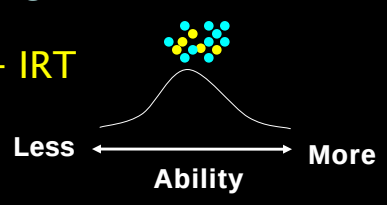
When Do Measurement Models Produce Diagnostic Information?

Tzur M. Karelitz
PACE Center, Tufts University

Jimmy de la Torre
Rutgers, The State University of New Jersey

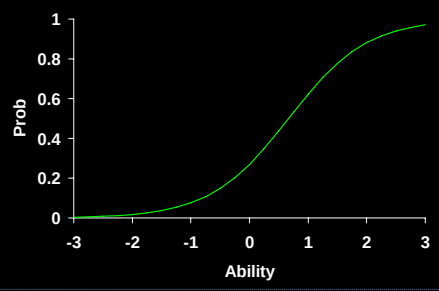
Frameworks for Designing & Analyzing Assessments

- Continuous Framework - IRT
  - Single ability continuum
  - Ordered values
  - Conjoint scaling



- Discrete Framework - CDM
  - Multiple binary skills
  - Class membership
  - Fixed item structure
    - linked to theory

Measurement Models



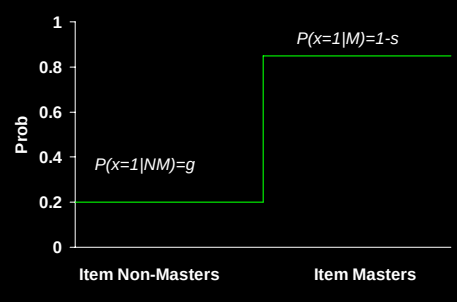
IRT Framework
2PL
(Birenbaum, 1968)

$$P(x_{ij}=1)=\frac{1}{1+\exp\{-a_j(\theta_i-b_j)\}}$$

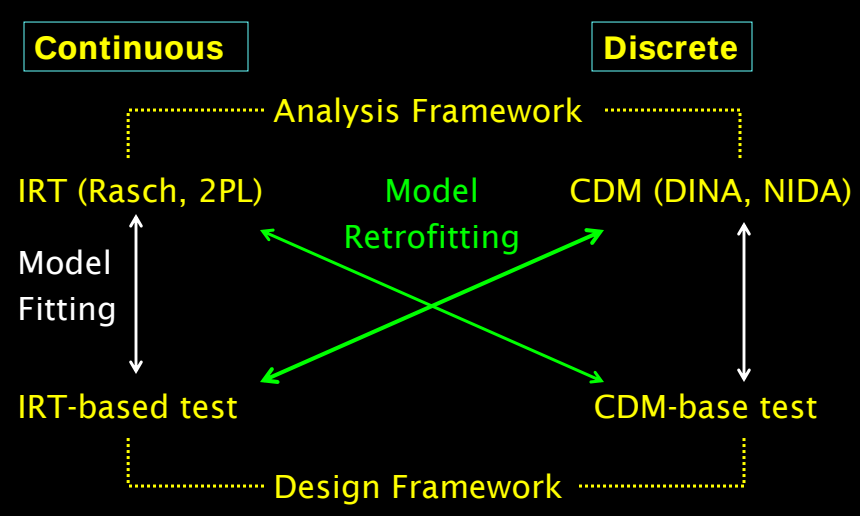
CDM Framework
DINA
Junker & Sijtsma, 2001

$$P(x_{ij}=1)=(1-s_j)^{\eta_{ij}}(g_j)^{(1-\eta_{ij})}$$

$$\eta_{ij}=\prod_{k=1}^{K}\alpha_{ik}^{Q_{jk}}$$



Model Fitting and Retrofitting



The Retrofitting Assumption

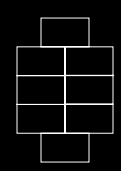
Models can extract diagnostic information irrespective of the framework used in constructing the assessment.

Diagnosticity

The extent to which the underlying structure of the data can be represented discretely.

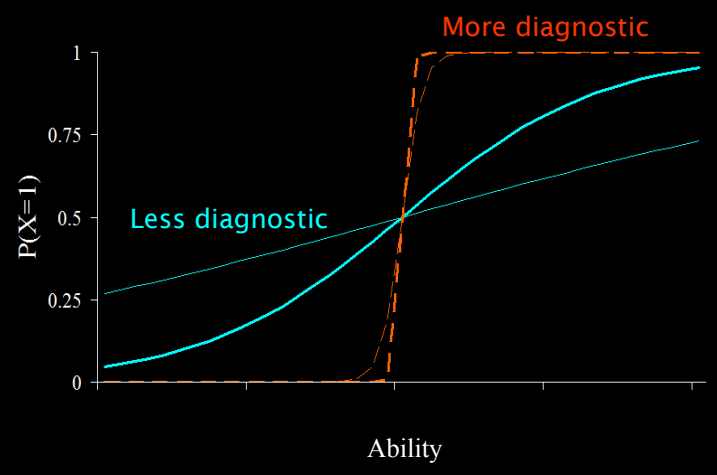
Less structure
More error

More structure
Less error



Less ← Diagnosticity → More

Manipulating diagnosticity



Study Goal

Evaluate the validity of the retrofitting assumption under different levels of diagnosticity.

1. Generate comparable data from IRT & CDM frameworks
2. Manipulate test diagnosticity
3. Fit and retrofit data with IRT & CDM models
4. Evaluate Model fit and Classification accuracy

Study Design

8 conditions:
2 x diagnosticity
2 x model match
2 x framework
100 data sets per condition

		Diagnosticity	
		Low	High
Model Match	Model Fitting	IRT-2PL CDM-DINA	
	Retrofitting	**IRT - DINA** CDM-2PL	

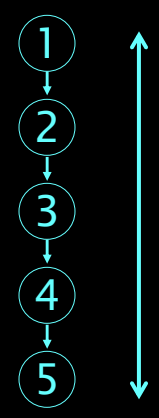
Challenges

- Different numerical representations
  - continuous ability vs. discrete skills
  - ordered values vs. semi-ordered classes
  - 1 dimension vs. multiple
- Different types of model parameters
  - Latent trait vs. latent class
  - Difficulty & Discrimination vs. Guessing & Slip

Study Assumption

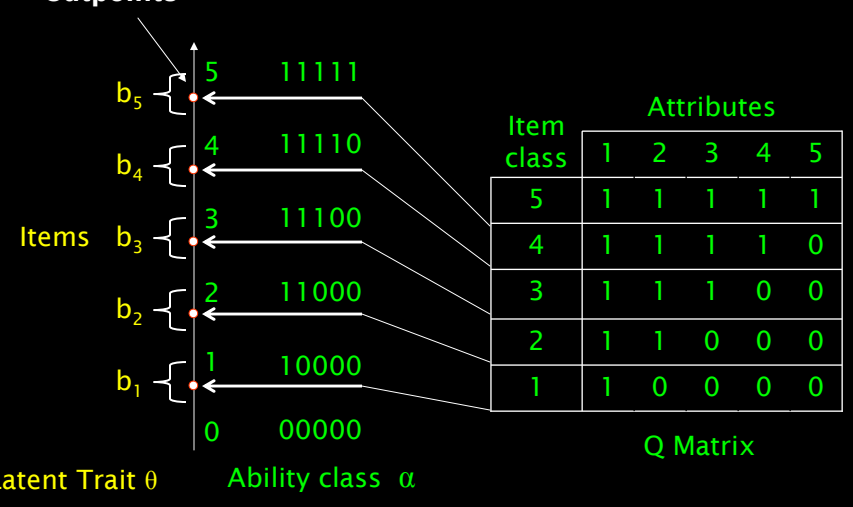
The underlying ability has a linear hierarchical structure

The mastery of each attribute is a prerequisite to the mastery of the next attribute



(Leighton, Gierl & Hunka, 2004)

Comparable Item & Ability Structures



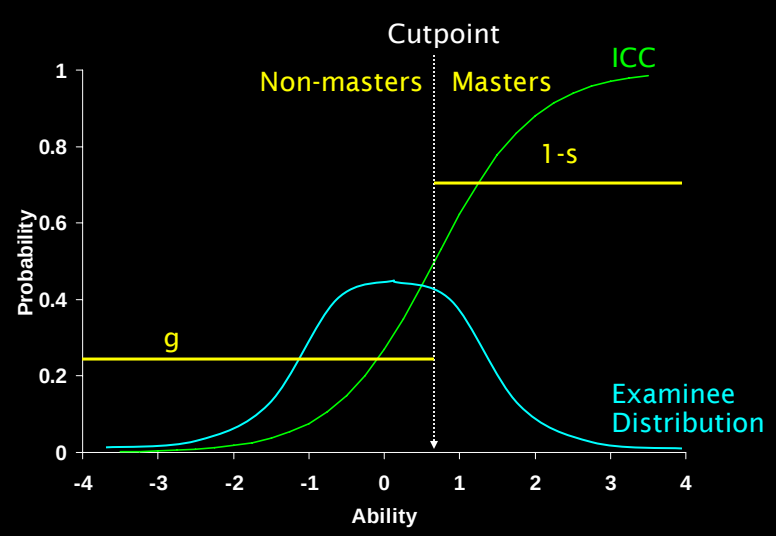
Latent Trait θ Ability class α

IRT CDM

Comparable Data sets

- Examinees
  - N = 5000, $\theta \sim N(0,1)$
  - Assign examinees to classes by cutpoints
- Items
  - K = 5 attributes, J = 25 items
  - Difficulties, b_j, distributed around cutpoints
  - Discriminations, a_j, manipulate diagnosticity
  - Convert 2PL parameters to $(1-s_j)$ and g_j using G-EMI:
    - Average ICC response probabilities for masters & non masters, weighted by examinee distribution

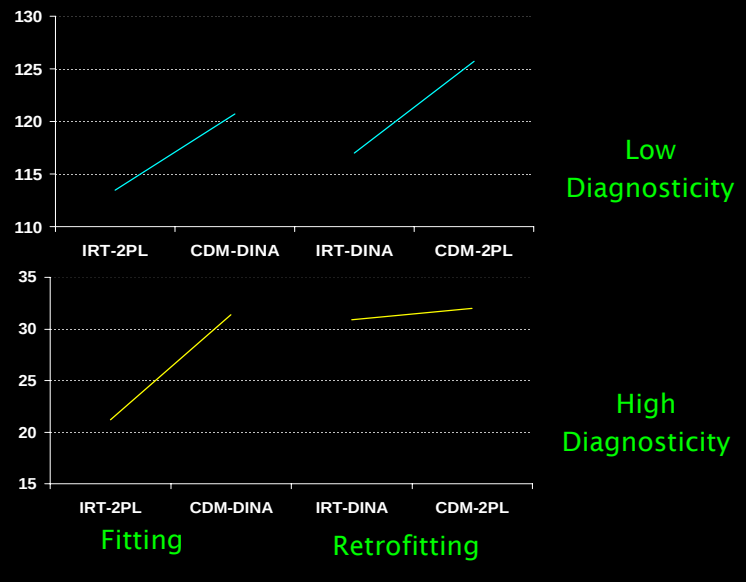
Converting ICC between frameworks



G-EMI indices

Results: Model Fit

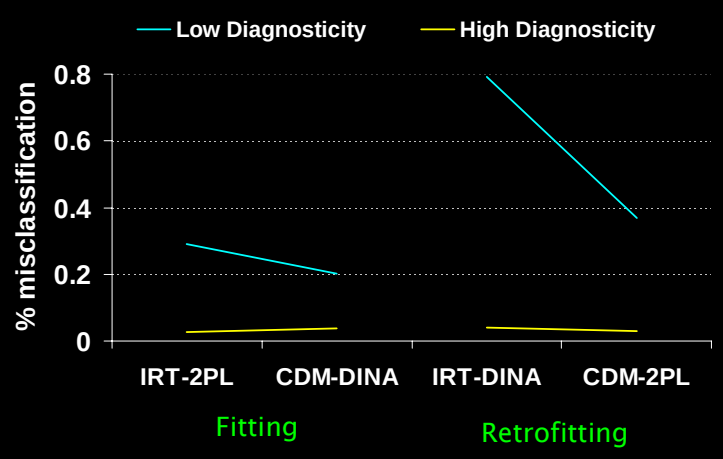
RMSD- Root Mean Squared Deviation



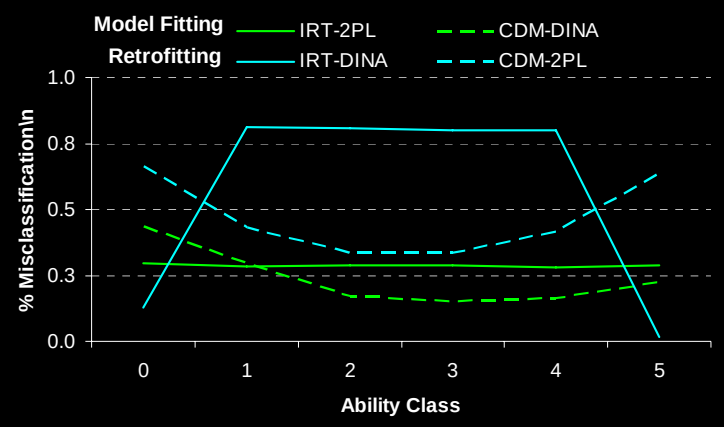
Low Diagnosticity

High Diagnosticity

Results: Ability class misclassification

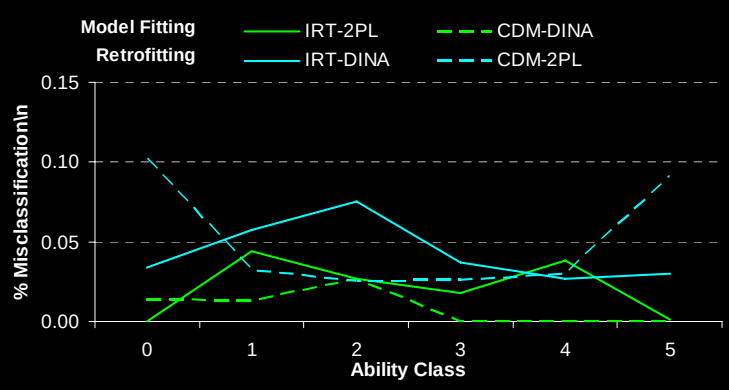


Results:
Ability class misclassification by class



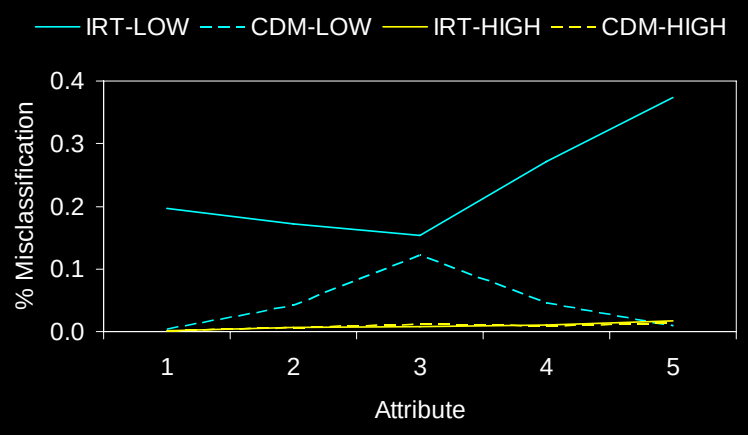
Low Diagnosticity

Results:
Vector misclassification by class



High Diagnosticity

Results:
DINA misclassification by attribute



Conclusions

- Highly diagnostic tests lead to accurate classification regardless of the frameworks.
- Low diagnosticity intensifies the negative effects of model mismatch.
- The validity of retrofitting CDM to IRT was not supported.
- Model fit indices may hide valuable information.