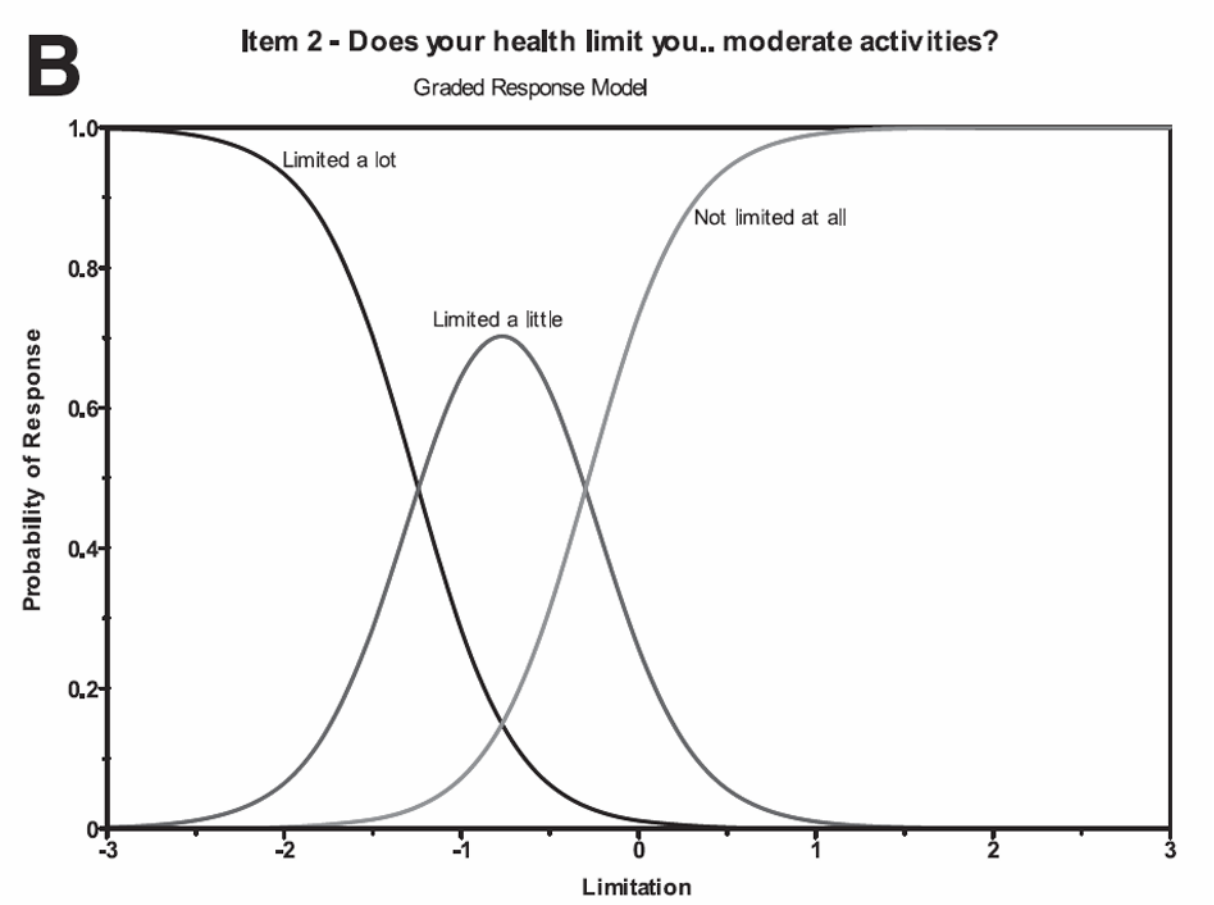
	Estimate	Standard Error	t-Statistic
Item 1	0.817	0.009	89.994
Item 2	0.900	0.006	150.430
Item 3	0.883	0.007	130.428
Item 4	0.893	0.006	160.859
Item 5	0.922	0.006	157.661
Item 6	0.808	0.009	86.793
Item 7	0.907	0.005	184.953
Item 8	0.962	0.003	276.161
Item 9	0.947	0.005	182.619
Item 10	0.771	0.015	50.003
Item 11	0.838	0.008	109.927
Item 12	0.725	0.009	76.944
Item 13	0.818	0.015	53.084
Item 14	0.772	0.013	60.284
Item 15	0.741	0.017	43.350

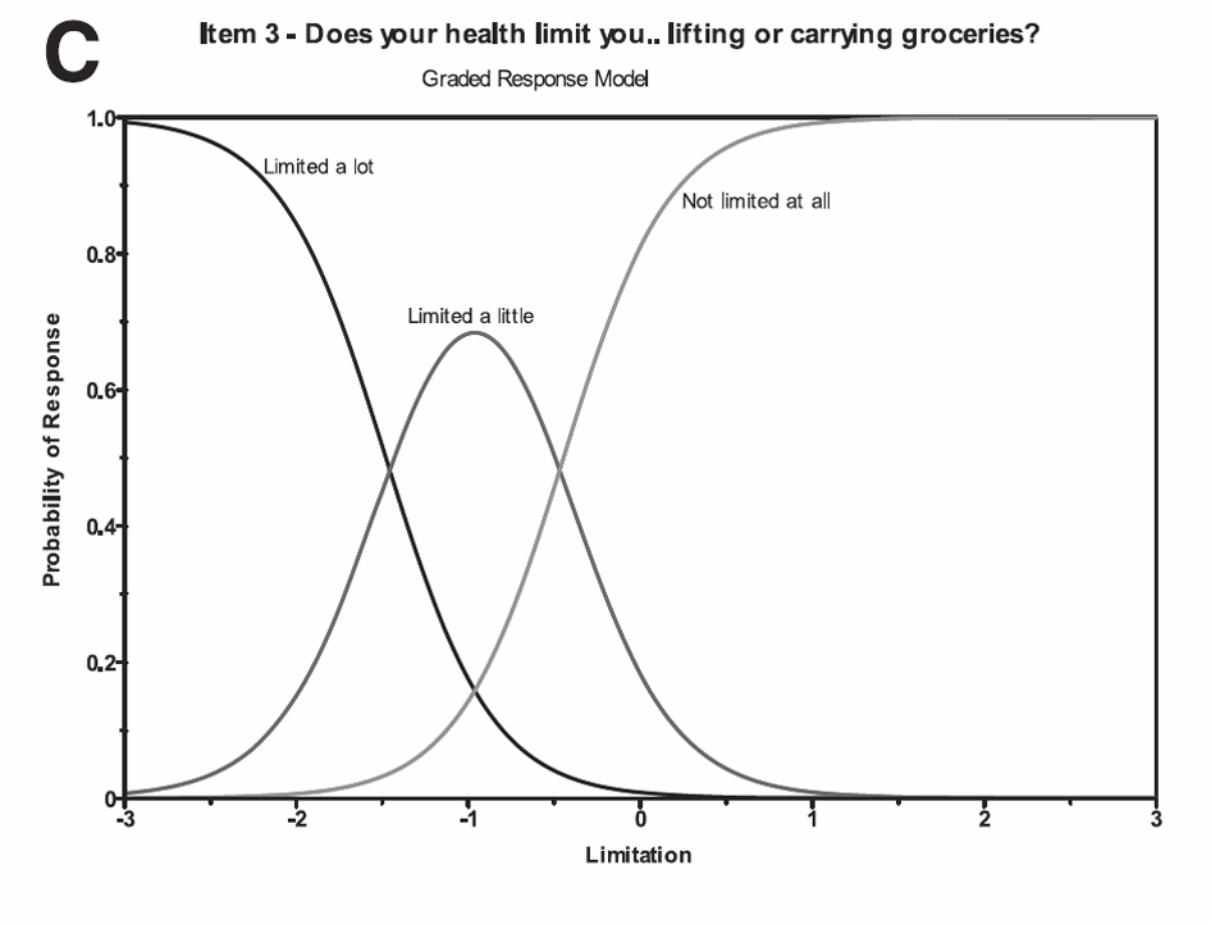
Comparative Fit Index = 0.95; Root Mean Square Error of Approximation = 0.12
(Cronbach's coefficent alpha = 0.94)

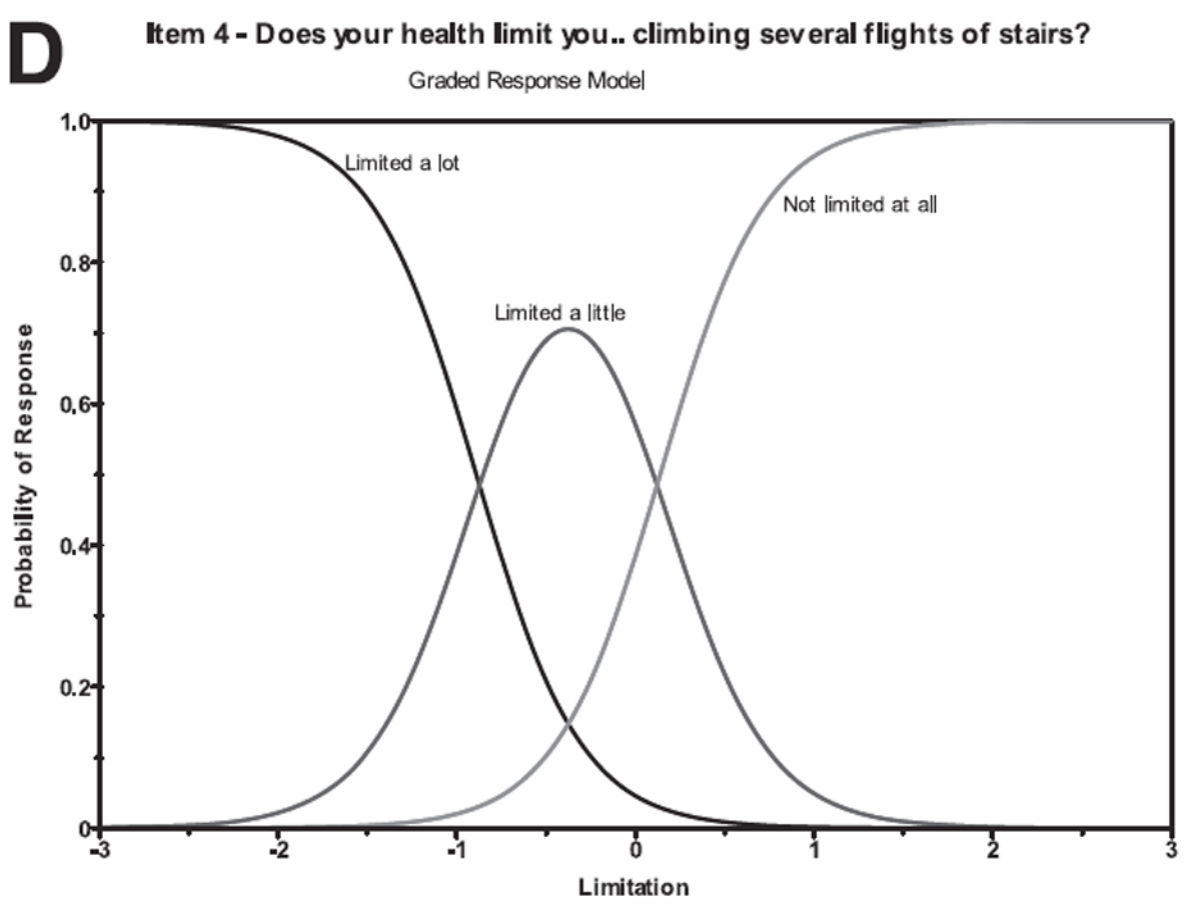
TABLE 2. Slope (Discrimination) and Category Threshold Estimates for 1-PL and 2-PL Models (in parentheses; n = 3223)

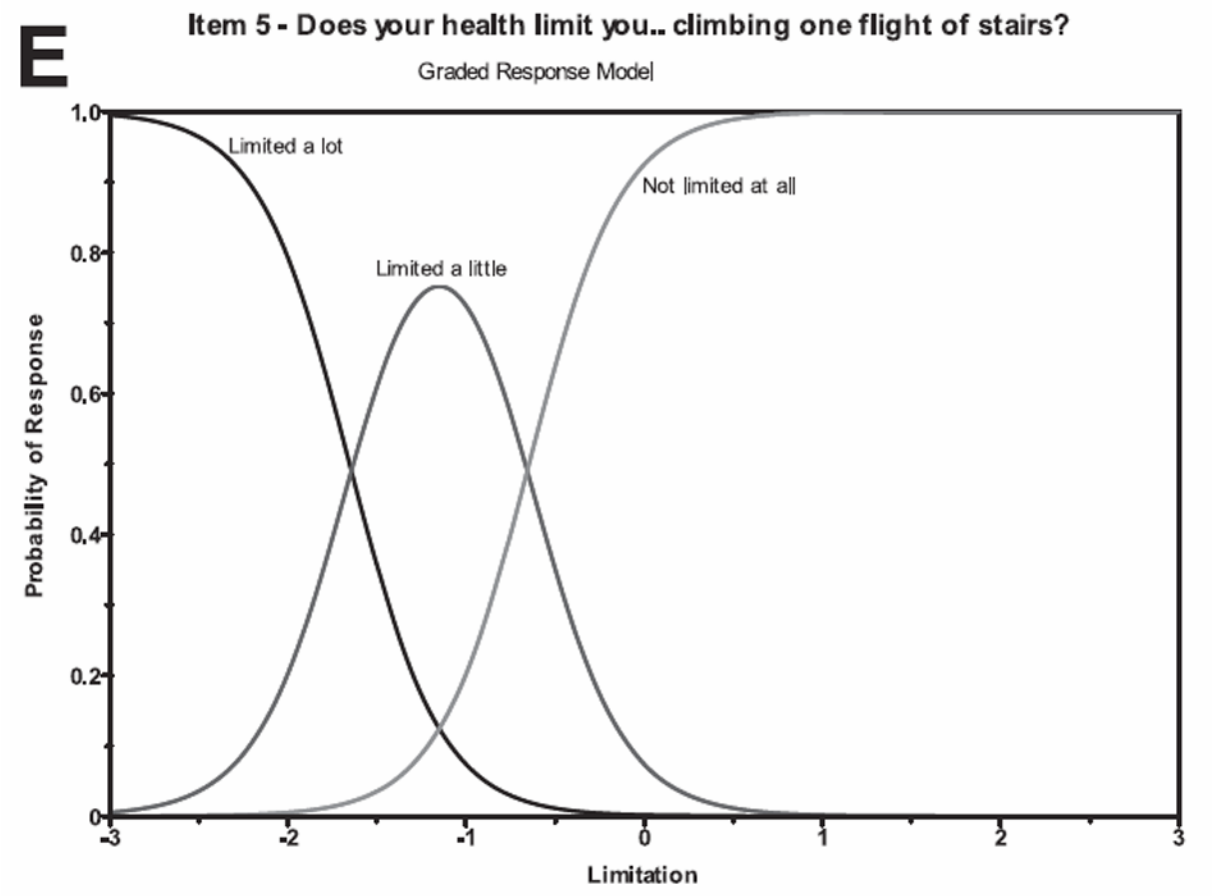
Item	Slope (Discrimination)	Category Threshold Estimates				
Item 1	2.73 (2.67)	−0.22 (−0.23)	0.83 (0.83)			
Item 2	2.73 (3.56)	−1.38 (−1.26)	−0.29 (−0.28)			
Item 3	2.73 (3.23)	−1.58 (−1.48)	−0.46 (−0.45)			
Item 4	2.73 (3.41)	−0.97 (−0.89)	0.17 (0.14)			
Item 5	2.73 (3.87)	−1.84 (−1.65)	−0.71 (−0.65)			
Item 6	2.73 (2.46)	−1.40 (−1.43)	−0.19 (−0.19)			
Item 7	2.73 (3.55)	−0.88 (−0.81)	0.01 (−0.01)			
Item 8	2.73 (4.16)	−1.27 (−1.12)	−0.49 (−0.45)			
Item 9	2.73 (4.24)	−1.92 (−1.70)	−1.07 (−0.97)			
Item 10	2.73 (2.26)	−2.48 (−2.66)	−1.55 (−1.62)			
Item 11	2.73 (2.74)	−2.46 (−2.44)	−1.33 (−1.30)	−0.71 (−0.70)	0.02 (0.01)	
Item 12	2.73 (1.88)	−2.19 (−2.58)	−1.28 (−1.49)	−0.51 (−0.59)	0.31 (0.35)	1.32 (1.55)
Item 13	2.73 (2.40)	−2.61 (−2.72)	−2.34 (−2.43)	−1.91 (−1.97)	−1.52 (−1.56)	
Item 14	2.73 (2.08)	−2.60 (−2.90)	−2.21 (−2.46)	−1.74 (−1.91)	−0.93 (−1.00)	
Item 15	2.73 (1.94)	−2.04 (−2.33)	−1.44 (−1.60)			


A
Item 1 - Does your health limit you.. vigorous activities?
Graded Response Model
Probability of Response
1.0
0.8
0.6
0.4
0.2
0
Limited a lot
Limited a little
Not limited at all
-3
-2
-1
0
1
2
3
Limitation

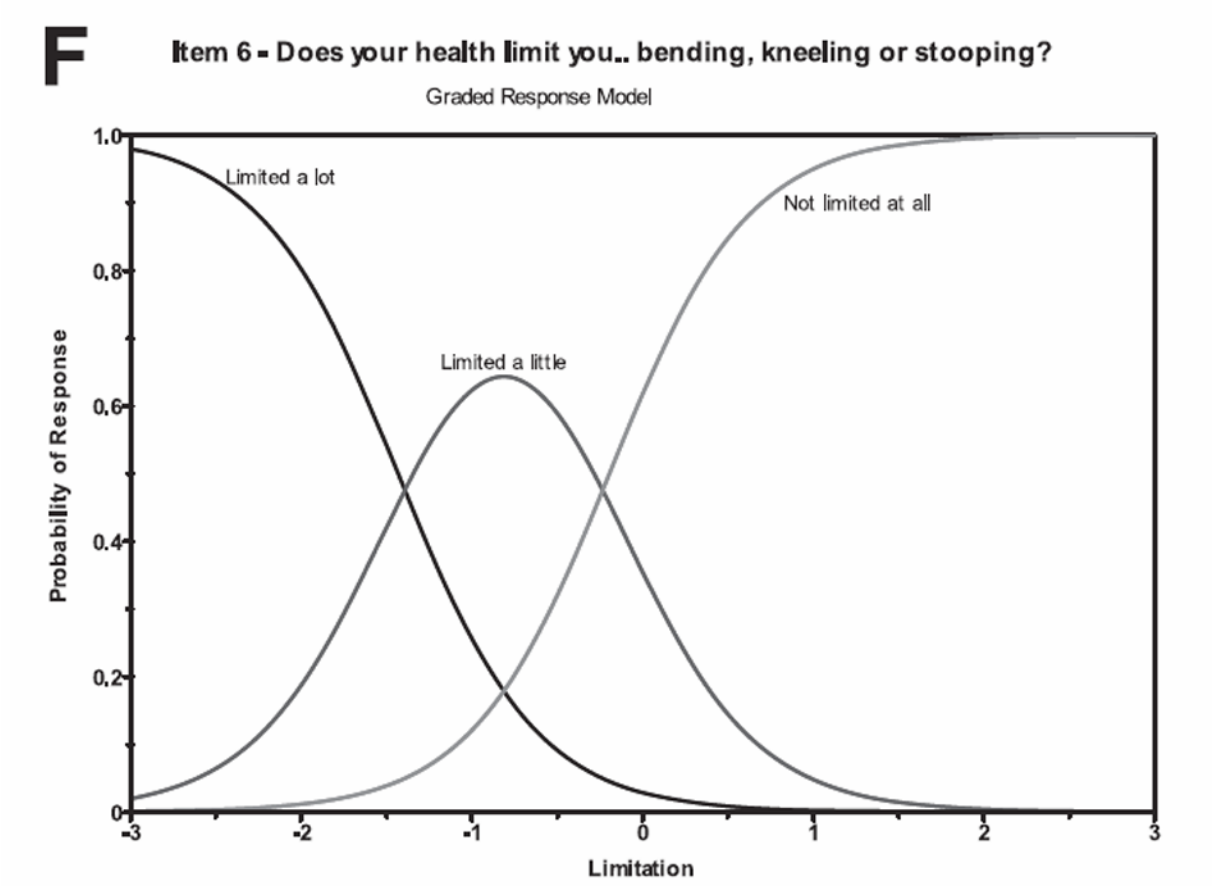

B
Item 2 - Does your health limit you.. moderate activities?
Graded Response Model
Limited a lot
Limited a little
Not limited at all
Probability of Response
1.0
0.8
0.6
0.4
0.2
0
-3
-2
-1
0
1
2
3
Limitation


C
Item 3 - Does your health limit you.. lifting or carrying groceries?
Graded Response Model
Limited a lot
Limited a little
Not limited at all
Probability of Response
1.0
0.8
0.6
0.4
0.2
0
-3
-2
-1
0
1
2
3
Limitation

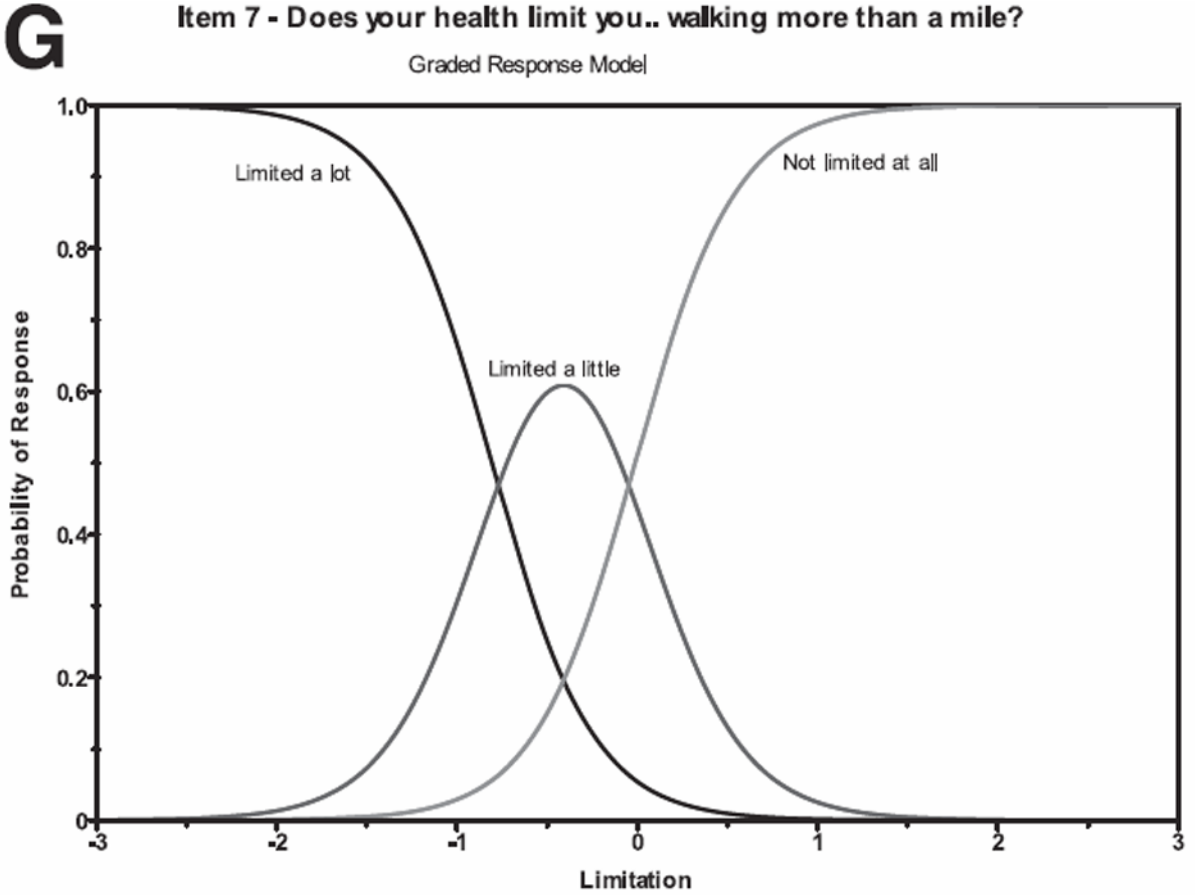

D
Item 4 - Does your health limit you.. climbing several flights of stairs?
Graded Response Model
Limited a lot
Limited a little
Not limited at all
Probability of Response
1.0
0.8
0.6
0.4
0.2
0
-3
-2
-1
0
1
2
3
Limitation


E
Item 5 - Does your health limit you.. climbing one flight of stairs?
Graded Response Model
Limited a lot
Limited a little
Not limited at all
Probability of Response
1.0
0.8
0.6
0.4
0.2
0
-3
-2
-1
0
1
2
3
Limitation

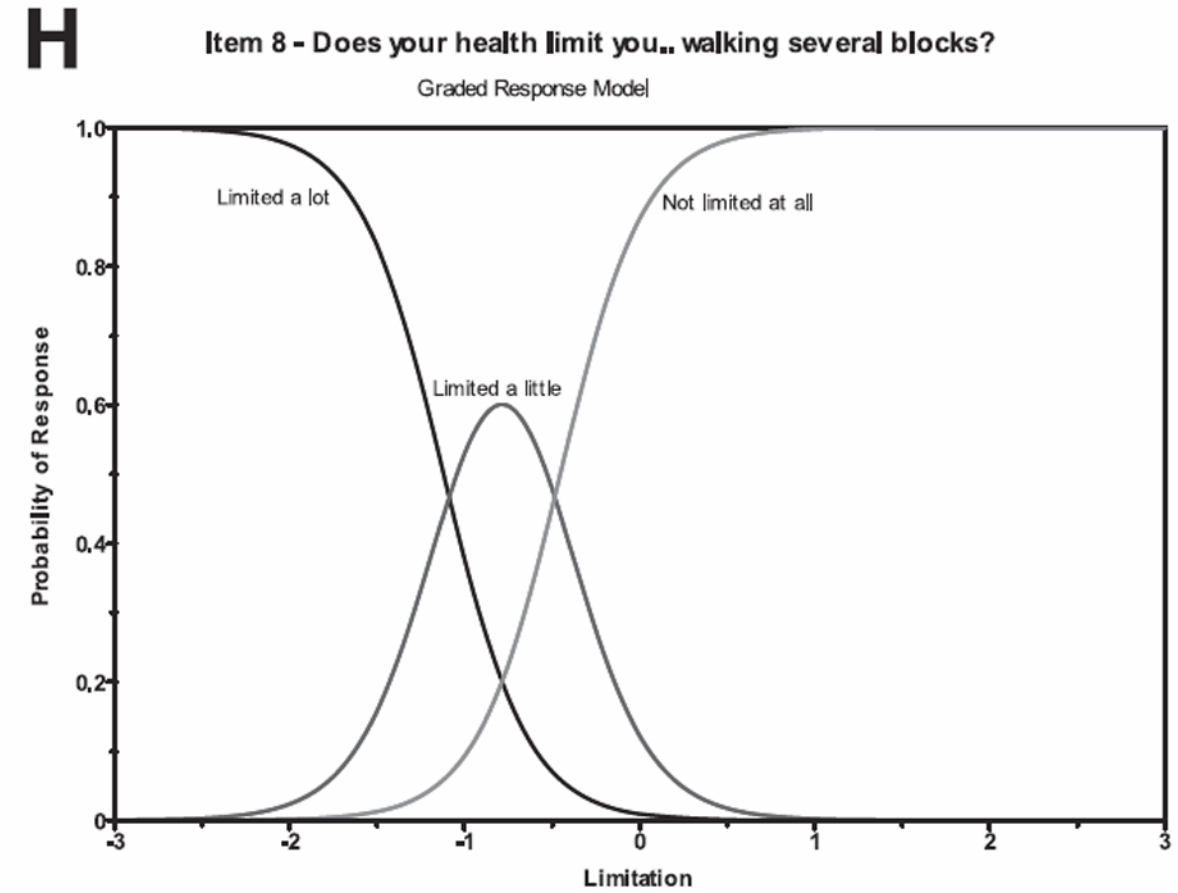
F

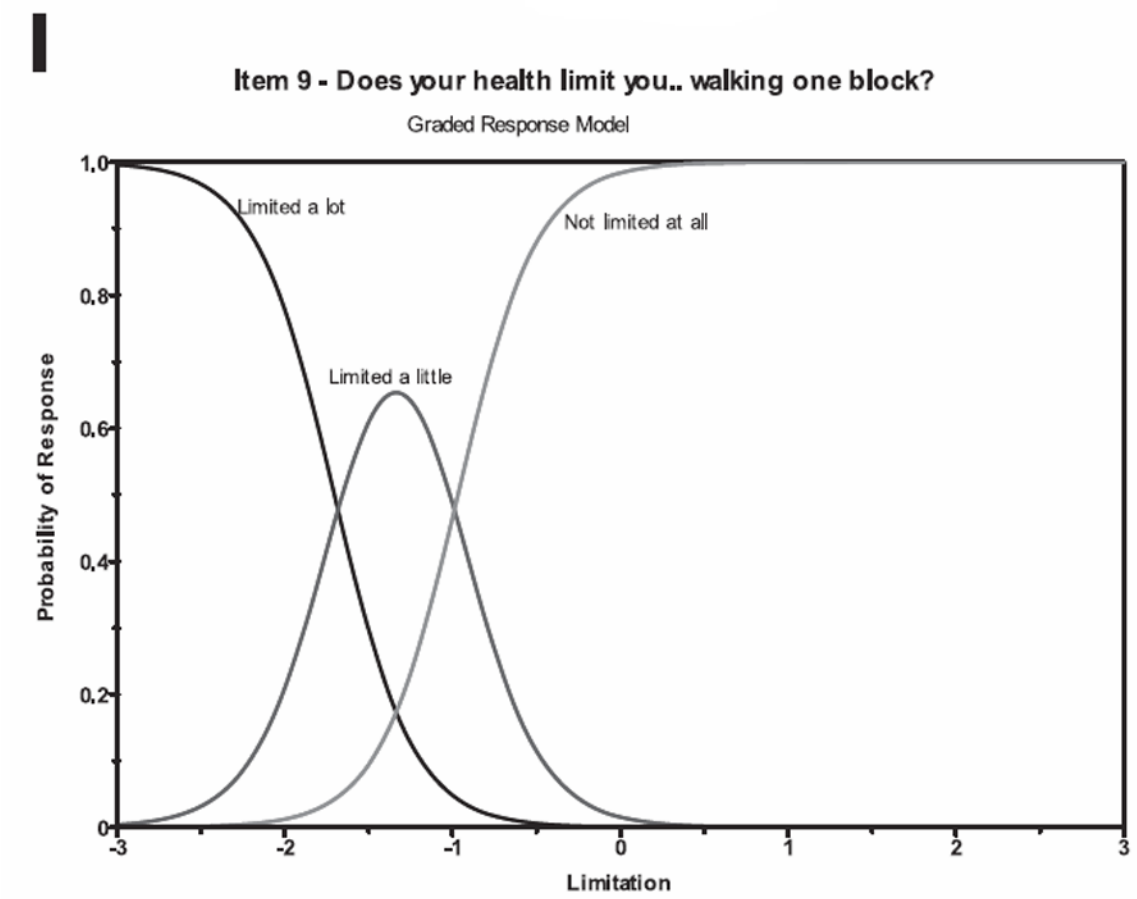

Item 6 - Does your health limit you.. bending, kneeling or stooping?
Graded Response Model
Limited a lot
Limited a little
Not limited at all
Probability of Response
1.0
0.8
0.6
0.4
0.2
0
-3
-2
-1
0
1
2
3
Limitation

G

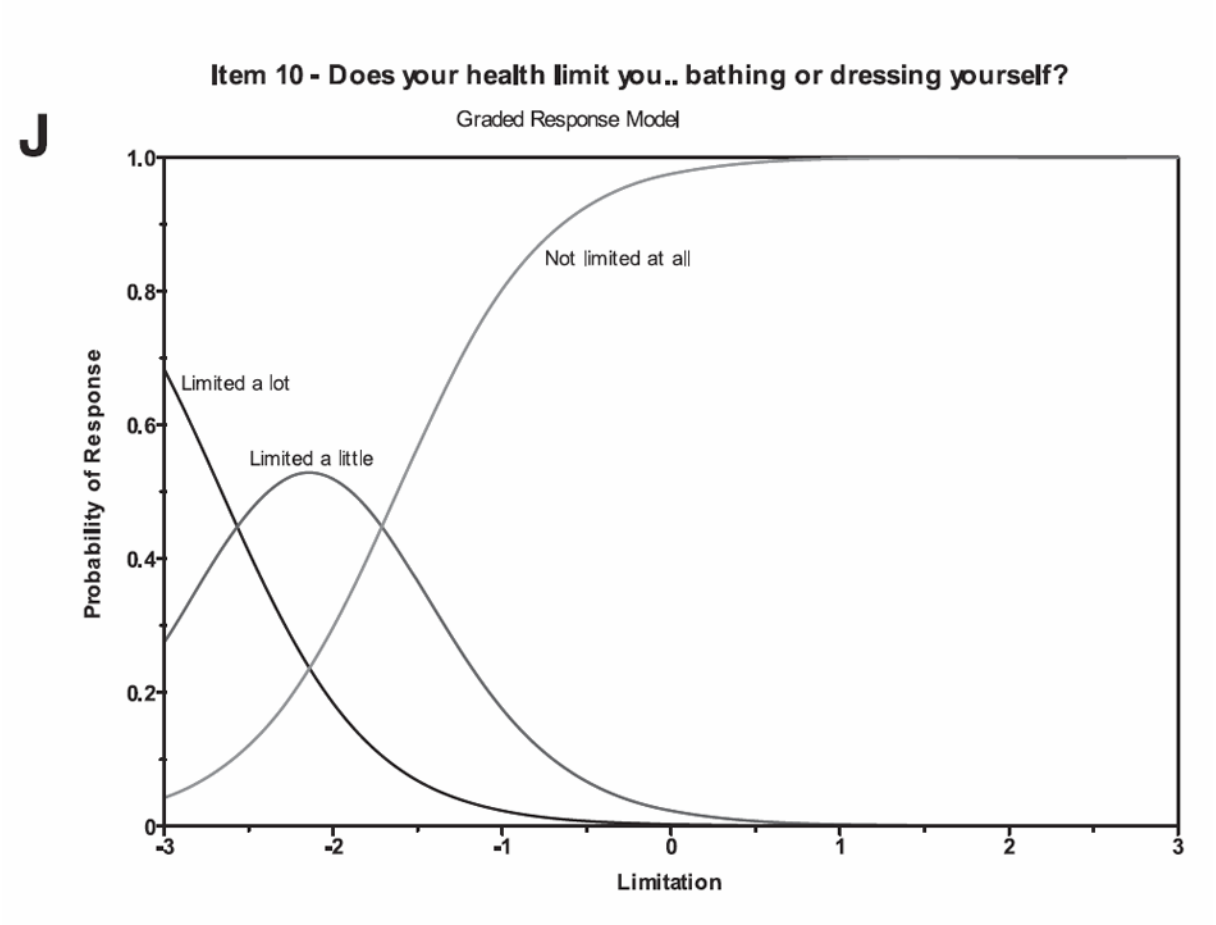

Item 7 - Does your health limit you.. walking more than a mile?
Graded Response Model
Limited a lot
Limited a little
Not limited at all
Probability of Response
1.0
0.8
0.6
0.4
0.2
0
-3
-2
-1
0
1
2
3
Limitation

H

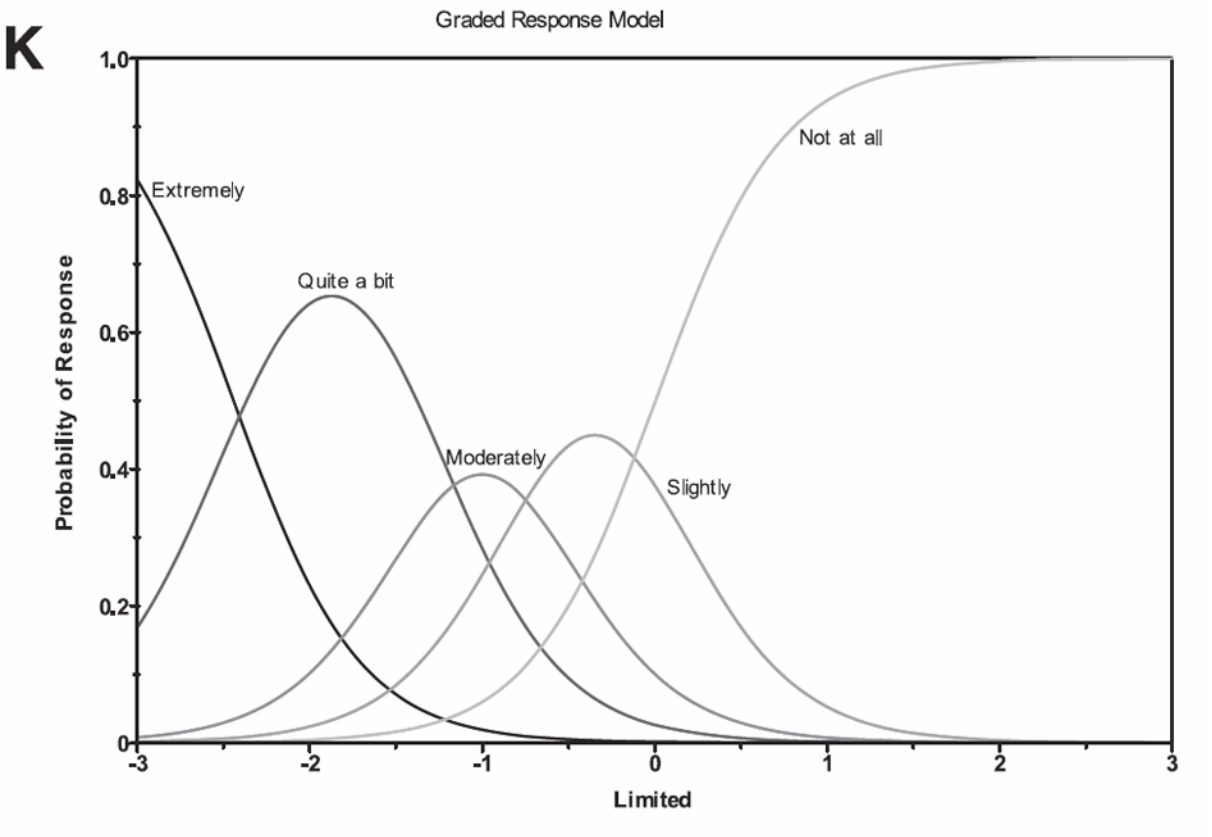

Item 8 - Does your health limit you.. walking several blocks?
Graded Response Model
Limited a lot
Limited a little
Not limited at all
Probability of Response
1.0
0.8
0.6
0.4
0.2
0
-3
-2
-1
0
1
2
3
Limitation


Item 9 - Does your health limit you.. walking one block?
Graded Response Model
Limited a lot
Limited a little
Not limited at all
Probability of Response
1.0
0.8
0.6
0.4
0.2
0
-3
-2
-1
0
1
2
3
Limitation

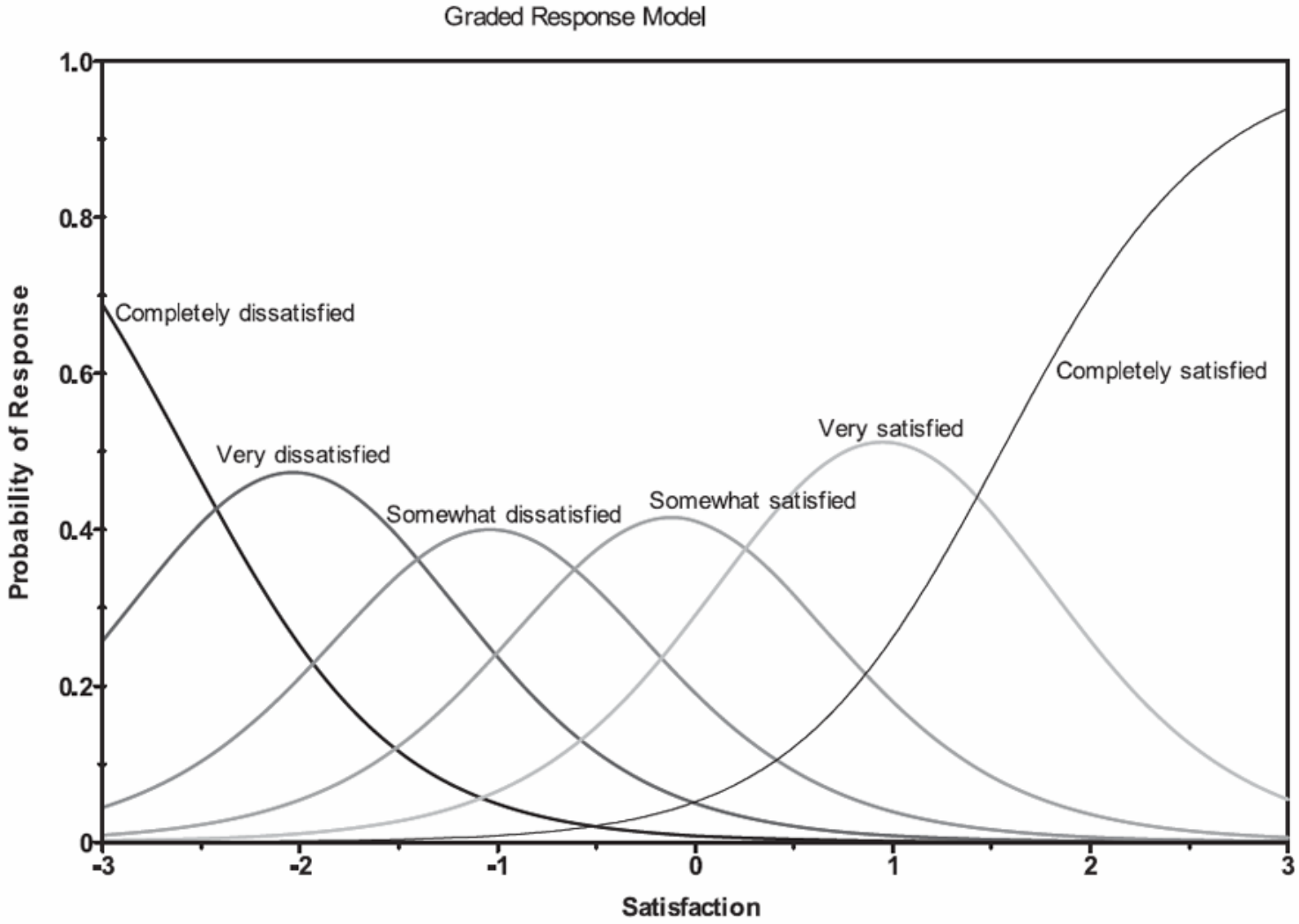
J


Item 10 - Does your health limit you.. bathing or dressing yourself?
Graded Response Model
Probability of Response
1.0
0.8
0.6
0.4
0.2
0
Not limited at all
Limited a lot
Limited a little
-3
-2
-1
0
1
2
3
Limitation

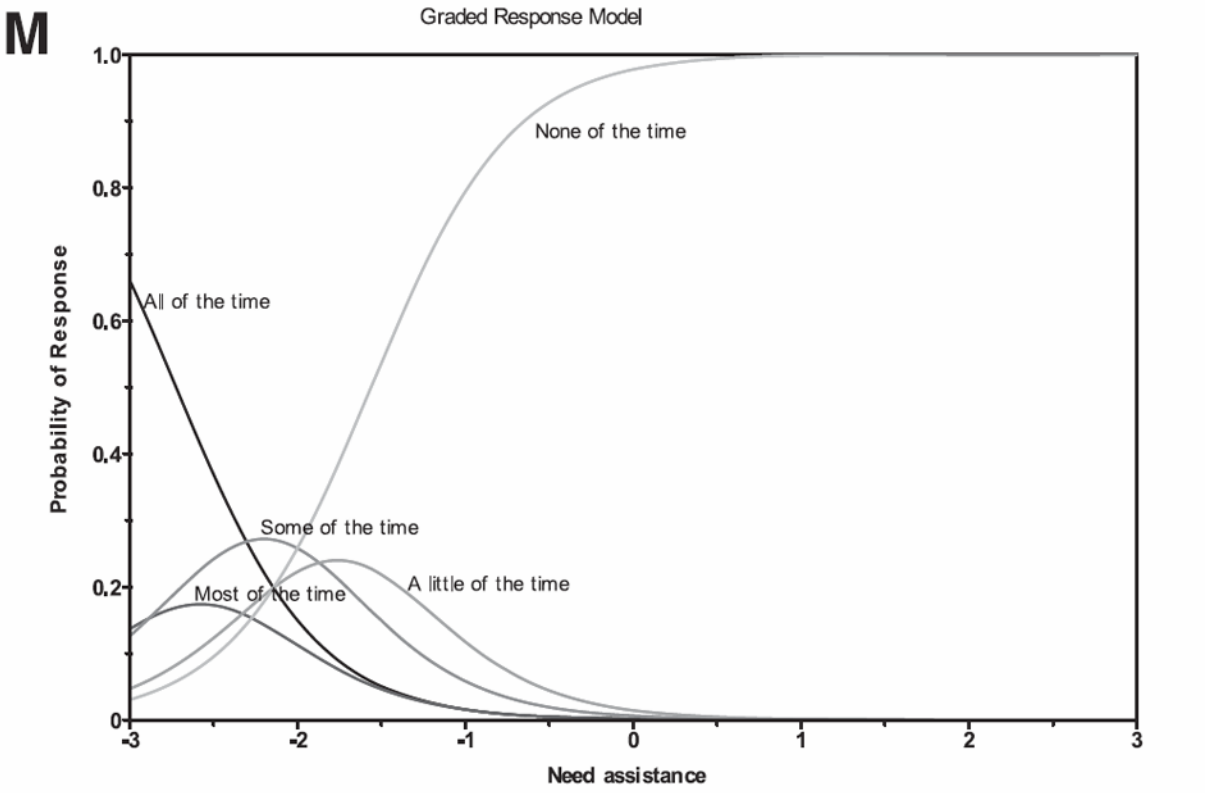
In the past 4 weeks, did health problems limit you in your everyday physical activities?

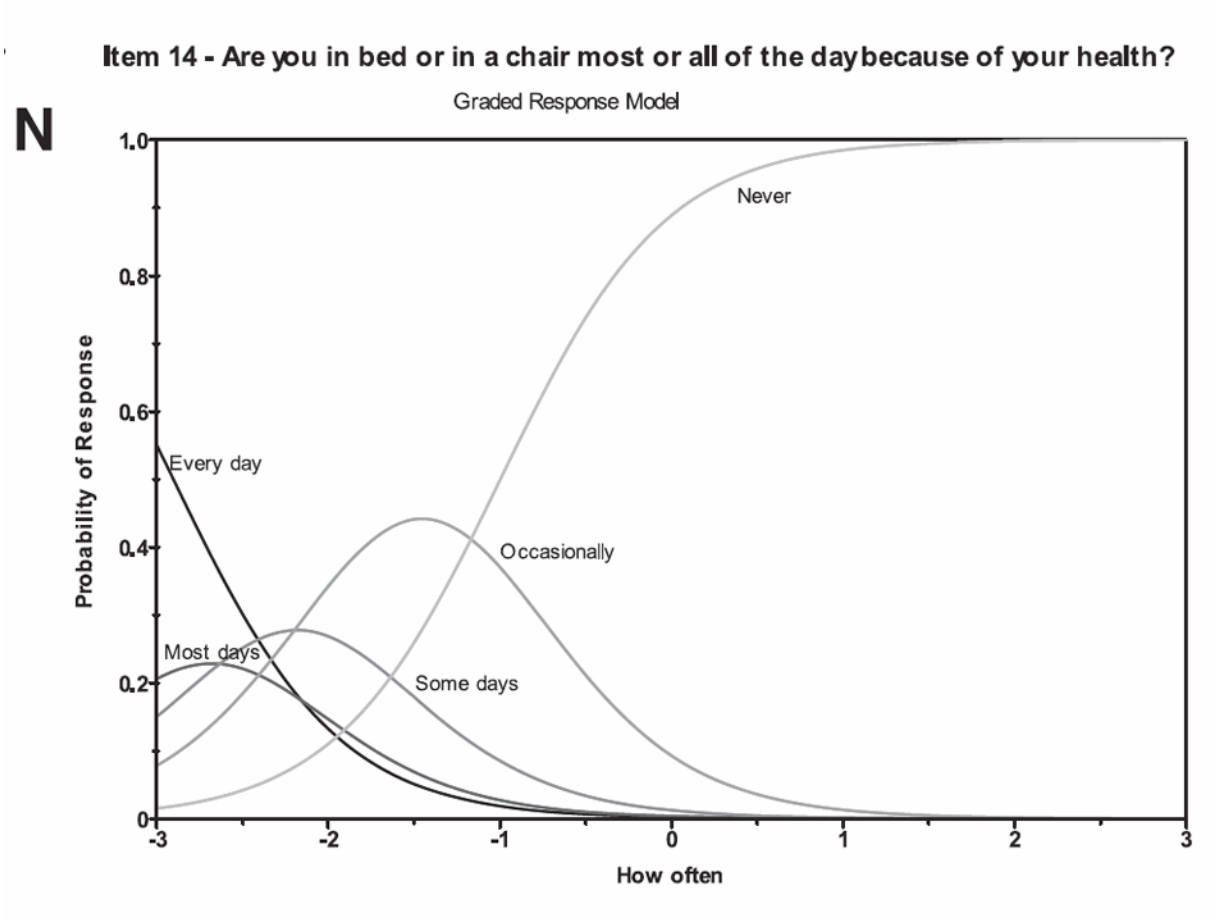


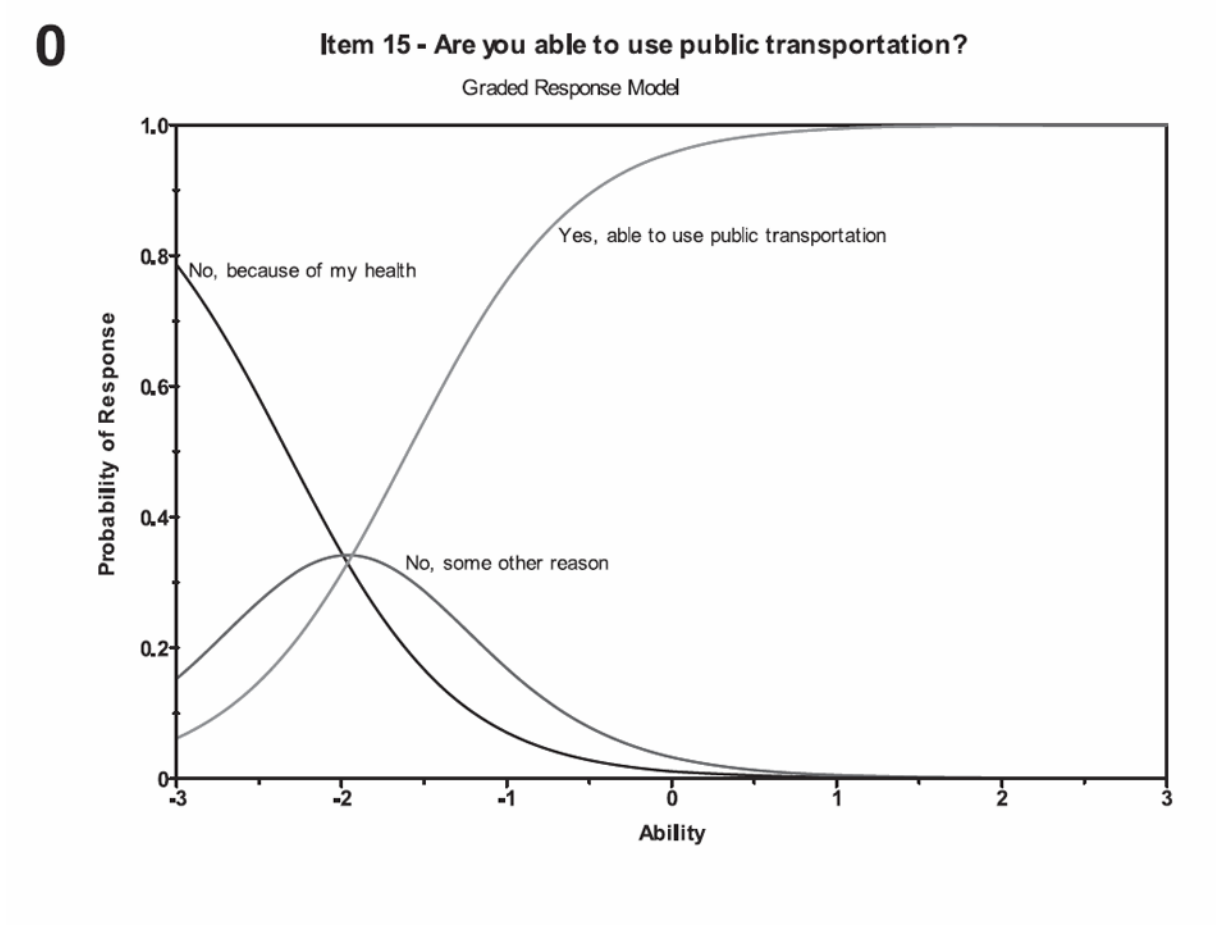
L **Item 12 - How satisfied are you with your physical ability to do what you want to do?**

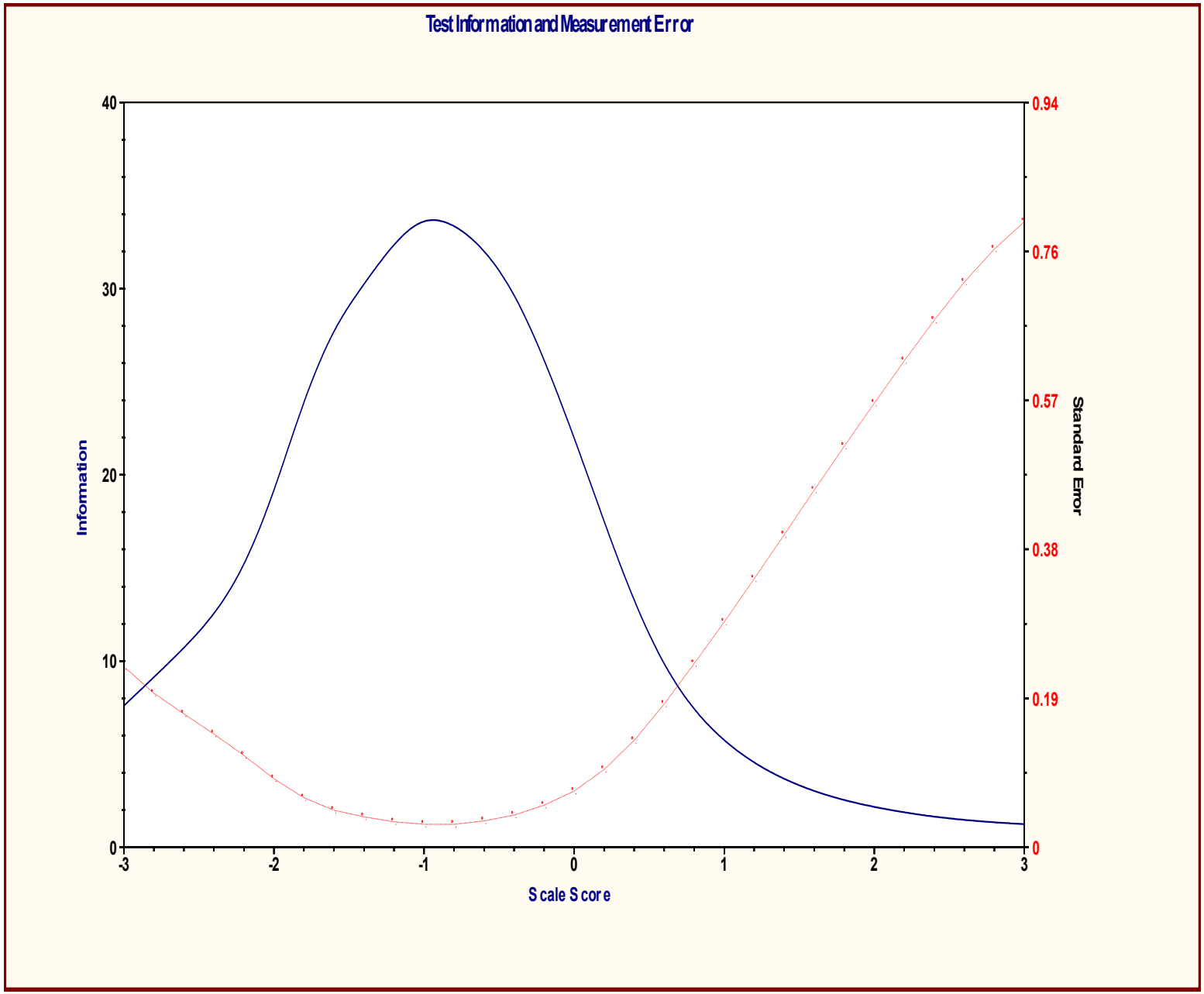


Item 13 - W hen you travel around your community.. need assistance because of your health?




Item 14 - Are you in bed or in a chair most or all of the daybecause of your health?
Graded Response Model
N
Probability of Response
1.0
0.8
0.6
0.4
0.2
0
Never
Every day
Occasionally
Most days
Some days
-3
-2
-1
0
1
2
3
How often


Item 15 - Are you able to use public transportation?
Graded Response Model
Probability of Response
1.0
0.8
0.6
0.4
0.2
0
No, because of my health
Yes, able to use public transportation
No, some other reason
-3
-2
-1
0
1
2
3
Ability


Test Information and Measurement Error
Information
Standard Error
Scale Score
40
30
20
10
0
0.94
0.76
0.57
0.38
0.19
0
-3
-2
-1
0
1
2
3

Item-scale correlation matrix

	Depress	Anxiety	Anger
Item #1	0.80*	0.20	0.20
Item #2	0.80*	0.20	0.20
Item #3	0.80*	0.20	0.20
Item #4	0.20	0.80*	0.20
Item #5	0.20	0.80*	0.20
Item #6	0.20	0.80*	0.20
Item #7	0.20	0.20	0.80*
Item #8	0.20	0.20	0.80*
Item #9	0.20	0.20	0.80*

*Item-scale correlation, corrected for overlap.

Item-scale correlation matrix

	Depress	Anxiety	Anger
Item #1	0.50*	0.50	0.50
Item #2	0.50*	0.50	0.50
Item #3	0.50*	0.50	0.50
Item #4	0.50	0.50*	0.50
Item #5	0.50	0.50*	0.50
Item #6	0.50	0.50*	0.50
Item #7	0.50	0.50	0.50*
Item #8	0.50	0.50	0.50*
Item #9	0.50	0.50	0.50*

*Item-scale correlation, corrected for overlap.

Evaluating Validity

Scale	Age	Obesity	ESRD	Nursing Home Resident
Physical Functioning	Medium (-)	Small (-)	Large (-)	Large (-)
Depressive Symptoms	?	Small (+)	?	Small (+)

Cohen effect size rules of thumb (d = 0.2, 0.5, and 0.8):
Small correlation = 0.100
Medium correlation = 0.243
Large correlation = 0.371
$r = d / [(d^2 + 4)^{.5}] = 0.8 / [(0.8^2 + 4)^{.5}] = 0.8 / [(0.64 + 4)^{.5}] = 0.8 / [(4.64)^{.5}]$
$= 0.8 / 2.154 = 0.371$.

Note: Often r's of 0.10, 0.30 and 0.50 are cited as small, medium, and large.

Change on SF-36 Physical Functioning Scale by Self-reported Retrospective Rating of Change

Interval	Lot Better (n = 21)	Little Better (n = 35)	Same (n = 252)	Little Worse (n = 113)	Lot Worse (n = 30)
12 months	4.99[a]	0.32[,b]	0.46[b]	-3.86[c]	-4.74[c]
6 months	4.08[a]	-0.58[b,c]	0.89[b]	-2.34[c]	-3.47[c]

Note: Cell entries in the same row that share a letter do not differ significantly ($p > 0.05$) from one another (Duncan's multiple range tests). SD of change was 7.74 for 12 months and 7.08 for 6 months.

Questions?

Powerpoint file posted at URL below (freely available for you to use, copy or burn):

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

Contact information:

drhays@ucla.edu 310-794-2294

For a good time call 867-5309 or go to:
http://twitter.com/RonDHays

Appendix 1: For more information

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.

Hays, R. D., Liu, H., Spritzer, K., & Cella, D. (2007). Item response theory analyses of physical functioning items in the Medical Outcomes Study. Medical Care, 45, S32-38.

Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Young, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., Cook, K., DeVellis, R., DeWalt, D., Fries, J. F., Gershon, R., Hahn, E. A., Pilkonis, P., Revicki, D., Rose, M., Weinfurt, K., Lai, J., & Hays, R. D. (2010). Initial item banks and first wave testing of the Patient-Reported Outcomes Measurement Information System (PROMIS) network: 2005-2008. Journal of Clinical Epidemiology, 63 (11), 1179-1194.

Appendix 2:
Item Response Theory (IRT)

IRT models the relationship between a person's response Y_i to the question (i) and his or her level of the latent construct θ being measured by positing

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

b_{ik} estimates how difficult it is for the item (i) to have a score of k or more and the discrimination parameter a_i estimates the discriminatory power of the item.

Appendix 3: Intraclass Correlation and Reliability

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

BMS = Between Ratee Mean Square
WMS = Within Mean Square
JMS = Item or Rater Mean Square
EMS = Ratee x Item (Rater) Mean Square

N = n of ratees
k = n of items or raters

Appendix 4: Confirmatory Factor Analysis Fit Indices

- Normed fit index: $\frac{\chi^2_{null} - \chi^2_{model}}{\chi^2_{null}}$

- Non-normed fit index: $\frac{\frac{\chi^2_{null}}{df_{null}} - \frac{\chi^2_{model}}{df_{model}}}{\left[\frac{\chi^2_{null}}{df_{null}} - 1\right]}$

- Comparative fit index: $1 - \frac{\chi^2_{model} - df_{model}}{\chi^2_{null} - df_{null}}$

RMSEA = SQRT (λ^2 – df)/SQRT (df (N – 1))