The Future of Outcomes Measurement: Item Banking, Tailored Short-Forms, and Computerized-Adaptive Assessment

David Cella

Richard Gershon

Jin Shei Lai

Michael Bass

Center on Outcomes, Research & Education (CORE)

Evanston Northwestern Healthcare and Northwestern University, Evanston, IL

Current Status: Fixed Instruments

- One instrument fits all?
- 30-50 questions
- All diseases use the same instrument (generic), or Each disease uses a different instrument (targeted)
- Coarse measure of all concepts

……Can item banks help?

Why Common Instruments (Item Banks)?

- Results comparable across studies
- Multiple brief-yet-accurate assessment options
- Common instruments may promote common language

…..But:

- Will common instruments stifle new approaches?

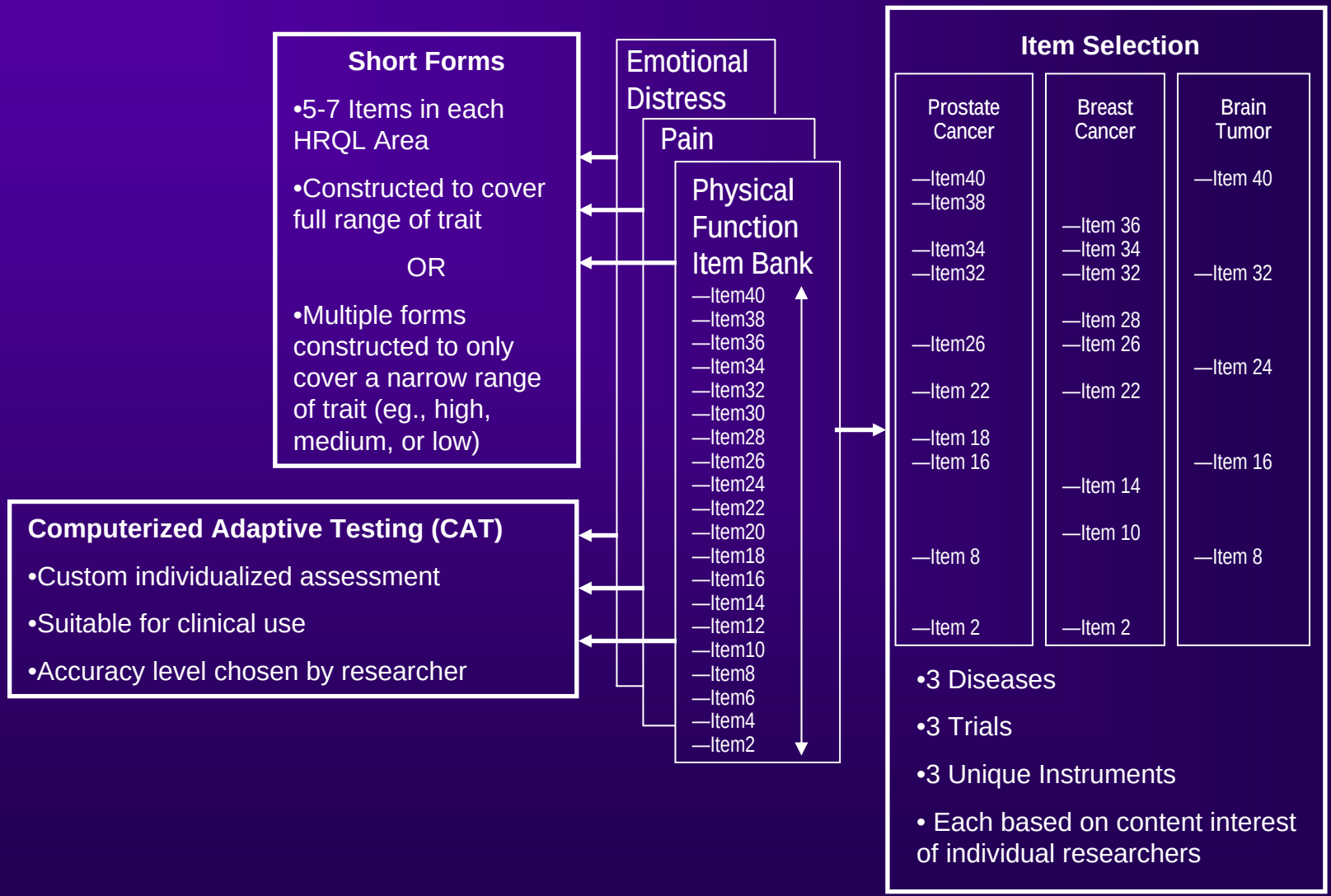
Outline

1. What Are Item Banks
2. Creating an Item Bank
3. Test Administration Options
4. Demonstrations of Working CAT
5. Examples From Others
6. Potential Cooperative Arrangements for the Future

1. What Are Item Banks

- Beyond “item pools”
- Items calibrated on a defined construct
- Require (and may promote) common language
- Enable customized short forms
- Enable Computerized Adaptive Testing (CAT)

Uses for Item Banks



Gershon et al, Exp Rev Pharmoecon Outcomes Res. (2003)

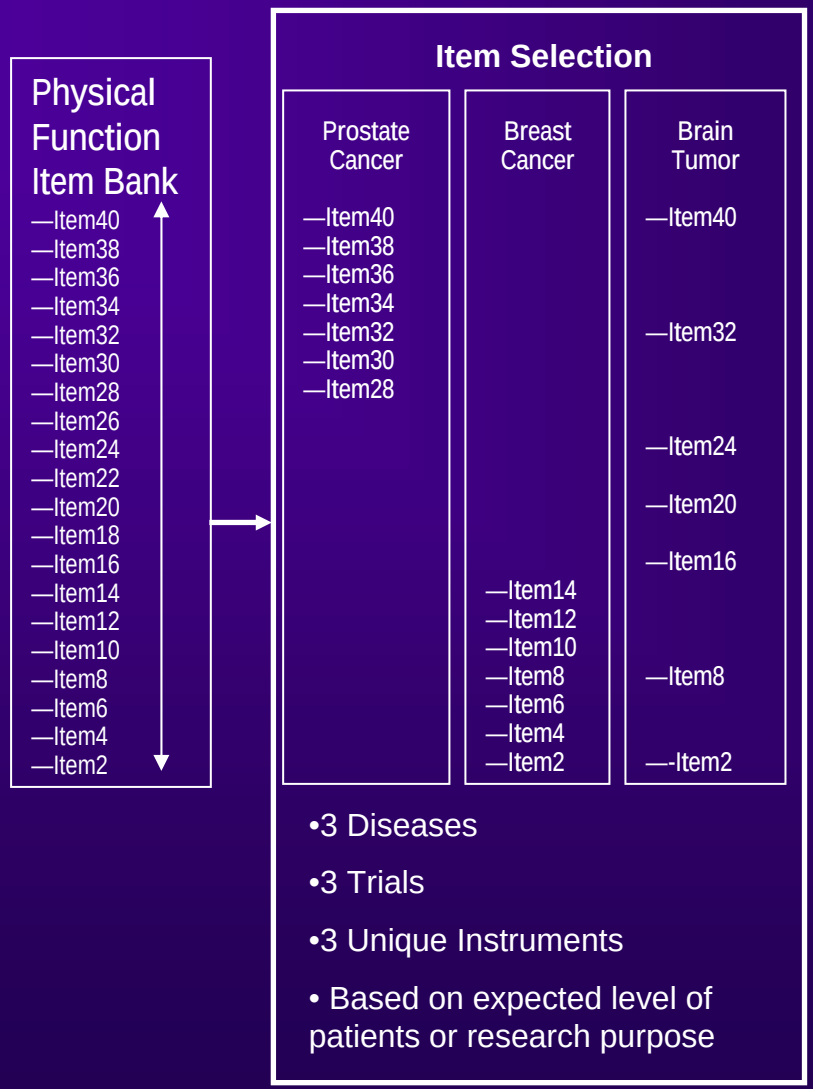
Short Form

- 5-7 Items in each HRQL Area
- Constructed to cover full range of trait

OR

- Multiple forms constructed to only cover a narrow range of trait (e.g., high, medium, or low)

Custom Instruments (multiple short forms)



EVANSTON NORTHWESTERN HEALTHCARE

Computerized Adaptive Testing (CAT)

- Custom individualized assessment
- Suitable for clinical use
- Accuracy level chosen by researcher

In Summary, Calibrated Item Banks Can be Used to:

- Create a standard static instrument
- Construct short forms
- Enable CAT
- Select items based on unique content interests and formulate custom short-form or full-length instruments

1. What Are Item Banks
2. Creating an Item Bank
3. Test Administration Options
4. Demonstrations of Working CAT
5. Examples from Others
6. Potential Cooperative Arrangements for the Future

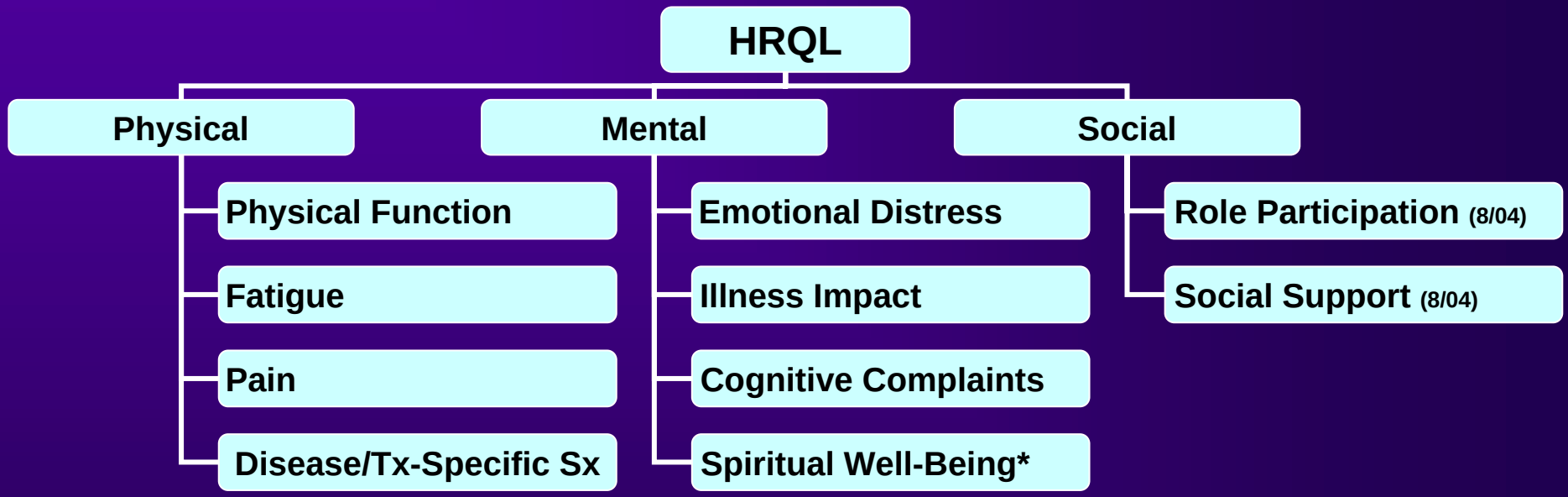
2. Creating an Item Bank

- Identify appropriate domains
- Assess pre-existing items and data sources (if any)
- Acquire, write, field test and analyze items
- Establish an operational Item Bank
- Implement CAT and short forms

Identify Appropriate Domains

- Literature Review
- Clinical input
  - Clinicians
  - Patients

WHO Based Model Guides Item Bank Development



*in progress

Assess Pre-existing Items and Available Data Sources

- Identify common items and rating scales
- Data analysis
  - Examine dimensionality
  - Examine item fit
  - Calibrate items on the continuum
- Examine construct deficiency
  - Statistical deficiency (gaps)
  - Clinical deficiency (gaps)

Acquire, Write, Field Test and Analyze Items

- Acquire and write new items (content validity)
- Field testing
  - CBT programming
  - Data collection
- Data analysis
  - Examine dimensionality
  - Examine Item Fit
  - Calibrate items on the continuum
- Evaluate item parameter equivalence across sub-groups

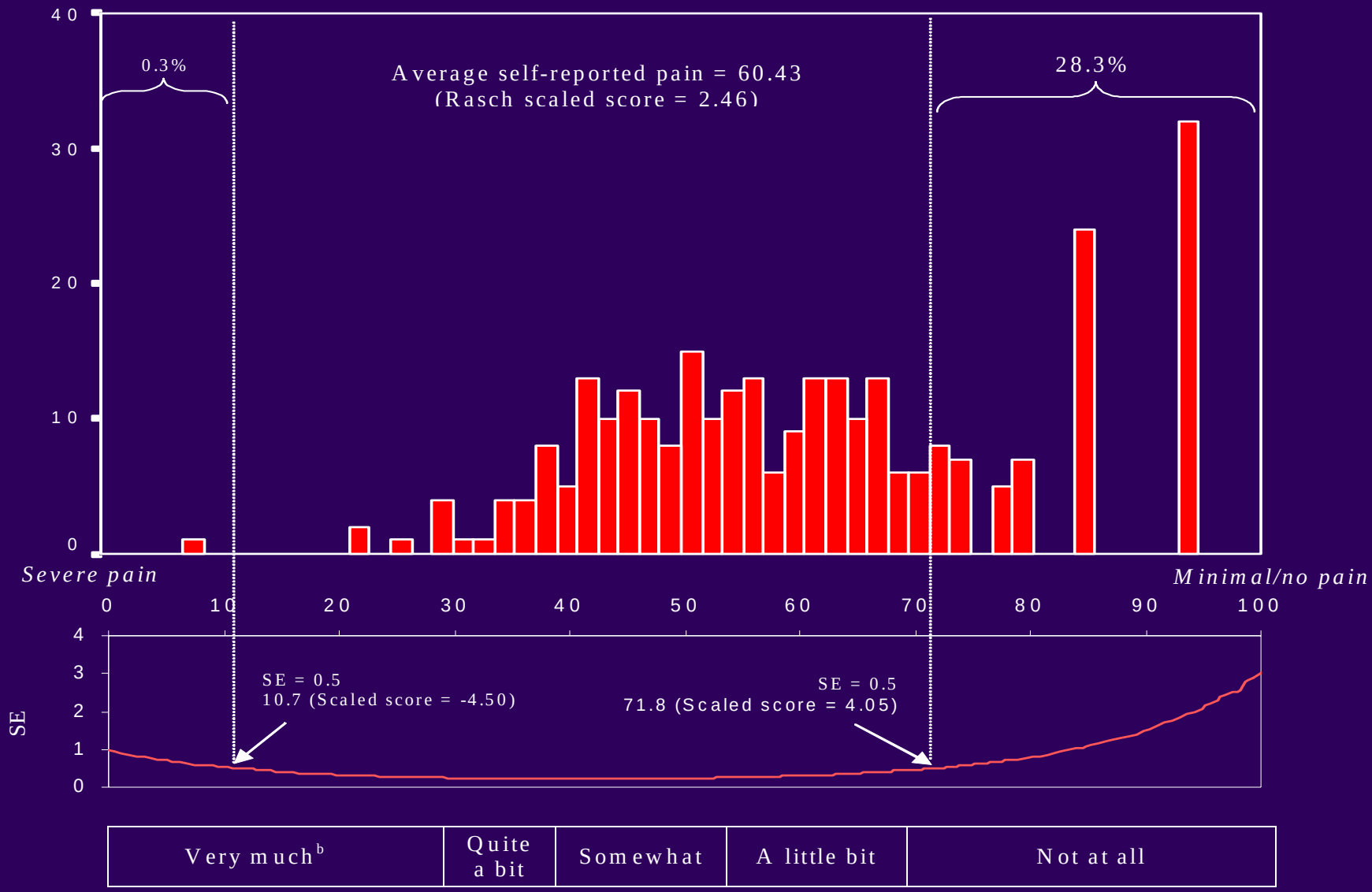
Establish an Operational Item Bank

- Psychometric results
- Clinical input

Implement CAT and Short Forms

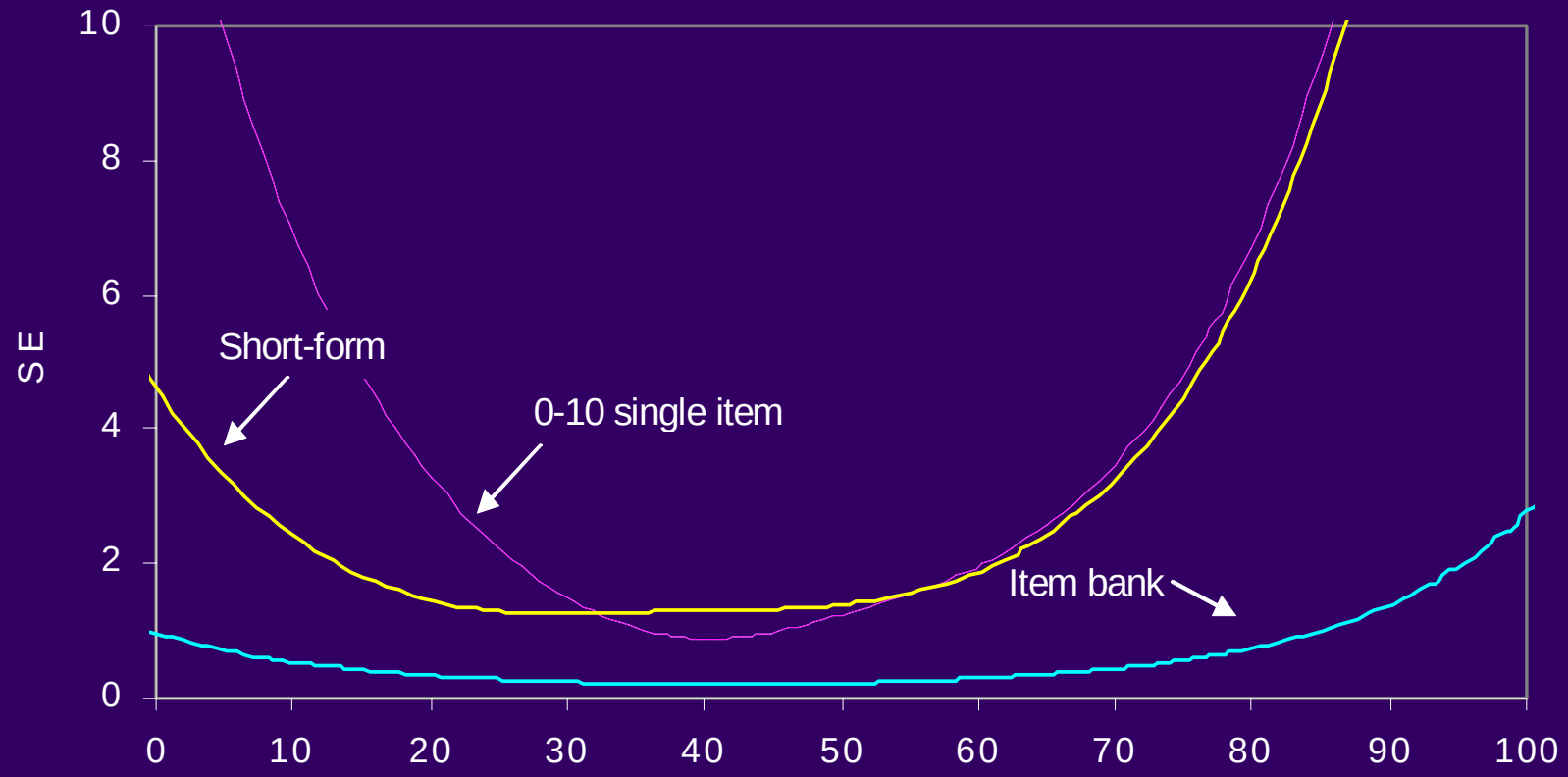
- Implement CAT
  - Establish parameters
  - Simulate across the continuum
  - Pilot test in clinical practice
- Create short forms
  - Test in target population(s)

Bank Precision Level along the Pain Scores Distribution



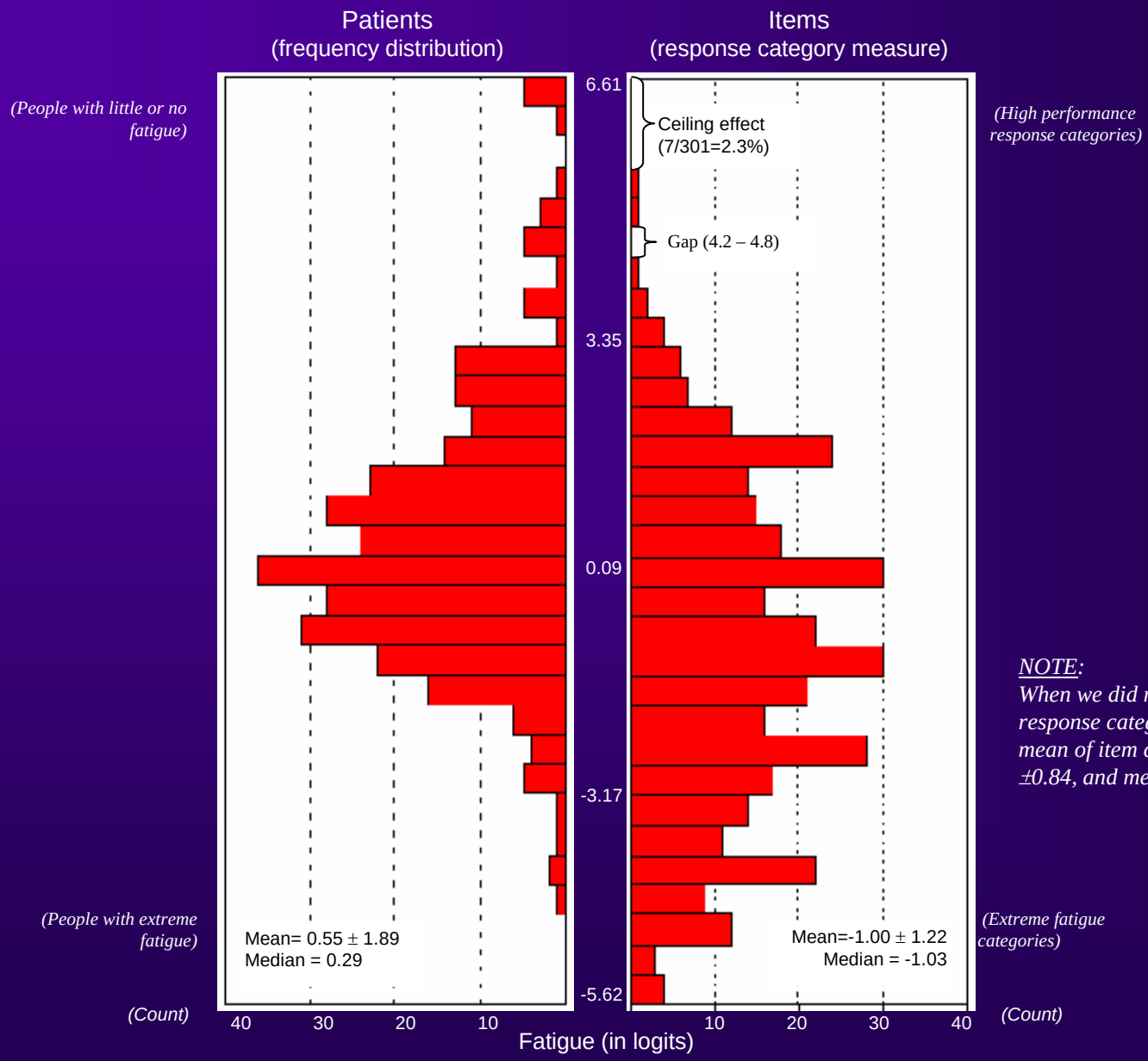
EVANSTON NORTHWESTERN HEALTHCARE

Comparing Precision of Item Bank, Short-form and 0-10 Numeric Rating Scale



EVANSTON NORTHWESTERN HEALTHCARE

Proficiency* of the Fatigue Item Bank



NOTE:
When we did not take individual response category into account, mean of item calibrations was –0.38 ±0.84, and median = -0.41

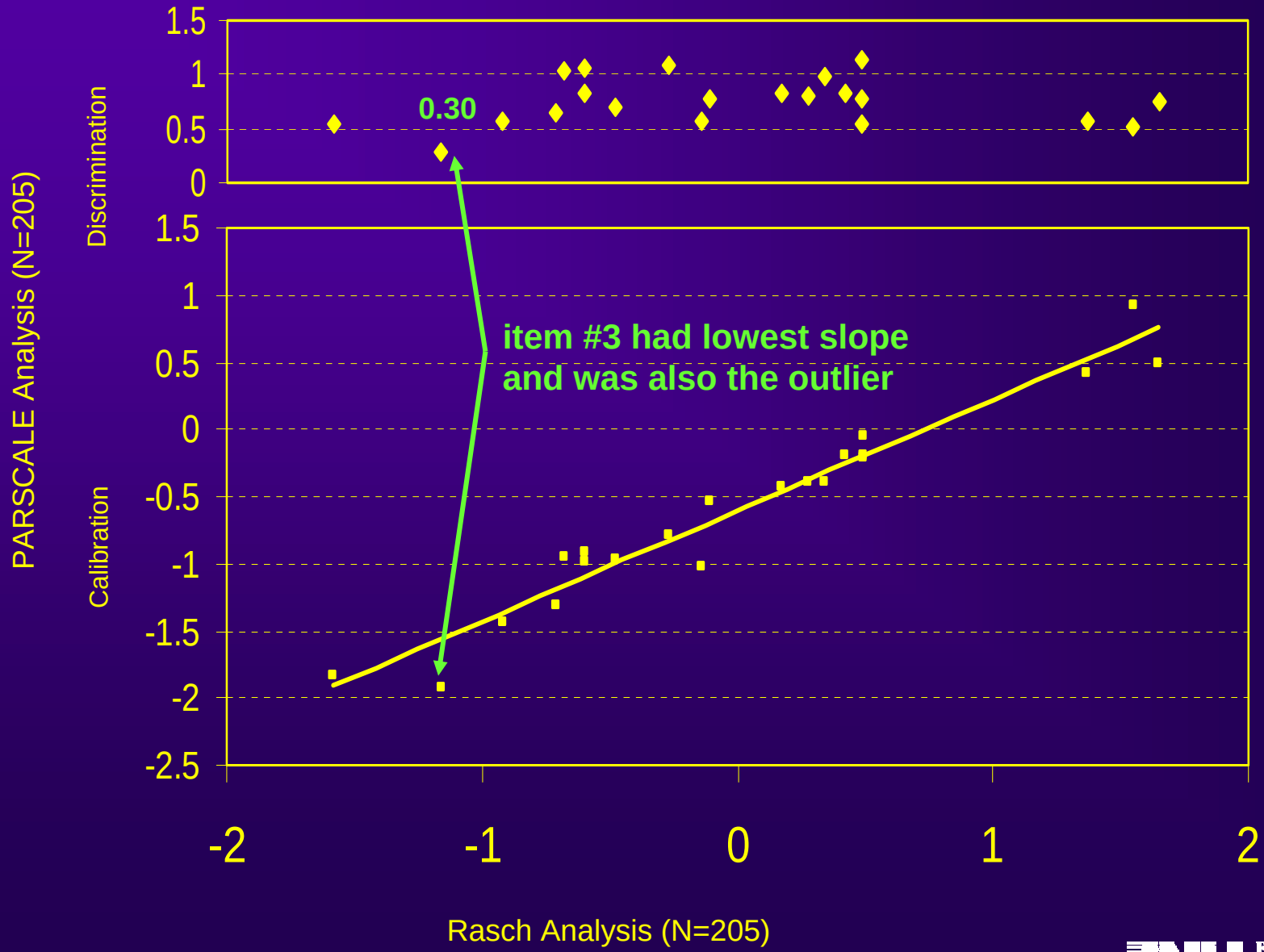
** Proficiency = the ability of items to capture the full range of the patients*

EVANSTON NORTHWESTERN HEALTHCARE

Typical Analysis Preparing a Bank

- Dimensionality
  - Factor Analysis & Item-total correlation
    - Positive and Negative Illness Impact are separate factors
  - IRT analysis (“positive” II bank)
    - 21 items were retained via results of 1) item-total correlation and repeated Rasch analyses.
    - Items are also analyzed using PARSCALE.

Rasch (WINSTEPS) versus 2-PL (PARSCALE)



EVANSTON NORTHWESTERN HEALTHCARE

Rasch Location	Rasch MnSq	Rasch ptbis	2-PL slope	2-PL location	
1.65	1.02	0.73	0.75	0.49	Because of my illness, I developed new interests
1.55	1.24	0.67	0.51	0.92	Because of my illness, I am pursuing new interests
1.37	0.99	0.74	0.56	0.43	Because of my illness, I established a new path for my life
0.49	1.21	0.61	0.55	-0.05	Because of my illness, I learned to ask others for help
0.49	0.77	0.75	1.14	-0.2	Because of my illness, I am more open to new ideas
0.49	0.88	0.71	0.78	-0.21	Because of my illness, I learned not to let hassles bother me the way they used to
0.42	0.97	0.74	0.83	-0.19	Because of my illness, I am more comfortable with who I am
0.34	0.92	0.77	0.99	-0.39	Because of my illness, my life is more meaningful
0.28	0.83	0.73	0.79	-0.38	Because of my illness, I learned to deal better with uncertainty
0.17	0.92	0.7	0.83	-0.44	Because of my illness, I am more willing to express my emotions
-0.11	1.03	0.66	0.78	-0.54	Because of my illness, I take fewer things for granted
-0.14	1.08	0.62	0.58	-1.02	Because of my illness, I am more able to accept the way things work out
-0.27	0.99	0.71	1.09	-0.78	Because of my illness, I discovered that I'm stronger than I thought I was
-0.48	1.05	0.67	0.7	-0.97	Because of my illness, relationships have become more meaningful
-0.6	0.84	0.67	1.05	-0.91	Because of my illness, I have more compassion for others
-0.6	0.84	0.73	0.82	-0.99	My illness has given me a greater appreciation for life
-0.68	0.86	0.68	1.04	-0.96	Because of my illness, I am able to appreciate each day more fully
-0.71	1.14	0.63	0.65	-1.31	My illness has helped me see what is really important in life
-0.92	1.27	0.56	0.56	-1.44	Because of my illness, I have learned to appreciate my physical health
-1.16	**1.36**	**0.47**	**0.3**	**-1.92**	**Because of my illness, I know I can handle difficult times**
-1.58	1.19	0.53	0.54	-1.83	Because of my illness, I know who I can count on in times of trouble

Tend to endorse 'not at all'

Tend to endorse 'Very Much'

$$\text{Corr}_{(\text{MnSq, slope})} = -0.84$$

EVANSTON NORTHWESTERN HEALTHCARE

1. What Are Item Banks
2. Creating an Item Bank
3. Test Administration Options
4. Demonstrations of Working CAT
5. Examples from Others
6. Potential Cooperative Arrangements for the Future

3. Test Administration Options

- Web
- Laptop
- PDAs
- Interactive Voice Response (IVR)
- Paper and Pencil

. . .and their implications for clinical adoption

A Comparison of Survey Modalities

	Web	Laptop	PDA	IVR	P & P
Immediate Feedback	*	*	*	*	
Individualized Assessment	*	*	*	*	
Access anytime	*		*	*	
Cost	?	?	?	?	?

P&P versus PDA (Walter et al)

	Berlin Mood Adjective Checklist		
staff	paper pencil 30 items	PDA 30 items	
preparation time	60 s	20 s	- 40 s
data entry	120 s	< 1 s	- 120 s
data bank organization & patient report	90 s	20 s	- 70s
			- 230 s

Savings: one working day per 100 questionnaires

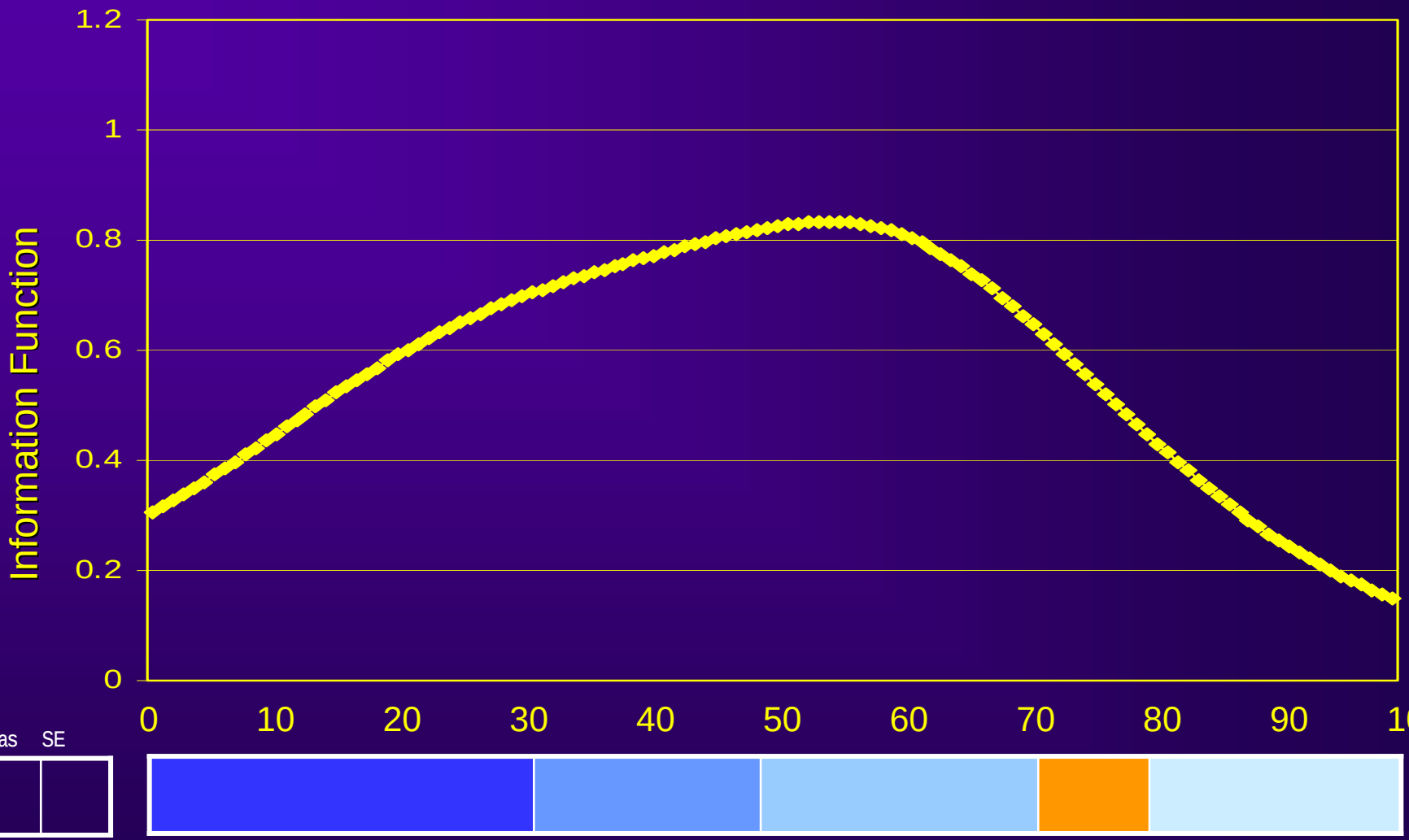
Comparison of paper and pencil assessment with PDA based assessment. Calculation ist based on 400 questionnaires per month and 8 questionnaires per patient at a particular time.

1. What Are Item Banks
2. Creating an Item Bank
3. Test Administration Options
4. Demonstrations of Working CAT
5. Examples from Others
6. Potential Cooperative Arrangements for the Future

4. Demonstrations of working CAT

- Simulated patient
- Fatigue CAT IVR demonstration

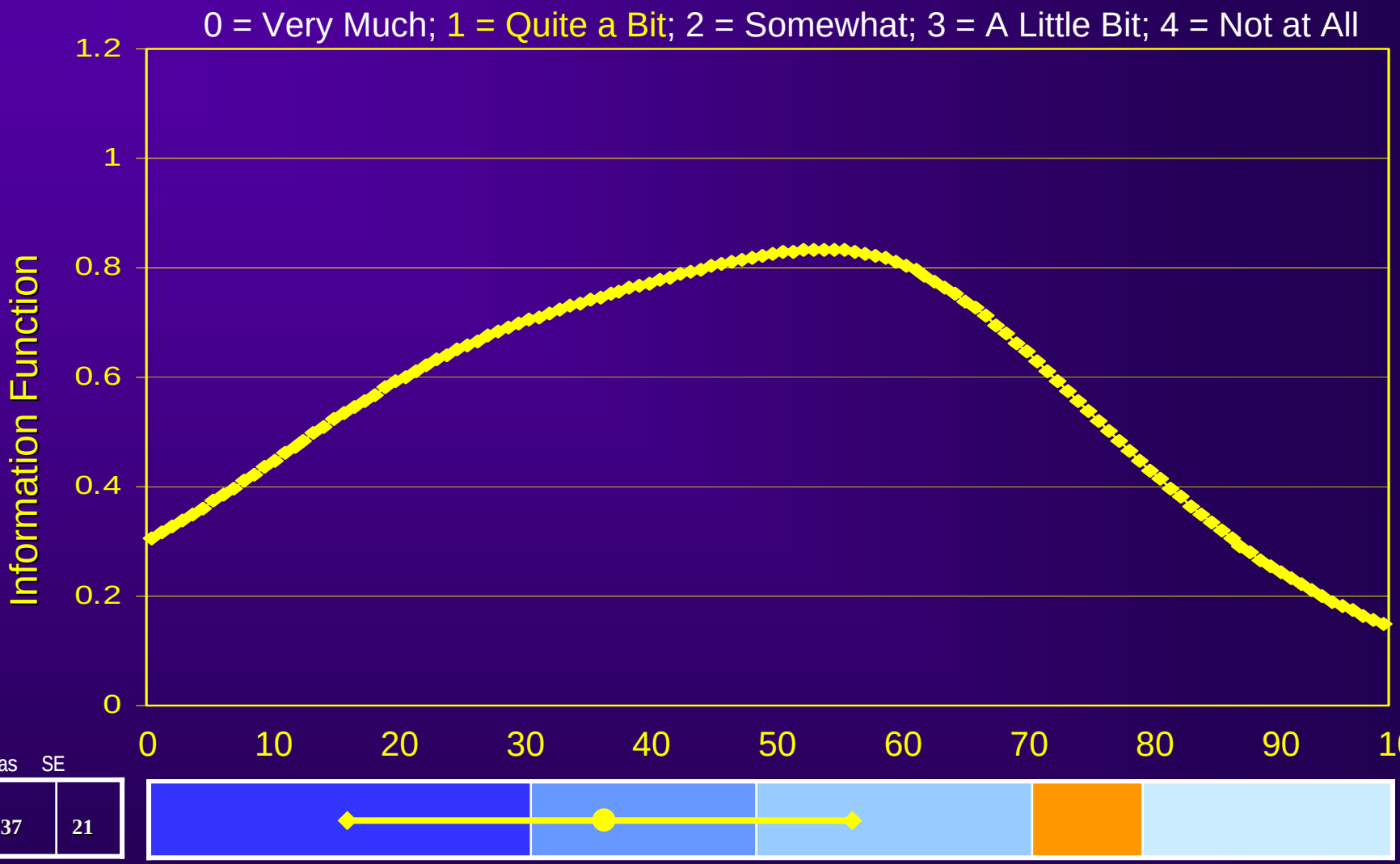
Simulation: GP1 – I have a lack of energy



Item	Meas	SE
1		

EVANSTON NORTHWESTERN HEALTHCARE

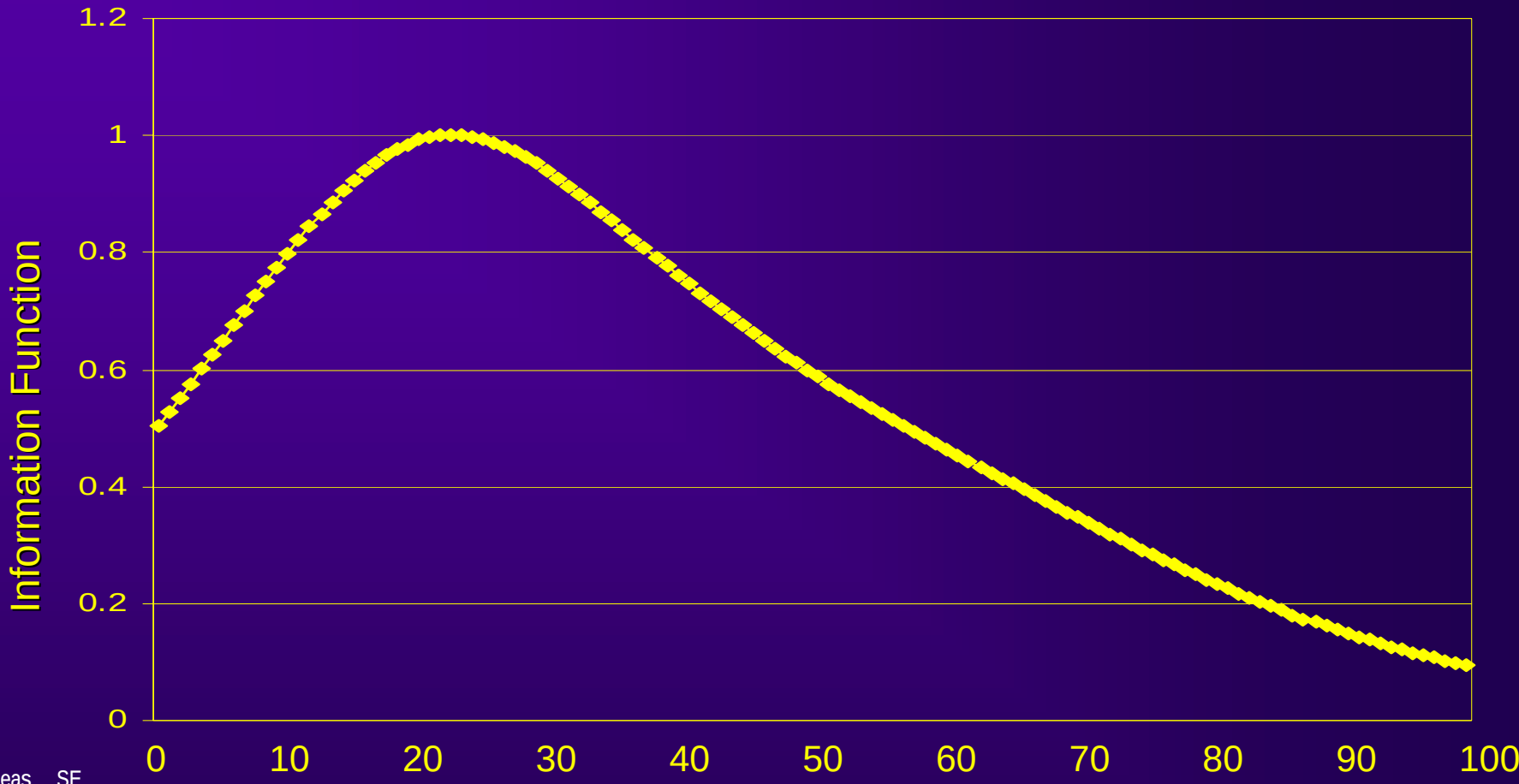
Simulation: GP1 – I have a lack of energy



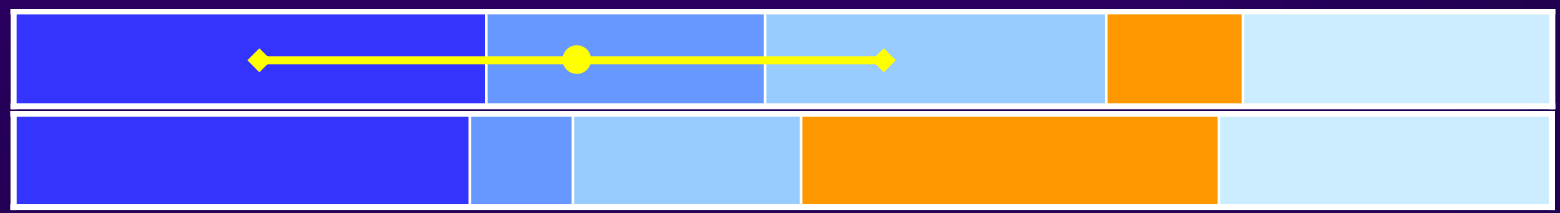
Item	Meas	SE
1	37	21

EVANSTON NORTHWESTERN HEALTHCARE

Simulation: F65 - I have had enough energy to enjoy life.



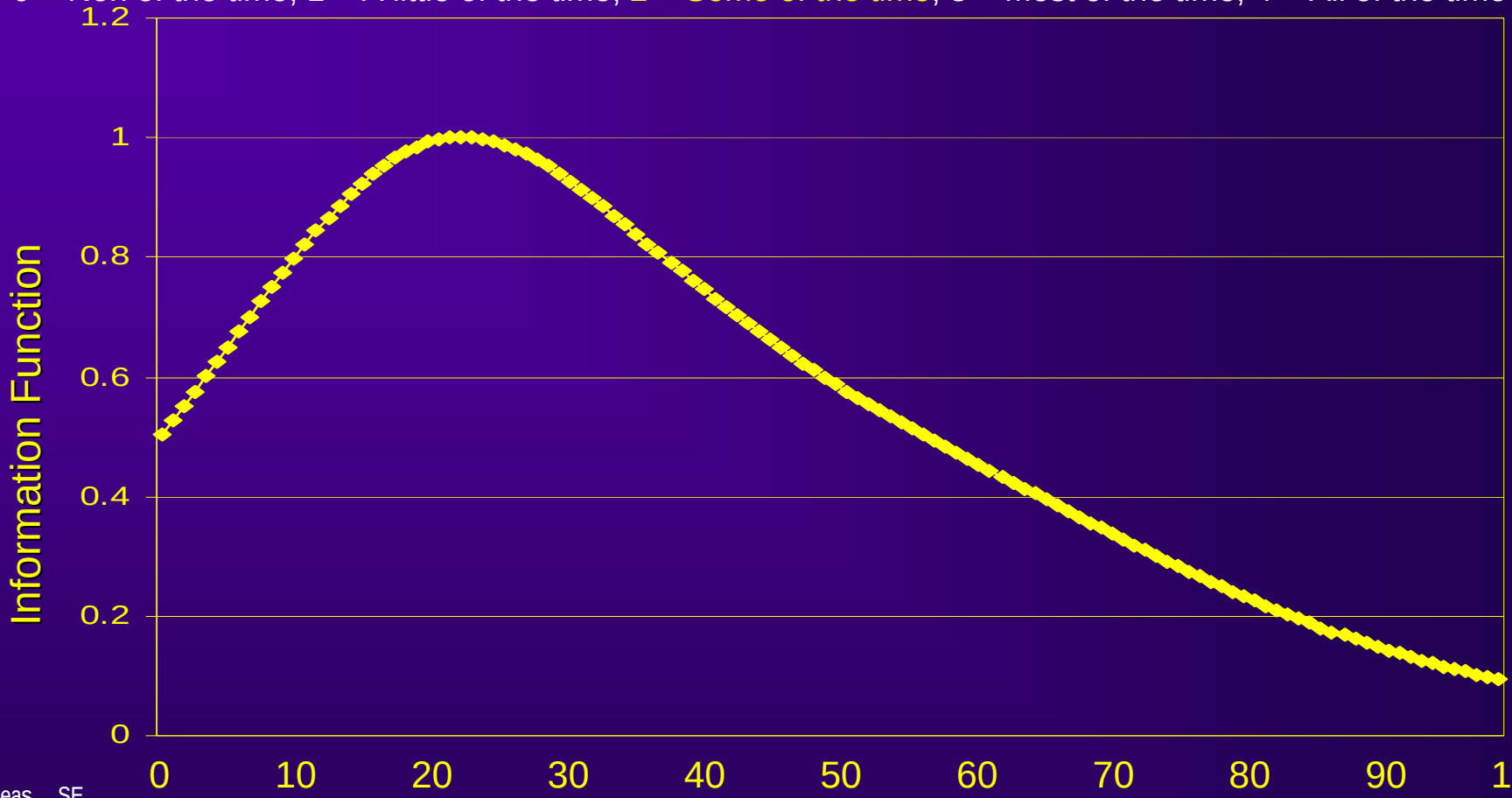
Item	Meas	SE
1	37	21
2		



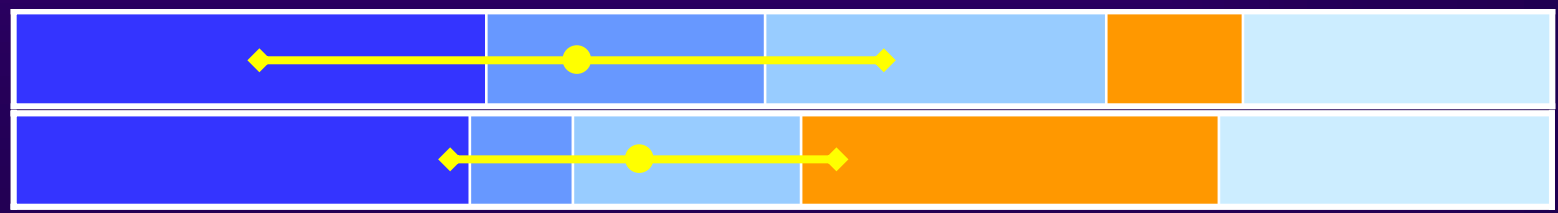
EVANSTON NORTHWESTERN HEALTHCARE

Simulation: F65 - I have had enough energy to enjoy life.

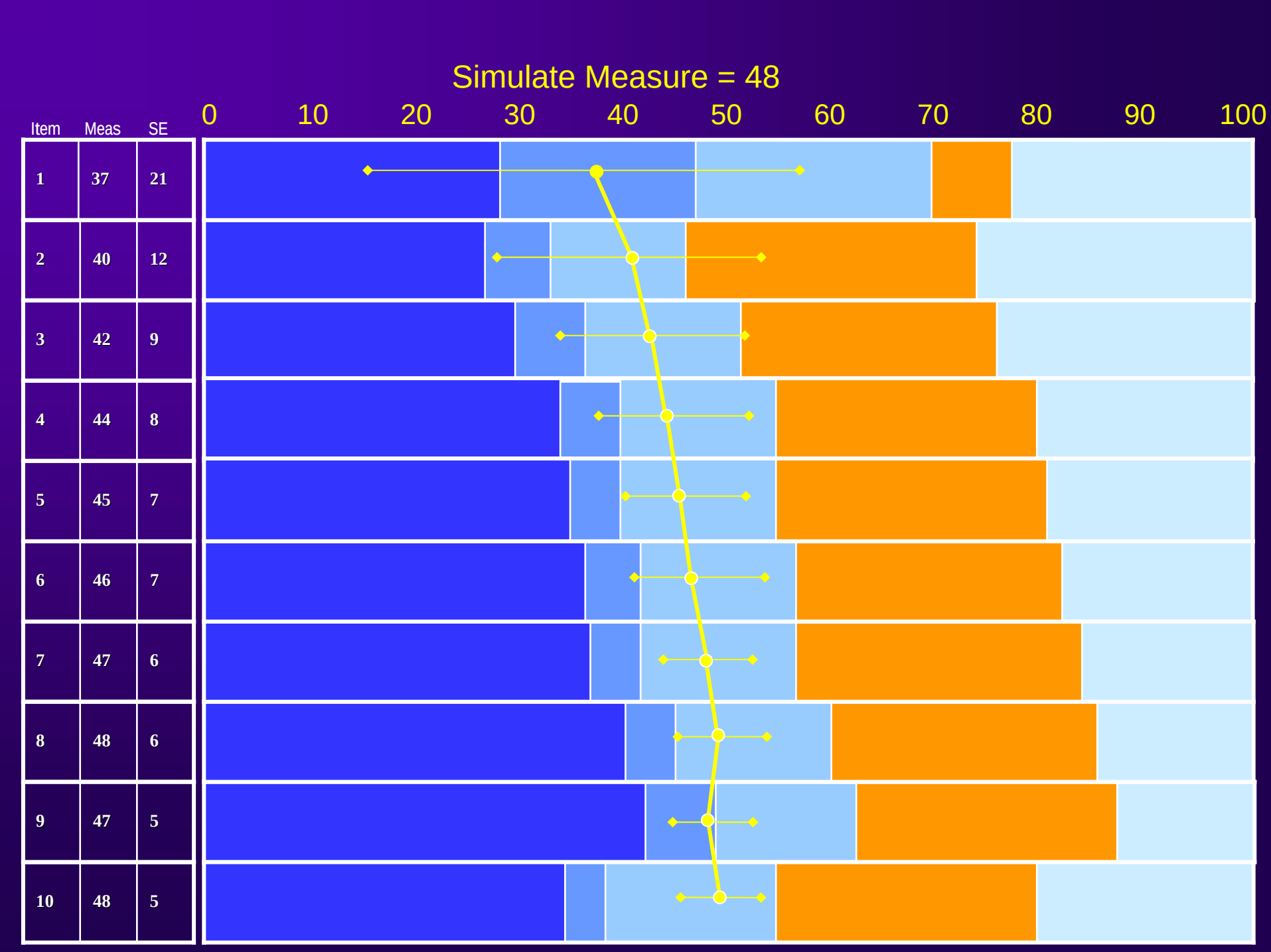
0 = Non of the time; 1 = A little of the time; 2 = Some of the time; 3 = Most of the time; 4 = All of the time

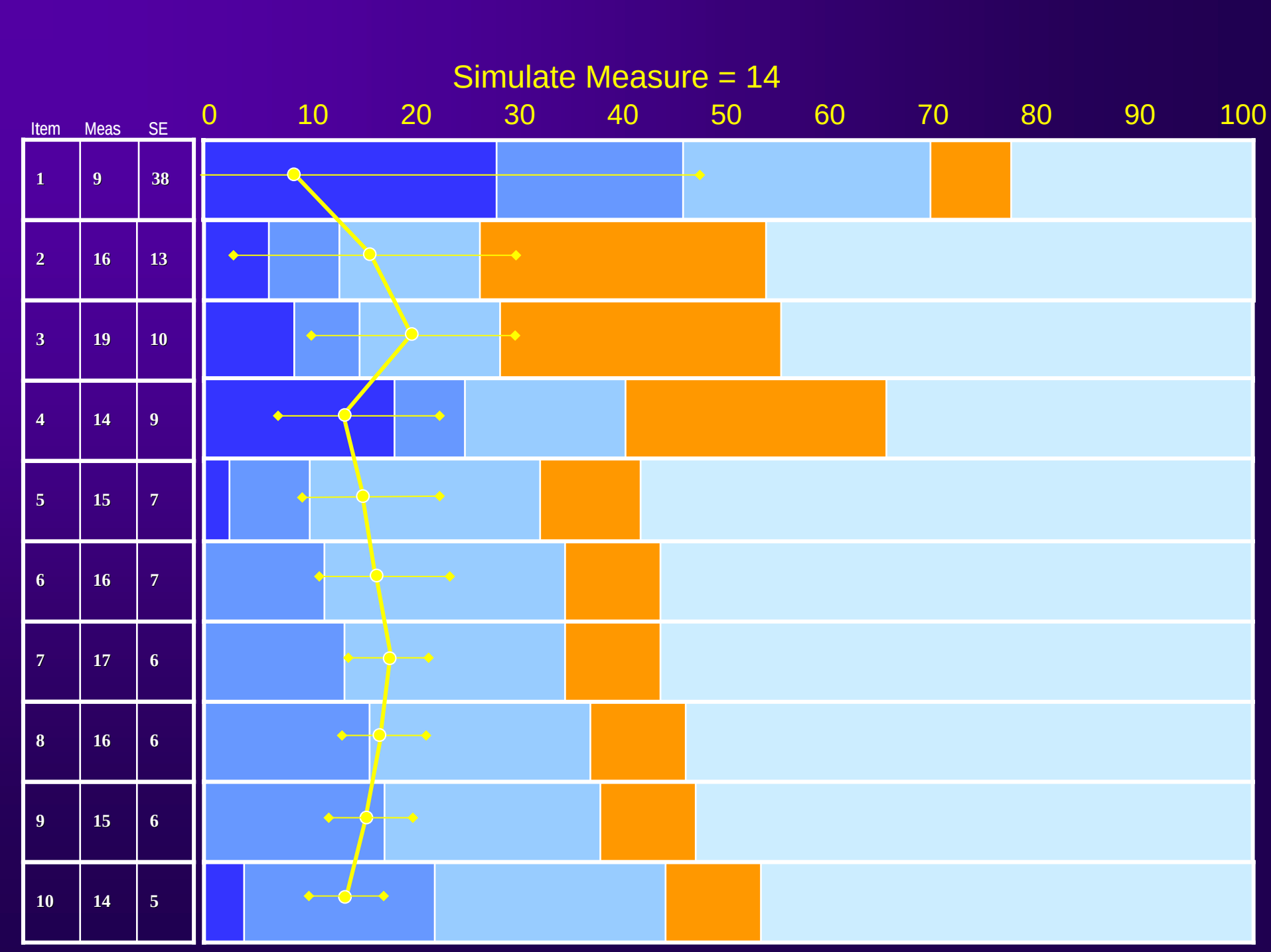


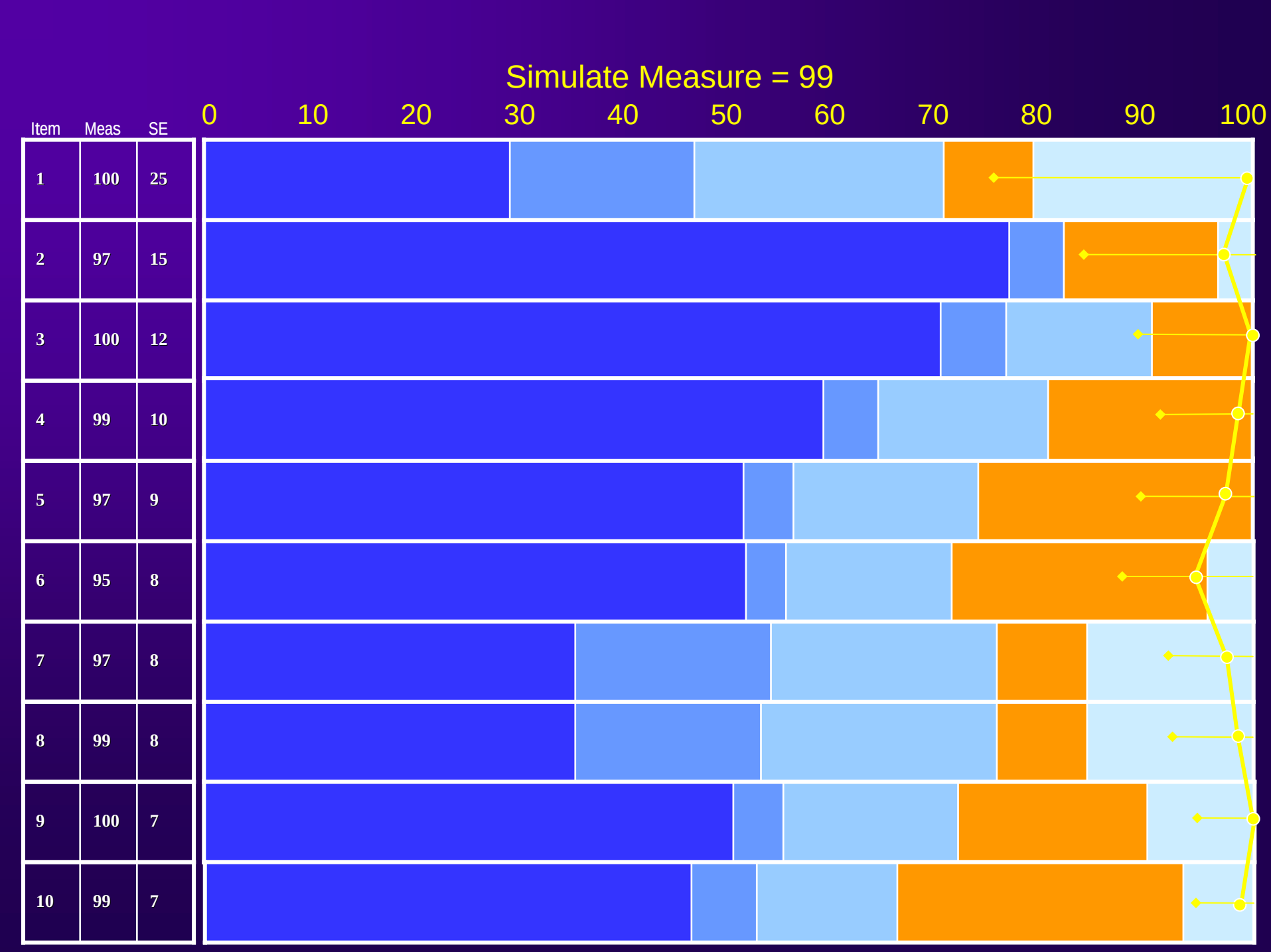
Item	Meas	SE
1	37	21
2	40	12



EVANSTON NORTHWESTERN HEALTHCARE


Simulate Measure = 48
0
10
20
30
40
50
60
70
80
90
100
Item
Meas
SE
1 37 21
2 40 12
3 42 9
4 44 8
5 45 7
6 46 7
7 47 6
8 48 6
9 47 5
10 48 5


Simulate Measure = 14
0
10
20
30
40
50
60
70
80
90
100
Item
Meas
SE
1
9
38
2
16
13
3
19
10
4
14
9
5
15
7
6
16
7
7
17
6
8
16
6
9
15
6
10
14
5


Simulate Measure = 99
0
10
20
30
40
50
60
70
80
90
100
Item
Meas
SE
1 100 25
2 97 15
3 100 12
4 99 10
5 97 9
6 95 8
7 97 8
8 99 8
9 100 7
10 99 7

Fatigue CAT IVR Demonstration (Actual Patient)

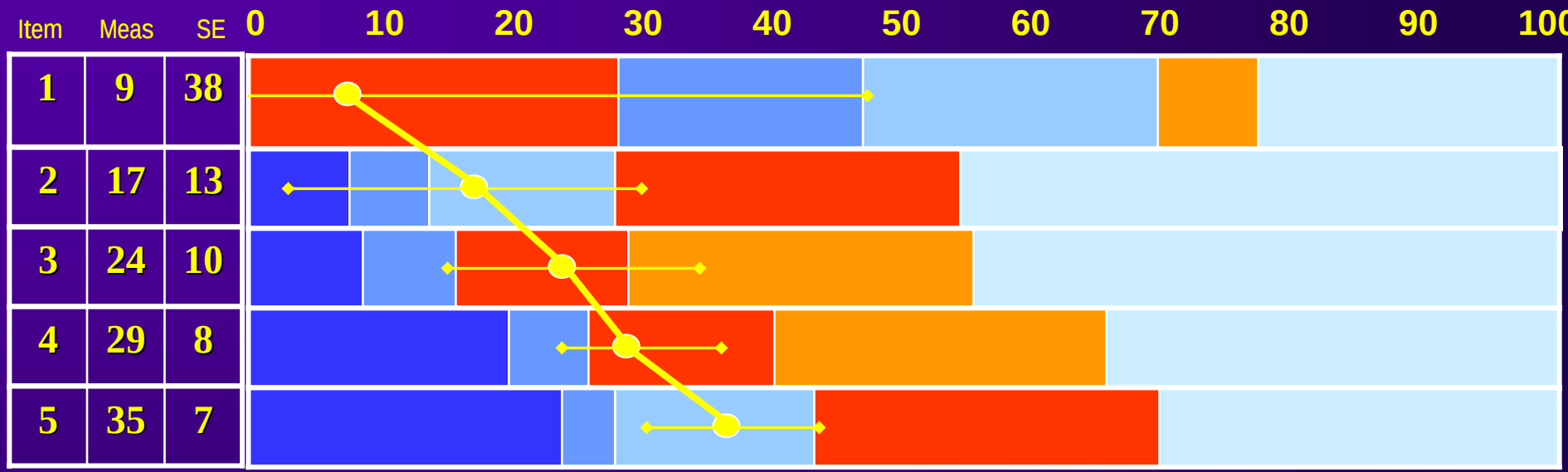
User Experience

C:\WINNT\System32\cmd.exe - CAT_IVR

```
C:\IVR\bin>CAT_IVR
Initializing line: 0
Initializing line: 1

Type 'quit' to end program.
```

IVR - 5 Item Measure = 35



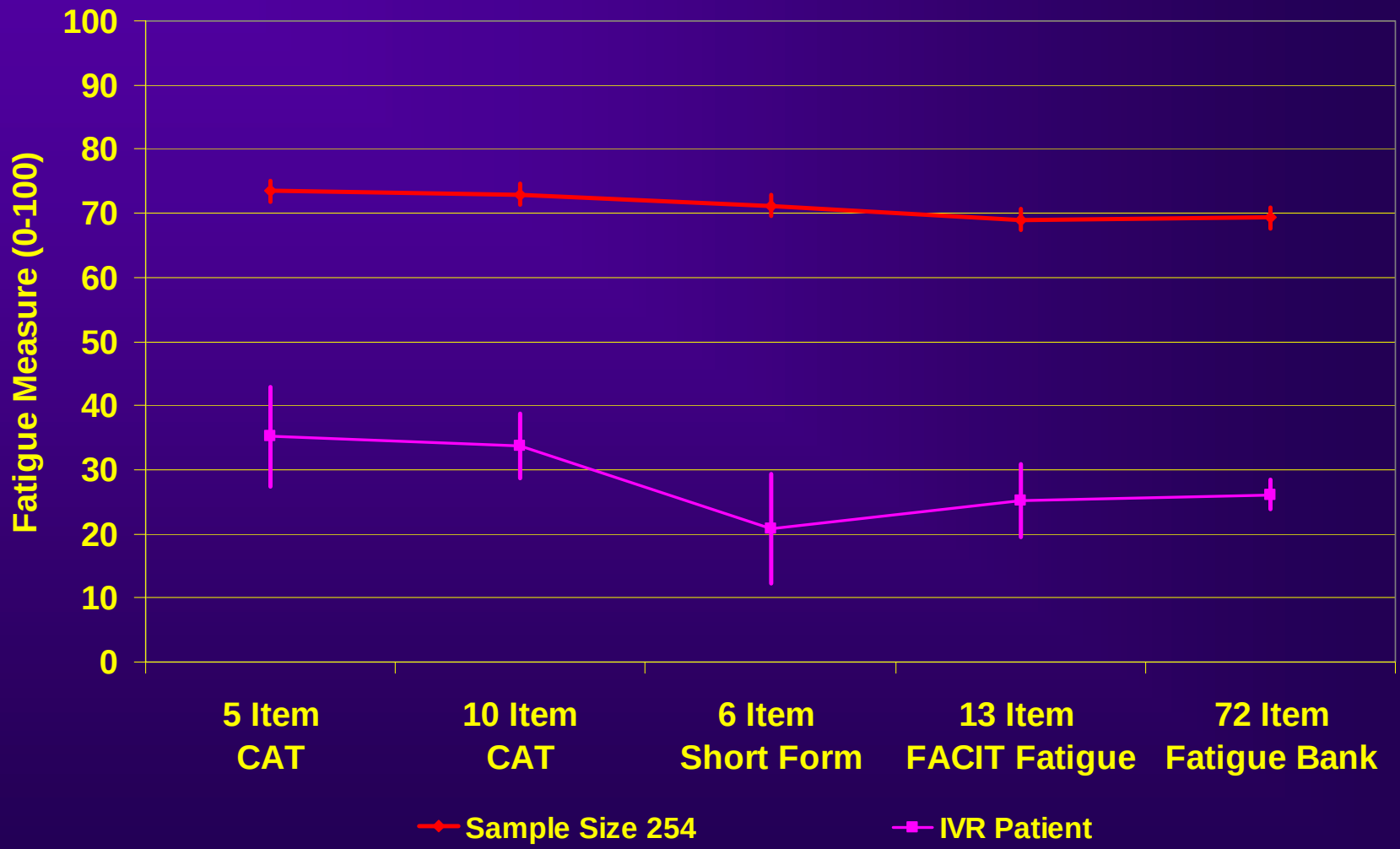
Question	Response
1. I have a lack of energy.	Very much
2. I have had enough energy to eat.	Most of the time
3. I have had enough energy to take a bath or shower.	Some of the time
4. I have had enough energy to read.	Some of the time
5. I have had enough energy to leave the house.	Most of the time

EVANSTON NORTHWESTERN HEALTHCARE

Fatigue Results

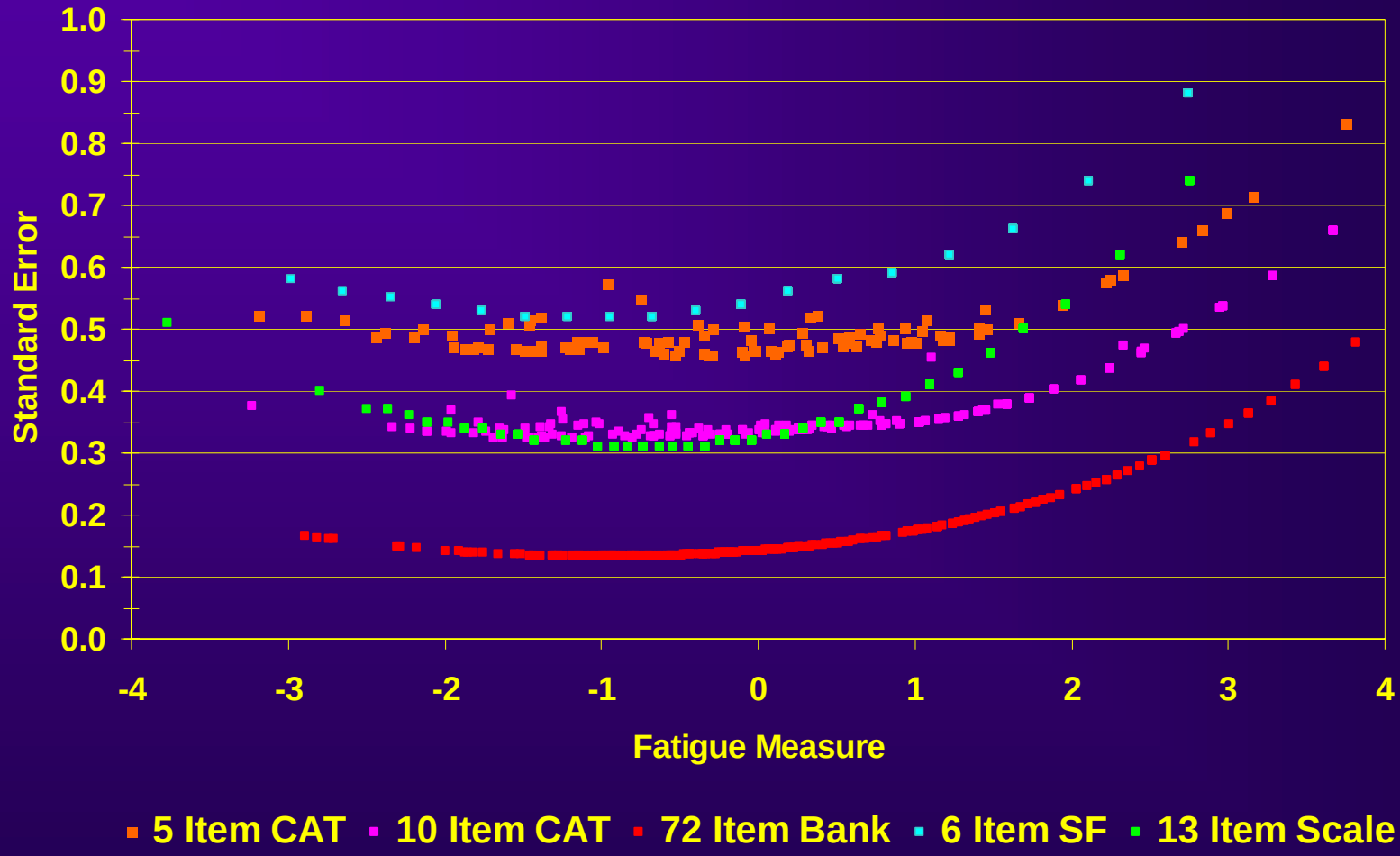
	5 Item CAT	10 Item CAT	6 Item Short Form	13 Item FACIT Fatigue	72 Item Fatigue Bank
Sample Average (n=254)	73.46	72.93	71.15	69.02	69.28
Standard Error of Mean	1.62	1.60	1.64	1.63	1.60
IVR Patient	35.14	33.70	20.77	25.23	26.07
Standard Error	7.70	5.12	8.62	5.69	2.28

Patient Estimate of Fatigue Scale



EVANSTON NORTHWESTERN HEALTHCARE

Fatigue Measure and Standard Error Comparision by Test Length



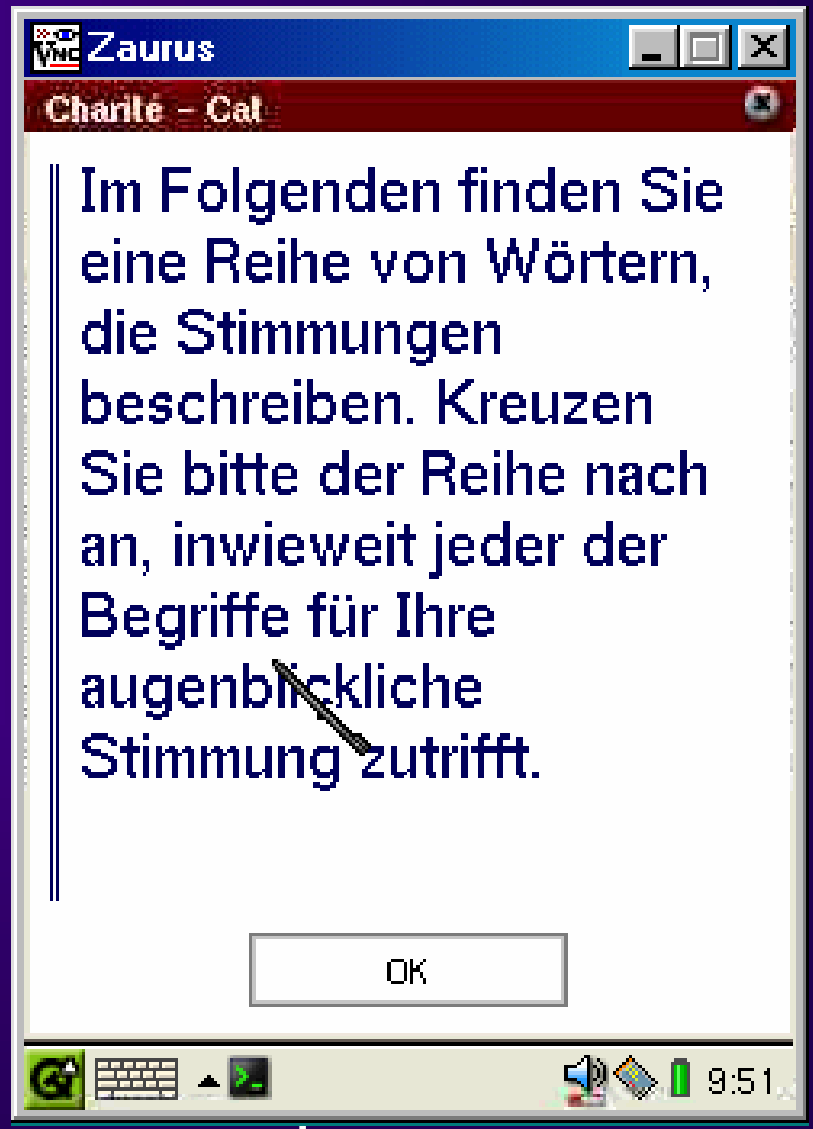
EVANSTON NORTHWESTERN HEALTHCARE

1. What Are Item Banks
2. Creating an Item Bank
3. Test Administration Options
4. Demonstrations of Working CAT
5. Examples from Others
6. Potential Cooperative Arrangements for the Future

Rose, Walter et al PDA Example

XML-RPC

LAN
WLAN



CAT-Engine

Computation of latent trait item selection

Server

Frontend

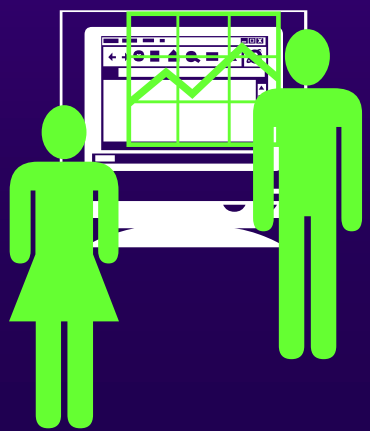
EVANSTON NORTHWESTERN HEALTHCARE

CAT PDA Demonstration

Workflow of PDA Based Assessment

PDA Assessment

data transfer to main system (cable or WLAN)



discuss results with psychologist or physician

EVANSTON NORTHWESTERN HEALTHCARE

ASSIST: Touch Screens at the Clinic

Used in clinical trials

- Large, colorful touch-screens at study site
- No special telecommunications arrangements necessary
- Support and deployment for many sites, many countries
- Validated (21 CFR Part 11compliance) for FDA submission
- High satisfaction ratings from site staff and patients

Assist TECHNOLOGIES

EVANSTON NORTHWESTERN HEALTHCARE

Typical Touch-Screen Question

Assist TECHNOLOGIES

EVANSTON NORTHWESTERN HEALTHCARE

1. What Are Item Banks
2. Creating an Item Bank
3. Test Administration Options
4. Demonstrations of Working CAT
5. Examples from Others
6. Potential Cooperative Arrangements for the Future

6. Potential Cooperative Arrangements for the Future

Thoughts on . . .

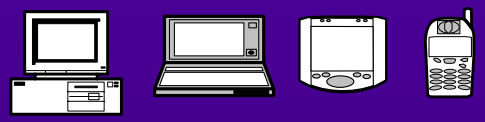
- A centralized bank repository
- How a public-private partnership might look

National Item Banks are Coming?

- National Institute of Neurological Disorders and Stroke (NINDS)
- NIH Roadmap: Patient Reported Outcome Measurement Information System (PROMIS)

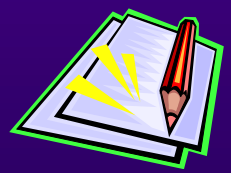
CORE Data Repository

Assessment Module



Web Laptop PDA IVR

Import Module



Paper and Pencil

Central Database

- Item Banks
- Study Protocols
- Users
- Patients
- Patient Reported Data

Report Module

Patients

Clinicians

Study Coordinators

EVANSTON NORTHWESTERN HEALTHCARE

How A Public-Private Partnership Might Look

- Who updates the PROMIS banks? Who pays for them?

- Non-profit licensing center or “delivery” company
  - Partnership with NIH
  - One or many centers
  - Nominal unit cost to support maintenance
  - Critical question: Does it have R&D staff, or depend upon NIH-funded continuation through R-01 pool, PAs, RFAs, etc?

- License to a for-profit test center with distribution rights
  - Use existing national sales network
  - Return to question: Who updates the banks?

- NIH Office of PRO Measurement?