Developing Tailored Instruments: Item Banking and Computerized Adaptive Assessment

Chih-Hung Chang, Ph.D.
Feinberg School of Medicine
Northwestern University

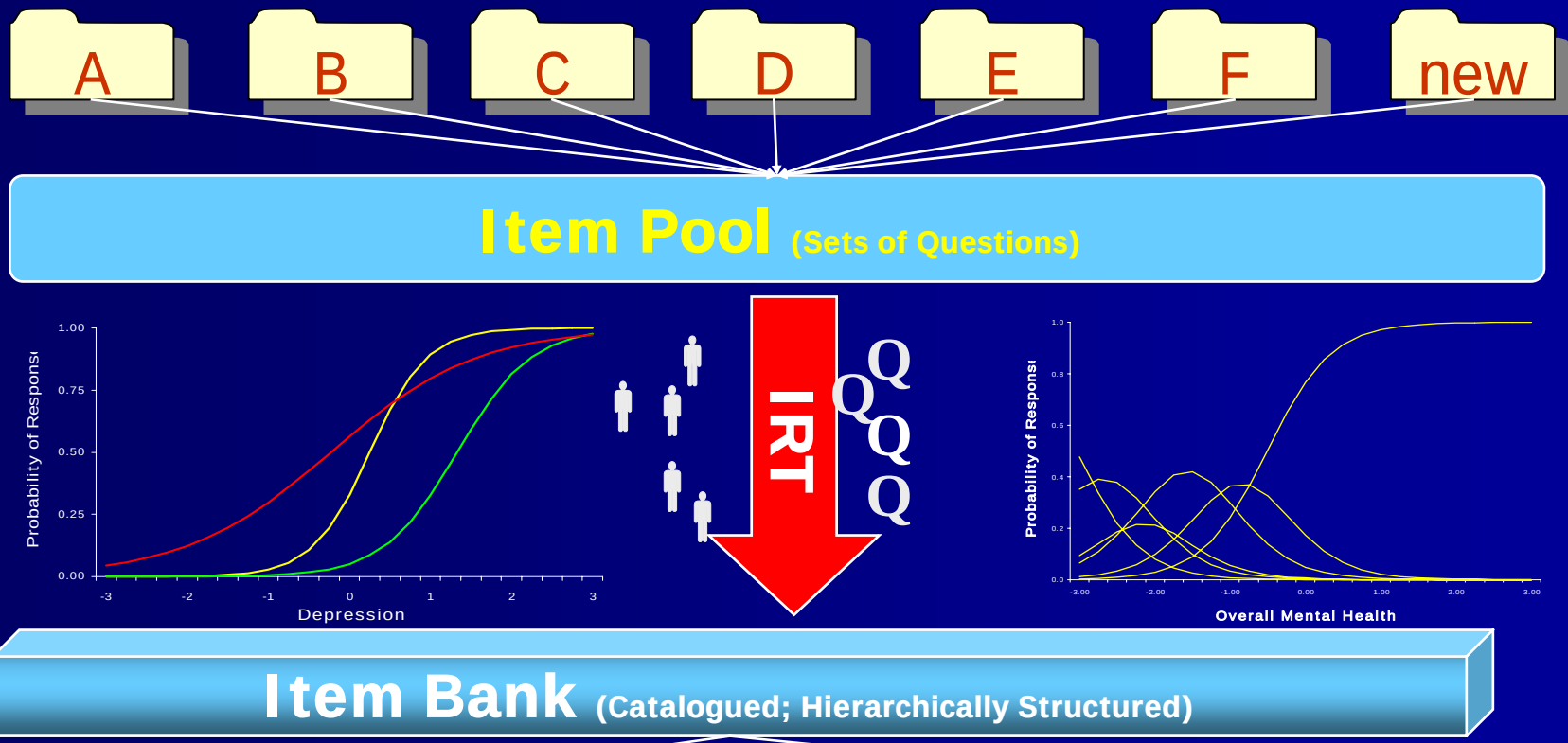
NCI-DIA Sponsored Advances in Health Outcomes Measurement Conference
June 23-25, 2004, Bethesda, MD

Acknowledgements

- R. Darrell Bock, Ph.D.
- Benjamin D. Wright, Ph.D.
- Robert D. Gibbons, Ph.D.
- David J. Weiss Ph.D.
- Jakob B. Bjorner, M.D., Ph.D.
- David Thissen, Ph.D.
- Bryce B. Reeve, Ph.D.

Slide prepared by Chih-Hung Chang, Ph.D.

Item Banking and CAT



Slide prepared by Chih-Hung Chang, Ph.D.

Principles of Adaptive Testing

- IRT pre-calibrated item bank
- Initial item selection
- Test scoring method
- Item selection during test administration
- Stopping rules

Item Bank

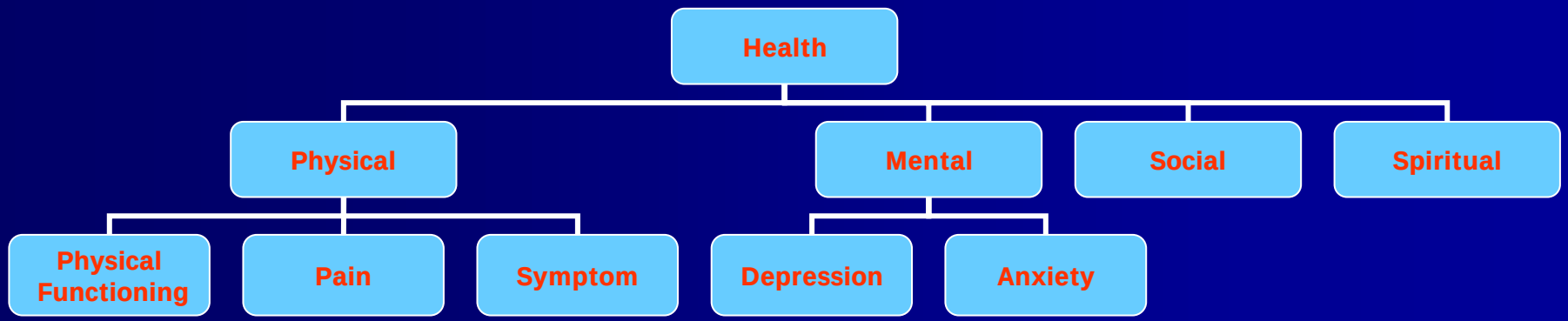
- IRT-calibrated items/questions
- Items cover entire latent continuum
- Items represent differing trait
- Items provide differing information
- Basis for tailored/adaptive testing
- Items can be selected to maximize precision and retain clinical relevance

Item Banking is Inter-disciplinary

- Psychometricians
- Information scientists
- Clinicians/healthcare providers
- Outcomes researchers
- Content experts
- ...

Approaches to Develop Item Banks

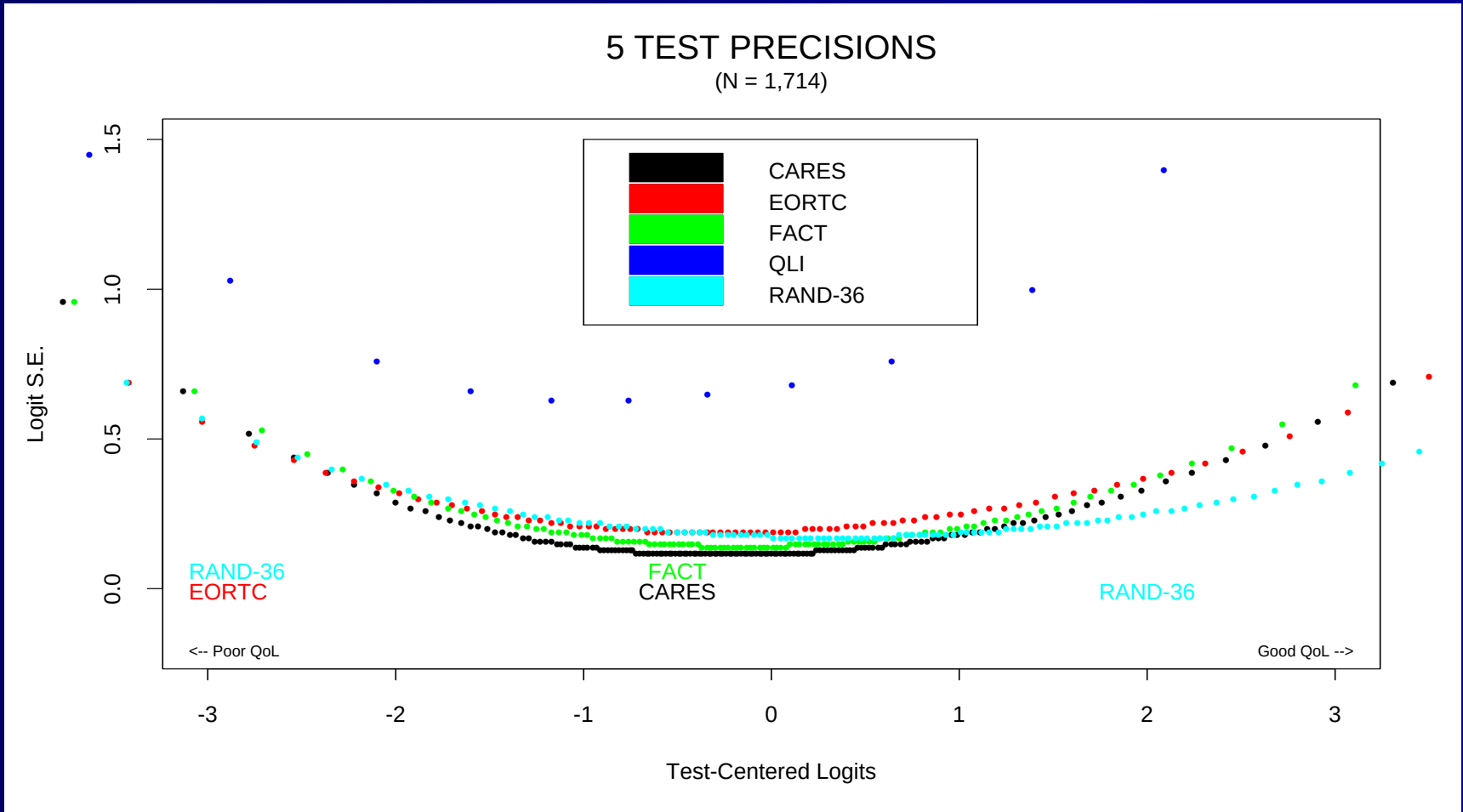
- Top-Down Approach
- Bottom-Up Approach



Data Collection Designs

- Random Groups
- Single group with counterbalance
- Common-item nonequivalent groups

Concurrent Calibration



Chang C-H, & Cella, D. (1997). Equating health-related quality of life instruments in applied oncology settings. *Physical Medicine and Rehabilitation: States of the Art Reviews, 11(2),* 397-406.

Independent vs. Overlapping Item Banks

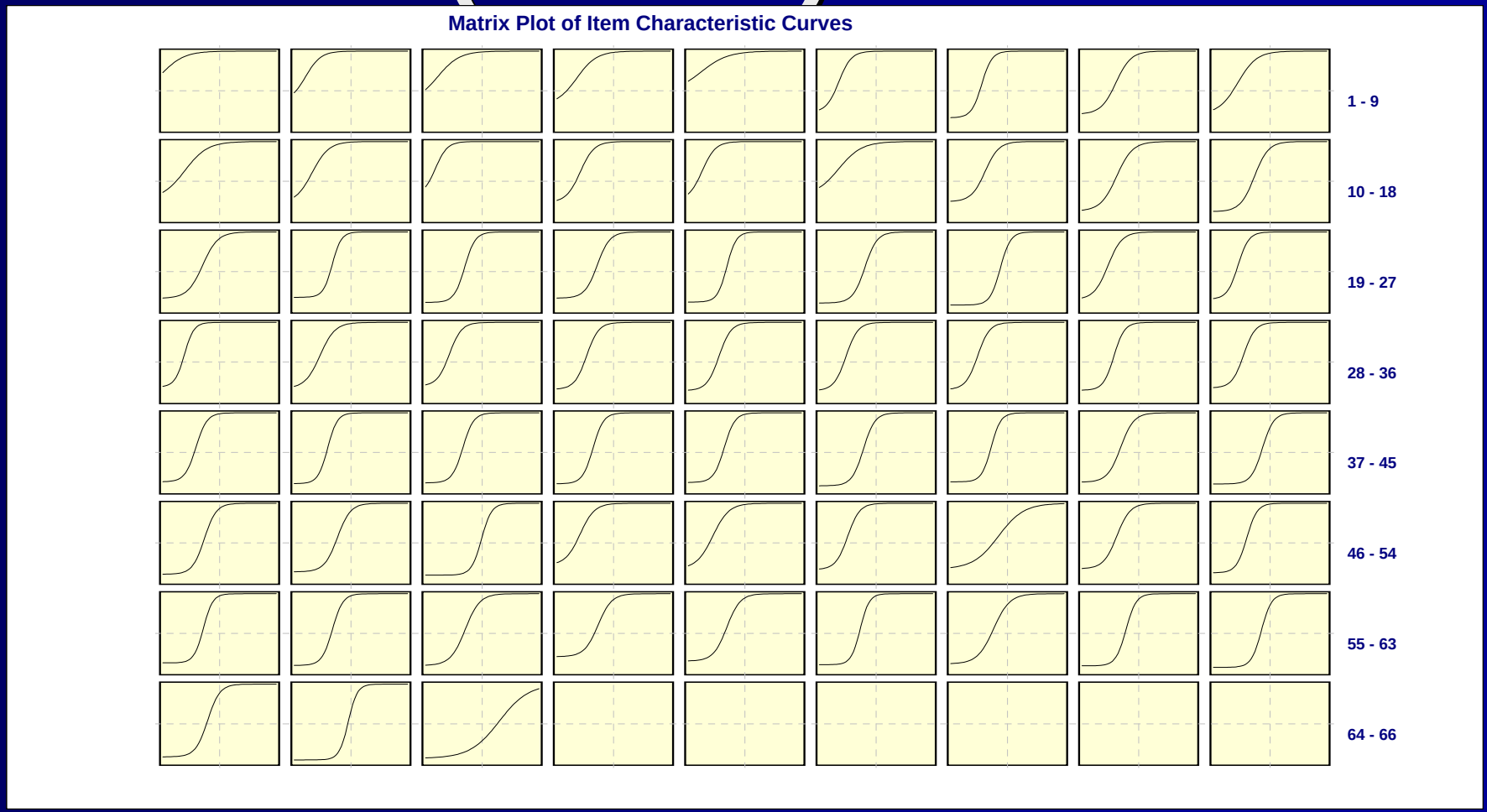
- Multi-form Structure Design (Multiple Assignment)

Physical	Mental	Social
X	X	
X		X
X	X	X

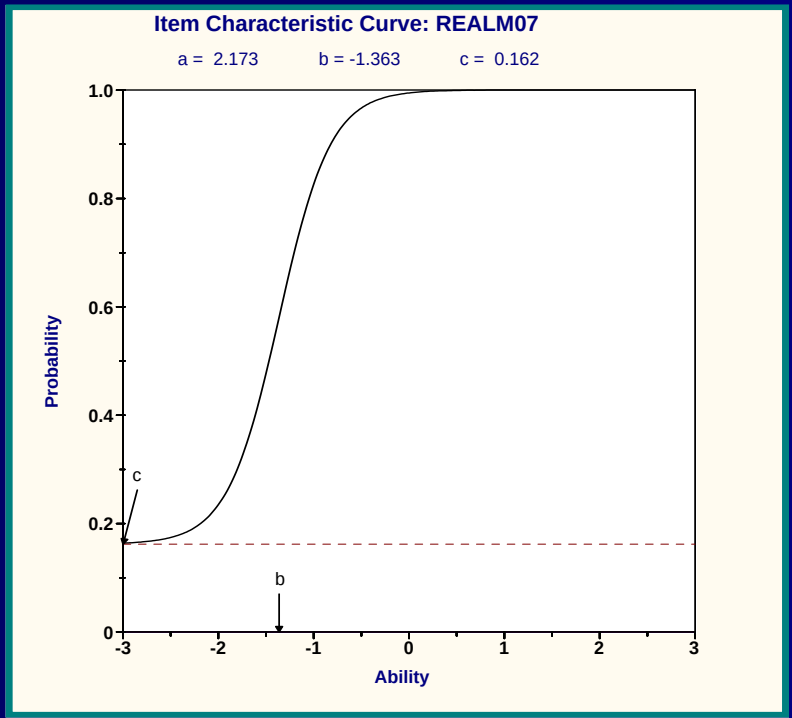
Why Short / Brief Forms?

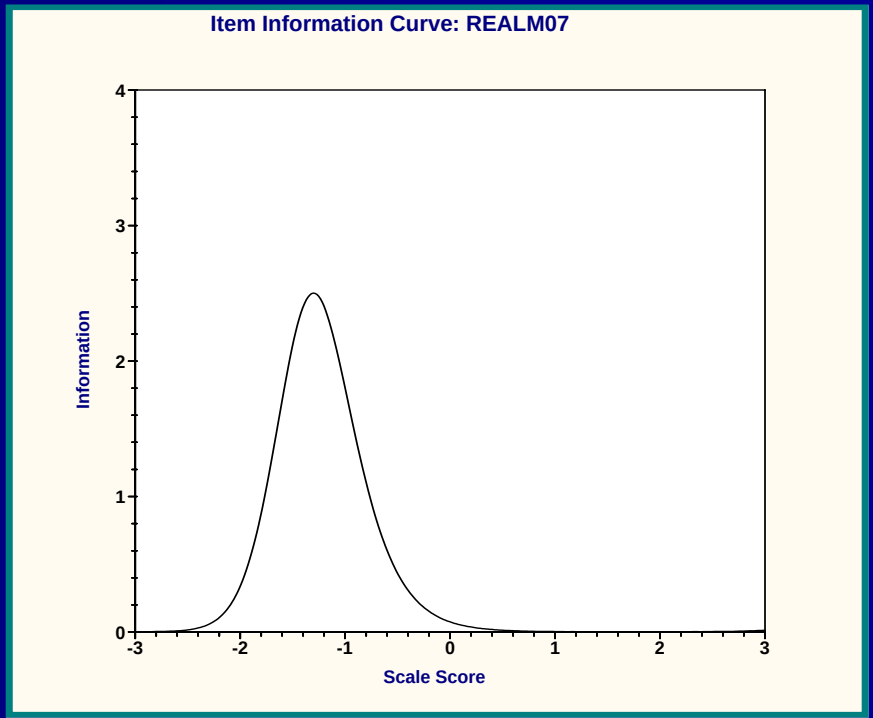
- Reduce respondent burden
- Clinical applications in busy clinics
- Large scale data collection
- Retain comparability with its long form
- Automation via CAT?

Item Reduction: 66-item Rapid Estimate of Adult Literacy in Medicine (REALM)



Item Characteristic Curve and Item Information Curve

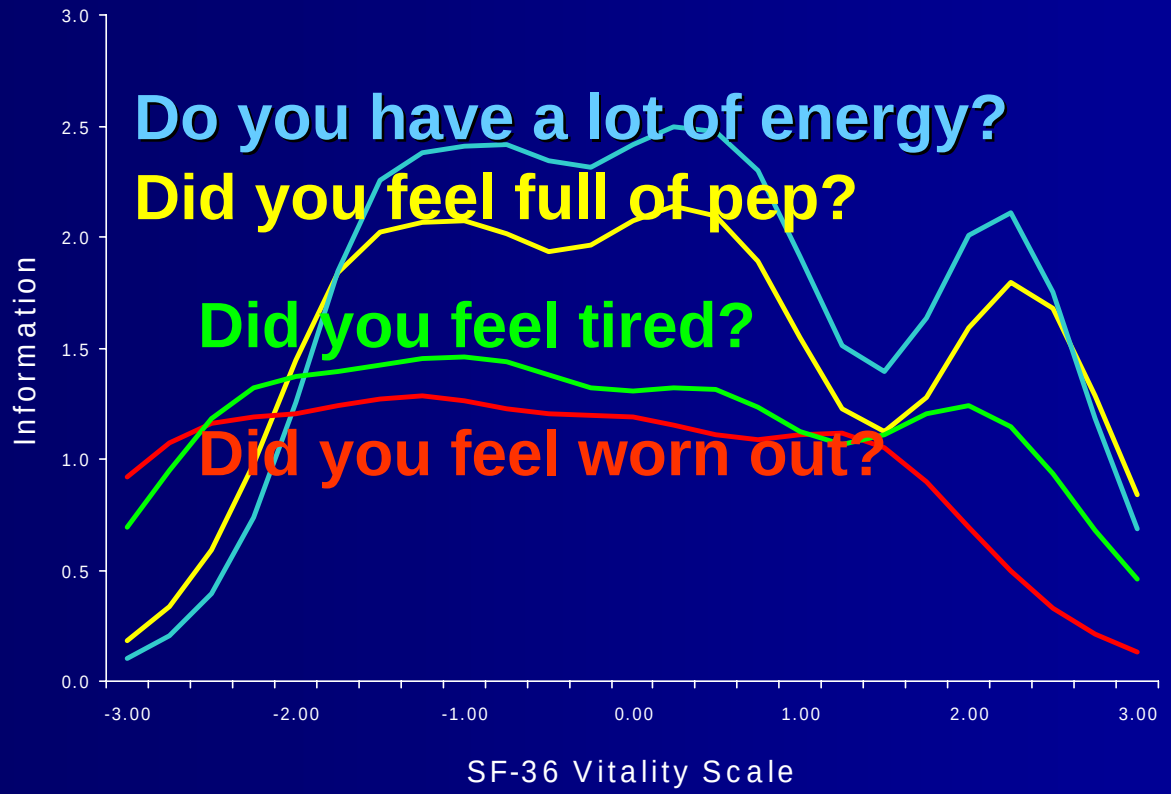




3-Parameter Model, Normal Metric **Item: 7**

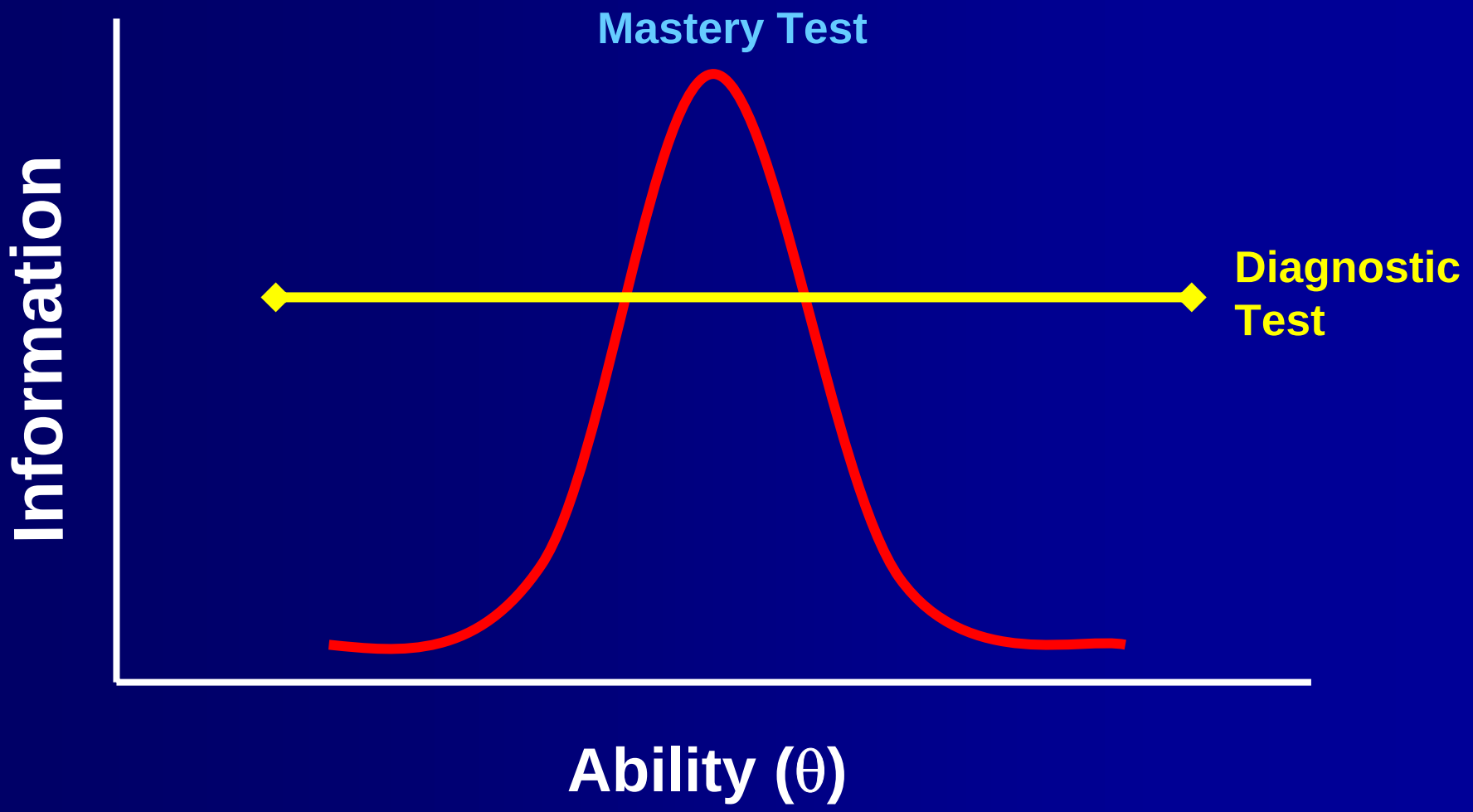
The parameter a is the item discriminating power, the reciprocal (1/a) is the item dispersion, b is an item location parameter and c the guessing parameter.

Content Redundancy



All of the time – Most of the time – Good bit of the time – Some of the time – Little of the time – None of the time

Two-Target Information Functions in Automated Test Design



Slide prepared by Chih-Hung Chang, Ph.D.

Content Deficiency

Multi-Dimensionality:
Full-Information Item Bi-Factor Analysis*

Item Content	Depression	SA	SRA	IR	PA
3. could not shake off the blues	0.70	0.14	0.00	0.00	0.00
6. felt depressed.	**0.78**	**0.34**	0.00	0.00	0.00
9. my life had been a failure	0.71	0.00	0.00	0.00	0.00
14. felt lonely	0.76	0.28	0.00	0.00	0.00
17. had crying spells	**0.68**	**0.49**	0.00	0.00	0.00
18. felt sad	**0.72**	**0.36**	0.00	0.00	0.00
1. bothered by things	0.50	0.00	-0.11	0.00	0.00
2. did not feel like eating	0.48	0.00	**0.30**	0.00	0.00
5. trouble keeping my mind	0.47	0.00	-0.03	0.00	0.00
7. everything I did was an effort	0.41	0.00	**0.61**	0.00	0.00
11. sleep was restless	0.55	0.00	-0.09	0.00	0.00
13. talked less than usual	0.46	0.00	0.10	0.00	0.00
20. could not get going	0.63	0.00	-0.01	0.00	0.00
15. people unfriendly	0.30	0.00	0.00	0.57	0.00
19. felt people disliked me	0.62	0.00	0.00	**0.52**	0.00
4.* felt I was as good as others	0.45	0.00	0.00	0.00	**0.71**
8.* felt hopeful about future	0.44	0.00	0.00	0.00	**0.61**
10. felt fearful	0.66	0.00	0.00	0.00	-0.17
12.* was happy	0.66	0.00	0.00	0.00	**0.41**
16.* enjoyed life.	0.67	0.00	0.00	0.00	**0.49**

*Gibbons RD, Hedeker DR. Full-Information Item Bifactor Analysis. Psychometrika 1992;57 (3):423-36.

Slide prepared by Chih-Hung Chang, Ph.D.

What kind of short form?

	Rarely or none of the time (less than 1 day)	**Some or a little of the time** (1-2 days)	**Occasionally or a moderate amount of time** (3-4 days)	**All of the time** (5-7 days)
1. I was bothered by things that usually don't bother me				

Question 1

0 I do not feel sad.
1 I feel sad
2 I am sad all the time and I can't snap out of it.
3 I am so sad or unhappy that I can't stand it.

Are you basically satisfied with your life? **True / False**

Other Concerns

- Response Categories
  - True / False
  - Not at all --- Very much
  - None of the time --- All of the time
- Time Frame
  - during the PAST month, including today
  - during the past 4 weeks
  - during the past week / past 7 days
  - during the past two weeks

Other Concerns

- Diversity of patient populations
  - Health Literacy
  - Language/Culture/Ethnicity

Adaptive Test

- An adaptive test is a tailored, individualized measure which involves selecting a set of test items for each individual that best measures the psychological characteristics of that person (Weiss, 1985)

Weiss DJ. Adaptive testing by computer. *J Consult Clin Psychol.* Dec 1985;53(6):774-789.

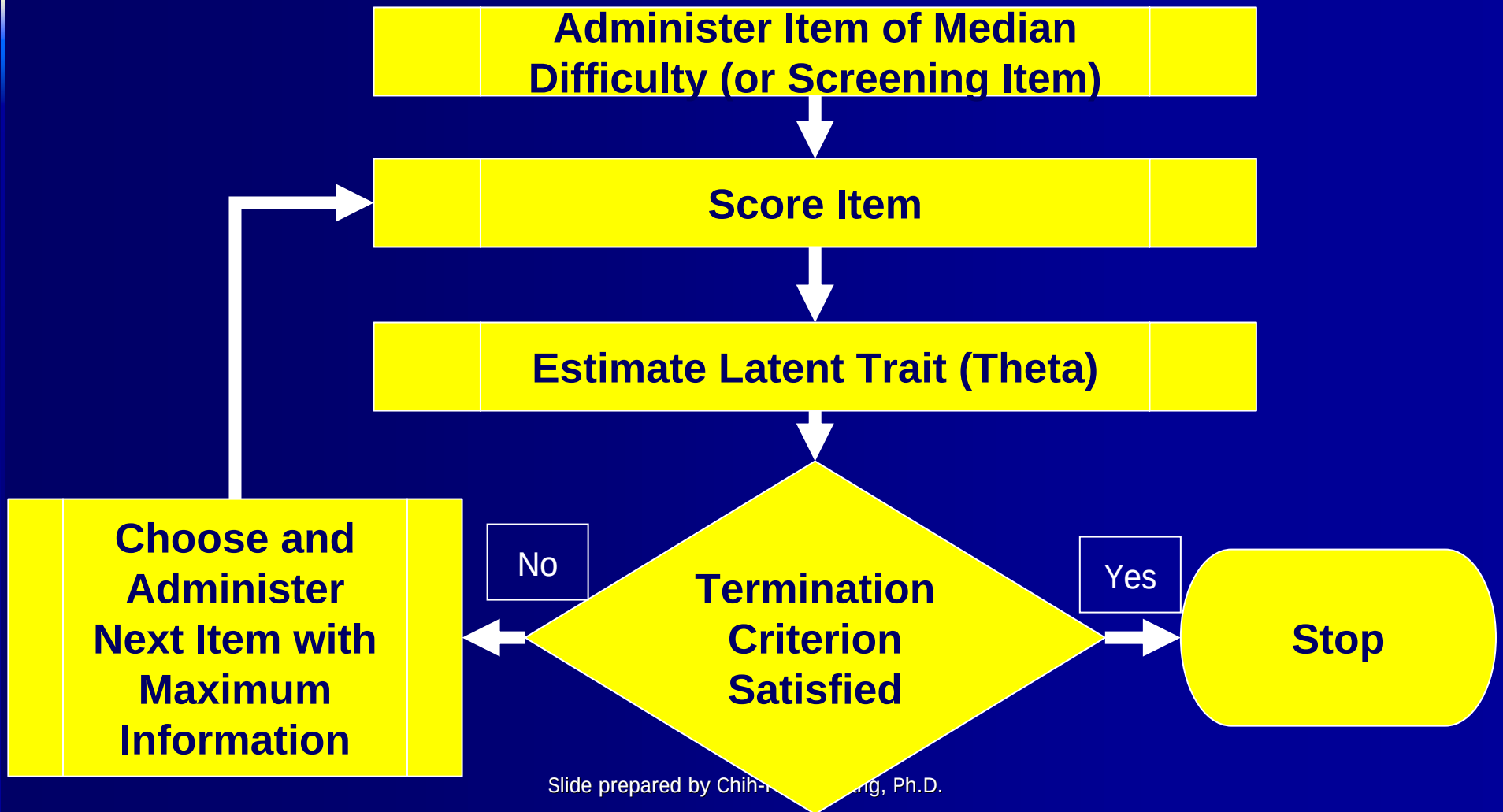
Computerized Adaptive Testing

- Adaptive testing selects questions based on previous responses
- Tailored item and test difficulties
- Eliminates floor and ceiling effects
- Require fewer questions to arrive at an accurate estimate
- Automate question administration, data recording, scoring, and prompt reporting
- Allows for immediate feedback

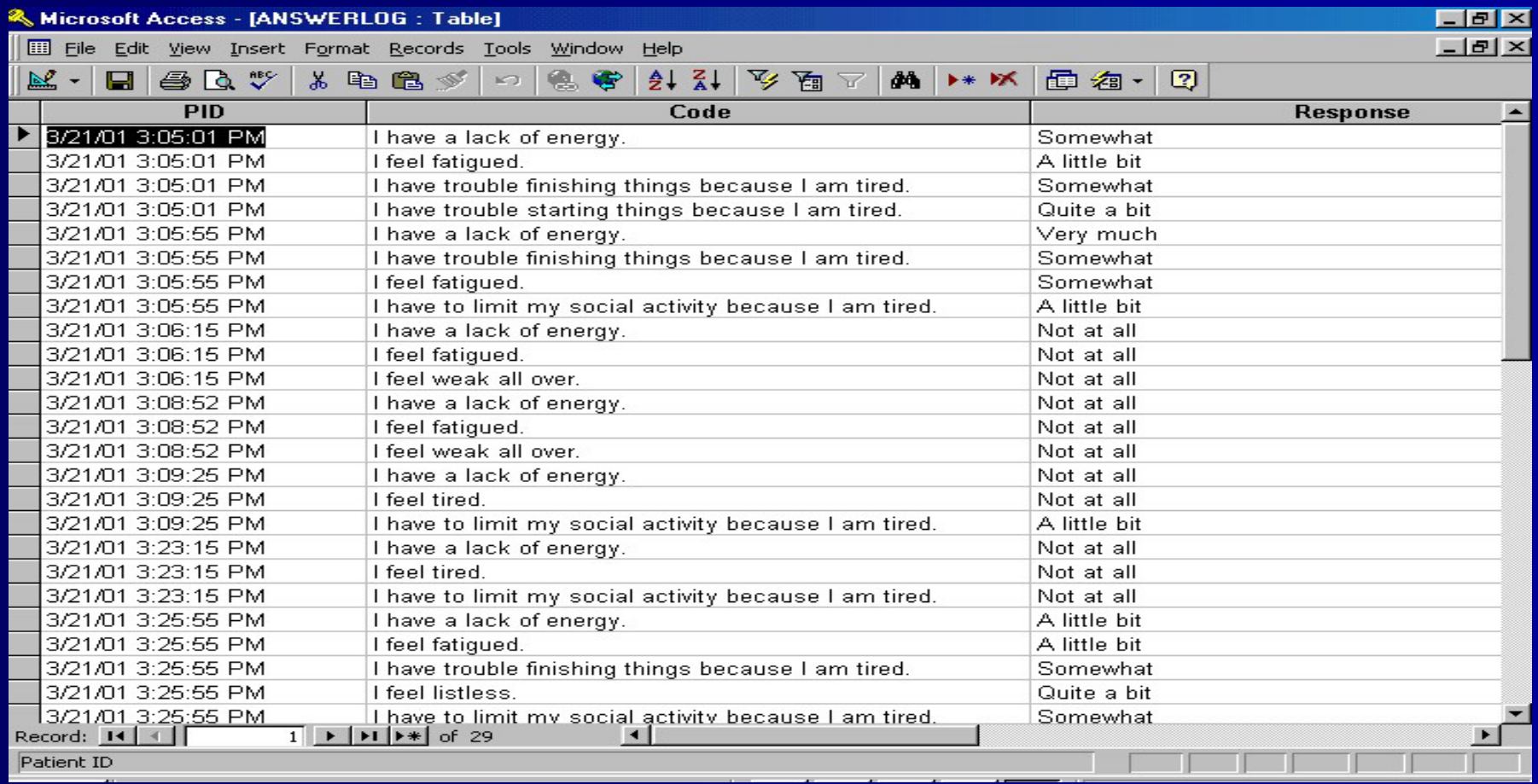
Principles of Adaptive Testing

- IRT pre-calibrated item bank
- Initial item selection
- Test scoring method
- Item selection during test administration
- Terminating rules

CAT Algorithm



CAT Answer Log



PID	Code	Response
3/21/01 3:05:01 PM	I have a lack of energy.	Somewhat
3/21/01 3:05:01 PM	I feel fatigued.	A little bit
3/21/01 3:05:01 PM	I have trouble finishing things because I am tired.	Somewhat
3/21/01 3:05:01 PM	I have trouble starting things because I am tired.	Quite a bit
3/21/01 3:05:55 PM	I have a lack of energy.	Very much
3/21/01 3:05:55 PM	I have trouble finishing things because I am tired.	Somewhat
3/21/01 3:05:55 PM	I feel fatigued.	Somewhat
3/21/01 3:05:55 PM	I have to limit my social activity because I am tired.	A little bit
3/21/01 3:06:15 PM	I have a lack of energy.	Not at all
3/21/01 3:06:15 PM	I feel fatigued.	Not at all
3/21/01 3:06:15 PM	I feel weak all over.	Not at all
3/21/01 3:08:52 PM	I have a lack of energy.	Not at all
3/21/01 3:08:52 PM	I feel fatigued.	Not at all
3/21/01 3:08:52 PM	I feel weak all over.	Not at all
3/21/01 3:09:25 PM	I have a lack of energy.	Not at all
3/21/01 3:09:25 PM	I feel tired.	Not at all
3/21/01 3:09:25 PM	I have to limit my social activity because I am tired.	A little bit
3/21/01 3:23:15 PM	I have a lack of energy.	Not at all
3/21/01 3:23:15 PM	I feel tired.	Not at all
3/21/01 3:23:15 PM	I have to limit my social activity because I am tired.	Not at all
3/21/01 3:25:55 PM	I have a lack of energy.	A little bit
3/21/01 3:25:55 PM	I feel fatigued.	A little bit
3/21/01 3:25:55 PM	I have trouble finishing things because I am tired.	Somewhat
3/21/01 3:25:55 PM	I feel listless.	Quite a bit
3/21/01 3:25:55 PM	I have to limit my social activity because I am tired.	Somewhat

Potential Problems with CAT in Health Outcomes Measurement

- Context effects
- Unbalanced content
- Time frame
- Response categories
- *Multidimensionality*

Developing, Maintaining, and Renewing Items in an Item Bank

- How to best calibrate existing items?
  - Model selection
  - Whose item parameters to use?
  - Standardization? Generic vs. disease-specific
- Item parameter drift
  - Anchor or Re-calibrate?
- How to write and best test new items?

Research Needed for Effective CAT implementation

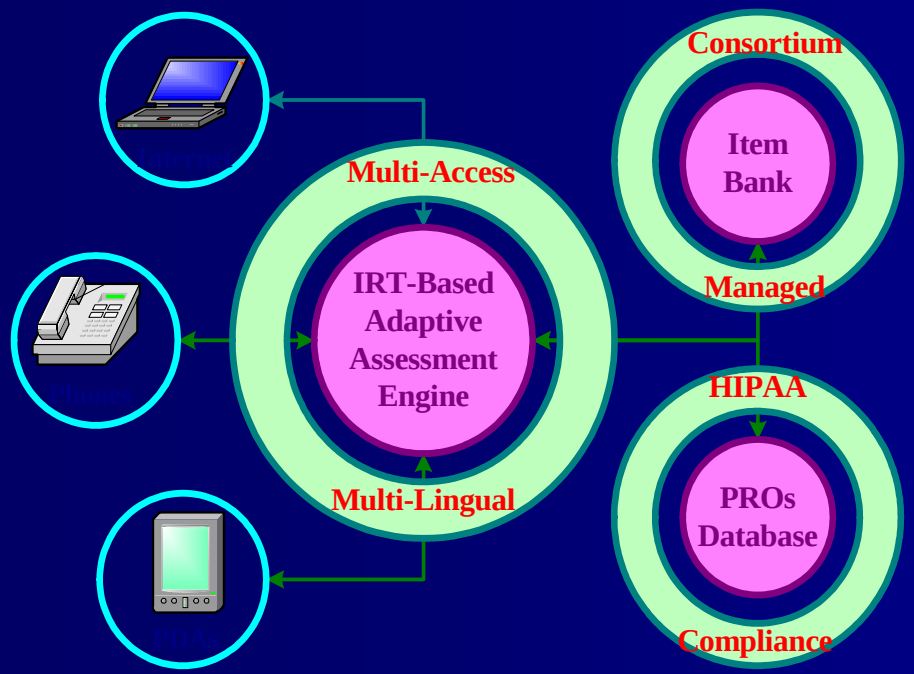
- Item production
- Item statistics
- Maintaining a valid bank of items for test construction
- Item exposure
- Cost-benefit considerations
- Fairness
- Behavior of examinees under different types of test administration conditions
- Delivery options

Current Patient-Reported Outcomes Measurement

- *Mostly lack of ...*
  - –ubiquitous accessibility
  - –multi-language support
  - –clinically-relevant PROs questions
  - –adaptive assessment capability
  - –adequate privacy protection

The PROsIT System

Patient-Reported Outcomes with Innovative Technologies
(R43 MH#####)

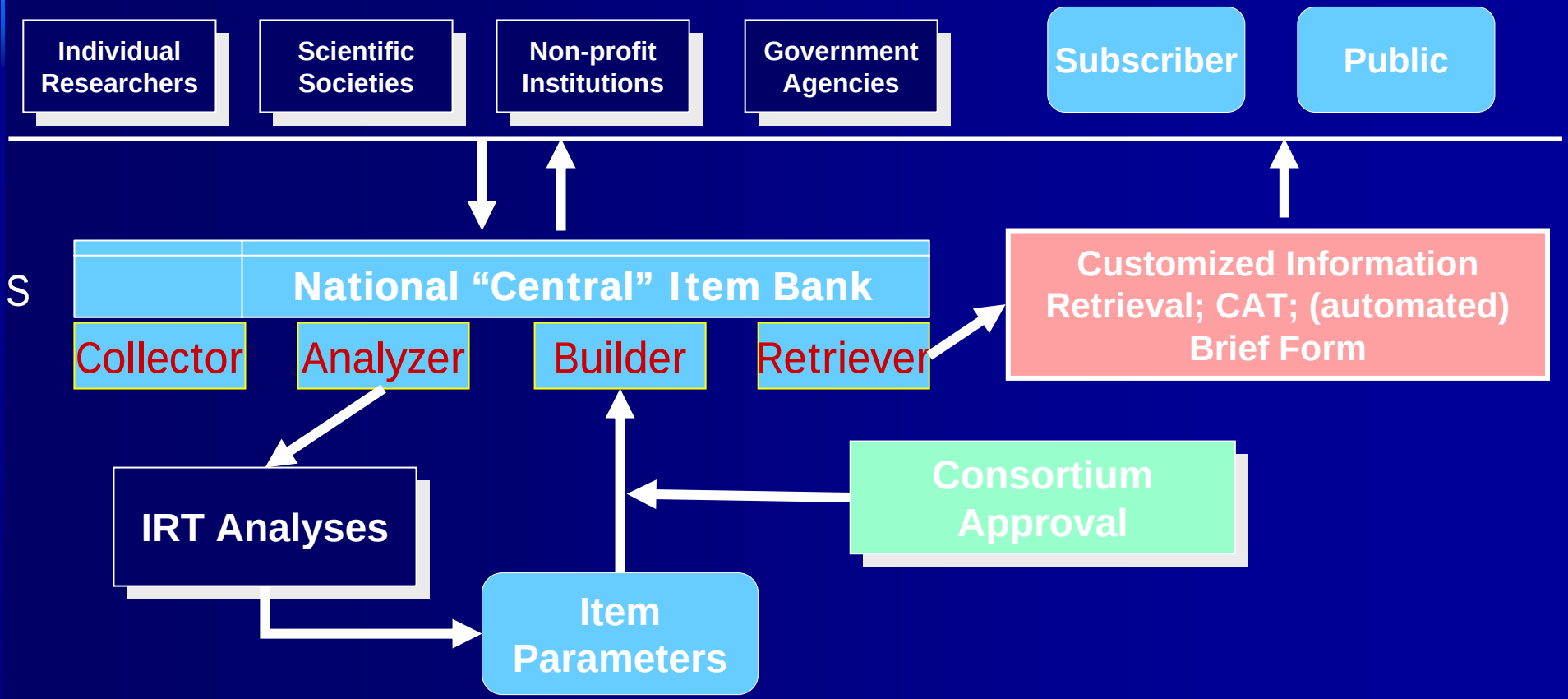


- Multi-platform
- Multi-lingual
- Item response theory based
- Item banks
- Computerized adaptive testing

Chang, C.-H., & Yang, D. (2003, April 15). Patient-Reported Outcomes Information Technology: The PROsIT™ System. *ISPOR CONNECTIONS, 9(2)*, 5-6.

Slide prepared by Chih-Hung Chang, Ph.D.

Infrastructure of a *National* "Central" Item Bank (NCIB)



PROMIS
SCC

Slide prepared by Chih-Hung Chang, Ph.D.

Contact Information

Chih-Hung Chang, Ph.D.
Buehler Center on Aging
Feinberg School of Medicine
Northwestern University
750 N. Lake Shore Drive, Suite 601
Chicago, IL 60611
Tel: (312) 503-4354
Fax: (312) 503-5868
E-mail: chchang@northwestern.edu