Computer Adaptive Testing and the Patient-Reported Outcomes Measurement Information System (PROMIS)*

Kitty S. Chan, PhD
Associate Professor
Department of Health Policy and Management
Johns Hopkins Bloomberg School of Public Health

November 1, 2011

*Acknowledgement: Many graphic slides were taken or adapted from a lecture for a similar course by Bryce B. Reeve, PhD, then Psychometrician and Program Director, Outcomes Research Branch, National Cancer Institute.
What is CAT?

CAT integrates the power of item response theory measurement framework and computer technology to administer a patient reported outcome (PRO) measure that select a question to administer based on a person’s response to previously administered questions.
What is CAT?

Hypothetical Example: Nutrition during Infancy

- What is the minimum daily nutrition requirements for your baby?
- What vitamins does your baby need?
- What are benefits of breastfeeding?
- What are the signs that your baby is ready to feed him/herself?
- When should you introduce solid foods?
- What are the signs that your baby is hungry?
What are CAT’s advantages

- Provide an accurate estimate of a person’s score with the minimal number of questions.
  - Questions are selected to match the health status of the respondent.

- CAT minimizes floor and ceiling effects.
  - People near the top or bottom of a scale will receive items that are designed to assess their health status.

- Before you have a CAT, you need an item bank
What is an “Item Bank”

- A large collection of items measuring a single domain

- The items have been evaluated and tested to ensure their relevance, clarity, and psychometric robustness
  - Items are selected to maximize precision and retain clinical relevance
  - Items in the same bank are linked on a common metric
How Do You Link Different Measures?

Different Linking Designs

- One group take two (or more) tests
- Two different but equivalent groups take two (or more tests)
- Tests given to two different groups, with common items internal or external to tests*
- Two groups take different tests, but common group of individuals take both tests

* More commonly used due to feasibility
The Idea Behind IRT Score Calibration and Linking

- Use one of the linking designs to bridge measures
- When bridged by common group or common (or anchor) items
  - Parameters for “new” or “different” items are linked on same scale
- Find any differentially functioning items
- With standard item parameters and after modeling differences in item functioning, scores should be calibrated
Graphic representation of Common Items Linking Design
How does CAT work?
In the past 7 days, I felt depressed.

Depressive Symptoms

Item Bank
(Validated & IRT-Calibrated Depression Items)
In the past 7 days, I felt depressed.

Depressive Symptoms

Item Bank
(Validated & IRT-Calibrated Depression Items)
In the past 7 days, I felt helpless.
In the past 7 days, I felt that nothing could cheer me up.

Depressive Symptoms

Item Bank
(Validated & IRT-Calibrated Depression Items)
NIH Roadmap Initiative: PROMIS

Patient-Reported Outcomes Measurement Information System

http://www.nihpromis.org
PROMIS: Goals

NIH Roadmap Initiatives: Re-Engineering the Clinical Research Enterprise

Improve assessment of self-reported symptoms and other health-related quality of life domains across many chronic diseases.

Advance the science and technology to:
- Facilitate the collection of standardized patient-reports of their health, functioning and well-being and
- Integrate their responses to inform decision-making in research and healthcare delivery
PROMIS accomplishes these goals by...

- Developing **Item Banks** for Patient Reported Outcome Domains

- Once Calibrated, items from item bank can be used to
  - Develop CAT for the domain
  - Develop short forms measures of domain
  - Select a pre-existing short form measure of domain

- Creating online gateway to these item banks, CATs and measure to enable their use for clinical research and practice ("Assessment Center")
PROMIS: Structure

PROMIS Domains for Item Banking
- **Core**: pain, fatigue, depression, anxiety, anger, physical function, social function, and overall general health
- **Additional**: sleep/wake function, cognitive function, sex functioning, illness impact
- Pediatric PROMIS

Continued Development and Validation

Neuro-QoL, a related NIH supported resource
PROMIS Item Banks

PROMIS Current

Physical Health

- Pain Behavior
- Pain Interference
- Fatigue
- Pain Intensity
- Physical Function
- Sleep Disturbance
- Sleep-Related Impairment
- Sexual Function

Adult

Physical Health Banks

- Pain Interference
- Fatigue
- Upper Extremity Function
- Mobility
- Asthma Impact

Pediatric
What Does PROMIS Measure?
PROMIS Item Banks

PROMIS Current Mental Health Banks

Mental Health

Adult
- Anxiety
- Depression
- Anger
- Illness Impact Negative
- Illness Impact Positive
- Applied Cognition—Concerns
- Applied Cognition—Ablities

Pediatric
- Anxiety
- Depression
- Anger
PROMIS Item Banks

PROMIS Current Social Health Banks

Social Health

Adult
- Ability to Participate in Roles & Activities
- Satisfaction With Roles & Activities
- Companionship
- Emotional Support
- Informational Support
- Instrumental Support
- Social Isolation

Pediatric
- Peer Relationships
PROMIS: Overall Picture
Neuro-QoL Framework (NINDS)
Psycho-metric Testing

Item Bank (IRT-calibrated items reviewed for reliability, validity, and sensitivity)

Short Form Instruments

New Items

Item Pool

Content Expert Review

Focus Groups

Cognitive Testing

Secondary Data Analysis

Questionnaire administered to large representative sample

Probabilty of Response

Theta

Information

Theta

CAT

Items from Instrument A

Items from Instrument B

Items from Instrument C

Items from Instrument A

Items from Instrument B

Items from Instrument C

New Items

Secondary Data Analysis

Cognitive Testing

Focus Groups

Content Expert Review

Questionnaire

Psycho-metric Testing

Item Bank

Short Form Instruments
Develop short forms from PROMIS Item Banks

Advantages:

– Select a set of items that are matched to the severity level of the target population.

– All scales built from the same item bank are linked on a similar metric.
PROMIS CAT Outperforms Legacy Questionnaires

4-item SF36/Vitality
4-item CAT
13-item FACIT-Fatigue
13-item CAT

Precision ↓

US General Population mean
Multiple Delivery Platforms

Dynamic Assessments

- Personal Interview
- Telephone Interview
- Self Administered
- Interactive Voice Response
- Internet Administered
- Hand-held Device
- TV

Source: John Ware (October 2001) talk to NIH
PROMIS Assessment Center

Goal: To enable administration of item banks of standardized patient-reported outcomes measures for use in clinical research, population surveillance, and clinical practice.
Assessment Center Features

- an online, dynamic application that will allow researchers to centralize all research activities
- includes features that promote instrument development, study administration, data management, and storage of statistical analysis results
- houses a library of instruments and items with an emphasis on health-related quality of life
Language Notes

Available Translations

- Most banks available in Spanish
- Some translations in other languages in progress
  - Chinese (Mainland, simplified)
  - Portuguese

See full set of available and in progress translations

http://www.nihpromis.org/measures/translations
Let’s Try it Out

http://www.assessmentcenter.net/ac1/
Thank you for your attention!

Contact Information:

kchan@jhsph.edu

Hampton House 633