

# Item Response Theory (IRT) Models for Questionnaire Evaluation: Response to Reeve

Ron D. Hays

*October 22, 2009, ~3:45-4:05pm*

<http://twitter.com/RonDHays>

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

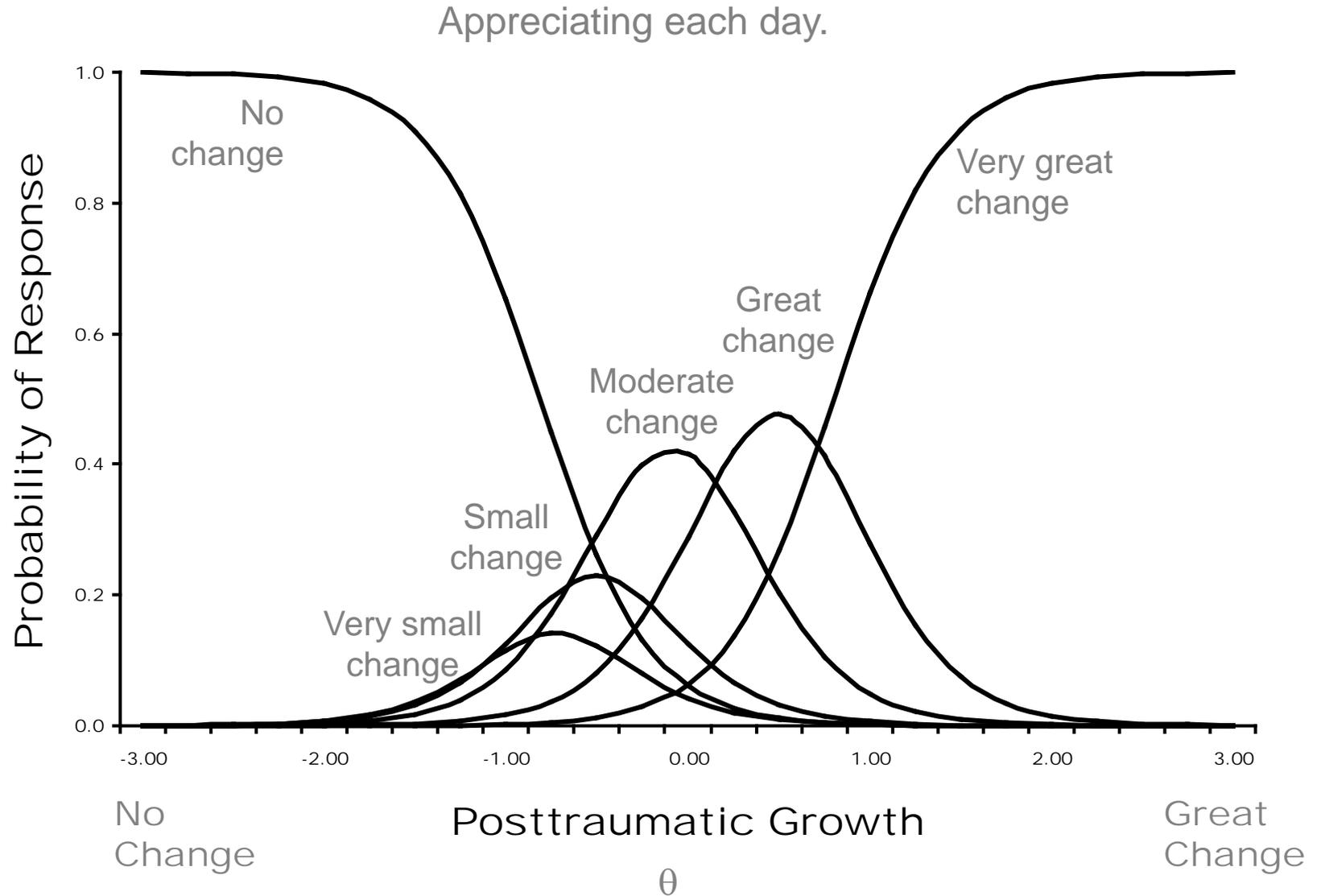
# Features of IRT with diagnostic utility

- Category response curves
- Information/reliability
- Differential item functioning
- Person fit
- Computer-adaptive testing

# Category Response Curves (CRCs)

- Reeve's Figure 7 showed that 2 of 6 response options are never most likely to be chosen
  - No, **very small**, **small**, moderate, great, very great change
- He suggests 1 or both of the response categories could be dropped or reworded to improve the response scale

# Figure 7



# Drop response options?

- No, very small, small, moderate, great, very great change



- No, moderate, great, very great change

# Reword?

- Might be challenging to determine what alternative wording to use so that the replacements are more likely to be endorsed.

# Keep as is?

- CAHPS global rating items
  - 0 = worst possible
  - 10 = best possible
- 11 response categories capture about 3 levels of information.
  - 10/9/8-0 or 10-9/8/7-0
- Scale is administered as is and then collapsed in analysis

# Information/Reliability

- For z-scores (mean = 0 and SD = 1):
  - Reliability =  $1 - SE^2 = \mathbf{0.90}$  (when SE = 0.32)
  - Information =  $1/SE^2 = \mathbf{10}$  (when SE = 0.32)
  - Reliability =  $1 - 1/\text{information}$
- Lowering the SE requires adding or replacing existing items with more informative items at the target range of the continuum.
  - But this is ...

# Easier said than done

- Limit on the number of ways to ask about a targeted range of the construct
- One needs to avoid asking the same item multiple times.
  - “I’m generally sad about my life.”
  - “My life is generally sad.”
- Local independence assumption
  - Significant residual correlations

# Item parameters (graded response model) for global physical health items in Patient-Reported Outcomes Measurement Information System

Item	A	b1	b2	b3	b4
Global01	7.37 (na)	-1.98 (na)	-0.97 (na)	0.03 (na)	1.13 (na)
Global03	7.65 (2.31)	-1.89 (-2.11)	-0.86 (-0.89)	0.15 ( 0.29)	1.20 ( 1.54)
Global06	1.86 (2.99)	-3.57 (-2.80)	-2.24 (-1.78)	-1.35 (-1.04)	-0.58 (-0.40)
Global07	1.13 (1.74)	-5.39 (-3.87)	-2.45 (-1.81)	-0.98 (-0.67)	1.18 ( 1.00)
Global08	1.35 (1.90)	-4.16 (-3.24)	-2.39 (-1.88)	-0.54 (-0.36)	1.31 ( 1.17)

Note: Parameter estimates for 5-item scale are shown first, followed by estimates for 4-item scale (in parentheses). na = not applicable

Global01: In general, would you say your health is ...? Global03: In general, how would you rate your physical health? Global06: To what extent are you able to carry out your everyday physical activities? Global07: How would you rate your pain on average? Global08: How would you rate your fatigue on average?

a = discrimination parameter; b1 = 1<sup>st</sup> threshold; b2 = 2<sup>nd</sup> threshold; b3 = 3<sup>rd</sup> threshold; b4 = 4<sup>th</sup> threshold

# Differential Item Functioning (DIF)

- Probability of choosing each response category should be the same for those who have the same estimated scale score, regardless of their other characteristics
- Evaluation of DIF
  - Different subgroups
  - Mode differences
  - Different response options

# Person Fit

- Large negative  $Z_L$  values indicate misfit.
- Person responded to 14 items in physical functioning bank ( $Z_L = -3.13$ )
  - For 13 items the person could do the activity (including running 5 miles) without any difficulty.
  - However, this person reported *a little difficulty* being out of bed for most of the day.

# Unique predictors of person misfit

- Less than high school education
- Non-white
- More chronic conditions

# Computer Adaptive Testing (CAT)

<http://www.nihpromis.org/>

- Patient-reported outcomes measurement information system (PROMIS) project
  - Item banks measuring patient-reported outcomes
  - Computer-adaptive testing (CAT) system

.

# PROMIS Banks (454 items)

<http://www.assessmentcenter.net/ac1/>

- Emotional Distress
  - Depression (28)
  - Anxiety (29)
  - Anger (29)
- Physical Function (124)
- Pain
  - Behavior (39)
  - Impact (41)
- Fatigue (95)
- Satisfaction with Participation in Discretionary Social Activities (12)
- Satisfaction with Participation in Social Roles (14)
- Sleep Disturbance (27)
- Wake Disturbance (16)

# Time to complete item

- Polimetrix panel sample
- 12-13 items per minute (automatic advance)
- 8-9 items per minute (next button)
  - 6 items per minute among UCLA Scleroderma patients

# CAT

- Context effects (Lee & Grant, 2009)
  - 1,191 English and 824 Spanish respondents to 2007 California Health Interview Survey
  - Spanish respondents self-rated health was worse when asked before compared to after questions about chronic conditions.

# Assessment Center/Q-Bank

Assessment Center - Mozilla Firefox

File Edit View History Bookmarks Tools Help

https://www.assessmentcenter.net/ac1/Default.aspx?SID=46798650-0A7F-4A87-85D7-7DC3FD73E2E

Getting Started Latest Headlines

Gmail - Inbox (10166... Text and Web - Goo... Anticlimactic feeling c... why low standardize... Mplus Discussion >> ... Assessment Center

### Differential Item Functioning (DIF) Information

Update

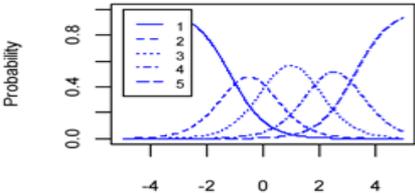
DIF Groups (Reference/Focal)	Uniform	Uniform Prob.	Non-Uniform	Non-Uniform Prob.
PROMIS Male - Female	1.22	0.2700	0.32	0.5695

### Calibration Sample

Description: PROMIS Promis Wave 1

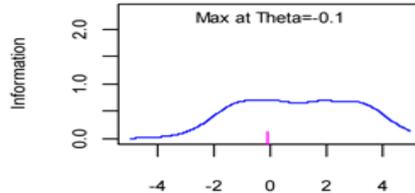
IRT Model: Graded Response Model

Category Response Function



Sample Population: PROMIS Wave 1 Full Bank

Item Information Function



Model Fit & Scalability Indices

	Value	Prob.
S-X <sup>2</sup>	73.87	0.5153
S-G <sup>2</sup>	79.41	0.3418
X <sup>2</sup>	33.00	0.0276
G <sup>2</sup>	49.90	0.0183
Loevinger H	0.00	
Scalability		
Location At Max Information	-0.10	

Threshold 1 is between the lowest scored response and the next highest response(s).

### IRT Parameters

Slope (a)	Guessing (c)	Threshold (CB) 1	Threshold (CB) 2	Threshold (CB) 3	Threshold (CB) 4
1.56467	0	-1.1758	0.1564	1.8111	3.2703

Done

www.assessmentcenter.net

Start Eudora - [... C:\Project... UltraEdit-3... Assessme... MailWasha... 2:12 PM

# Assessment Center/Q-Bank

Assessment Center - Windows Internet Explorer

https://www.assessmentcenter.net/ac1/Default.aspx?SID=7A84B49D-2878-4428-8BEB-2CDA152

File Edit View Favorites Tools Help

McAfee SiteAdvisor

hp Yahoo! Web Search Bookmarks Settings HP Access HP Solutions HP Create HP Smart Deals

Google Search Mail Calendar Plus Bookmarks Check Translate dr.ron...

Favorites Google Calendar (2) Google Calendar iGoogle (5) iGoogle (4) iGoogle (3) iGoogle (2) iGoogle Redirect (2) iGoogle Redirect

Assessment Center

**PROMIS Assessment Center**<sup>SM</sup>

Studies Instruments Set-up Administration Resources My Account Help Logout

Current Study: Frightening

[Study Content \(Frightening\)](#) >> Instrument Characteristics

### Instrument Statistics

Instrument: PROMIS Bank-Pain Behavior 1.0 Update Back

Calibration Sample	Promis Wave 1
Sample Description	Promis Wave 1 Full-Bank
Scaling Model Used For Calibration	Graded Response Model
Total Number of Items	39

	Conditional Reliability									Min	Max
Scale Score	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0	36.2	76.3
SE	133.20	15.40	1.70	.20	.10	.10	.10	.50	5.30		
Reliability	.00	.00	.00	.97	.99	1.00	.99	.72	.00		

Done Internet 100%

start Sprint SmartView Assessment Center - ... Microsoft PowerPoint ... PROMIS 9:38 AM

# Thank you!

