Combining Scaling and Classification: A Psychometric Model for Scaling Ability and Diagnosing Misconceptions

Laine Bradshaw
The University of Georgia

Cognition and Assessment SIG Business Meeting
April 30, 2013

THE UNIVERSITY OF GEORGIA
COLLEGE OF EDUCATION

- Sample Multiple Choice Item
  - Ways to psychometrically model this item
- Scaling Individuals and Classifying Misconceptions (SICM) model
  - New psychometric model developed through my dissertation
- Brief results from empirical data analysis
  - Force Concepts Inventory

THE UNIVERSITY OF GEORGIA
COLLEGE OF EDUCATION

Example Multiple Choice Item

Which of the following operations correctly shows how to find the area, in inches, of a rectangle that is 3 feet long and 8 inches wide?

(a) 36 in. x 8 in. Correct!
(b) 8 in. x ¼ in. Confusion with converting units
(c) 36 in. + 36 in. + 8 in. + 8 in. Confuses area with perimeter
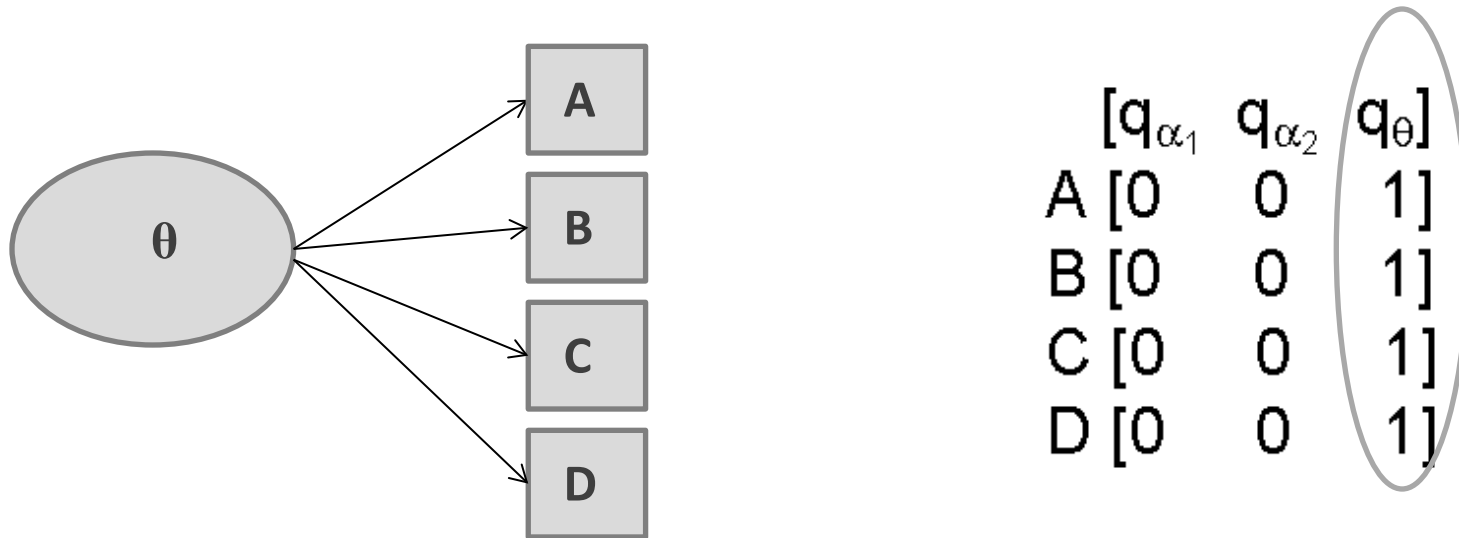(d) ¼ in. + ¼ in. + 8 in. +8 in. Confusion with converting units and area

- Not uncommon to find items written like this one
  - Incorrect options align to common student conceptions or errors
    - In science (e.g., Hestenes, et al., 1992; Sadler, 1998; Sadler, et al., 2010)
    - In statistics (e.g., Garfield, 1998; Khazanov, 2009)
    - In general, test design practices seek "plausible" distractors

- How do we statistically capitalize on this rich information?

Psychometric Choices

- The psychometric model chosen should
  - Reflect empirical theories of the domain-specific science
  - Provide types of information that teachers and students seek
- Common choices
  - Practical model: CTT total scores and subscores
  - Research settings: NR IRT model (Bock, 1972)
    - Capture the unique information in the item response
    - Item response is a function of an overall continuous math ability (θ)
- Alternate choice
  - Scaling Individuals and Classifying Misconceptions (SICM) model was tailored for this kind of item
  - Item response is a function of:
    - An overall continuous math ability (θ)
    - Two categorical misconceptions (α):
      - α_1 = confuses area with perimeter
      - α_2 = difficulty with multiplicative comparisons needed to make conversions among units
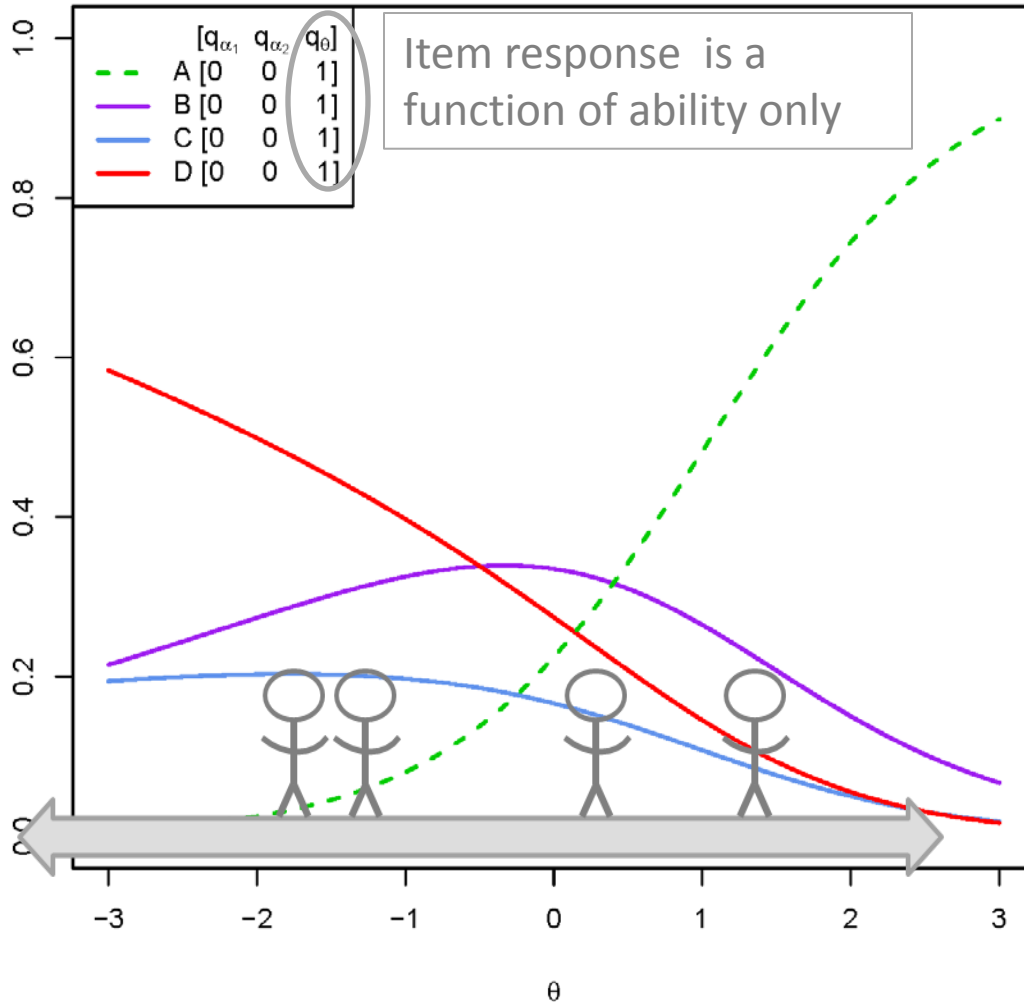   - Four possible misconception patterns(α)

NR IRT Item Response Function

- IRT methods scale examinees by locating them along a **single continuum** according to an overall ability

- Item response is a function of student's overall continuous math ability (θ)

$$\begin{array}{c c c c} & [q_{\alpha_1} & q_{\alpha_2} & q_{\theta}] \\ A & [0 & 0 & 1] \\ B & [0 & 0 & 1] \\ C & [0 & 0 & 1] \\ D & [0 & 0 & 1] \end{array}$$

NR IRT Item

NR 2−PL IRT

	[q_{α_1}	q_{α_2}	q_θ]
A	[0	0	1]
B	[0	0	1]
C	[0	0	1]
D	[0	0	1]

Item response is a function of ability only

This information is useful for:

- Comparing students' abilities (for scholarship or awards)

- Tracking growth on an (assumed) interval level

What about the errors or misconceptions?

What about the errors, or misconceptions?

- Frequently, CTT methods are used
  - Subscores for the number of times a student selects an incorrect alternative aligned to a misconception
- Problems
  - Small number of items per misconception
  - Are item responses independent conditional on ability alone?

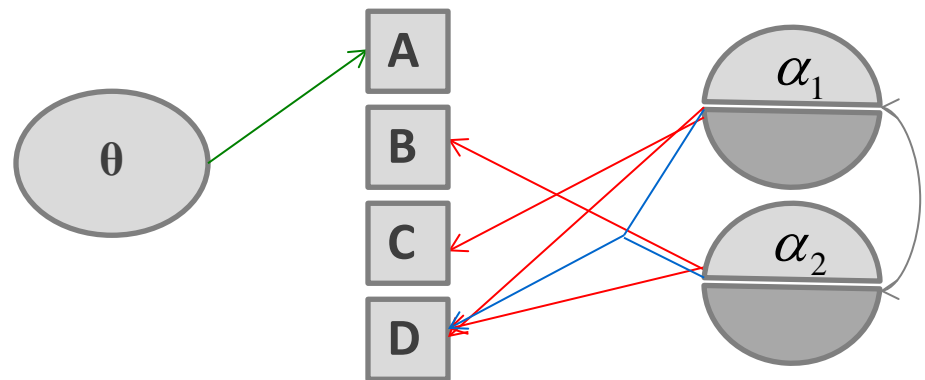# What about the errors, or misconceptions?

- The SICM model offers an alternate solution
  - Harnesses practicality of diagnostic classification models
    - Provide more reliable multidimensional feedback with small number of items
    - How? Use categorical latent variables
  - Includes misconceptions as a part of the item response function
   - Models misconceptions as a latent variable

SICM Model Item Response Function

- In the SICM Model, the item response is a function of

  1. Ability, as in the NR IRT model

     - Continuous trait

  2. Attributes, as in diagnostic classification models (DCMs)

     - Categorical traits
     - Attributes are defined as misconceptions instead of skills or abilities

- SICM model specifications:

$$
\begin{array}{c@{\quad}ccc}
 & [q_{\alpha_1} & q_{\alpha_2} & q_\theta] \\
A & [0 & 0 & 1] \\
B & [0 & 1 & 0] \\
C & [1 & 0 & 0] \\
D & [1 & 1 & 0]
\end{array}
$$

SICM Model Item

SICM Model: $\alpha = [\,0\;0\,]$

	$[q_{\alpha_1}\; q_{\alpha_2}\; q_\theta]$
A	[0 0 1]
B	[0 1 0]
C	[1 0 0]
D	[1 1 0]

SICM Model: $\alpha = [\,0\;1\,]$

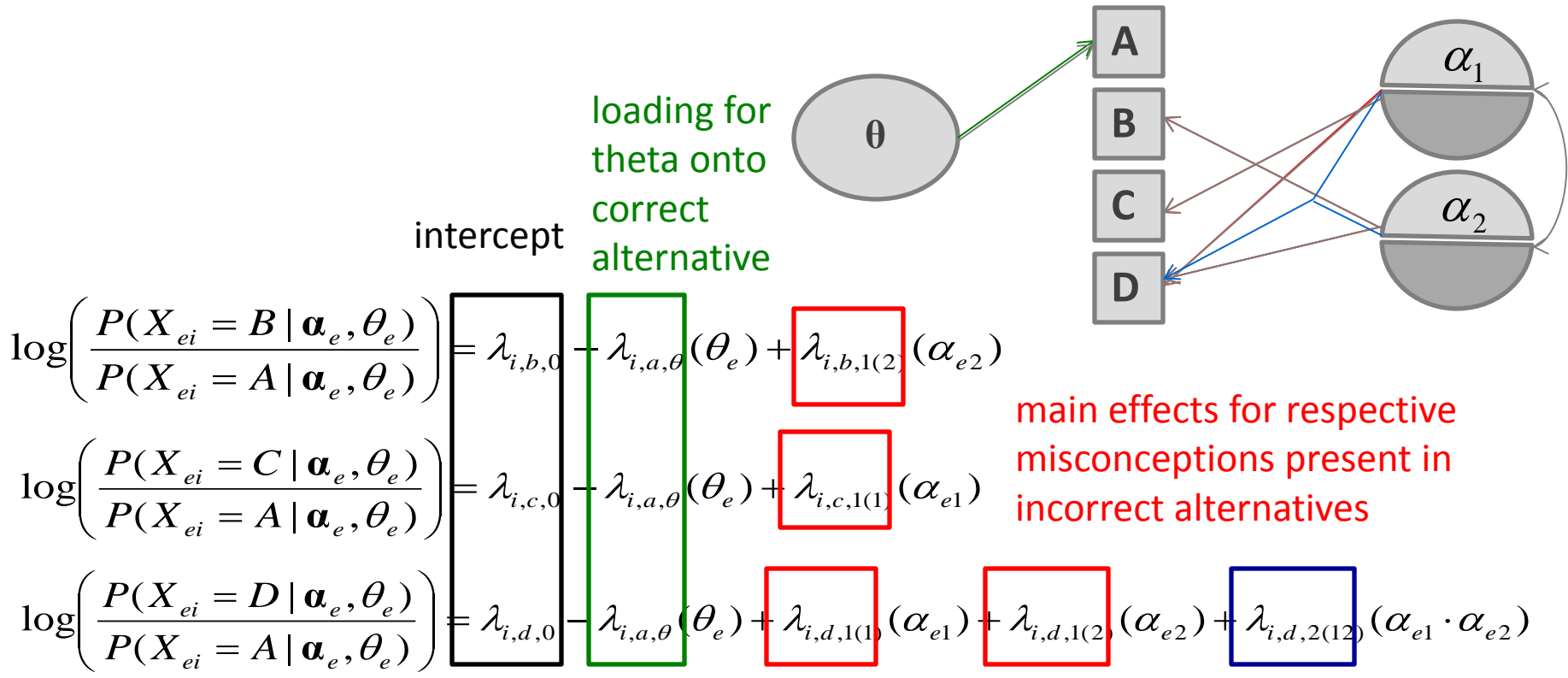SICM Model: $\alpha = [\,1\;0\,]$
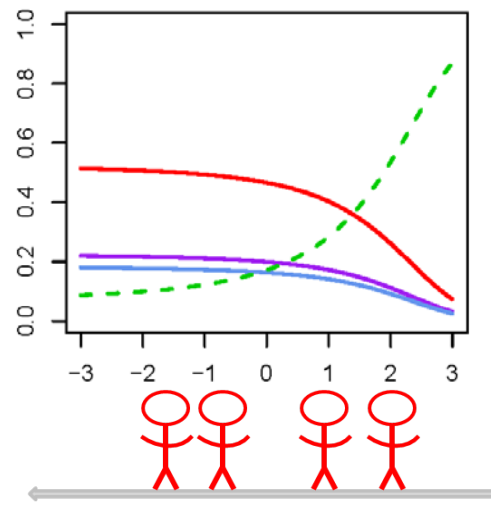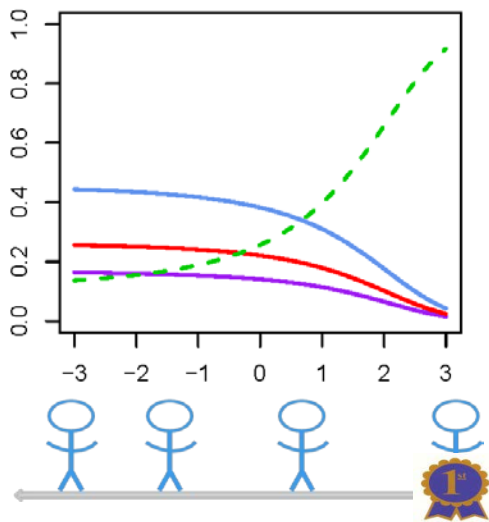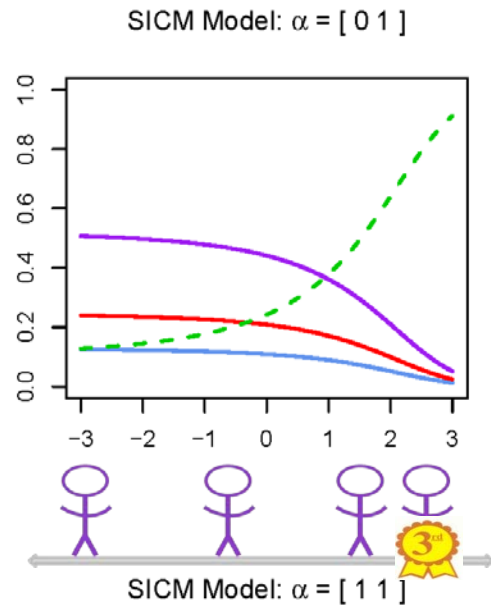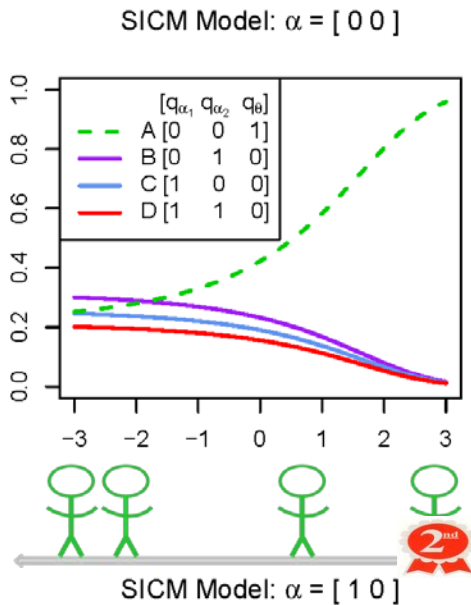
SICM Model: $\alpha = [\,1\;1\,]$

- SICM model estimates can be used for:
  - Classifying examinees according to misconceptions to tailor instruction or remediation

	Area Misconception	Unit Conversion Misconception
(red)	✓	✓
(purple)	✗	✓
(blue)	✓	✗
(green)	✗	✗

  - Comparing examinees' abilities for ranking or accountability purposes

SICM Model Item Response Function for Example Item

loading for theta onto correct alternative

intercept

main effects for respective misconceptions present in incorrect alternatives

$$\log\left(\frac{P(X_{ei}=B\mid\boldsymbol{\alpha}_e,\theta_e)}{P(X_{ei}=A\mid\boldsymbol{\alpha}_e,\theta_e)}\right) = \lambda_{i,b,0} + \lambda_{i,a,\theta}(\theta_e) + \lambda_{i,b,1(2)}(\alpha_{e2})$$

$$\log\left(\frac{P(X_{ei}=C\mid\boldsymbol{\alpha}_e,\theta_e)}{P(X_{ei}=A\mid\boldsymbol{\alpha}_e,\theta_e)}\right) = \lambda_{i,c,0} + \lambda_{i,a,\theta}(\theta_e) + \lambda_{i,c,1(1)}(\alpha_{e1})$$

$$\log\left(\frac{P(X_{ei}=D\mid\boldsymbol{\alpha}_e,\theta_e)}{P(X_{ei}=A\mid\boldsymbol{\alpha}_e,\theta_e)}\right) = \lambda_{i,d,0} + \lambda_{i,a,\theta}(\theta_e) + \lambda_{i,d,1(1)}(\alpha_{e1}) + \lambda_{i,d,1(2)}(\alpha_{e2}) + \lambda_{i,d,2(12)}(\alpha_{e1}\cdot\alpha_{e2})$$

interaction between two misconceptions in D

Model is identified by setting the parameters in the baseline category (A) to zero and by standardizing the continuous predictor (θ).

Example Data Analysis

Force Concepts Inventory (FCI)

- Test that seeks to identify misconceptions students have about Newtonian force concepts

- Careful test construction efforts to write incorrect options to be reflective of student misconceptions

- One of the most widely administered tests in physics education

- Purpose: illustrate SICM model's use to scale ability and provide categorical misconception feedback
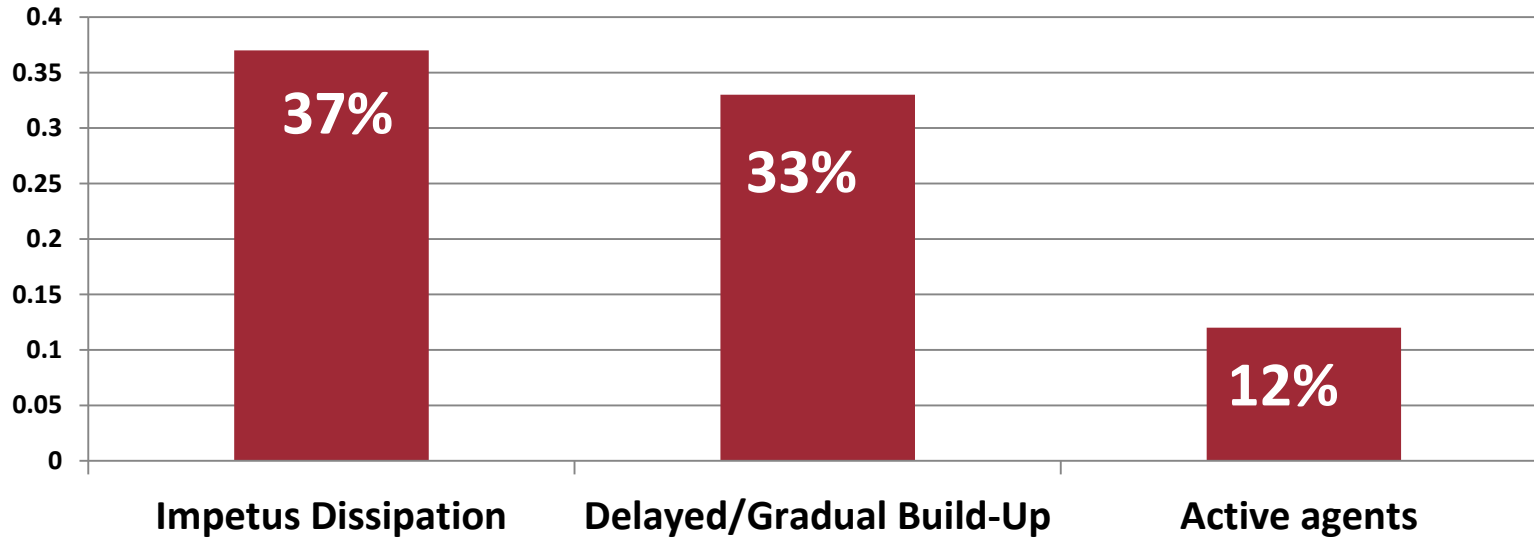
Force Concepts Inventory (FCI)

- 30 item test
- First 3 misconceptions measured by at least 5 items were included in the SICM model
  - Misconception 1: *impetus dissipation*
  - Misconception 2: *gradual/delayed impetus build-up*
  - Misconception 3: *only active agents exert force*
- Each misconception measured by 6 items
  - Measured by 10, 7, and 6 options, respectively
- Data: 10,039 high school students enrolled in a physics class

Results: Some Highlights

- Relative model-data fit
  - SICM model was a more parsimonious model that the NR-IRT model

- Diagnostic quality of incorrect options
  - For average ability examinees, possessing a misconception increased the probability of selecting an aligned incorrect option on average by 10.8%, 10.5%, and 29% for Misconceptions 1-3, respectively

Results: Some Highlights
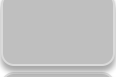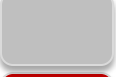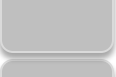
- Misconception prevalence

Results for Students J and K

α_1

Item	J	K
12	No Misconception	Misconception
13	Misconception	Misconception
14	No Misconception	No Misconception
23	No Misconception	Misconception
24	Misconception	No Misconception
27	No Misconception	No Misconception
CTT	2	3
SICM	.09	.11

Misconception

No Misconception

Results for Students J and K

α_1

Item	J	K
12	No Misconception	Misconception
13	Misconception	Misconception
14	No Misconception	No Misconception
23	No Misconception	Misconception
24	Misconception	No Misconception
27	No Misconception	No Misconception
CTT	2	3
SICM	.09	.11

α_2

Item	J	K
8	No Misconception	Misconception
10	No Misconception	Misconception
21	No Misconception	Misconception
23	No Misconception	No Misconception
26	No Misconception	Misconception
27	No Misconception	No Misconception
CTT	0	4
SICM	.19	.75

α_3

Item	J	K
15	Misconception	No Misconception
16	Misconception	No Misconception
17	No Misconception	No Misconception
18	No Misconception	No Misconception
28	Misconception	Misconception
30	No Misconception	No Misconception
CTT	3	1
SICM	.69	.04

θ

	J	K
Total Score	8	7
SICM	-0.08	-0.09
SICM	[001]	[010]

Misconception

No Misconception

Diagnosis	J	K
Cut-off 2	[101]	[110]
Cut-off 3	[001]	[110]
Probabilistic	[001]	[010]

Simulation Study

Simulation Study Design

Characteristics		Value or Interval		
Test	Number of Items	30, 60		
Sample	Number of Examinees	3000; 10,000		
Measurement Model		Low	High	
	Sampling interval for intercepts	(-.5, .5)		
	Sampling interval for α main effects	(.75, 1.25)	(1.75, 2.25)	
	Sampling interval for λ_θ	(.25, .75)	(1.0, 1.5)	
	Sampling interval for two-way interaction effects	(0.5, 1)		
	Higher-order interactions	0		
Structural Model	Number of Attributes (Misconceptions)	3,6		
	Tetrachoric Correlation among Attributes	0.25,0.50		
	Distribution of Continuous Trait	$\mathbf{N}(0,1)$		

64 Conditions

50 Replications for Each Condition

Concluding Remarks

Concluding Remarks

- SICM model addresses a growing demand for assessment systems: to gain more feedback about what students do not understand

- Ranking individuals and providing diagnostic feedback are two "commonly co-occurring" purposes of a test that may be viewed as "fundamentally antithetical purposes" in commonly used testing paradigms (Wainer et al., 2001, pg. 342)

- SICM model diagnostic feedback complements traditional measures of overall ability
  - Reliably measure and statistically account for misconceptions

Questions? Comments?

laineb@uga.edu

Thank you!