

# **Longitudinal Diagnostic Classification Models: Theory and Applications**

Matthew Madison, University of Georgia  
Madeline Schellman, Navy at Pearson

# Workshop Overview

**Access session materials at**  
**[www.matthewmadison.com/workshop](http://www.matthewmadison.com/workshop)**

# Workshop Overview

- Preliminaries: LCDM framework and estimation in R (1:15 – 2:00)
- Break 1 (2:00 – 2:05)
- Longitudinal DCMs I (single group) (2:05 – 3:00)
- Break 2 (3:05 – 3:15)
- Longitudinal DCMs II (multigroup) (3:15 – 4:15)
- Break 3 (4:15 – 4:20)
- Q&A, discussion, evaluations (4:20 – 5:00)

# A Little About Matt

- From Columbia, South Carolina
- Research interests: DCMs, measurement of growth
- Hobbies: sports (basketball, golf), true crime



# A Little About Madeline

- From Marietta, GA
- Research interests: DCMs, mathematics education, item experiments, diagnostic growth percentiles
- Hobbies: MMA, Broadway shows, reading



# A Little About You

- Hometown/country
- Graduate program or professional position
  - Background and research interests
- Why you want to learn DCMs and longitudinal DCMs

# DCM Preliminaries

LCDM Framework and Estimation in R

# Conceptual Example

- Suppose an MMA coach wants to test basic MMA ability
- Using traditional psychometric approaches, the coach could obtain an ability estimate for each fighter
  - » Classical test theory: obtain a test score (# items correct)
  - » Item response theory: obtain a latent ability estimate
- Both these approaches are normative in nature, only provide information on students' location on a continuum

# Conceptual Example

- Traditional testing procedures measure an **overall** ability in an area with a **continuous** latent variable:



**MMA Ability**

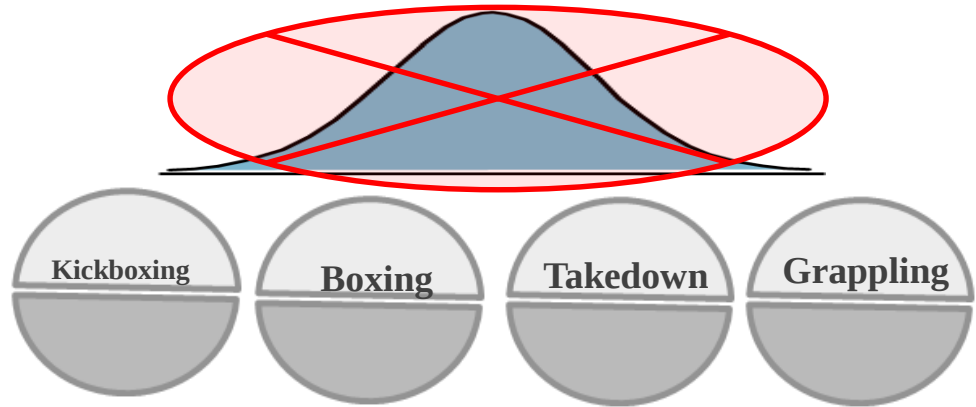
# Conceptual Example

- Using traditional, norm referenced approaches, the result is a weak ordering
  - » Ordering is weak because of error in estimates
  - » Madeline < Thug Rose < Amanda Nunes
- Questions that traditional psychometrics cannot answer:
  - » Why is Madeline so low?
    - How do we help? What specific skills does she need to work on?
  - » How much ability is “enough” to pass?
    - What score is proficiency, or mastery?
  - » What skills have the fighters mastered, not mastered?

# MMA Ability Example

- Instead of measuring an overall MMA ability, let's break MMA down into a set of skills or *attributes*:

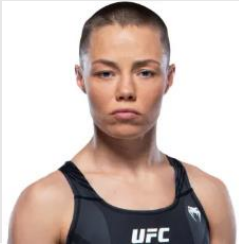
- Kickboxing
- Boxing
- Takedown
- Grappling



- These attributes are categorical (not continuous).
- Often, they are dichotomous:
  - Model places examinees into two categories (e.g., Proficiency vs Non-proficiency)

# Diagnostic Classification Models

- DCMs use item responses to place students into groups according to proficiency or non- proficiency of multiple attributes

| Fighter                                                                             | Kickboxing | Boxing | Takedown | Grappling |
|-------------------------------------------------------------------------------------|------------|--------|----------|-----------|
|    | ✓          | ✗      | ✗        | ✓         |
|   | ✓          | ✗      | ✗        | ✓         |
|  | ✓          | ✓      | ✓        | ✓         |

# Diagnostic Classification Models

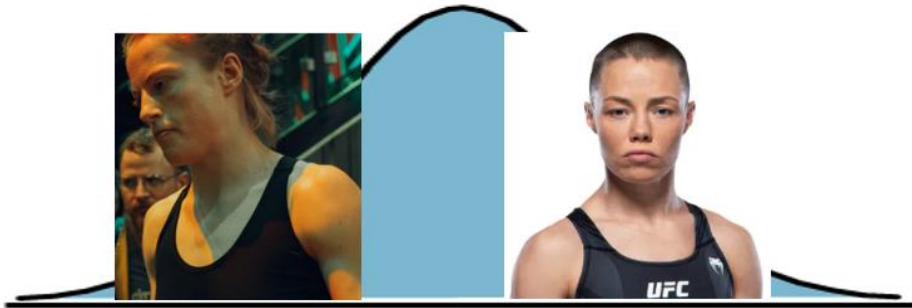
- Questions from before:
  - » Why is Madeline so low?
    - Because she can't box and she doesn't have good takedowns. She needs to work on these skills.
  - » What skills are the fighters proficient, not proficient?
    - DCM provides these classifications directly.

# When are DCMs appropriate?

- Short answer: not always
- Successful implementation depends on:
  - 1) domain-specific theories/content experts.
    - What are the set of attributes we are trying to measure?
    - Can items be written to measure these attributes?
  - 2) alignment of purpose between assessment and model
    - Is classification the purpose?

# When are DCMs Not Appropriate?

- When you want to scale and desire a finer level of granularity
- DCMs do not distinguish within classes (limitation)
  - Thug Rose has better kickboxing than Madeline, but they both possess the skill



| Fighter                                                                              | Kickboxing | Boxing | Takedown | Grappling |
|--------------------------------------------------------------------------------------|------------|--------|----------|-----------|
|   | ✓          | ✗      | ✗        | ✓         |
|  | ✓          | ✗      | ✗        | ✓         |

# Conceptual Foundation Summary

- DCMs are psychometric models designed to classify
  - Diagnostic and criterion-referenced interpretations
- DCMs provide valuable information with more feasible data demands than other psychometric models
  - Higher reliability than IRT/MIRT models
  - Naturally accommodate multidimensionality
  - Complex items structures possible
  - Alignment of assessment goals and psychometric model

# Questions? Comments?



# Statistical Foundation

- Latent class models use responses to probabilistically put respondents into latent classes
- DCMs are confirmatory latent class models
  - » Latent classes specified a priori as attribute profiles
  - » Q-matrix specifies item-attribute structure (confirmatory)
  - » Person parameters are attribute proficiency probabilities

# DCMs as Latent Class Models

- DCMs are constrained latent class models

**Observed Data:** Probability of observing examinee  $r$ 's vector of item responses to all  $I$  items

**Measurement Component:**  
Product of Conditional Item Response Probabilities (Item Responses are Independent)

$$P(X_r = x_r) = \sum_{c=1}^C v_c \prod_{i=1}^I \pi_{ic}^{x_{ir}} (1 - \pi_{ic})^{1-x_{ir}}$$

**Structural component:**  
Proportion of examinees in each class

# Statistical Foundation

- Latent class models (LCMs) use responses to probabilistically put respondents into latent classes
- DCMs are confirmatory latent class models
  - » Latent classes specified a priori as attribute profiles
  - » Q-matrix specifies item-attribute structure (confirmatory)
  - » Person parameters are attribute proficiency probabilities
- DCMs are the categorical latent trait analogue of IRT models
  - » Item parameters model relationships between attribute proficiency and responses

# Item Response Function Basics

- 2-PL item response function

- » Discrimination ( $a$ )
- » Difficulty ( $b$ )
- » Ability ( $\theta$ )

$$\begin{aligned}\text{logit}(X_i = 1) &= a_i(\theta - b_i) \\ &= a_i\theta - a_ib_i \\ &= \boxed{\psi_i + \delta_i\theta}\end{aligned}$$

- DCM

- » Intercept ( $\lambda_0$ )
- » Main effects ( $\lambda_1$ )
- » Attribute proficiency status ( $\alpha$ )

$$\text{logit}(X_i = 1) = \boxed{\lambda_{i,0} + \lambda_{i,1}\alpha}$$

- $\alpha$  is either 0 or 1
- $\lambda_0$  is log-odds of a correct response for non-proficient
- $\lambda_1$  is the increase in log-odds of a correct response for proficient

# Item Response Function Basics

$$P(X_r = x_r) = \sum_{c=1}^C v_c \prod_{i=1}^I \pi_{ic}^{X_{ir}} (1 - \pi_{ic})^{1-X_{ir}}$$

$$\text{logit}(X_i = 1) = \lambda_{i,0} + \lambda_{i,1}\alpha$$

$$\pi_{ic} = \frac{\log(\text{logit})}{1 + \log(\text{logit})}$$

$$\pi_{ic} = \frac{\log(\lambda_{i,0} + \lambda_{i,1}\alpha)}{1 + \log(\lambda_{i,0} + \lambda_{i,1}\alpha)}$$

| $\alpha$ | Logit                           |
|----------|---------------------------------|
| 0        | $\lambda_{i,0}$                 |
| 1        | $\lambda_{i,0} + \lambda_{i,1}$ |

$$\pi_{ic} = \frac{\log(\lambda_{i,0})}{1 + \log(\lambda_{i,0})}$$

$$\pi_{ic} = \frac{\log(\lambda_{i,0} + \lambda_{i,1})}{1 + \log(\lambda_{i,0} + \lambda_{i,1})}$$

# Development of DCMs

- Over the past couple of decades, numerous DCMs have been developed
- Each DCM makes assumptions about how mastered attributes combine/interact to produce an item response
- We're going to focus on one, the Log-linear Cognitive Diagnosis Model (LCDM; Henson, Templin, & Willse, 2009)

# Item Response Function Basics

- Consider an item that measures two attributes

$$\text{logit}(X_i = 1) = \lambda_{i,0} + \lambda_{i,1(1)}\alpha_1 + \lambda_{i,1(2)}\alpha_2 + \lambda_{i,1(1)}\lambda_{i,1(2)}\alpha_1\alpha_2$$

Intercept      Main Effect (A1)      Main Effect (A2)      Interaction (Between A1 and A2)

| $\alpha_1$ | $\alpha_2$ | Logit                                                                                    |
|------------|------------|------------------------------------------------------------------------------------------|
| 0          | 0          | $\lambda_{i,0}$                                                                          |
| 0          | 1          | $\lambda_{i,0} + \lambda_{i,1(2)}$                                                       |
| 1          | 0          | $\lambda_{i,0} + \lambda_{i,1(1)}$                                                       |
| 1          | 1          | $\lambda_{i,0} + \lambda_{i,1(1)} + \lambda_{i,1(2)} + \lambda_{i,1(1)}\lambda_{i,1(2)}$ |

$$\text{logit}(X_i = 1) = \lambda_{i,0} + \sum_{a=1}^A \lambda_{i,1(a)}\alpha_{ca}q_{ia} + \sum_{a=1}^A \sum_{a'>1}^A \lambda_{i,2(a,a')}\alpha_{ca}\alpha_{ca'}q_{ia}q_{ia'} + \dots$$

# LCDM Interaction Plots

- Play with excel file: “[Example LCDM Items.xlsx](#)”
- Manipulate item parameters and see how logits and item response probabilities are affected

# Statistical Foundations Summary

- DCMs are confirmatory latent class models
- DCMs are categorical latent trait IRT models
- Person parameters ( $\alpha$ ): attribute proficiency probabilities
- Item parameters ( $\lambda$ ): proficiency  $\sim$  responses
- LCDM item response function
  - Logistic regression with binary predictors
  - Reference-coded ANOVA model
  - Parameters indicate increases in probability correct for different proficiency statuses
  - Many DCMs that place constraints on item parameters

# DCM Inputs and Outputs

- Inputs:
  - Item response data ( $N \times I$  matrix)
  - Q-matrix ( $I \times A$  matrix)
  - Type of DCM you want to estimate, plus options
- Outputs
  - Item parameter estimates
  - Person parameter estimates (proficiency probabilities)
  - Attribute parameter estimates (proficiency proportions, correlations)
  - Model fit statistics

# Estimation in R

- Several packages estimate DCMs
  - **CDM**
  - GDINA
  - mirt
  - blatant
  - hmcdm
  - npcd
  - PACER
- Other software programs as well
  - Mplus, flexMIRT, Latent Gold, mdltn

# Estimation in R

- In the Examples folder, open “example0.R”
- Two attributes, 10 items,  $N = 1000$

# Break

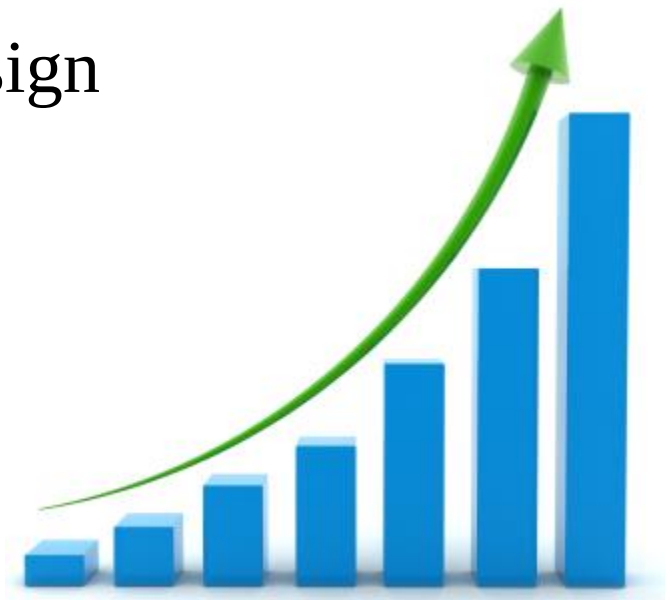
- Take a quick 5-minute break
- Next up: Longitudinal DCMs I



# **Longitudinal DCMs I (single group)**

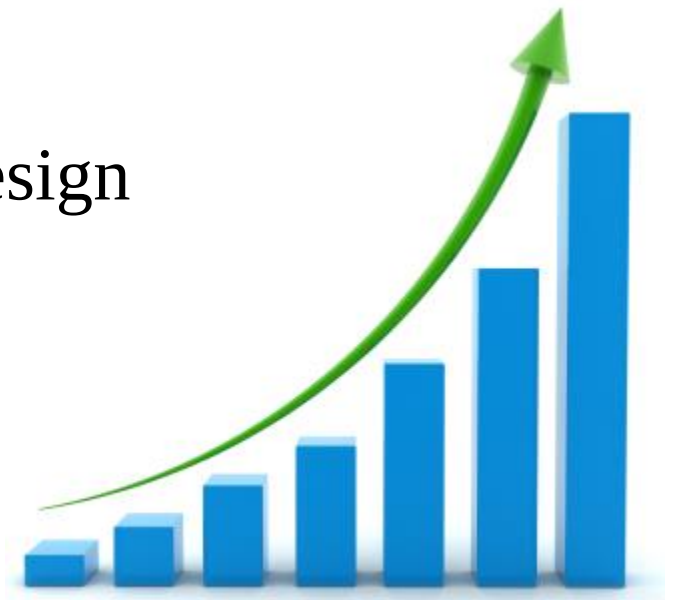
# Assessing Growth in Education

- Recent shift in educational policy
  - » Meeting state standards → measuring student progress
- Need to quantify learning
- Single-group pre-test/post-test design
  - » Examinees take pre-test
  - » Intervention (e.g., instruction)
  - » Examinees take post-test
- Goal: assess growth in students



# Assessing Growth in Education

- Recent shift in educational policy
  - » Meeting state standards → measuring student progress
- Need to evaluate instructional methods and materials
- Is Method A >> Method B?
- Control group pre-test/post-test design
  - » Compare methods
  - » Examine differential growth



# Growth and Item Response Models

- Classical Test Theory (CTT)
  - » Difference in pre-test and post-test score = Gain Score
  - » Gain score is interpreted as quantification of learning
  - » Can suffer from low reliability
- Item Response Theory (IRT)
  - » Difference in pre-test and post-test ability = growth
  - » Longitudinal IRT models
  - » Lack diagnostic feedback, reliability

# Growth and DCMs

- Diagnostic classification models (DCMs)
  - » Classify on categorical latent traits: *attributes*
  - » Attribute mastery or non-mastery

| Time Point | Student                                                                            | Addition | Subtraction | Multiplication | Division |
|------------|------------------------------------------------------------------------------------|----------|-------------|----------------|----------|
| Pre-test   |   | ✗        | ✗           | ✗              | ✗        |
| Post-test  |  | ✓        | ✓           | ✓              | ✗        |

## Pre-test

## Post-test

# Growth and DCMs

- Diagnostic classification models (DCMs)
  - » Classify on categorical latent traits: *attributes*
  - » Attribute mastery or non-mastery
- Growth in a DCM framework
  - » Individual: non-mastery  $\leftrightarrow$  mastery
  - » Group: 30% at pre-test  $\rightarrow$  80% at post-test
  - » Previous work modeled testing occasions separately

# Latent Class Models

- DCMs are constrained latent class models

**Observed Data:** Probability of the observed item responses

**Measurement Component:**  
Models how latent classes respond to items

The diagram shows the equation for the probability of observed data in a Latent Class Model. The equation is  $\mathbb{P}(\mathbf{X}_r = \mathbf{x}_r) = \sum_{c=1}^C \nu_c \prod_{i=1}^I \pi_{ic}^{X_{ir}} (1 - \pi_{ic})^{1-X_{ir}}$ . The terms are grouped into three colored ovals: a green oval around  $\mathbb{P}(\mathbf{X}_r = \mathbf{x}_r)$ , a blue oval around  $\sum_{c=1}^C \nu_c$ , and a red oval around  $\prod_{i=1}^I \pi_{ic}^{X_{ir}} (1 - \pi_{ic})^{1-X_{ir}}$ . A green arrow points from the green oval to the 'Observed Data' box. A blue arrow points from the blue oval to the 'Structural component' box. A red arrow points from the red oval to the 'Measurement Component' box.

$$\mathbb{P}(\mathbf{X}_r = \mathbf{x}_r) = \sum_{c=1}^C \nu_c \prod_{i=1}^I \pi_{ic}^{X_{ir}} (1 - \pi_{ic})^{1-X_{ir}}$$

**Structural component:**  
Models how classes are related, class proportions

# Latent Class Models

- There are longitudinal latent class models
  - » Latent transition analysis (LTA)

$$\mathbb{P}(\mathbf{X}_r = \mathbf{x}_r) = \sum_{c_1=1}^C \sum_{c_2=1}^C \cdots \sum_{c_T=1}^C \underbrace{\nu_{c_1}}_{\text{Structural component}} \underbrace{\tau_{c_2|c_1} \tau_{c_3|c_2} \cdots \tau_{c_T|c_{T-1}}}_{\text{Transition Probabilities}} \underbrace{\prod_{t=1}^T \prod_{i=1}^I \pi_{ic_t}^{x_{rit}} (1 - \pi_{ic_t})^{1-x_{rit}}}_{\text{Measurement Component}}$$

**Structural component:**  
Proportion of examinees in each class at Time Point 1

**Transition Probabilities:**  
Probabilities of transition between latent classes from time to time

**Measurement Component:**  
Product of Conditional Item Response Probabilities (Across time points)

# Transition DCM

- Transition Diagnostic Classification Model (TDCM)
  - » Combines LTA with the LCDM as measurement model
  - » General form ( $T$  time points):

$$\mathbb{P}(\mathbf{X}_r = \mathbf{x}_r) = \sum_{c_1=1}^C \sum_{c_2=1}^C \cdots \sum_{c_T=1}^C \nu_{c_1} \tau_{c_2|c_1} \tau_{c_3|c_2} \cdots \tau_{c_T|c_{T-1}} \prod_{t=1}^T \prod_{i=1}^I \pi_{ic_t}^{x_{rit}} (1 - \pi_{ic_t})^{1-x_{rit}}$$

- » Pre-test/post-test form (2 times points w/ invariance):

$$\mathbb{P}(\mathbf{X}_r = \mathbf{x}_r) = \sum_{c_1=1}^C \sum_{c_2=1}^C \nu_{c_1} \tau_{c_2|c_1} \prod_{t=1}^2 \prod_{i=1}^I \pi_{ic_t}^{x_{rit}} (1 - \pi_{ic_t})^{1-x_{rit}}$$

# Measurement Invariance

- Reflects the notion that measurement is happening the same way across groups or over time
- In a longitudinal psychometric analysis, person estimates should be determined the exact same way
- For the TDCM, we want that proficiency is being determined the same way over time
  - » Otherwise, change in proficiency status is meaningless

$$\mathbb{P}(\mathbf{X}_r = \mathbf{x}_r) = \sum_{c_1=1}^C \sum_{c_2=1}^C \nu_{c_1} \tau_{c_2|c_1} \prod_{t=1}^2 \prod_{i=1}^I \pi_{ic}^{x_{rit}} (1 - \pi_{ic})^{1-x_{rit}}$$

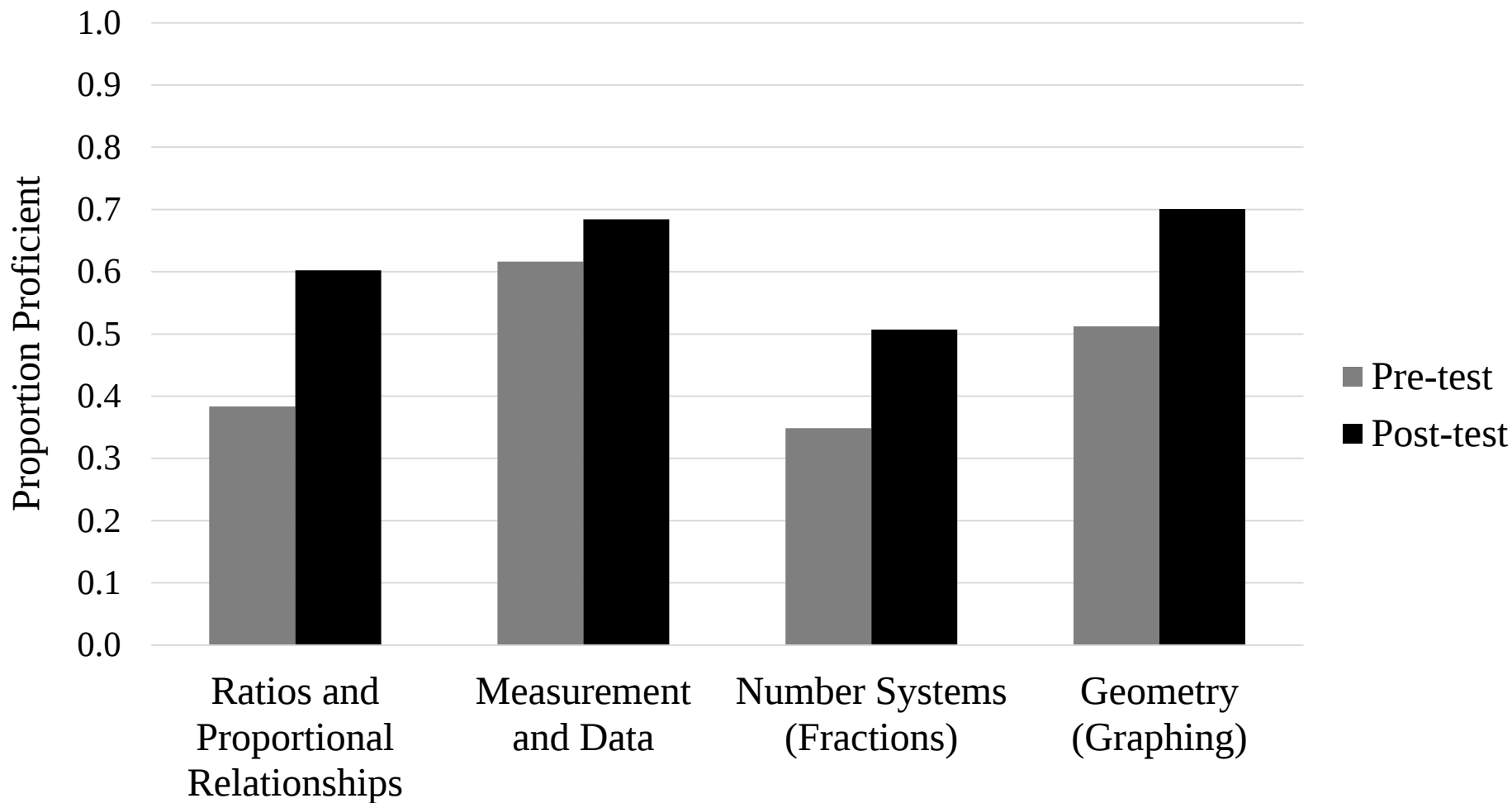
# Empirical Data Analysis

- Mathematics assessment measuring four attributes
  - » Ratios and proportional relationships
  - » Measurement and data
  - » Number system (fractions)
  - » Geometry (graphing)
- Sample of 879 middle school students
- Same 21 items at pre- and post-test

# Data Analysis Results

- Item parameters:
  - » Median intercept: -1.82 (14% chance or correct response)
  - » Median discrimination = .55 (masters – non-masters)
- Attribute Correlations:
  - » Pre-test: ranged from .58 to .84
  - » Post-test: ranged from .73 to .90
  - » Indicated multidimensionality

# Data Analysis Results



# Data Analysis Results

| Attribute                             | Pre-test | Post-test | Growth | <i>p</i> -value | Odds Ratio    |
|---------------------------------------|----------|-----------|--------|-----------------|---------------|
| Ratios and proportional relationships | .384     | .602      | .219   | <.001           | 2.43 (medium) |
| Measurement and data                  | .616     | .684      | .068   | <.001           | 1.35 (small)  |
| Numbers and operations (fractions)    | .348     | .507      | .159   | <.001           | 1.92 (medium) |
| Geometry (graphing)                   | .512     | .701      | .189   | <.001           | 2.23 (medium) |

# Data Analysis Results

- Individual Attribute Transition Matrices

| Ratios and proportional relationships |                 |             |
|---------------------------------------|-----------------|-------------|
|                                       | Non-proficiency | Proficiency |
| Non-proficiency                       | .52             | .48         |
| Proficiency                           | .21             | .79         |

| Measurement and data |                 |             |
|----------------------|-----------------|-------------|
|                      | Non-proficiency | Proficiency |
| Non-proficiency      | .59             | .41         |
| Proficiency          | .31             | .69         |

| Numbers and operations (fractions) |             |         |
|------------------------------------|-------------|---------|
|                                    | Non-mastery | Mastery |
| Non-proficiency                    | .57         | .43     |
| Proficiency                        | .16         | .84     |

| Geometry (graphing) |                 |             |
|---------------------|-----------------|-------------|
|                     | Non-proficiency | Proficiency |
| Non-proficiency     | .42             | .58         |
| Proficiency         | .19             | .82         |

# Estimation in R

- In the Examples folder, open “example1.R”
- Four attributes, 10 items,  $N = 1500$

# Estimation in R

A group of biology educators were interested in the progression of students' proficiency in foundational biology concepts (mitosis and meiosis) using an innovative blended learning instructional approach. They developed three forms of a test measuring both concepts and administered the test at three different points during the semester. Each form had ten items, with the forms sharing three common items (Items 3, 5, 7). They had a total of 875 Biology 101 students take all three tests.

- In the Examples folder, open “example2.R”
- Two attributes, 10 items,  $N = 875$
- Different Q-matrices at each time point
- Items 3, 5, 7 are common items across the tests

# **Longitudinal DCMs II (multigroup)**

# TDCM Applied Study

- National Assessment of Educational Progress, 2011
  - » 23% of students w/o disabilities scored below *basic*
  - » 65% of students with disabilities scored below *basic*
- Basic: operations with rational numbers and real world problem solving
- Research suggests that students with disabilities struggle with problems embedded in text
- University of Kentucky (Bottge, et. al) focus on improving mathematics instruction and assessment for students with disabilities

# Enhanced Anchored Instruction

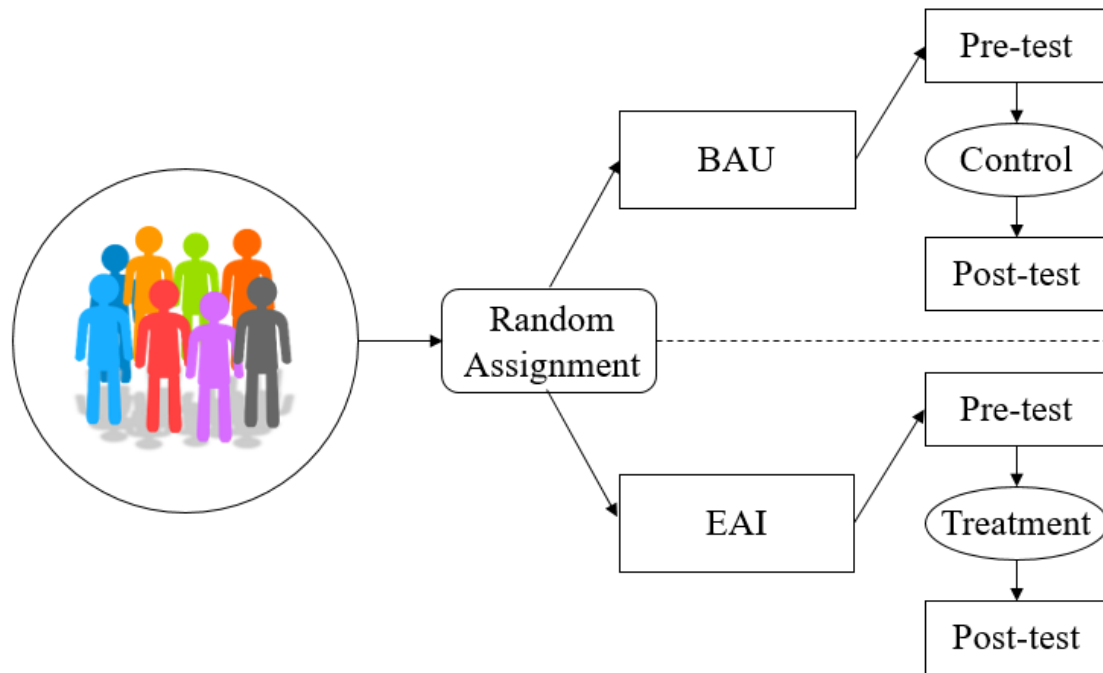
- Acronym (EAI)
- Instructional method developed to improve problem solving
- Realistic problems are embedded in interesting contexts
  - » 8-15 minute videos
  - » Hands-on applications
- Include modules and interactive media
  - » Help students visualize abstract concepts

# EAI Study Design

- Overall research questions:
  - » To what extent does EAI improve students' problem solving abilities?
  - » How well does it work compared to traditional instruction?
- Team developed an assessment targeting multiple areas
  - » 21-item problem-solving test
  - » Measured four Common Core *strands*
    - Ratios and proportional relationships
    - Measurement and data
    - Numbers systems (fractions)
    - Geometry (graphing)

# EAI Study Design

- Business as Usual (BAU,  $N = 456$ ) vs. EAI ( $N = 423$ )
- Sample of 6-8 graders



- Is EAI better than traditional instruction?

# EAI Study Design

- Items not initially designed with a DCM in mind
  - » But since items were prospectively written for each strand, well-suited for a DCM analysis
- Methodological research questions:
  - » How do we analyze growth with a general DCM?
    - Prompted development of TDCM
  - » How do we analyze intervention effects in a DCM?
    - Multiple-group TDCM

# Analysis

- Evaluate intervention effect
  - » Use multiple-group TDCM
  - » Analyze growth for BAU and EAI
  - » Compare growth
- How does growth for EAI students compare to growth for BAU students?
  - » For which attributes does EAI appear to be better than BAU?
  - » Where can EAI be improved?

# Analysis

- To analyze the intervention effect, extend TDCM to multiple groups
  - » Attribute proficiency proportions and transitions depend on group membership ( $g$ )

$$\mathbb{P}(\mathbf{X}_r = \mathbf{x}_r | G = g) = \sum_{c_1=1}^C \sum_{c_2=1}^C \boxed{v_{c_1|g} \tau_{c_2|c_1,g}} \prod_{t=1}^2 \prod_{i=1}^I \pi_{ic}^{x_{rit}} (1 - \pi_{ic})^{1-x_{rit}}$$

# Analysis

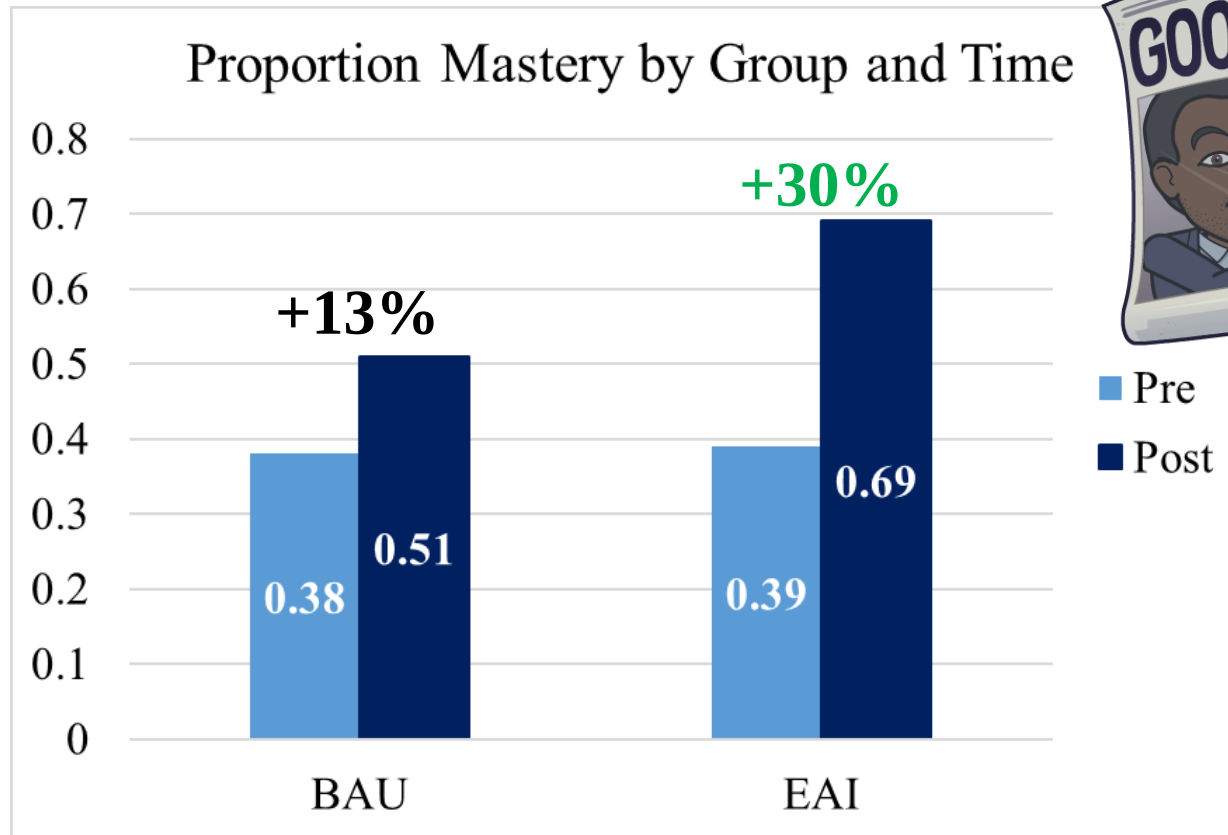
- Let  $\delta_{g_1}$  and  $\delta_{g_2}$  be the attribute mastery growth for both groups, respective
- To test for differential growth, can use a Wald test:

$$z = \frac{|\hat{\delta}_{g_1} - \hat{\delta}_{g_2}|}{se(\hat{\delta}_{g_1} - \hat{\delta}_{g_2})},$$

- $z$  follows a standard normal distribution
- Similar to a “difference-in-difference” test
- Use Hochberg (1988) multiple tests correction to control type I error rates

# Results: Differential Growth

- Attribute 1: Ratios and Proportions



# Results: Differential Growth

| Attribute                             | Growth (EAI) | Growth (BAU) | Difference | <i>p</i> -value |
|---------------------------------------|--------------|--------------|------------|-----------------|
| Ratios and proportional relationships | .31          | .13          | .18 (.05)  | <.001           |
| Measurement and data                  | .07          | .07          | .00 (.04)  | .929            |
| Number Systems (fractions)            | .18          | .13          | .05 (.04)  | .597            |
| Geometry (graphing)                   | .26          | .11          | .15 (.05)  | .008            |

# Results: Conditional Growth

- Proficiency transitional probabilities
  - » Were EAI non-proficients more likely to transition to proficiency than BAU non-proficients?
  - » Ratios and proportional relationships

| BAU                      |   |                                                    |
|--------------------------|---|----------------------------------------------------|
| Post-test                |   |                                                    |
| <hr/>                    |   |                                                    |
| 0                      1 |   |                                                    |
| <hr/>                    |   |                                                    |
| Pre-test                 | 0 | <div><div>.63</div><div><div>.37</div></div></div> |
|                          | 1 | <div><div>.27</div><div>.73</div></div>            |

| EAI      |   |           |     |
|----------|---|-----------|-----|
|          |   | Post-test |     |
|          |   | 0         | 1   |
| Pre-test | 0 | .41       | .59 |
|          | 1 | .14       | .86 |

# Alternative Analysis

- Can use gain scores and repeated measures MANOVA to analyze this data
  - » Interaction effects test differential growth
- Could also use a longitudinal IRT model
- Why use a DCM?
  - » Two main reasons

# Interpretive Value

- Effect size measures
  - » TDCM: difference in groups' growth in attribute mastery
  - » Gain scores: average difference in groups' gain scores
  - » Longitudinal IRT: difference in groups' growth in ability
- Effect sizes:

| Attribute              | TDCM     | Gain Scores | Long IRT |
|------------------------|----------|-------------|----------|
| Ratios and proportions | .18 (72) | .49         | .51      |



# Reliability

- Refers to the consistency of the growth estimates
  - » Are they trustworthy?

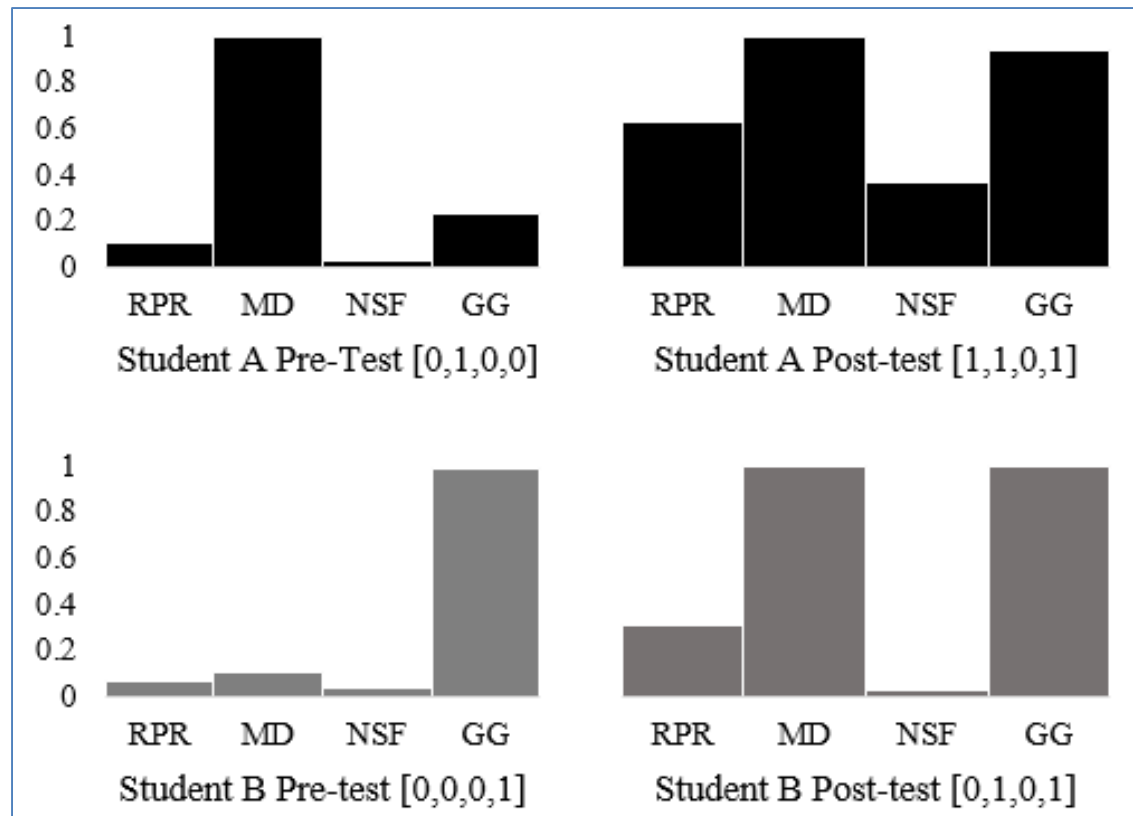
| Attribute              | TDCM | Gain Scores | Long IRT |
|------------------------|------|-------------|----------|
| Ratios and proportions | .915 | .488        | .550     |
| Measurement and data   | .968 | .651        | .566     |
| Number Systems         | .928 | .603        | .631     |
| Geometry (graphing)    | .911 | .619        | .638     |

*Beautiful*



# Interpretive Value

- Individual estimates
- Two examinees (Student A and B):
  - » Pre-test = 9/21
  - » Post-test = 12/21
  - » Gain Score = +3



# Conclusions

- Analyzed data from a randomized controlled trial
- Evaluated efficacy of EAI
  - » Used TDCM, which has benefits over traditional analysis
  - » Analysis suggests EAI better than traditional instruction
- Overall, EAI was successful
  - » But...could be improved for number systems and measurement and data
- The TDCM provides a meaningful measure of growth



# Estimation in R

The same group of biology education researchers wanted to gauge the degree to which the blended learning approach could be successful for other foundational biological concepts. To evaluate the efficacy of the blended learning approach, they conducted a quasi-experiment where instructors (and the students in their respective classes) were either assigned to the control group (traditional instruction;  $N = 1057$ ) or the treatment group (blended learning;  $N = 1104$ ). The pre- and post-test were identical and consisted of 20 items measuring four attributes (respiration, reproduction, DNA structure, protein synthesis). The test was administered at the beginning and end of the semester.

- In the Examples folder, open “example3.R”
- Four attributes, 20 items
- Two time points
- Two groups: control (1057) and treatment (1104)

# Break

- Take a quick 5-minute break
- Next up: Q&A and discussion



# Wrapping Up

- We discussed longitudinal DCM advancements
  - TDCM is an extension that combines LTA with the LCDM
  - Provides criterion-referenced estimates of growth
  - Can be more meaningful and interpretable
- All estimated in R
  - Much simpler than other software

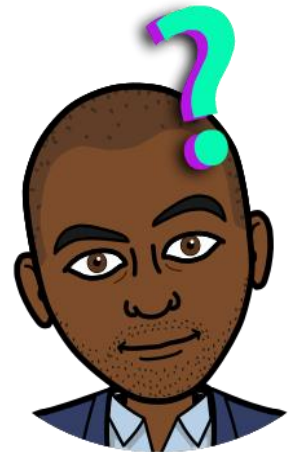
# Questions/Discussion

- Open forum for questions and discussion regarding:
  - » Experiences with DCMs items, tests, models, estimation, software, etc.
  - » Ideas within additional material
    - Further explanation/questions
  - » Ideas outside of this workshop material so far
    - Is there a topic you had hoped we would include?



# Questions/Discussion

- Other questions?
  - » Email us: [mjmadison@uga.edu](mailto:mjmadison@uga.edu) and [madeline.schellman@pearson.com](mailto:madeline.schellman@pearson.com)
  - » **Join DCMNET** ([www.matthewmadison.com](http://www.matthewmadison.com))
    - Network for DCM research and practice ( $\approx$  SEMNET)
    - Access to DCM experts
  - » Join the diagnostic measurement SIGMIE
  - » Access the DCM [database](#)



# Last Favor

- **We would love your feedback**

1. NCME Evaluation



2. Click [here](#) for a brief feedback form on the TDCM package