Patient Reported Outcomes Measurement Information System (PROMIS)

National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) U01 grant

Ron D. Hays, Ph.D., UCLA

September 29, 2006

GIM/HSR Research Seminar Series

12:02pm-12:59 pm

NIH Director Elias A. Zerhouni, MD

"There is a pressing need to better quantify clinically important symptoms and outcomes that are now difficult to measure. Clinical outcome measures, such as x-rays and lab tests, have minimally immediate relevance to the day-to-day functioning of patients with chronic diseases such as arthritis, multiple sclerosis, and asthma, as well as chronic pain conditions.

Often, the best way patients can judge the effectiveness of treatments is by perceived changes in symptoms. One main goal of the PROMIS initiative is to develop a set of publicly available computerized adaptive tests for the clinical research community."

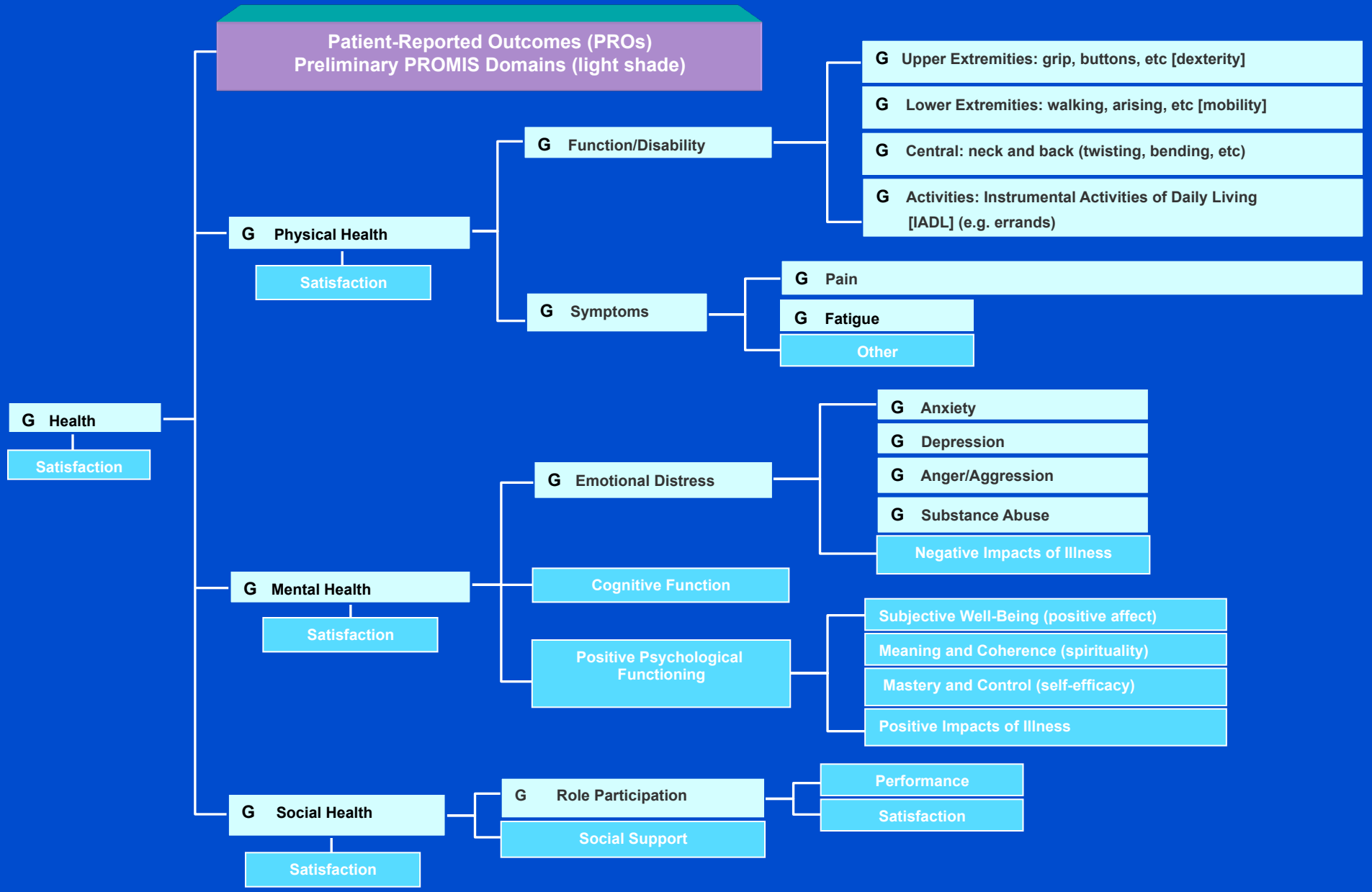
NIH RM04-011: Dynamic Assessment of Patient-Reported Chronic Disease Outcomes

Translation arm of re-engineering clinical research enterprise

Chronic diseases and their treatment affect "quality of life"

Improved assessment of "subjective" clinical outcomes

- Self-reported symptoms and other health-related quality of life domains



ADL – Activities of Daily Living
IADL – Instrumental Activities of Daily Living
G – Global Item

Initial PROMIS Domains

- Physical functioning (4)
- Pain (3)
- Fatigue (2)
- Social/role participation (2)
- Emotional distress
  - Anxiety
  - Depression
  - Anger
  - Alcohol abuse

Objectives

Develop and test a large bank of items measuring health-related quality of life

Create a publicly available, adaptable and sustainable system allowing clinical researchers access to a

* common item repository

* computer adaptive testing (CAT) platform

for efficient assessment across a range of chronic diseases

File Edit View Favorites Tools Help

Back Search Favorites

Address http://www.nihpromis.org Go Links

PROMIS

Patient-Reported Outcomes Measurement Information System

Dynamic Tools to Measure Health Outcomes From the Patient Perspective

Log-In User ID: Password: | Forgot Password Register

Search

Advanced Search Page

What Is PROMIS? Overview

The Patient-Reported Outcomes Measurement Information System (PROMIS) initiative establishes a collaborative relationship between NIH and individual research teams through a cooperative agreement (U01) mechanism. The broad objectives of the PROMIS network are to:

- Develop and test a large bank of items measuring patient-reported outcomes
- Create a computerized adaptive testing system that allows for efficient, psychometrically robust assessment of patient-reported outcomes in clinical research involving a wide range of chronic diseases
- Create a publicly available system that can be added to and modified periodically and that allows clinical researchers to access a common repository of items and computerized adaptive tests.

The network will collaborate on the collection of self-reported data from diverse populations of individuals with a variety of chronic diseases, using agreed-upon methods, modes, and questionnaires.

What would you like to measure in your study?

- pain
- fatigue
- emotional distress
- physical function
- social role participation

Click below for:

- Domain definitions
- item content review
- item characteristics

or select a box on the right and proceed

Internet

Collaborative Agreement

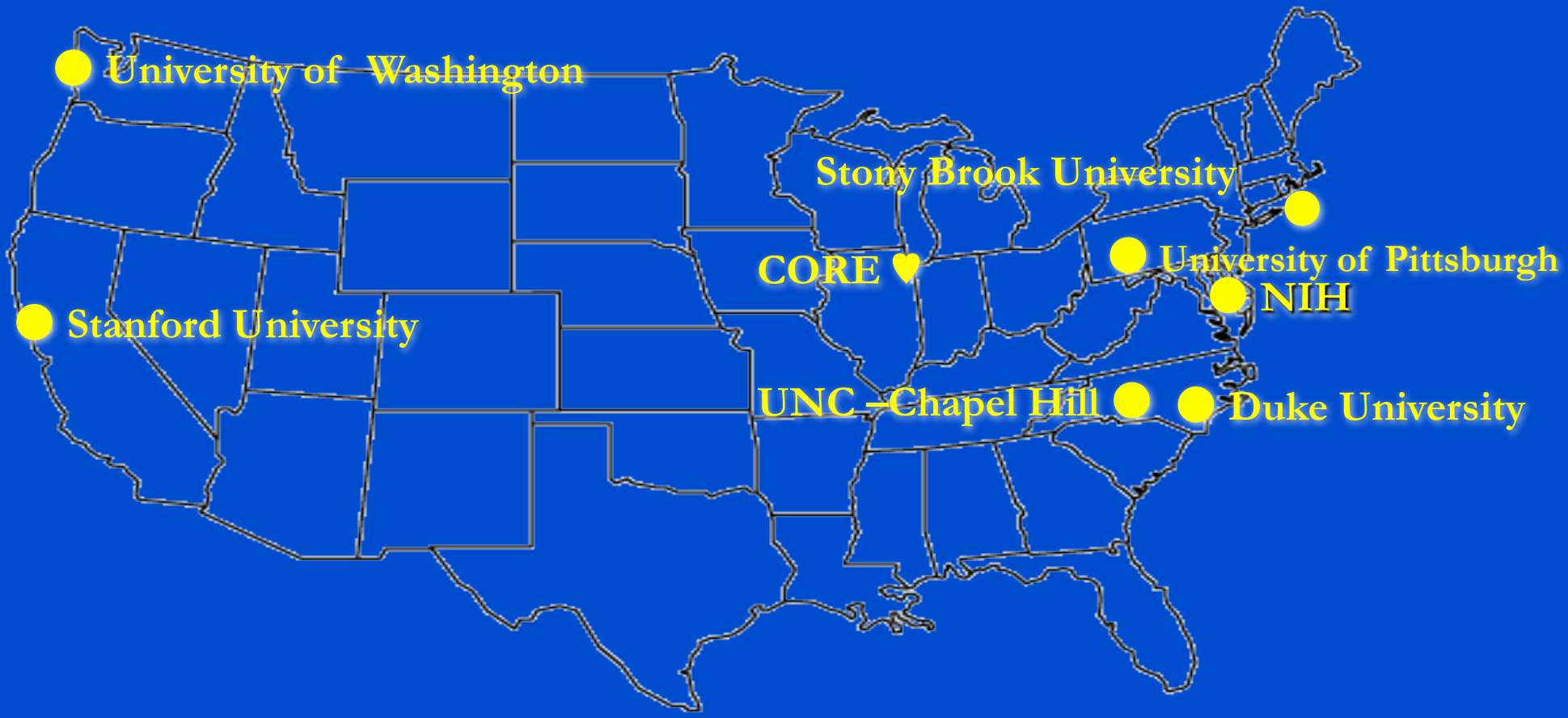
Steering Committee

6 Primary Research Sites

Statistical Coordinating Center

Scientific Advisory Board

PROMIS Network



Primary Research Sites

Duke (Kevin Weinfurt, evaluation committee, participation and data quality committee, use in clinical trials, cancer supplement)

Stanford (Jim Fries, physical function, domain hierarchy)

Stony Brook (Arthur Stone, fatigue, ecological momentary assessment)

UNC (Darren DeWalt, social/role participation, pediatrics, literacy)

University of Pittsburgh (Paul Pilkonis, emotional distress, sleep)

University of Washington (Dagmar Amtmann, pain, universal access)

Statistical Coordinating Center

Northwestern (David Cella and cast)

UCLA (Hays, Liu, Reise, Spritzer, Morales)

Other consultants (e.g., Dennis Revicki)

Westat

What **PROMIS** Promises

P recision
R epository
O utcome tools
M ethodologies
I nterpretability
S oftware

Precision

- Fewer items needed for equal precision
  - Making assessment briefer
- Error is understood at the individual level
  - Enabling individual assessment

Items from Instrument A

Items from Instrument B

Items from Instrument C

New Items

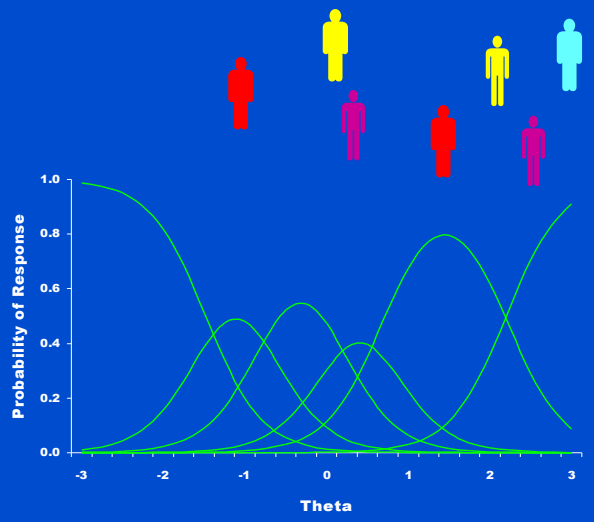
Item Pool

Content Expert Review

Focus Groups

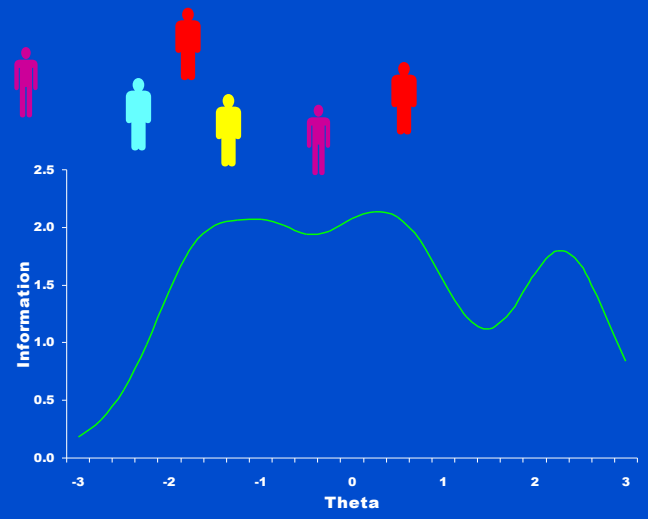
Cognitive Testing

Secondary Data Analysis

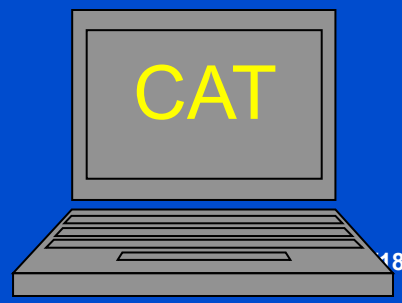


Questionnaire
administered to large representative sample

Item Response Theory (IRT)

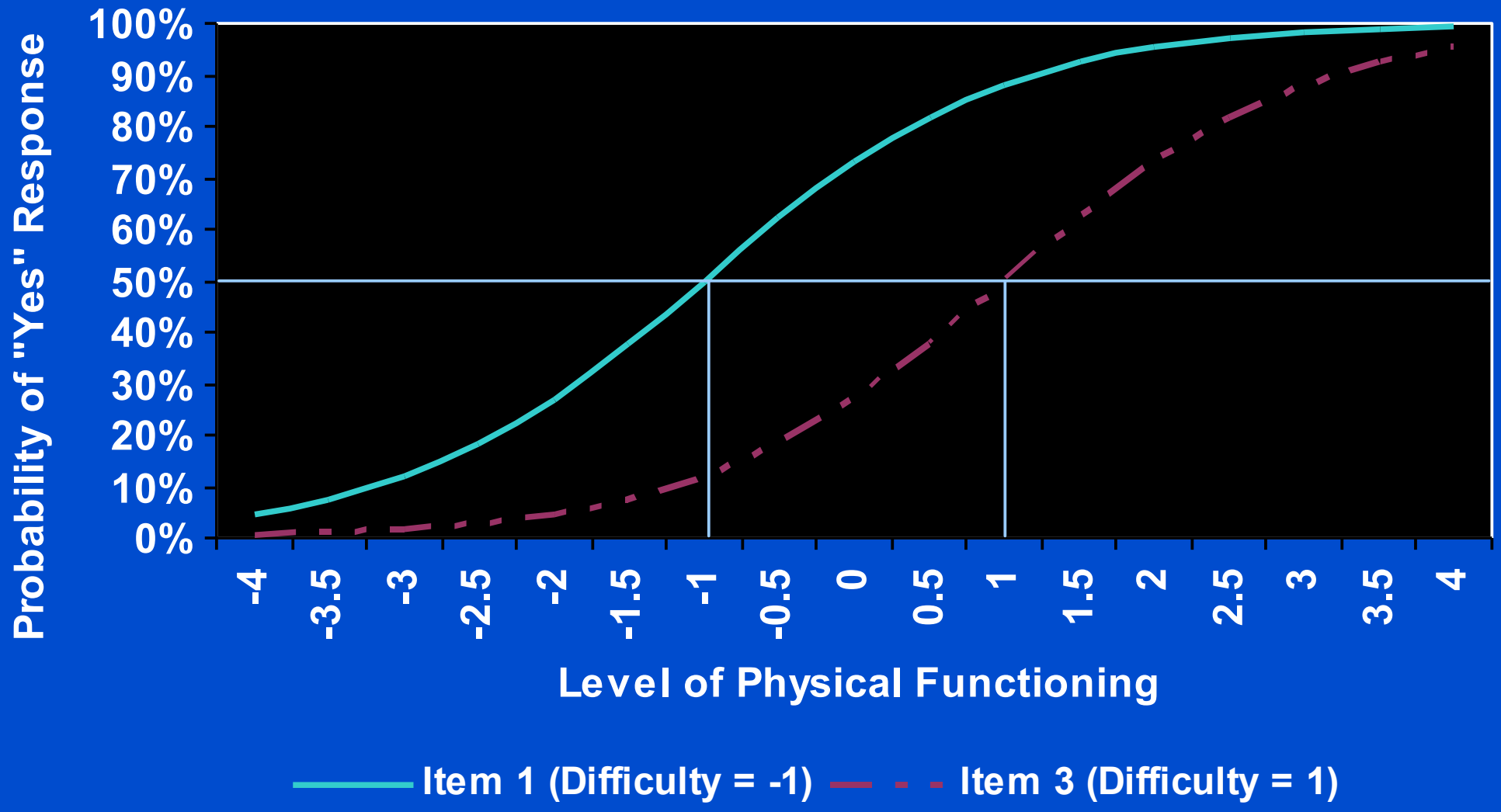


Item Bank
(IRT-calibrated items)

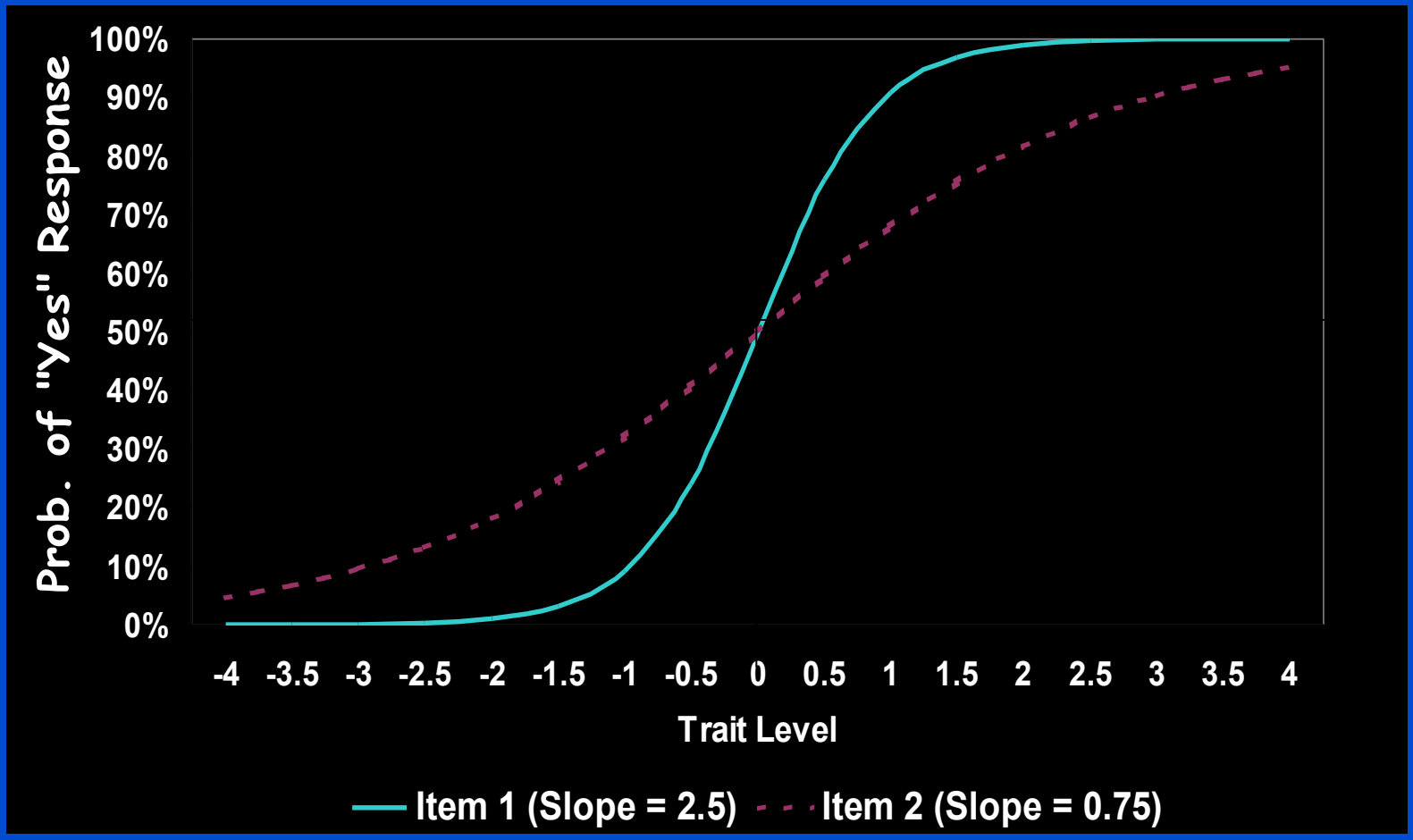


Short Form Instruments

Item Characteristic Curves (able to climb flight of stairs versus run a mile)

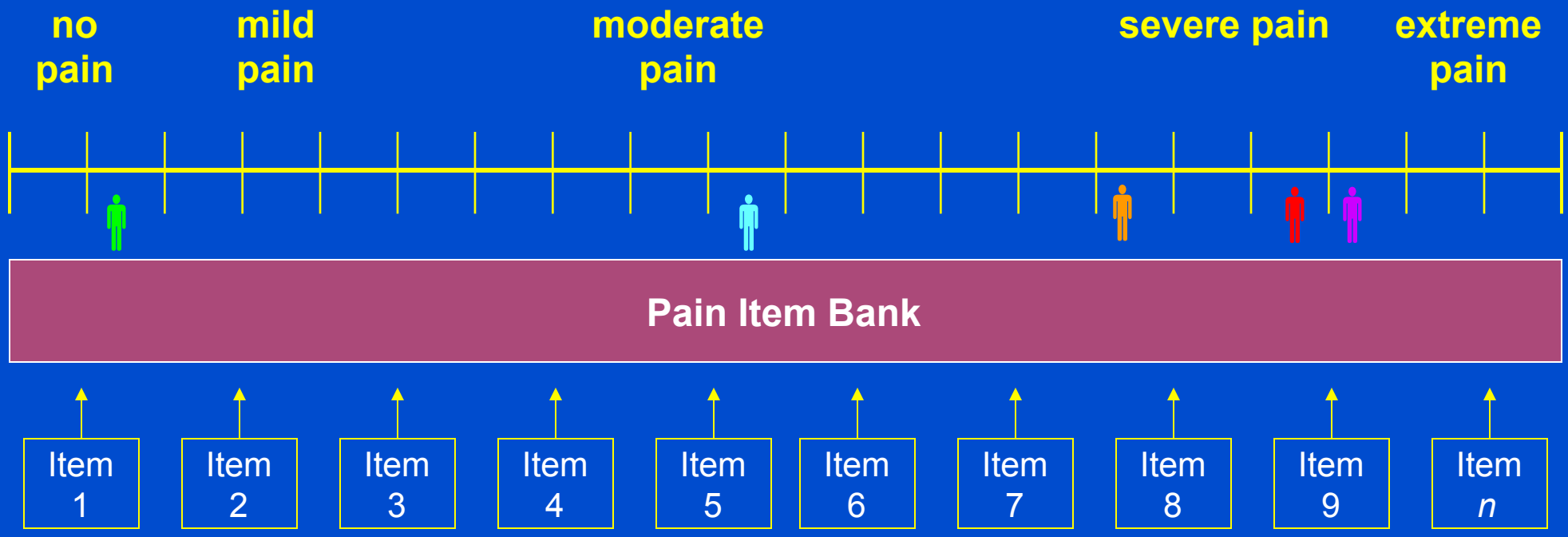


Item Characteristic Curves (2-Parameter Model)

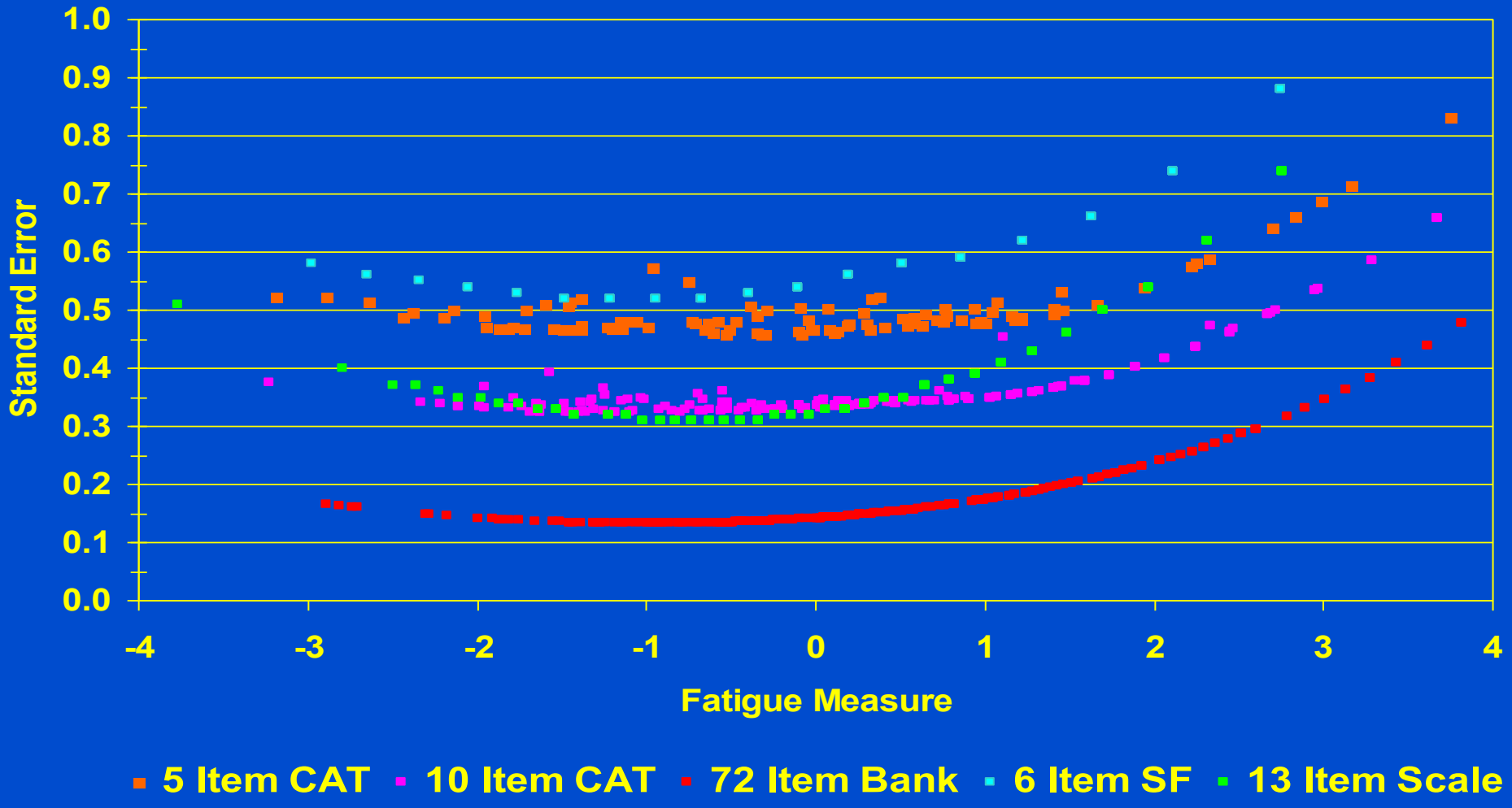


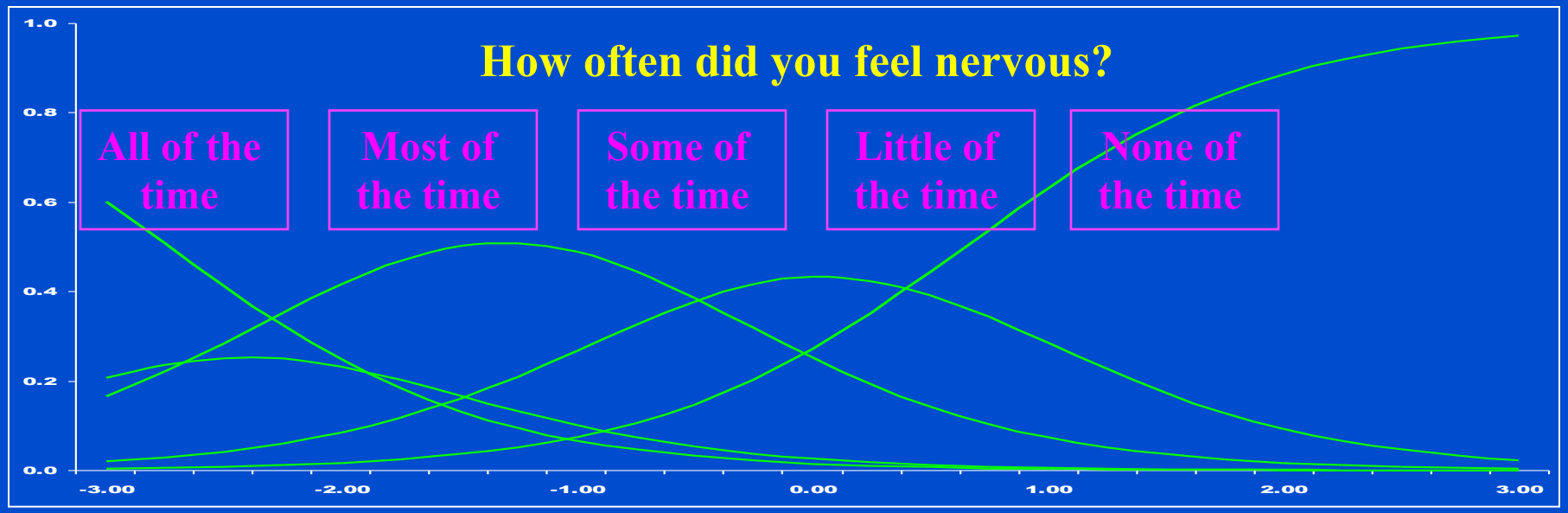
Item Banks

An item bank is comprised of a large collection of items measuring a single domain (e.g., pain).



Fatigue Measure and Standard Error Comparison by Test Length



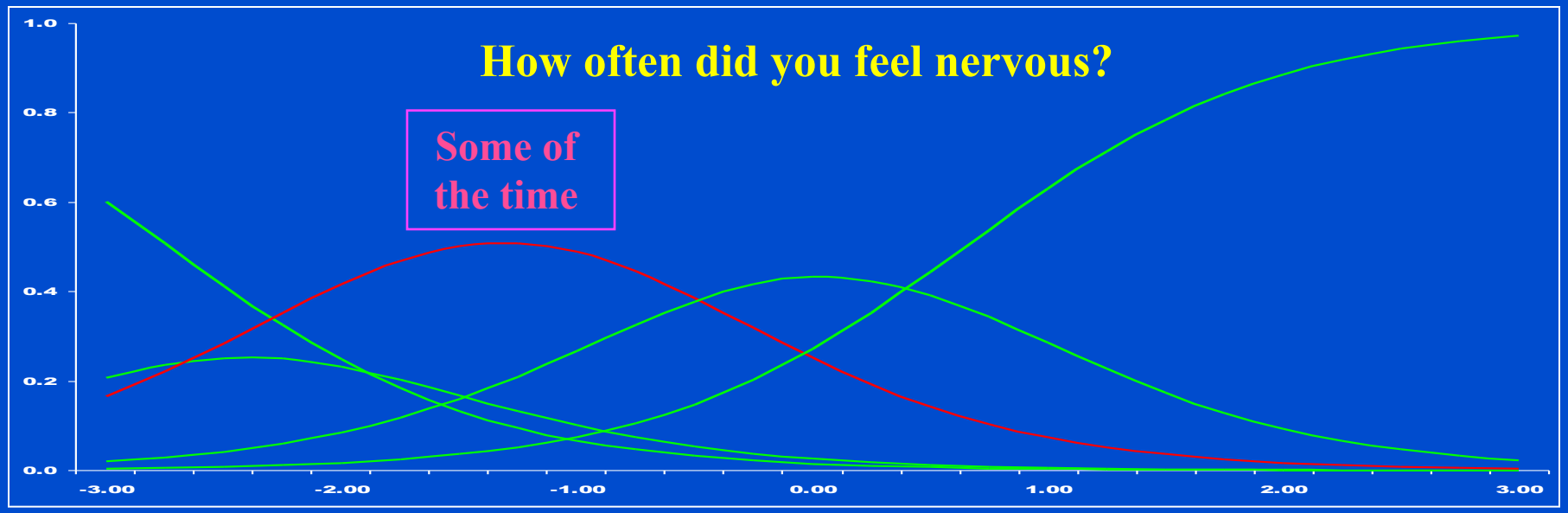


-3 -2 -1 0 1 2 3

Severe high moderate low very low

Emotional Distress

Item Bank
(IRT-calibrated emotional distress items)

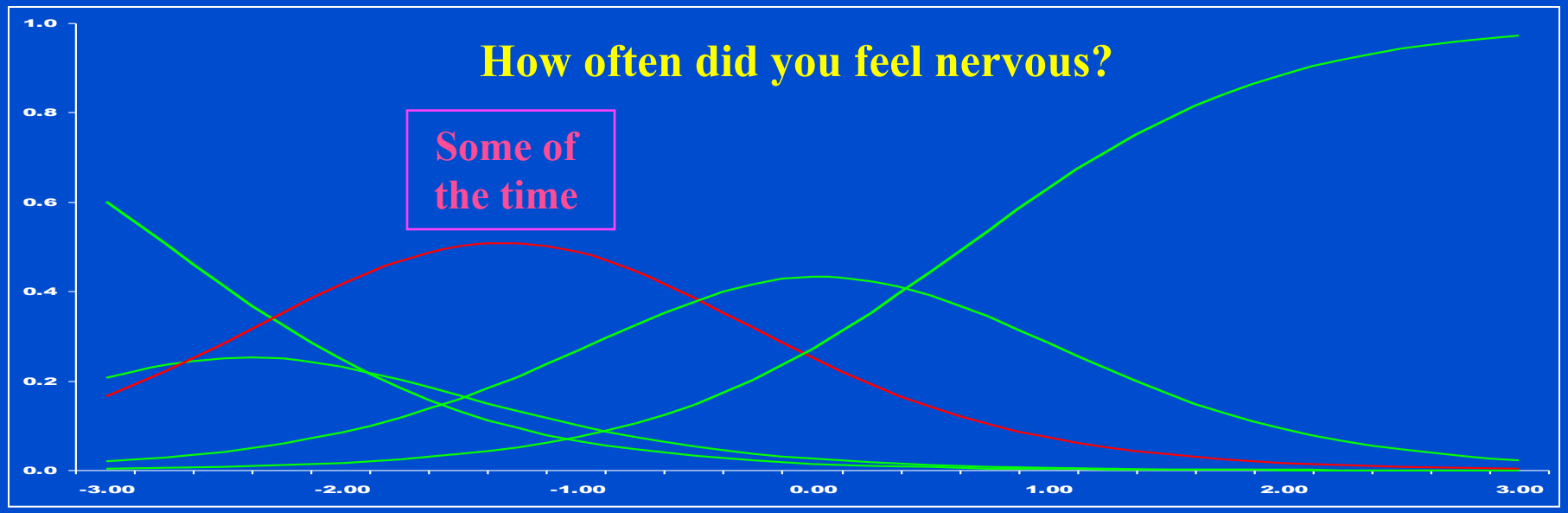


-3 -2 -1 0 1 2 3

Severe high moderate low very low

Emotional Distress

Item Bank
(IRT-calibrated emotional distress items)

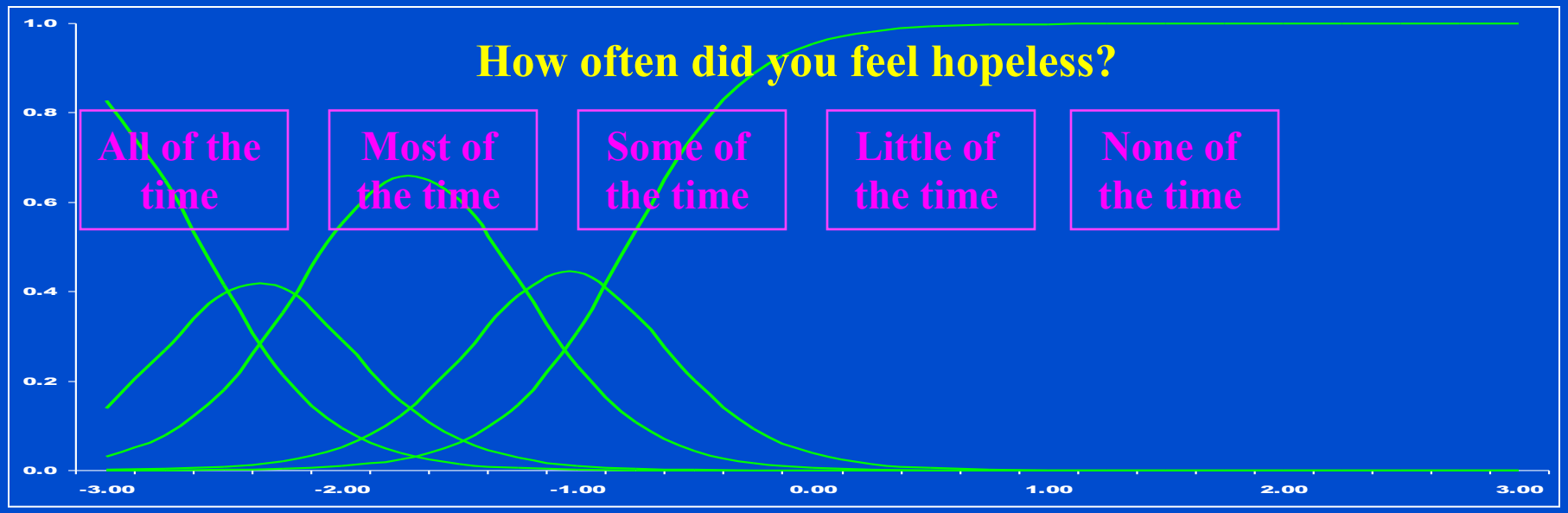


-3 -2 -1 0 1 2 3

Severe high moderate low very low

Emotional Distress

Item Bank
(IRT-calibrated emotional distress items)

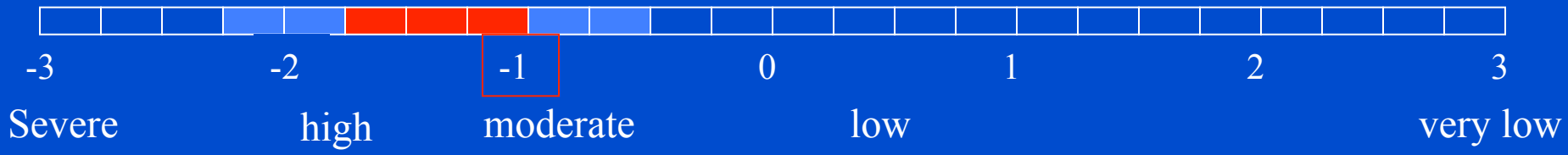


-3 -2 -1 0 1 2 3

Severe high moderate low very low

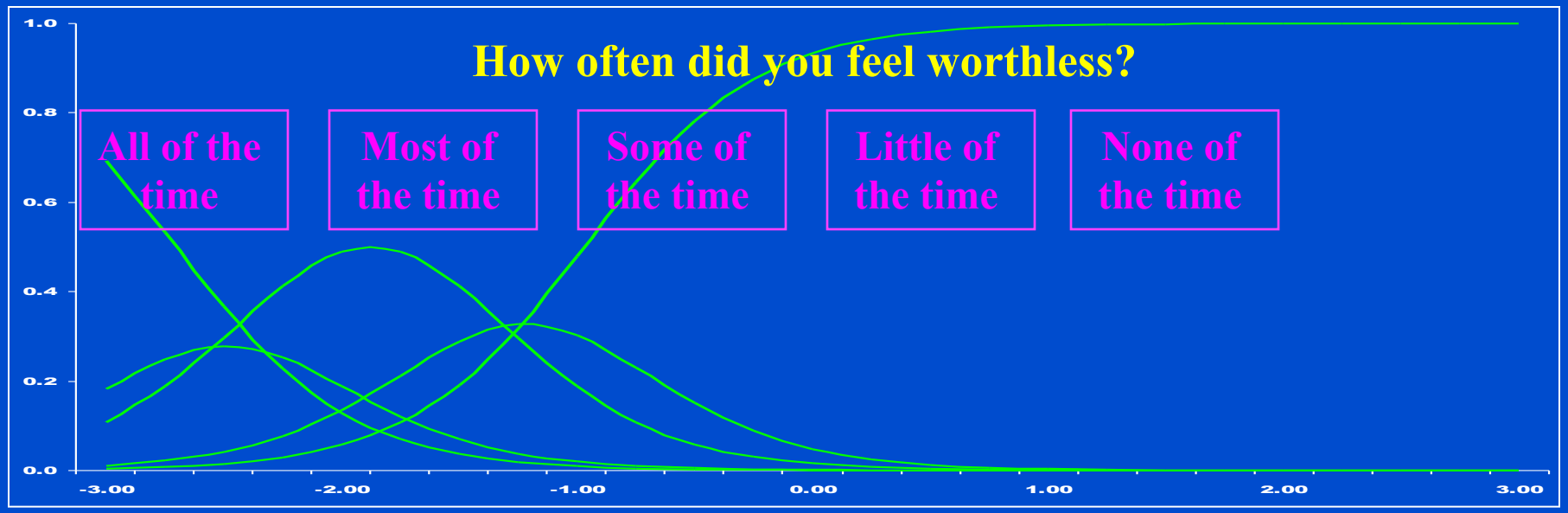
Emotional Distress

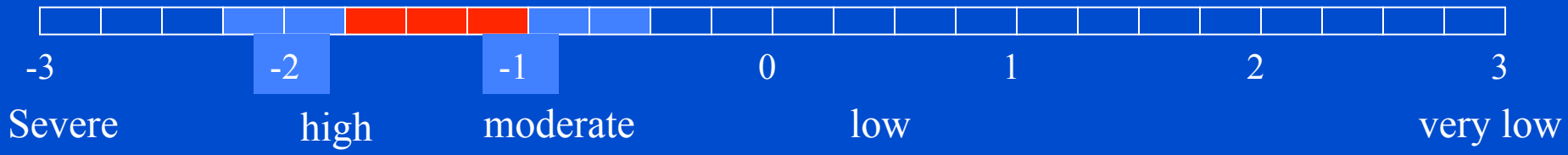
Item Bank
(IRT-calibrated emotional distress items)



Emotional Distress

Item Bank
(IRT-calibrated emotional distress items)





Emotional Distress

Item Bank
(IRT-calibrated emotional distress items)

-3 -2 -1 0 1 2 3

Severe high moderate low very low

Emotional Distress

Item Bank
(IRT-calibrated emotional distress items)

-3 -2 -1 0 1 2 3

Severe high moderate low very low

Target in on emotional distress score

Item Bank
(IRT-calibrated emotional distress items)

Repository

Repository: PROMIS Item Library

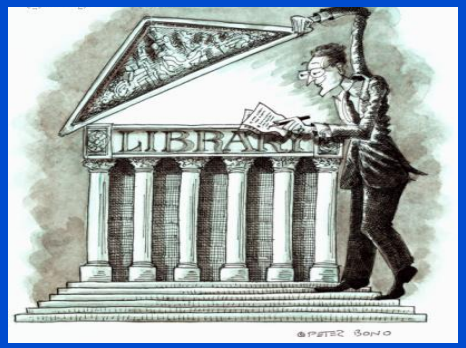
Literature searches and investigator contributions to the PROMIS domains

Relational database of more than 7,000 items

Catalog characteristics of items including

- Context
- Stem
- Response options
- Time frame
- Instrument of origin (if applicable)
  - Intellectual property status
  - Track modifications to items

Outcome Tools

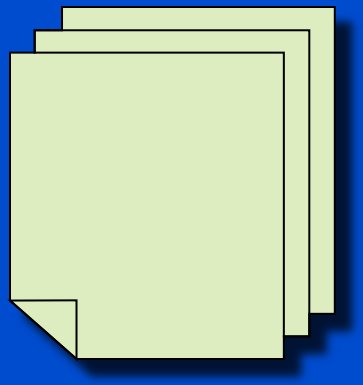


Item Library

CAT

Pick-a-PRO

Build-a-PRO



Methodologies

Methodologies

Qualitative Item Review

- Expert item review of 6,871 items
- 26 focus groups; over 120 patients interviewed; over 700 surveys

Analysis Plan

- Classical test theory and IRT analyses

Sampling Plan

- 11,500 people; 951 items; minimum n = 500/item

Interpretation

Interpretation

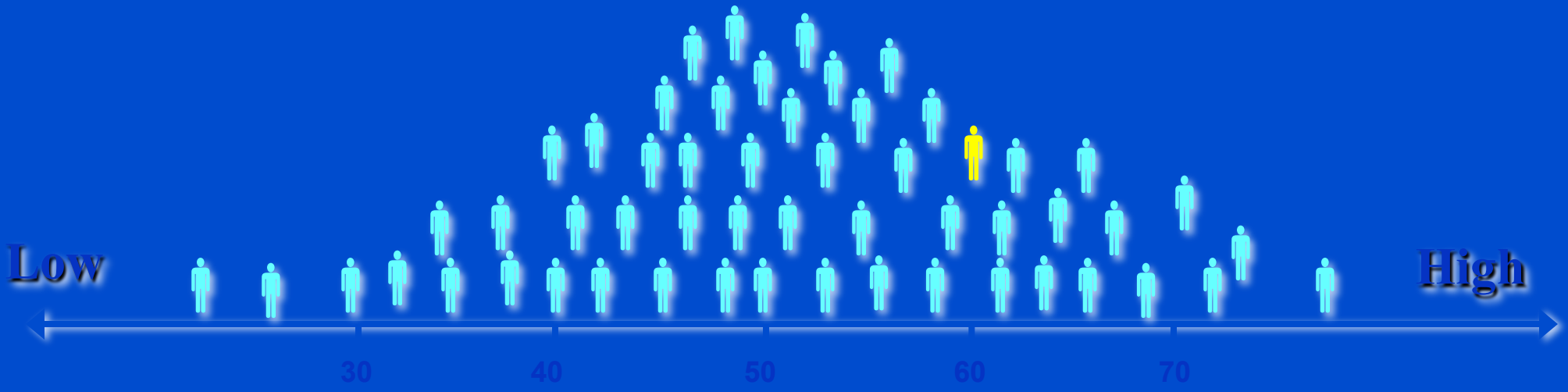
Interpretation Aids

M = 50, SD = 10

T = (z * 10) + 50

Example of high fatigue

Fatigue Score=60



This patient's fatigue score is **60**, significantly worse than average (50). People who score **60** on fatigue tend to answer questions as follows:

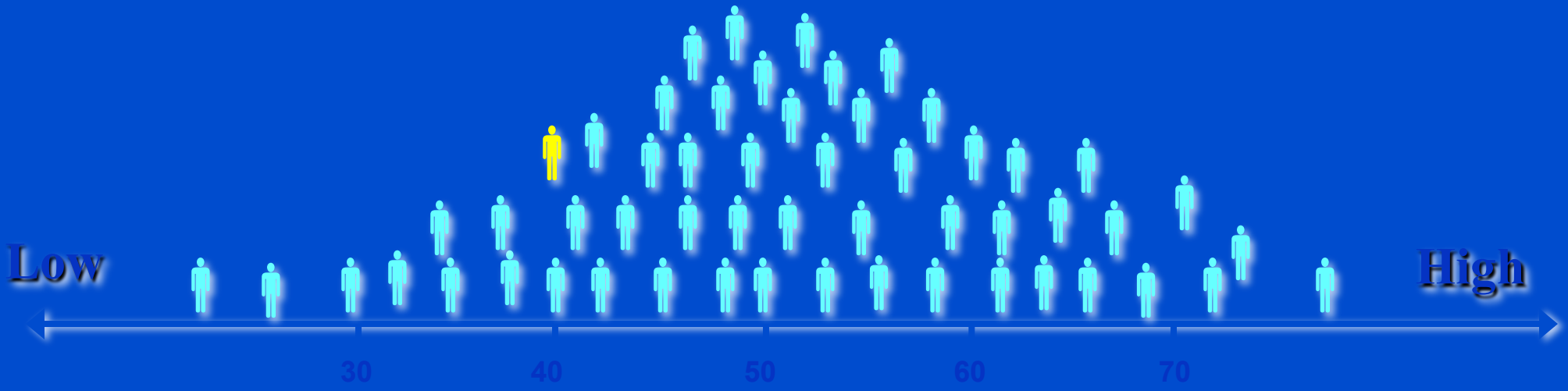
…"I have been too tired to climb one flight of stairs: VERY MUCH

…"I have had enough energy to go out with my family: A LITTLE BIT

Click here if you would like to see this patient's individual answers

Example of low fatigue

Fatigue Score=40



This patient's fatigue score is **40**, significantly better than average (50). People who score **40** on fatigue tend to answer questions as follows:

…"I have been too tired to climb one flight of stairs: SOMEWHAT
…"I have had enough energy to go out with my family: VERY MUCH

Click here if you would like to see this patient's individual answers

Software

Software

Item Library

Domains and item banks

Survey Construction

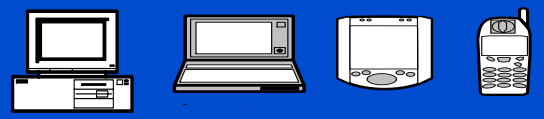
CAT
Pick-a-PRO
Build-a-PRO

Item Banks

Study Protocols

Clinical/PRO Data

Survey Module



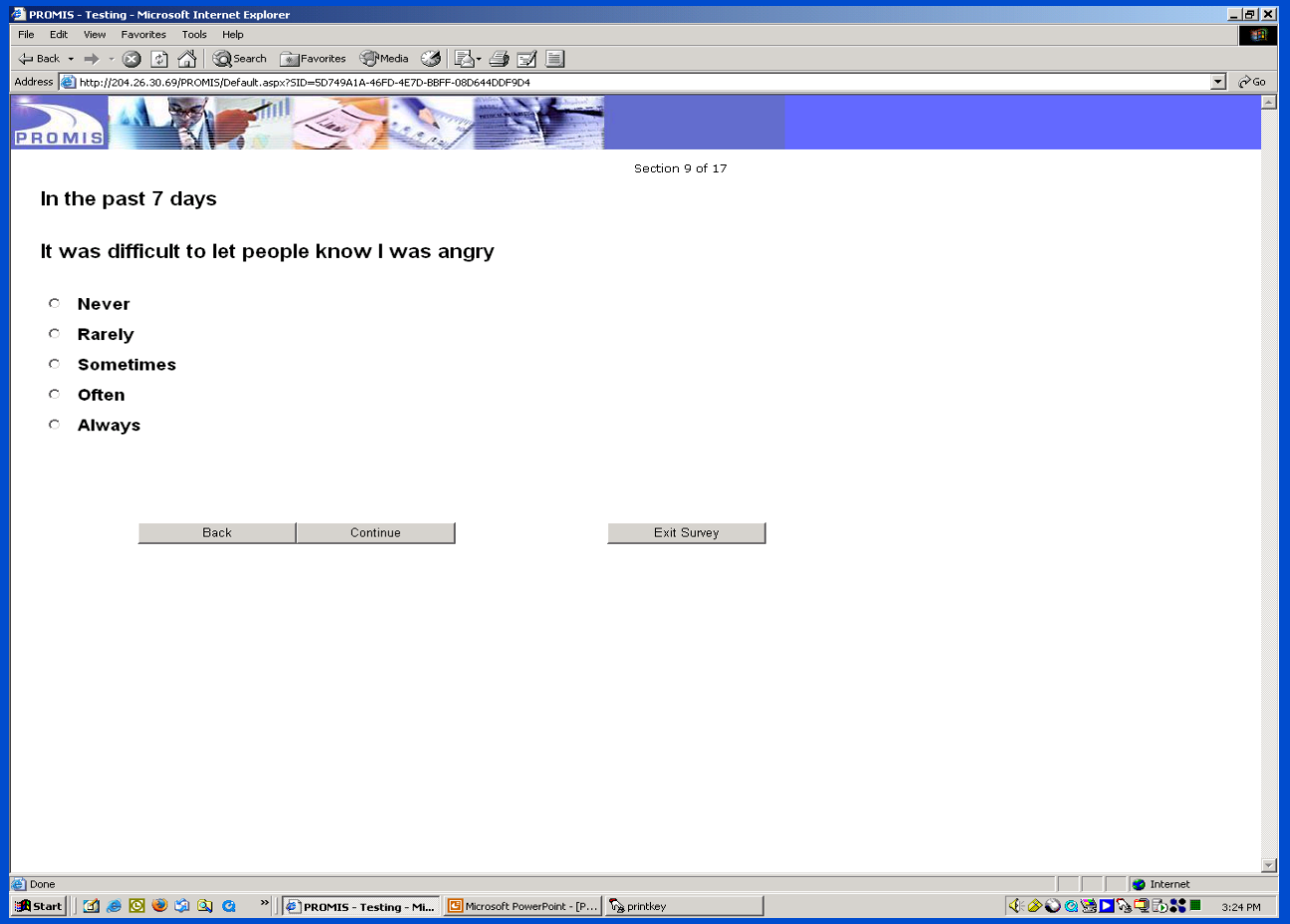
Web Laptop PDA IVR

Report Module

Patients Clinicians

Coming 2007-2008 to www.nihpromis.org

Web-based administration: Emotional distress (Anger) item



Four possible administration formats

- Automatic advance, not allowed to go back
- Automatic advance, allowed to go back
- Click to continue after response, not allowed to go back
- Click to continue, allowed to go back

Evaluated administration format

806 participants in Polimetrix PollingPlace registry

56 items measuring the performance of social/role activities

- Items rated on a 5-point frequency scale ranging from ”never” to “always”

56 items measuring satisfaction with social/role activities

- Items rated on a 5-point extent scale ranging from “not at all” to “very much”

Analysis Plan

Examined differences in:

- Time spent
- Mean domain scores
- Variance in scores
- Reliability

Administration format conclusions

Use automatic advance rather than continue button

Use back button to guard against accidental key entry

- Response time cost was minimal
- No effect on missing data or scores

PROMIS Sampling

Administer a large number of items to a range of population subgroups (general population and clinical) to permit the estimation of item parameters for item banks in five health-related quality of life domains.

Number of items

1013 items

- 784 items: 14 item banks (56 items per bank) in 5 domains
- 167 legacy items
- 62 demographic Items

146-202 items administered to any one respondent

n = 11,500 overall (500 observations per item minimum)

Target Polimetrix General Population Demographics

Gender	Male 50%
	Female 50%
Age	18-29 = 20%
	30-44 = 20%
	45-59 = 20%
	60-74 = 20%
	75+ = 20%
Ethnicity (match the general population)	Black = 12.3% Hispanic or Latino = 12.5%
Education	Min 25% High school graduate or less

Samples

	N
General population (Polimetrix)	7,500
Cancer (Duke, Polimetrix)	1000
Heart Disease (Duke, Polimetrix)	500
Rheumatoid arthritis (Stanford)	500
Osteoarthritis (Stanford, Polimetrix)	500
Psychiatric (Pittsburgh, Polimetrix)	500
Spinal Cord Injury (Polimetrix)	500
Chronic Obstructive Pulmonary Disease (Polimetrix)	500
TOTAL	11,500

Full Bank Administration

Domain	Sub-Domain	Items	Form	Sample	N
Emotional Distress	Anxiety	56	A	Gen. Pop. 1	500
	Depression	56			
	Anger/Aggression	56	B	Gen. Pop. 2	500
	Alcohol Abuse	56			
Physical Function	Part I	56	C	Gen. Pop. 3	500
	Part II	56			
	Part III	56	G	Gen. Pop. 7	500
Fatigue	Impact	56	D	Gen. Pop. 4	500
	Experience	56			
Social Role	Impact	56	E	Gen. Pop. 5	500
	Experience	56			
Pain	Interference	56	F	Gen. Pop. 6	500
	Quality	56			
	Behavior	56	G	Gen. Pop. 7	(above)

Block Administration

Item Number →		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		54	55	56
Sample Size ↓																				
General Pop. 8	250	H	H	H	H	H	H	H												
General Pop. 9	250				I	I	I	I	I	I	I									
General Pop. 10	250								J	J	J	J	J	J	J					
Heart Disease	250	H	H	H	H	H	H	H												
	250								J	J	J	J	J	J	J					
Cancer	250				I	I	I	I	I	I	I									
	250											K	K	K	K	K				

Between Banks

Block Administration Strategy

Item Number →										
Sample Size ↓		1	2	3	4	5	6	7	8	9
General Pop. 8	250	H	H	H	H	H	H	H		
General Pop. 9	250				I	I	I	I	I	I
General Pop. 10	250								J	J
Heart Disease	250	H	H	H	H	H	H	H		
	250								J	J
Cancer	250				I	I	I	I	I	I
	250									

Item 5 is administered in Block Forms H and I

to General Population Sample 8

and General Population Sample 9

as well as Heart Disease

and Cancer Samples

Types of Analyses

- Classical Test Theory Statistics
- IRT Model Assumptions
- Model Fit
- Differential Item Functioning
- Item Calibration

Classical Test Theory Statistics

- Out of range
- Item frequencies and distributions
- Inter-item correlations
- Item-scale correlations
- Internal consistency reliability

IRT Model Assumptions

- (Uni)dimensionality
- Local independence
- Monotonicity

Sufficient Unidimensionality

- Confirmatory factor models
  - One factor
  - Bifactor (general and group factors)

Local Independence

- After controlling for dominant factor(s), item pairs should not be associated.
  - Look at residual correlations (> 0.20)

Monotonicity

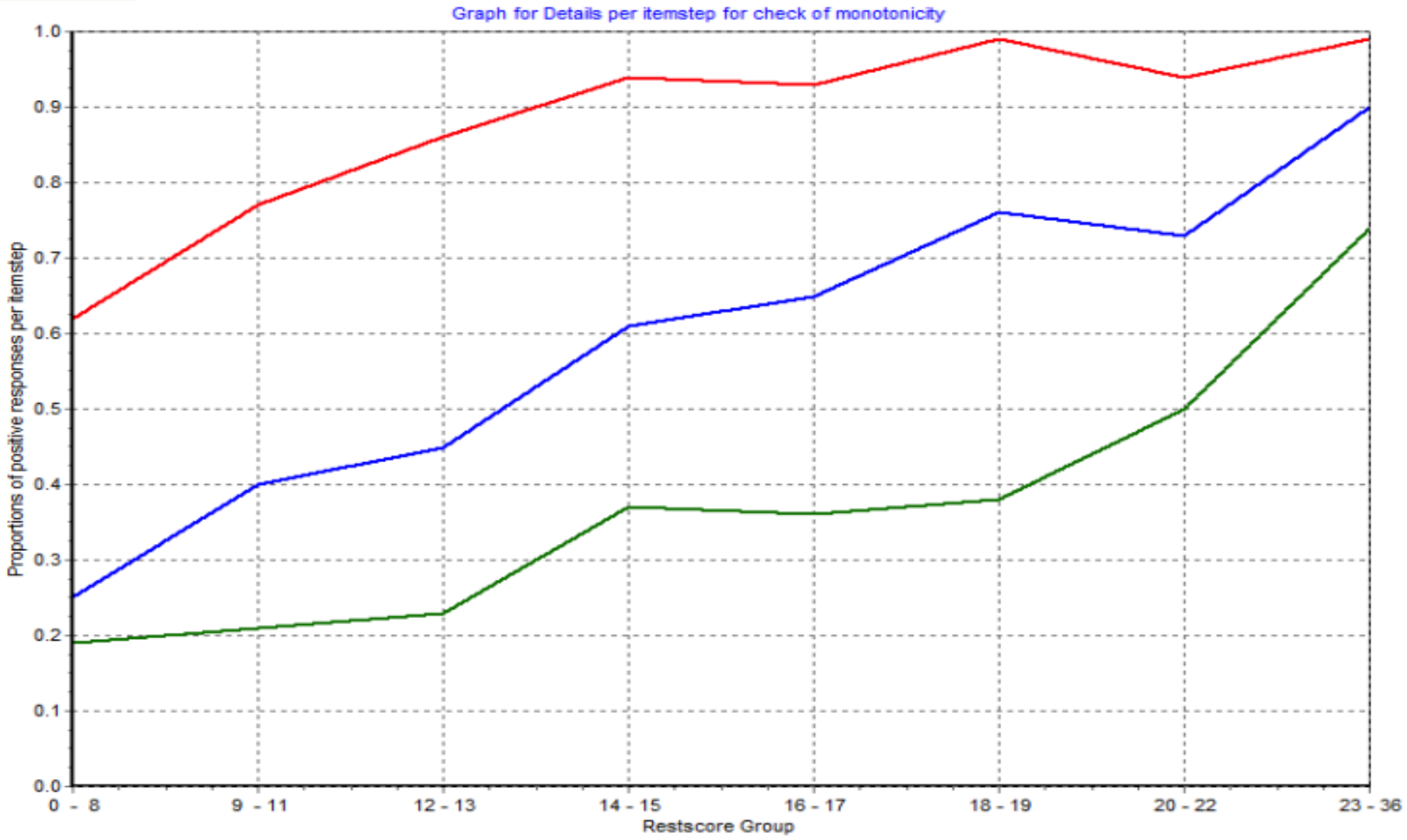
- Probability of selecting a response category indicative of better health should increase as underlying health increases.
- Item response function graphs with
  - y-axis: proportion positive for item step
  - x-axis: raw scale score minus item score

Step 5: View Results from TEST.out

Output items

Graph for Details per itemstep for check of monotonicity

Item1



Notes...

Full Report...

Save this part...

Print this part...

Back to STEP 2

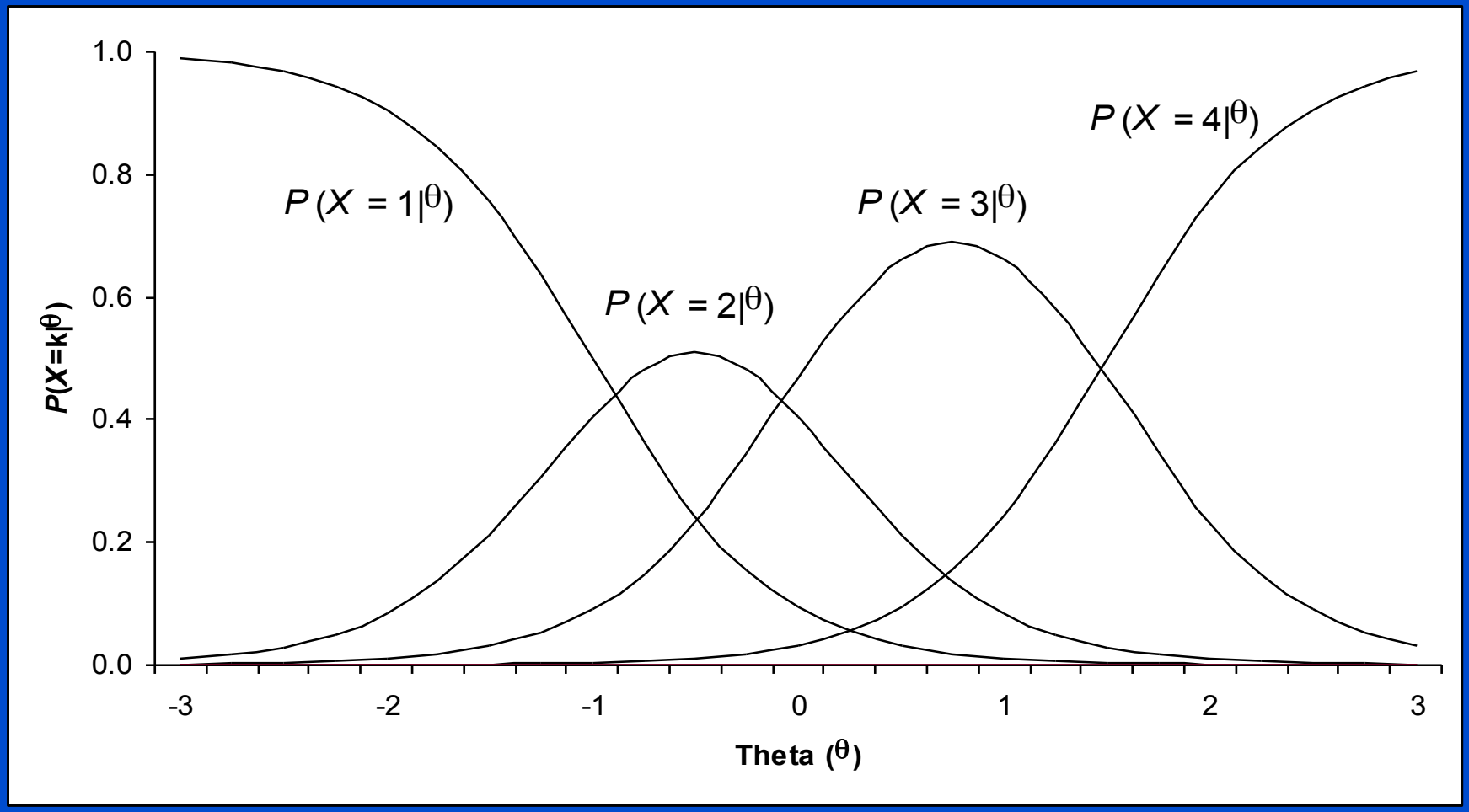
Back to STEP 4

Done

Help

start | Inbox - Outlook Express | Yahoo! Mail - bruin45... | mspguide.doc - Micro... | mspwin5 | 12:05 PM

Category Response Curves for Samejima's Graded Response Model



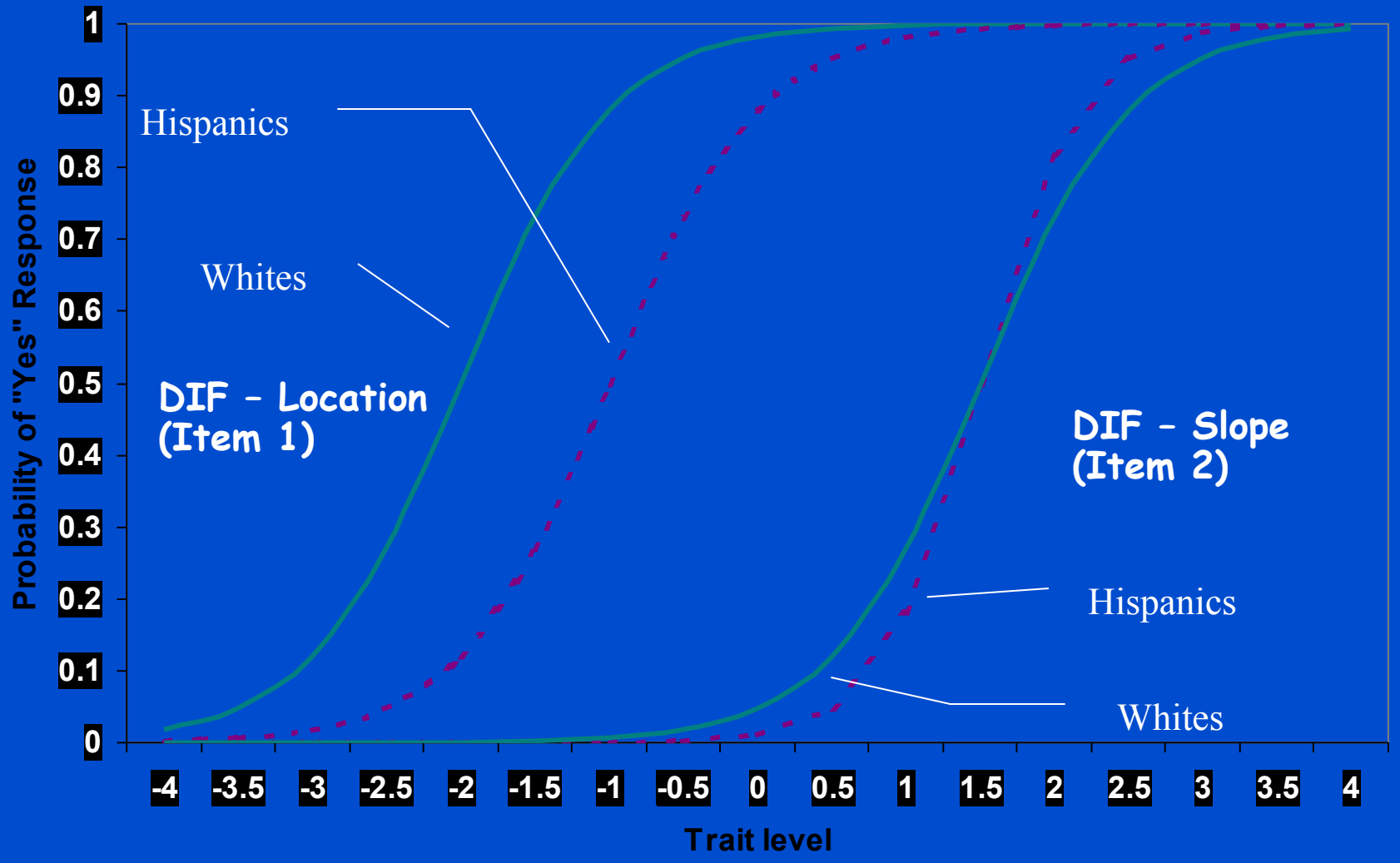
Model Fit

- Compare observed and expected response frequencies by item and response category
- Items that do not fit and less discriminating items identified and reviewed by content experts

Differential Item Functioning

- Uniform DIF
  - Threshold parameter
- Non-uniform DIF
  - Discrimination parameter
- Gender, race/ethnicity, age, education, disease

Dichotomous Items Showing DIF
(2-Parameter Model)



Item Calibration

- Item parameters (threshold, discrimination)
- Mean differences for studied disease groups

Questions?

Public website: http://www.nihpromis.org/

Peer-reviewed manuscripts, e.g.:

Hays, R. D. et al. (in press). Item response theory analyses of physical functioning items in the Medical Outcomes Study. Medical Care.

Reeve, B. B., et al. (submitted). Psychometric evaluation and calibration of health-related quality of life items banks: Plans for the Patient-Reported Outcome Measurement Information System (PROMIS)

Acknowledgements

“Slides” in this presentation were lifted or adapted from PROMIS presentations by David Cella, Richard Gershon, Bryce Reeve, and Jim Fries.

Appendices

Pre-Application Meeting for the RFA-RM-04-011: Dynamic Assessment of Patient-Reported Chronic Disease Outcomes
Monday, January 26, 2004
Deborah N. Ader, Ph.D. and Lawrence J. Fine, M.D., Dr.PH
Total Running Time: 02:40:08

http://videocast.nih.gov/PastEvents.asp?c=4&s=151