

Developing Tailored Instruments: Item Banking and Computerized Adaptive Assessment

Jakob B. Bjorner

QualityMetric Incorporated, Lincoln, RI
Health Assessment Lab, Waltham, MA

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

Three Steps to CAT

Examples from PRO

1. Item Bank Development

Fisher 1997, McHorney & Cohen 2000, Ware et al 2000, Gardner et al 2002, McHorney 2002, Bjorner et al 2003, Bode et al 2003, Lai et al 2003

2. CAT Simulations

Ware et al 2000, Gardner et al 2002, Bjorner et al 2003, Fliege et al 2003, Rose et al 2003, Ware et al 2003

3. Functioning CAT

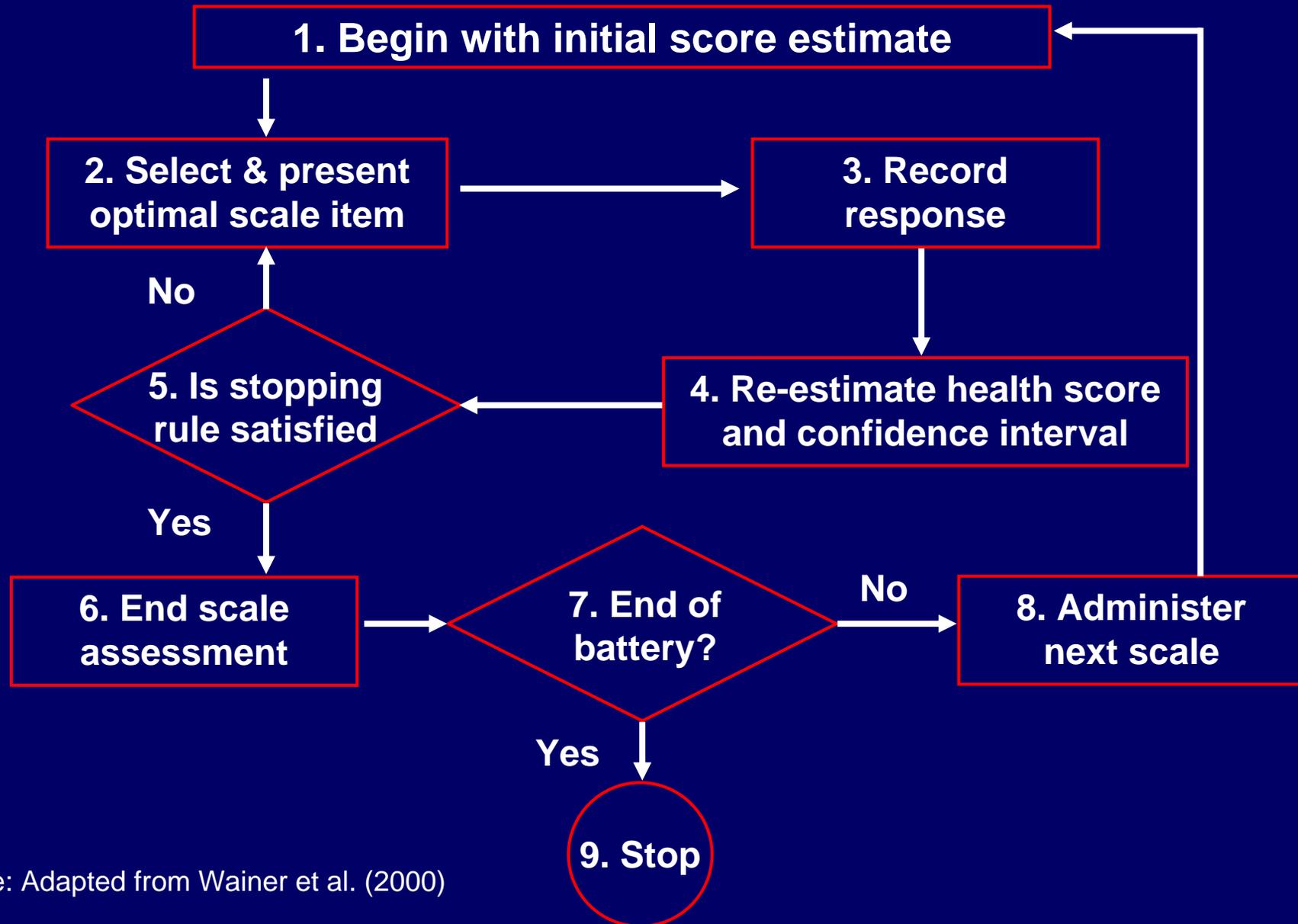
Ware et al 2000, Rose et al 2003, Ware et al 2003

CAT at QualityMetric

The Dynamic Health Assessment (DYNHA[®]) System

- **Eight Generic concepts**
 - Physical function, role function (physical), bodily pain, general health, vitality, social function, role function (emotional), mental health
- **Two disease-specific concepts**
 - Headache
 - Osteoarthritis

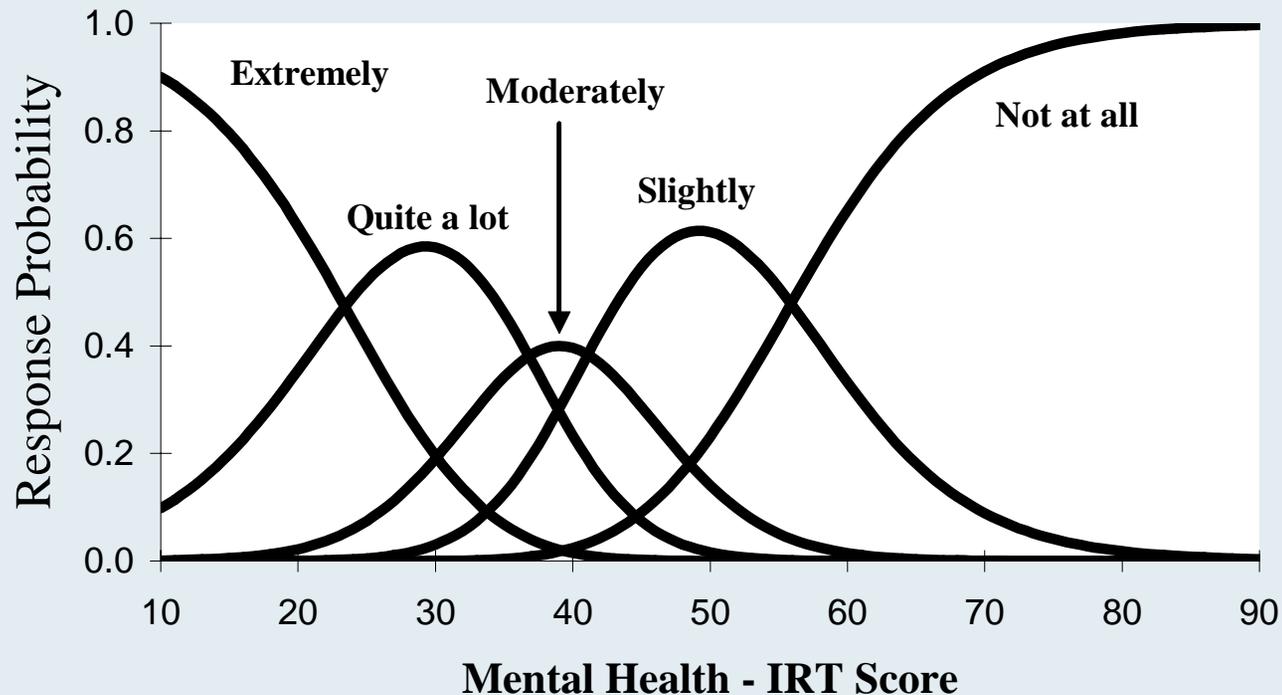
Logic of Dynamic Health Assessments



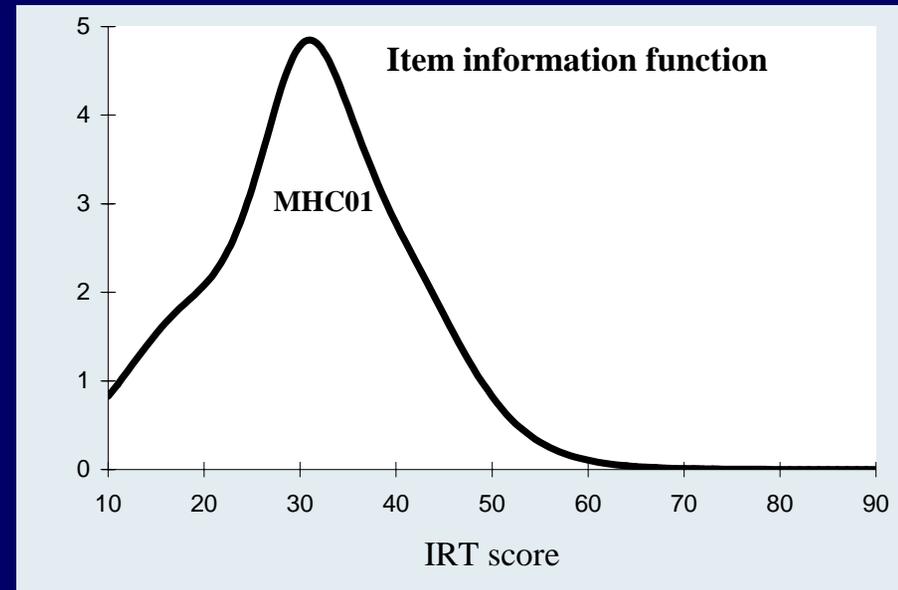
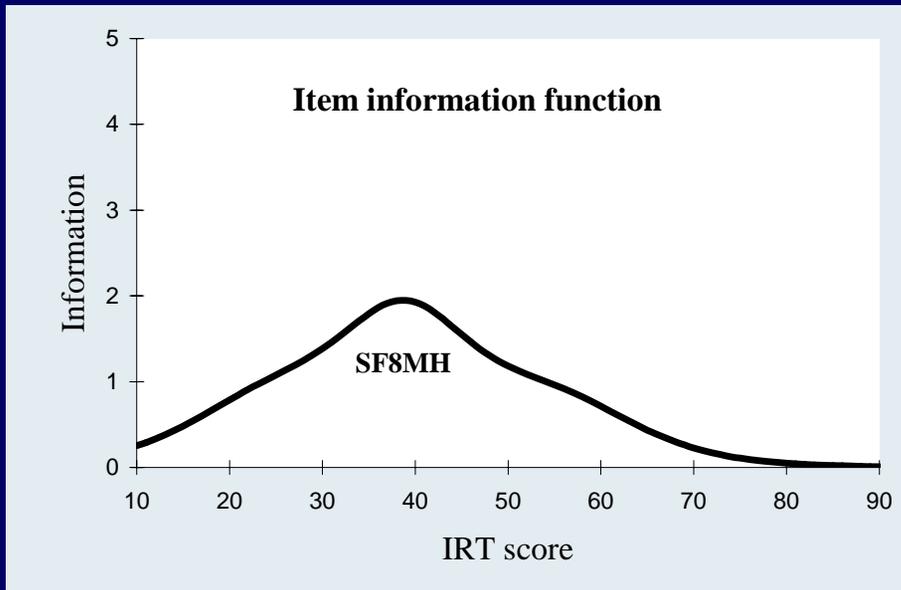
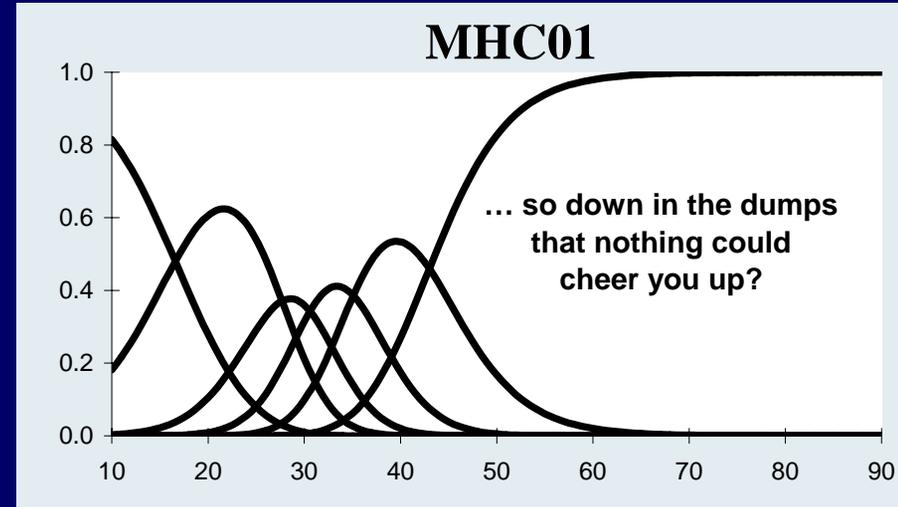
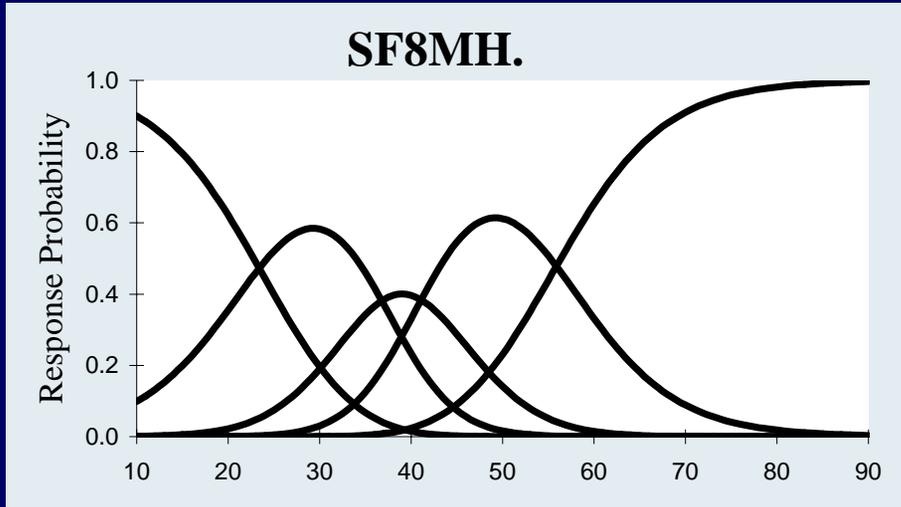
Source: Adapted from Wainer et al. (2000)

IRT Model

SF8MH. During the past 4 weeks, how much have you been bothered by emotional problems (such as feeling anxious, depressed or irritable)?

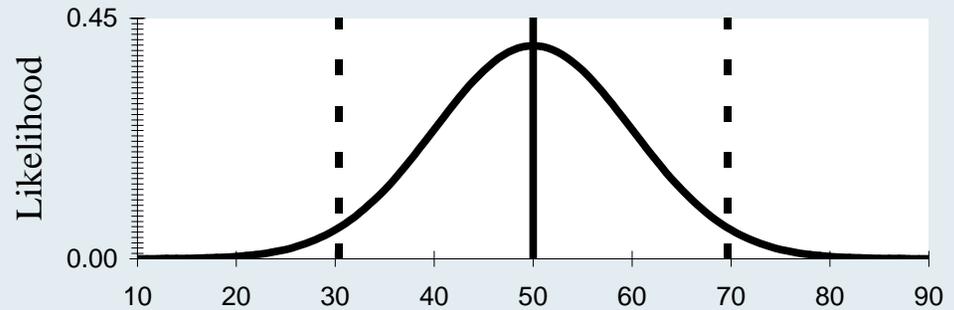


Item Information

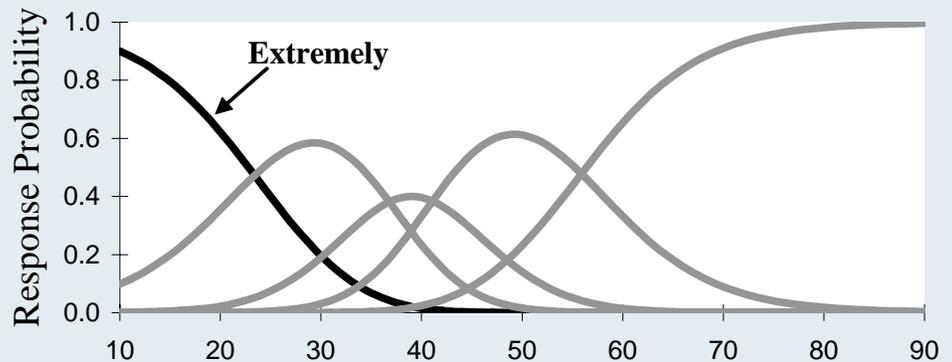


IRT Scoring in a CAT (EAP Approach)

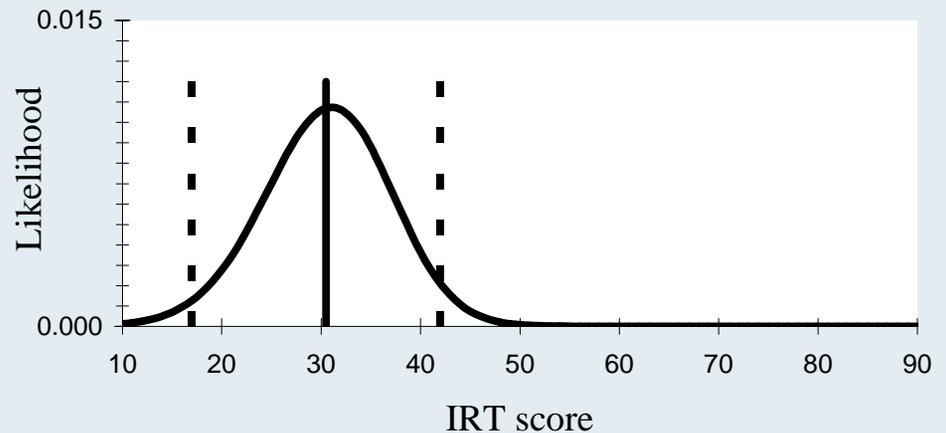
Prior distribution



SF8MH

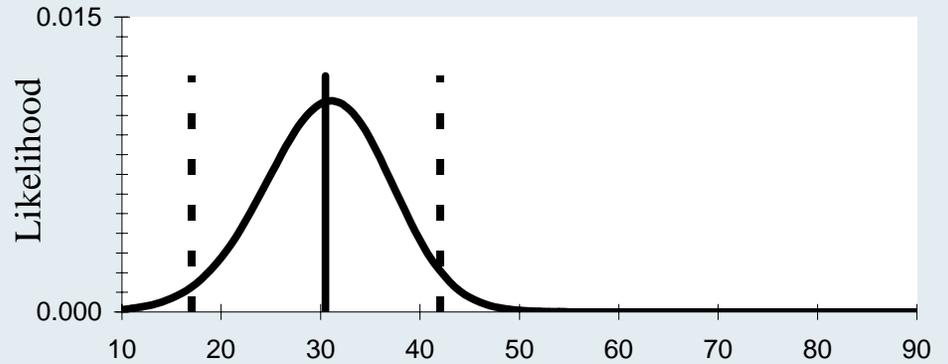


Posterior distribution

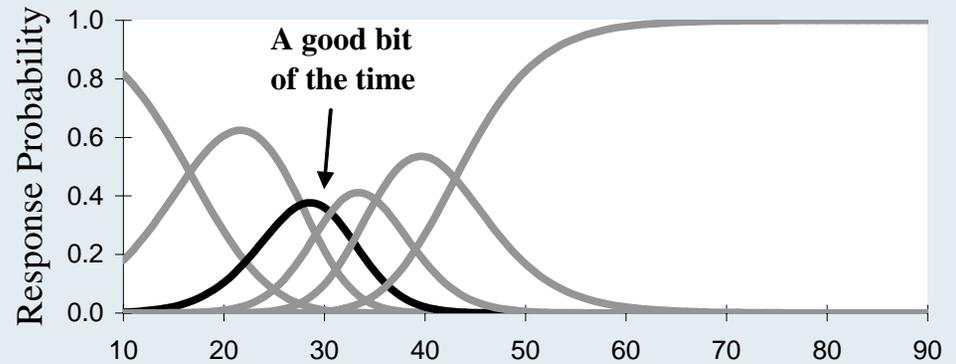


IRT Scoring in a CAT (EAP Approach)

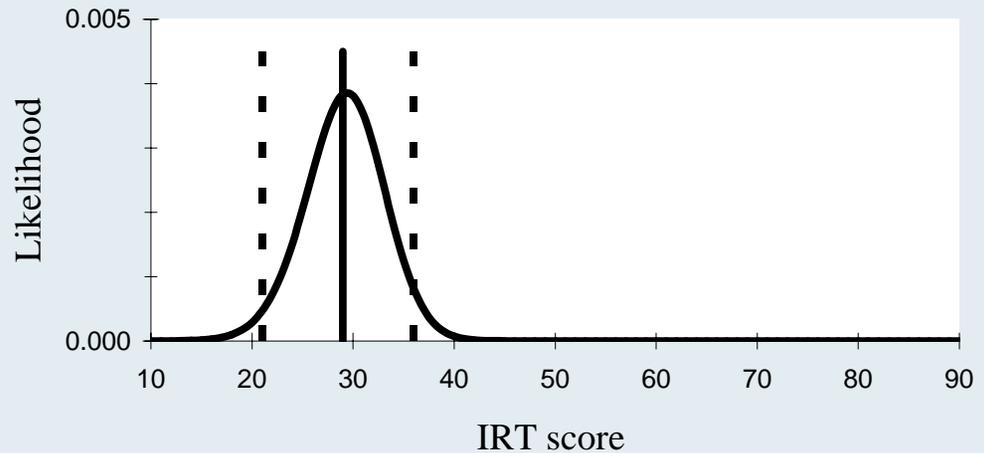
Posterior distribution
↓
Prior Distribution



MHC01



Posterior distribution



Item Bank Development

- Define and delimit construct
- Develop items and/or include items from existing instruments
- Collect data
- Analyze data
 - Estimate item parameters
 - Check model fit
- Set the metric
- (Perform CAT simulations to evaluate item selection rules and stopping rules)

Example: the SF Mental Health Item Bank

- **Mental health item bank V1**
 - Items from the Mental Health Inventory (MHI)
 - Data from the Medical Outcomes Study
- **Mental health item bank V2**
 - Additional items from standard mental health instruments
 - New data collected from the US general population

MHI Subdomains

- **Anxiety**
- **Depression**
- **Behavioral/emotional control**
- **Positive well-being**
- **Loneliness/Belonging**

Check of the IRT Model 1: Dimensionality and Local Independence

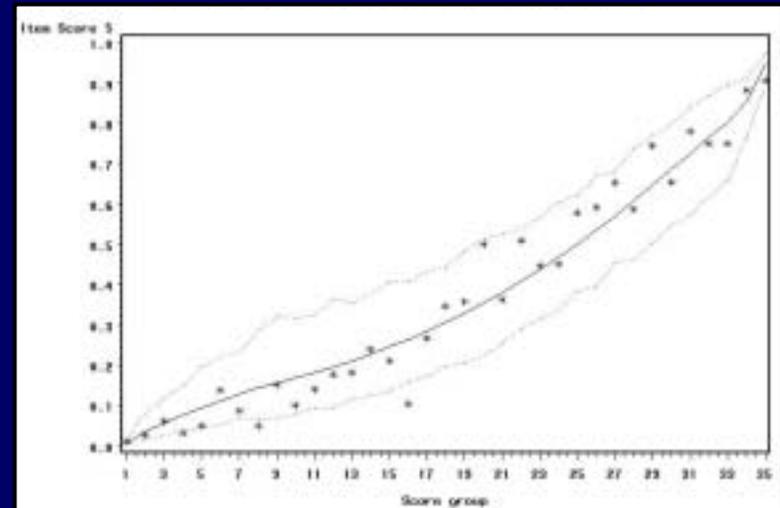
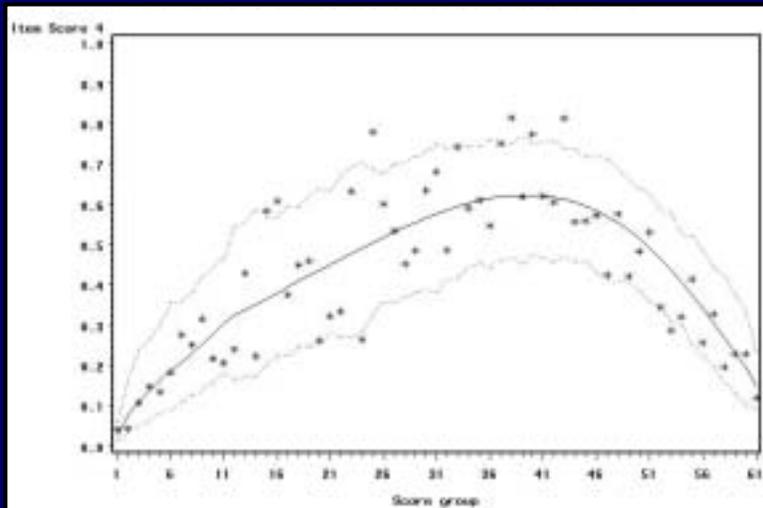
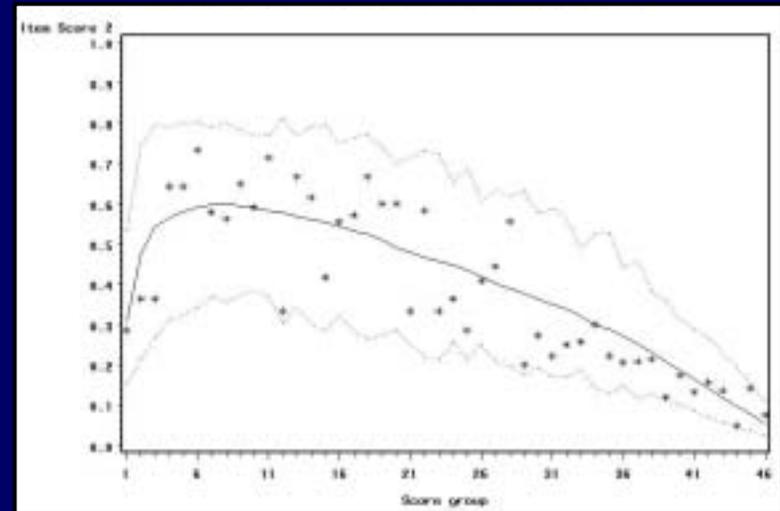
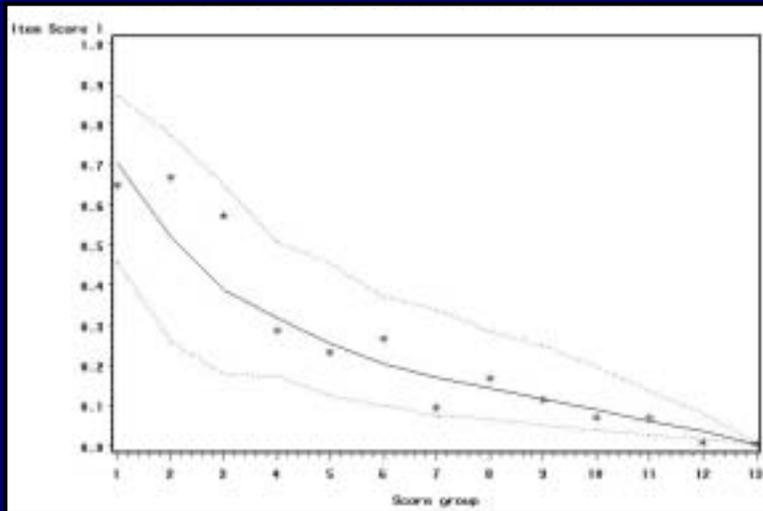
- **Used factor analysis for categorical data**
 - **Eigenvalue and confirmatory factor analyses**
- **Performed global and multigroup analyses**
- **The first factor explained 66% of the variance**
- **All items loaded strongly on first factor**
- **Correlations between factors based on the 5 subdomains was high**
 - **.81-.96 between Anxiety, Depression, Control, and Well-being**
 - **.71-.86 between Loneliness and other factors**
- **High residual correlations on three items on loneliness**

Decision to fit a unidimensional model, but exclude three loneliness items

Check of the IRT Model 2: Analyses of ICCs (also called Category Response Functions)

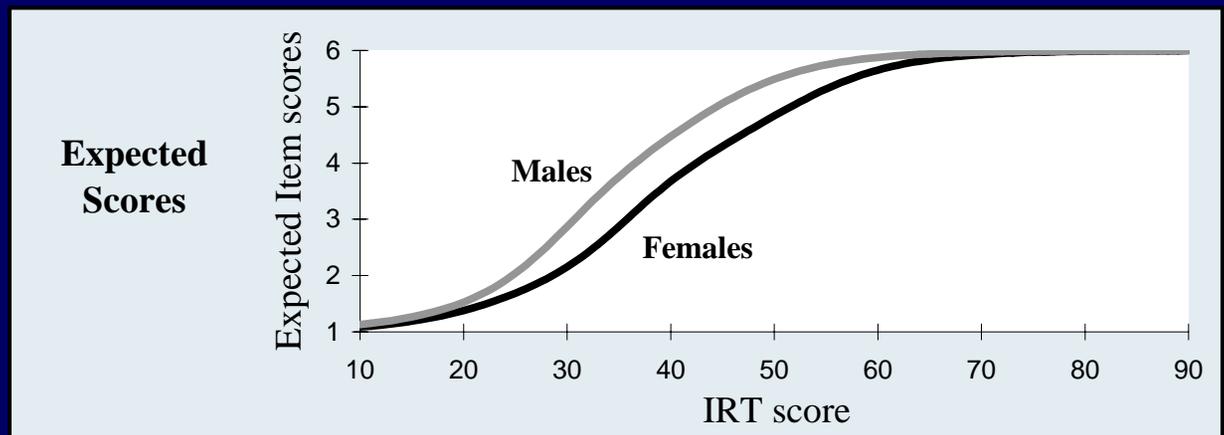
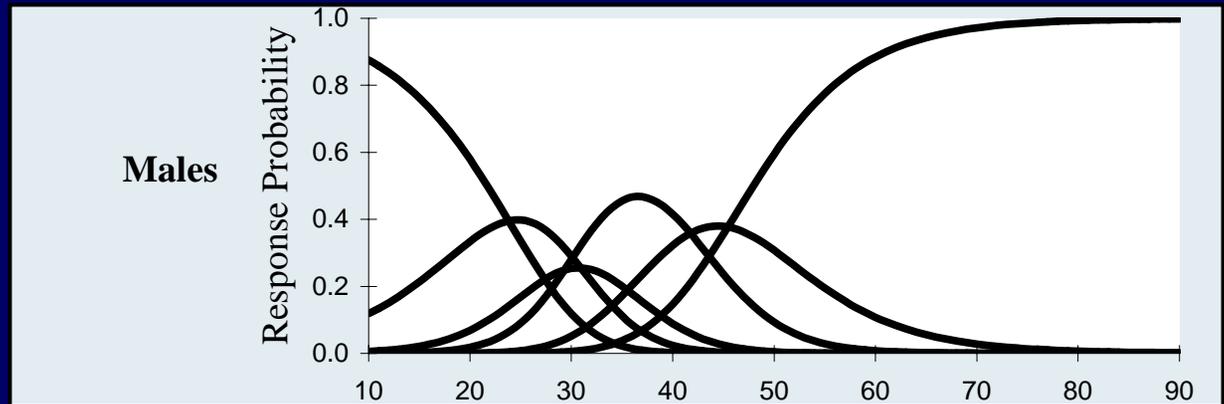
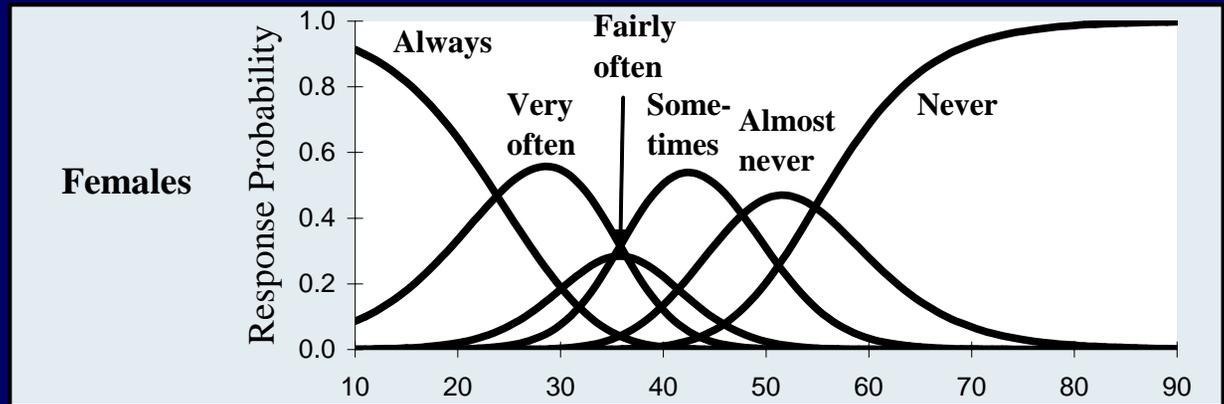
- Initial visual inspection using non-parametric IRT (Kernel Smoothing)
- Used the Generalized Partial Credit Model
- Evaluated fit through a method suggested by Orlando and Thissen
- Some item categories had to be collapsed to allow estimation
- The fit of the items seemed acceptable

Example of ICC Fit Analyses: SF8MH



How often have you felt like crying during the past month?

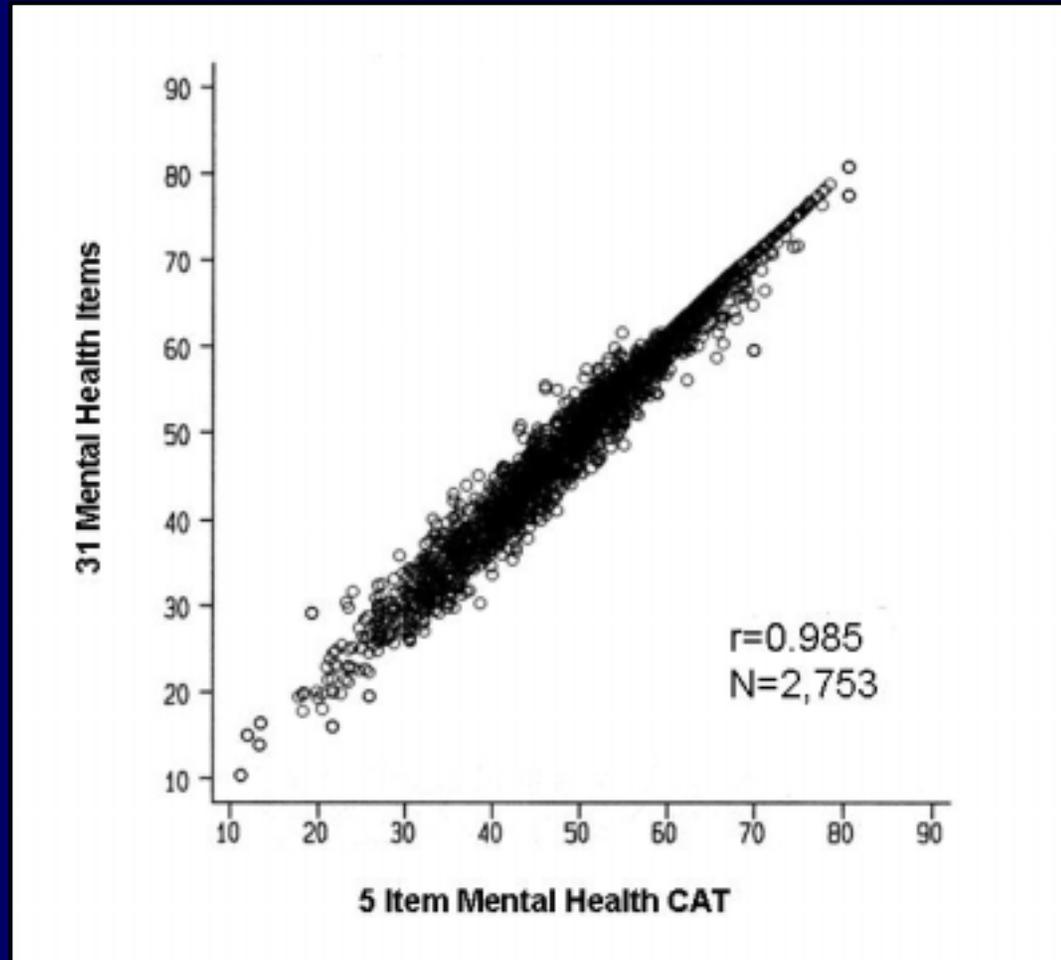
Check of the IRT Model 3: Test of DIF



Setting the Metric

- **Five items from the bank**
- **Sample from the US adult general population**
- **Define metric so:**
 - **mean score of the sample is 50**
 - **standard deviation is 10**
- **Develop benchmarks for interpretation**
- **Develop cut-off scores for possible clinical condition (e.g. depression)**
- **Procedures for calibrating the IRT scores to the metric of traditional scales**

Example of CAT Simulations



Final Mental Health CAT

- **Item selection rules**
 - Maximum information
 - Aim for representation of all 4 subdomains (Anxiety, Depression, Control, Well-being)
- **Stopping rules**
(can vary based on application)
 - Depend on score:
 - high precision required for people with poor mental health
 - For scores around clinical cut point:
 - assessment is continued
- **Most assessments use 3-7 items**

Considerations for a CAT in Practical Use

- **Integration into the flow of research/clinical practice**
- **Integration into existing systems**
- **Usability for the patient**
 - Visual appearance
 - Ease of use
- **Data security and back-up**
- **Quality control**
- **Feedback to respondents/clinicians**
- **Have fixed short form available as alternative**

Differences Between Patient Reported Outcomes and Educational Tests

1. Types of items

- Mostly rank scale, rarely Yes/No
- Visual Analogue Scales or Counts

2. Item exposure

- Do we need to keep items secret?

3. Generation of items

- Items can stay in the pool
- How many ways can you ask questions on pain?

Conclusions

- **Item banking:**
 - Done successfully by several research teams
 - Statistical/psychometric machinery solid, but can be further developed for PROs
 - Linking of banks from different research groups would be valuable
- **CAT:**
 - Five years of CAT application in practice have demonstrated feasibility
 - Integration into practical daily routines is crucial and achievable
 - CAT is ready for primetime in medical research and clinical practice