Multidimensional computer adaptive testing with PROMIS item banks.

Michael Bass, MS, Scott Morris, PhD, Richard Neapolitan, PhD

PROMIS

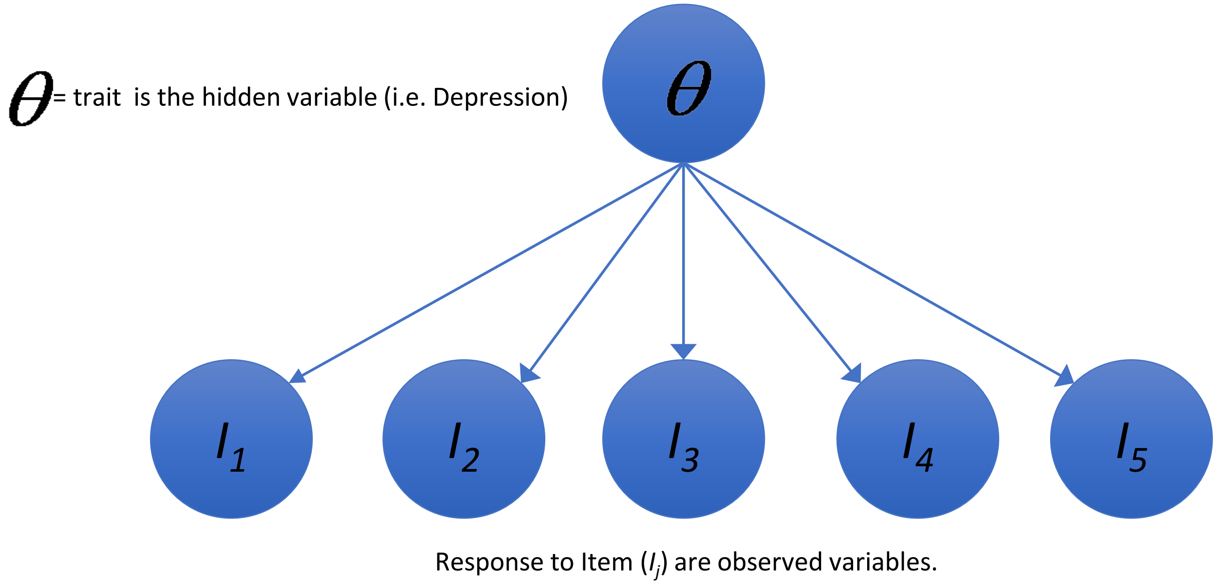
HealthMeasures

TRANSFORMING HOW HEALTH IS MEASURED

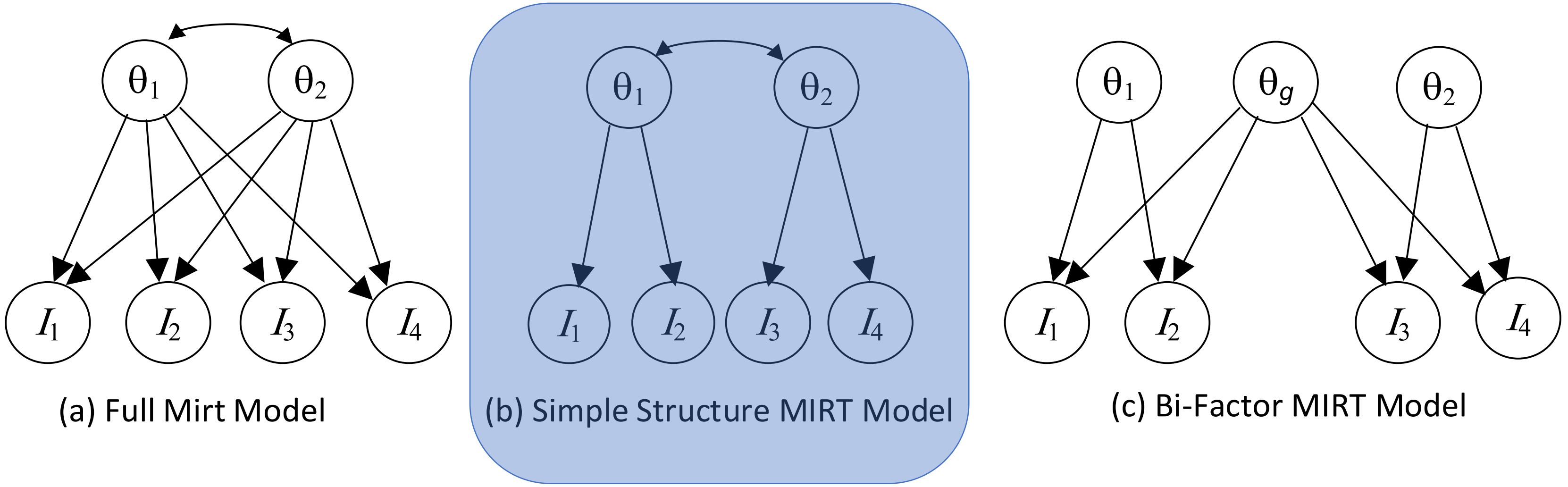
Disclosure

I disclose that neither I nor my partner have relevant financial relationships with commercial interests.

Unidimensional Structural Model for PROMIS Banks



Available Multidimensional IRT models



PROMIS

HealthMeasures
TRANSFORMING HOW HEALTH IS MEASURED

Calibrating PROMIS Items

- Wave 1 Sample
- Markov Chain Monte Carlo methods
- IRTPRO

Model Comparison

	# Prm	BIC	Relative Full MIRT
Full MIRT	556	213713.6	
Simple Structure	398	213751.6	38.03
Single Factor	395	217868.0	4154.4
Bi-Factor	474	212713.4	-1000.17

2 Parameter Logistic Model (2PL)

- What is the probability of a test-taker with the ability θ to correctly respond to item $_j$ category $_k$.

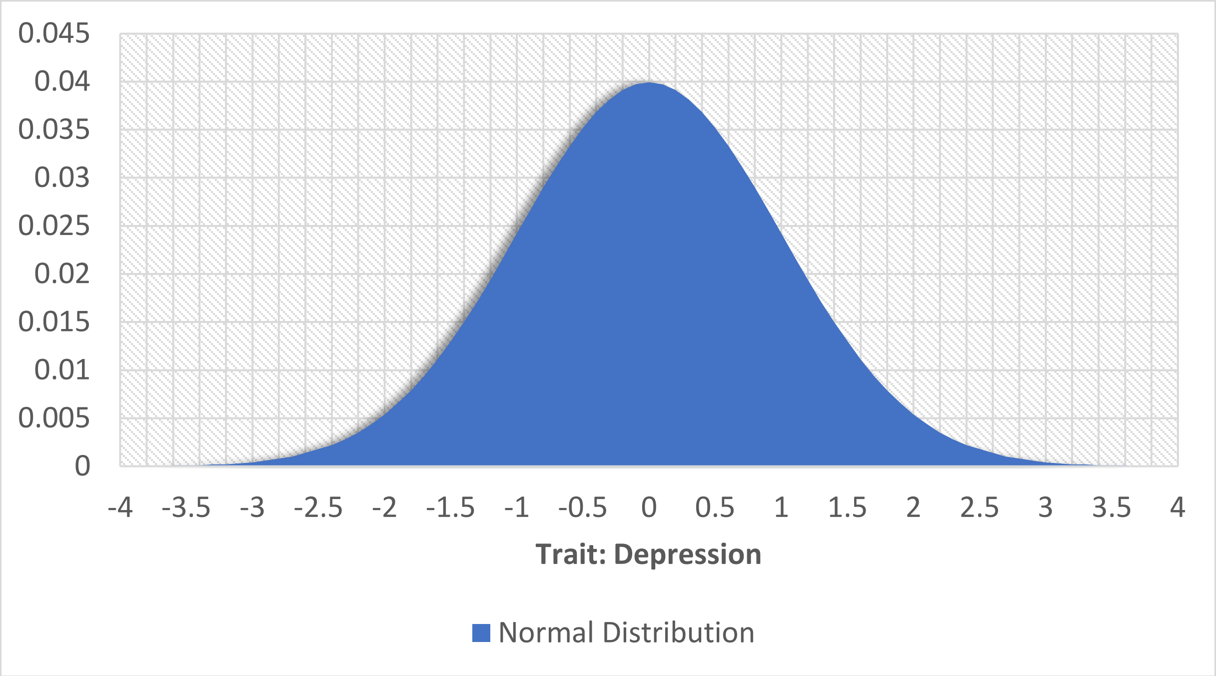
- Unidimensional $$P^*_{\theta jk} = \frac{1}{1+\exp\left[-Da_j(\theta - b_{jk})\right]}$$

- Multidimensional $$P^*_{\theta jk} = \frac{1}{1+\exp\left[-Da_j(\theta - b_{jk})\right]}$$

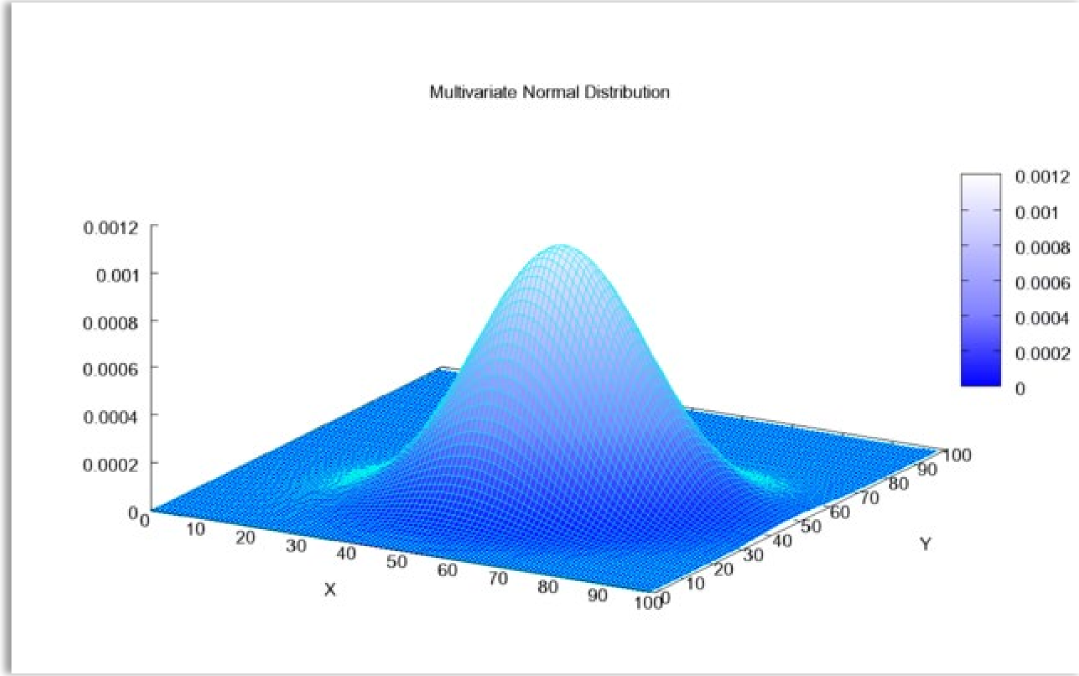
$$\begin{bmatrix} a_j^x & a_j^y & a_j^n \end{bmatrix}$$

$$\begin{bmatrix} \Theta^x - b_{jk} \\ \Theta^y - b_{jk} \\ \Theta^n - b_{jk} \end{bmatrix}$$

Initial Distribution



Unidimensional



Multidimensional

Approach - Simulations

- Monte Carlo simulation designed
- 2 sets of calibrations
  - 1 for generating responses
  - 1 for administering CAT
- True Trait Levels
  - uniformly across trait continuum
  - empirical distribution
    - Clinical and General population
- Administer all items in the banks
- Retrospectively apply stopping rule and outcome criteria

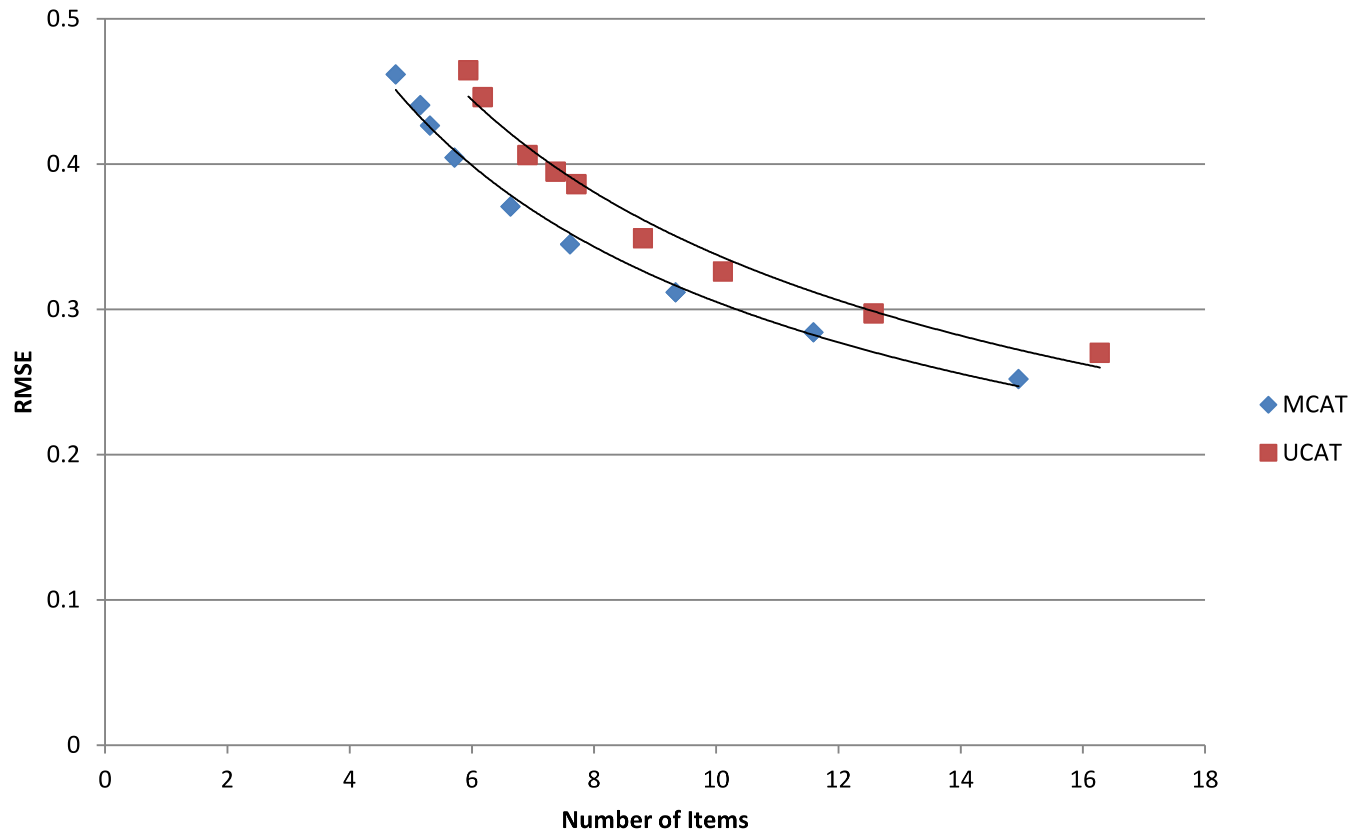
Simulation Outcome Criteria

- Item count to reach the specified standard error (SE)
- Root Mean Square (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left[\left(\hat{\theta}_{Anx,i} - \theta_{Anx,i}\right)^2 + \left(\hat{\theta}_{Dep,i} - \theta_{Dep,i}\right)^2\right]}{N}}$$

where $\hat{\theta}_{t,i}$ and $\theta_{t,i}$ represent the EAP estimate and true value of trait *t* for respondent *i*, and *N* is the number of simulated respondents.

RMSE vs Item count



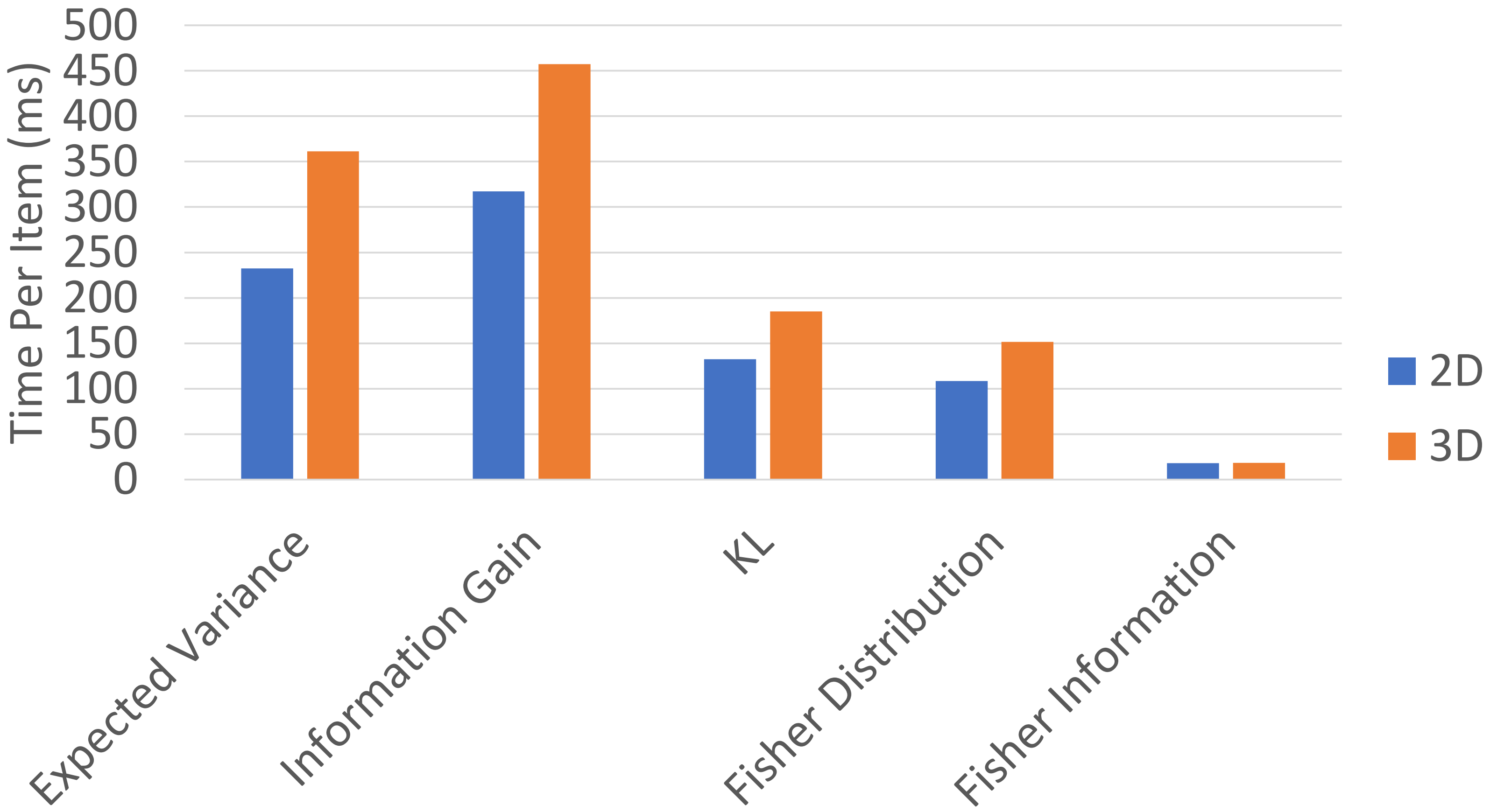
PROMIS

HealthMeasures
TRANSFORMING HOW HEALTH IS MEASURED

Item Selection - Methods

- Maximize Determinant of *Fisher Information* Matrix
- Maximize Posterior-Weighed Average of FI (*Fisher Distribution*)
- Minimize Sum of Posterior Variance (*Expected Variance*)
- Maximize Kullback-Leibler Information (*KL*)
- Maximize Change in Shannon Entropy (*Information Gain*)

Item Selection - Computation Time



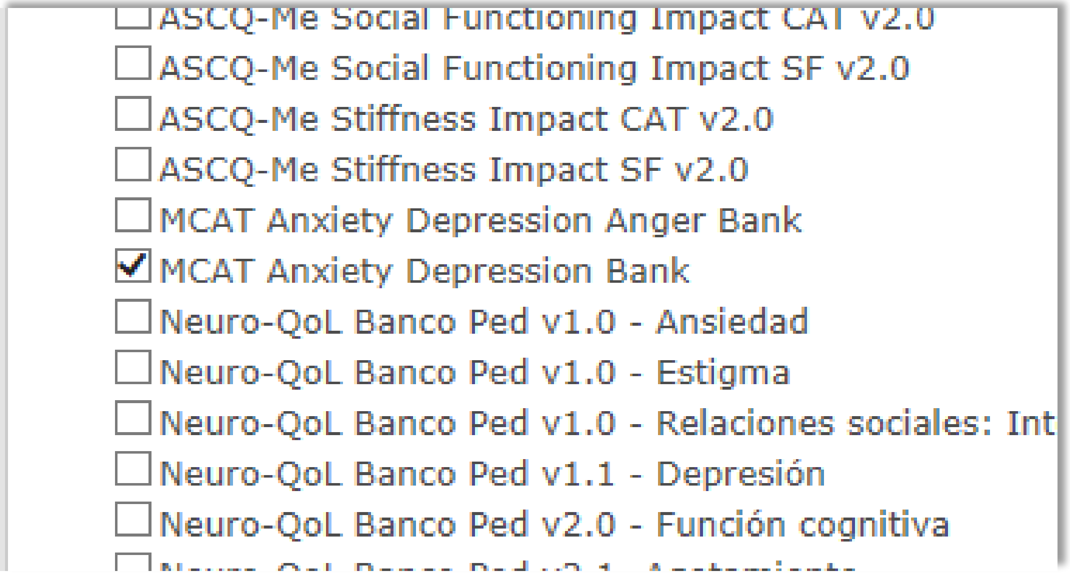
Item Selection - Conclusions

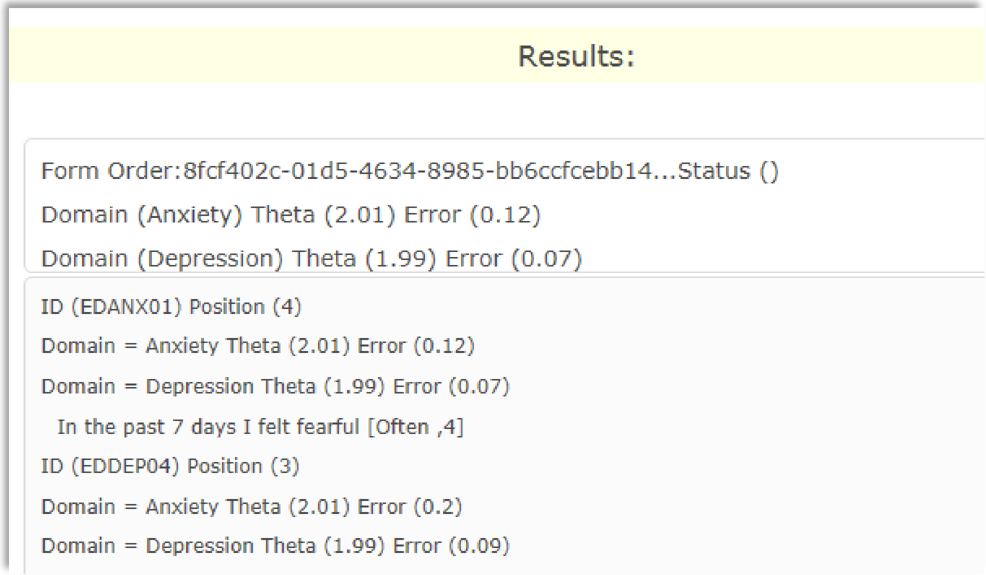
- All item selection methods perform similarly
  - Never more than one additional item (on average)
- Small but consistent differences
  - Fisher Information slightly less efficient
  - KL Information among most efficient
- Small improvements in precision may not justify substantial increase in computational complexity (time to select next item)
  - Likely to become worse with higher dimensionality
- KL Information provided good balance of high efficiency with only modest increase in computation time

Stopping rules

- All variations based on a predictive standard error reduction approach
  - Sum of Variance method
  - Sum of SD method
  - Sum of Variance Above Threshold
- All methods demonstrated gains in efficiency, but
- MCAT more sensitive than UCAT when there is poor item-person mapping

Operationalized MCAT in AC API





PROMIS

HealthMeasures
TRANSFORMING HOW HEALTH IS MEASURED

Performance Consideration

Domains	State Size (kb)	DB query (tick intervals)
1	64	9
2	512	1011
3	8192	48775

**Need a mechanism to keep state in memory

Conclusions

- Multidimensionality complicates stopping rules more than item selections methods
- MCAT is more sensitive to item-person mapping. How well do you know your cohort.
- In general, modest efficiency gains but need to consider number of dimensions and the impact on computational performance (i.e. lag)

PROMIS

HealthMeasures
TRANSFORMING HOW HEALTH IS MEASURED

Future direction

- On the fly MCAT
  - Domain covariances with unidimensional calibrations
- Stopping rules from educational testing
- Bifactor model (high level of dimensions simplify to 2D computations)

Acknowledgment

- Research reported in this publication was supported in part by the National Library of Medicine of the National Institutes of Health under Award Number R01LM011962.

PROMIS

HealthMeasures
TRANSFORMING HOW HEALTH IS MEASURED

Questions

Thank you!