Diagnostic Classification Models: Psychometric Issues and Statistical Challenges

Jonathan Templin
Department of Educational Psychology
The University of Georgia
Talk Overview

- **Diagnostic Classification Models (DCMs)**
  - The Log-linear Cognitive Diagnosis Model (LCDM)
  - Modeling philosophy and modeling strategy

- **Properties of DCMs**
  - Reliability
  - Estimation
  - Similarity to other psychometric/statistical models

- **Current and future research**
  - Attribute hierarchies
  - Longitudinal models
CONCEPTUAL FOUNDATIONS OF DIAGNOSTIC CLASSIFICATION MODELS
• A diagnosis is a **decision** that is made based on information from data (typically test items)

• Within psychological testing, providing a test score gives the information that is used for a diagnosis
  - The score is *not* the diagnosis

• For this talk, a diagnosis is by its nature a **discrete** status
  - Diagnoses are *Classifications*
    - Diagnostic classification models (DCMs; Rupp & Templin, 2008; Rupp, Templin, & Henson, 2010)
      - Categorical latent variable models that directly classify respondents
Diagnostic classification models (DCMs) have been called many different things (i.e., Rupp & Templin, 2008)

- Cognitive diagnosis models
- Skills assessment models
- Cognitive psychometric models
- Latent response models
- Restricted (constrained) latent class models
- Multiple classification models
- Structured located latent class models
- Structured item response theory
• Diagnostic decisions come from comparing observed behaviors to two parts of the psychometric model:

1. Item/variable information (item parameters)
   - How respondents with different diagnostic profiles perform on a set of test items
   - Helps determine which items are better at discriminating between respondents with differing diagnostic profiles

2. Respondent information pertaining to the base-rate or proportion of respondents with diagnoses in the population
   - Provides frequency of diagnosis (or diagnostic profile)
   - Base rates include associations among attributes
EVOLUTIONARY LINEAGES IN DCMS
DCMs have ancestral lineages that link them to (at least) three fields:

**Mathematical Psychology**
- Knowledge Spaces (e.g., Doignon & Falmagne, 1985)

**Clustering/Classification**
- Rule Space Methods (e.g., Tatsuoka, 1983)
- Attribute Hierarchy Method (e.g., Leighton, Gierl, Hunka, 2004)

**Item Response Theory**
- Mastery Model (Macready & Dayton, 1977)
- Restricted Latent Class Models (Haertel, 1989)
- DINA Model (Junker & Sijtsma, 2001)
- General Latent Trait Model (Embretson, 1984)
- General(ized) DCMs

- Multicomponent Latent Trait Model (Whitely, 1980)
- Linear Logistic Test Model (Fischer, 1973)
- Reparameterized Unified Model(s) (Hartz, 2002)
- Newer Clustering Methods (e.g., Douglas, Junker, Nugent, ...)

- Bayesian Inference Networks (e.g., Almond & Mislevy, 1999)
- Knowledge Spaces (e.g., Doignon & Falmagne, 1985)
The general latent class model defines the probability of observing a response vector $\mathbf{x}_r$ from respondent $r$ as:

$$P(\mathbf{x}_r) = \sum_{c=1}^{C} \eta_c \prod_{i=1}^{I} \pi_{ic}^{x_{ri}} (1 - \pi_{ic})^{(1-x_{ri})}$$

- Mixture of conditionally independent Bernoulli distributions
  - $\eta_c$ is the probability a respondent is from class $c$ ($c=1,\ldots,C$)
  - $\pi_{ic}$ is the probability that a respondent from class $c$ correctly responds to the $i^{th}$ item ($i=1,\ldots,I$)

- Exploratory technique: unknown number of classes a priori
DCMs are Confirmatory LCA Models

• DCMs are confirmatory LCA models
  ➢ Most defined for a set of $A$ dichotomous attributes ($\alpha$)
    • Attributes are either possessed ($\alpha = 1$) or not ($\alpha = 0$)
    • DCM attributes can have more than two levels
  ➢ DCMs are LCA models with $2^A$ latent classes
    • Each possible combination of attribute possession
    • i.e., a test measuring 3 dichotomous attributes has 8 latent classes

• LCA structural model parameters ($\eta_c$)
  ➢ Becomes attribute association model
    • Number of parameters can be further reduced
      (e.g., de la Torre & Douglas, 2004)
    • Structural hypotheses can be tested

• LCA measurement model parameters ($\pi_{ic}$)
  ➢ Items measure only some attributes (so-called Q-matrix indicator)
  ➢ Equated for classes with equivalent status of measured attributes
### Example DCM Constraints

#### Latent Classes

#### Attribute Patterns

#### Latent Class Item Parameters (Same Color = Same Value)

<table>
<thead>
<tr>
<th>(c)</th>
<th>(x)</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[0000]</td>
<td>(\pi_{1,1})</td>
<td>(\pi_{2,1})</td>
<td>(\pi_{3,1})</td>
</tr>
<tr>
<td>2</td>
<td>[0001]</td>
<td>(\pi_{1,2})</td>
<td>(\pi_{2,2})</td>
<td>(\pi_{3,2})</td>
</tr>
<tr>
<td>3</td>
<td>[0010]</td>
<td>(\pi_{1,3})</td>
<td>(\pi_{2,3})</td>
<td>(\pi_{3,3})</td>
</tr>
<tr>
<td>4</td>
<td>[0011]</td>
<td>(\pi_{1,4})</td>
<td>(\pi_{2,4})</td>
<td>(\pi_{3,4})</td>
</tr>
<tr>
<td>5</td>
<td>[0100]</td>
<td>(\pi_{1,5})</td>
<td>(\pi_{2,5})</td>
<td>(\pi_{3,5})</td>
</tr>
<tr>
<td>6</td>
<td>[0101]</td>
<td>(\pi_{1,6})</td>
<td>(\pi_{2,6})</td>
<td>(\pi_{3,6})</td>
</tr>
<tr>
<td>7</td>
<td>[0110]</td>
<td>(\pi_{1,7})</td>
<td>(\pi_{2,7})</td>
<td>(\pi_{3,7})</td>
</tr>
<tr>
<td>8</td>
<td>[0111]</td>
<td>(\pi_{1,8})</td>
<td>(\pi_{2,8})</td>
<td>(\pi_{3,8})</td>
</tr>
<tr>
<td>9</td>
<td>[1000]</td>
<td>(\pi_{1,9})</td>
<td>(\pi_{2,9})</td>
<td>(\pi_{3,9})</td>
</tr>
<tr>
<td>10</td>
<td>[1001]</td>
<td>(\pi_{1,10})</td>
<td>(\pi_{2,10})</td>
<td>(\pi_{3,10})</td>
</tr>
<tr>
<td>11</td>
<td>[1010]</td>
<td>(\pi_{1,11})</td>
<td>(\pi_{2,11})</td>
<td>(\pi_{3,11})</td>
</tr>
<tr>
<td>12</td>
<td>[1011]</td>
<td>(\pi_{1,12})</td>
<td>(\pi_{2,12})</td>
<td>(\pi_{3,12})</td>
</tr>
<tr>
<td>13</td>
<td>[1100]</td>
<td>(\pi_{1,13})</td>
<td>(\pi_{2,13})</td>
<td>(\pi_{3,13})</td>
</tr>
<tr>
<td>14</td>
<td>[1101]</td>
<td>(\pi_{1,14})</td>
<td>(\pi_{2,14})</td>
<td>(\pi_{3,14})</td>
</tr>
<tr>
<td>15</td>
<td>[1110]</td>
<td>(\pi_{1,15})</td>
<td>(\pi_{2,15})</td>
<td>(\pi_{3,15})</td>
</tr>
<tr>
<td>16</td>
<td>[1111]</td>
<td>(\pi_{1,16})</td>
<td>(\pi_{2,16})</td>
<td>(\pi_{3,16})</td>
</tr>
</tbody>
</table>
RECENT DEVELOPMENTS IN LATENT VARIABLE DCMS
Development of Psychometric Models

• Over the past several years, numerous latent variable DCMs have been developed
  ➢ Focus will be on DCMs that use latent variables for attributes

• Each DCM makes assumptions about how mastered attributes combine/interact to produce an item response
  ➢ Compensatory/disjunctive/additive models
  ➢ Non-compensatory/conjunctive/non-additive models

• With many models different models, analysts have been unsure which model would best fit their purpose
  ➢ Difficult to imagine all items following same assumptions
Recent developments have produced very general diagnostic models:

- General Diagnostic Model (GDM; von Davier, 2005)
- Loglinear Cognitive Diagnosis Model (LCDM; Henson, Templin, & Willse, 2009)
- Generalized DINA Model (de la Torre, 2011)

The general DCMs provide great flexibility:

- Subsume most latent variable DCMs
- Allow for both additive and non-additive relationships between attributes and items
- Sync with other psychometric models allowing for greater understanding of modeling process
  - Greater links to more general modeling methods
General Form of the LCDM

• The LCDM specifies the probability of a correct response as a function of a set of attributes and a Q-matrix:

\[
\pi_{ic} = P(X_{ri} = 1 \mid \alpha_r = \alpha_c) = \frac{e^{\lambda_i^T h(q_i, \alpha_r)}}{1 + e^{\lambda_i^T h(q_i, \alpha_r)}}
\]

• Where:

\[
\lambda_i^T h(q_i, \alpha_r) = \lambda_{i,0} + \sum_{u=1}^{K} \lambda_{i,1,(u)}(\alpha_{ru} q_{iu}) + \sum_{u=1}^{K} \sum_{v > u}^{K} \lambda_{i,2,(u,v)}(\alpha_{ru} \alpha_{rv} q_{iu} q_{iv}) + \ldots
\]

Latent Class Item Parameter
Logit($X_{ri} = 1 \mid \alpha_r$)
Intercepts
Main Effects
Two-Way Interactions
Higher Interactions
Imagine we obtained the following estimates for an item measuring two attributes:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Effect Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{i,0}$</td>
<td>-2</td>
<td>Intercept</td>
</tr>
<tr>
<td>$\lambda_{i,1,(1)}$</td>
<td>2</td>
<td>Simple Main Effect of Attribute 1</td>
</tr>
<tr>
<td>$\lambda_{i,1,(2)}$</td>
<td>1</td>
<td>Simple Main Effect of Attribute 2</td>
</tr>
<tr>
<td>$\lambda_{i,2,(1,2)}$</td>
<td>0</td>
<td>Interaction of Attributes 1 and 2</td>
</tr>
</tbody>
</table>
## LCDM Predicted Logits and Probabilities

### LCDM Logit Function

<table>
<thead>
<tr>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>LCDM Logit Function</th>
<th>Logit</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>$\lambda_{i,0} + \lambda_{i,1,(1)}<em>(0) + \lambda_{i,1,(2)}</em>(0) + \lambda_{i,2,(1,2)}<em>(0)</em>(0)$</td>
<td>-2</td>
<td>0.12</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>$\lambda_{i,0} + \lambda_{i,1,(1)}<em>(0) + \lambda_{i,1,(2)}</em>(1) + \lambda_{i,2,(1,2)}<em>(0)</em>(1)$</td>
<td>-1</td>
<td>0.27</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>$\lambda_{i,0} + \lambda_{i,1,(1)}<em>(1) + \lambda_{i,1,(2)}</em>(0) + \lambda_{i,2,(1,2)}<em>(1)</em>(0)$</td>
<td>0</td>
<td>0.50</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>$\lambda_{i,0} + \lambda_{i,1,(1)}<em>(1) + \lambda_{i,1,(2)}</em>(1) + \lambda_{i,2,(1,2)}<em>(1)</em>(1)$</td>
<td>1</td>
<td>0.73</td>
</tr>
</tbody>
</table>

### Logit Response Function

![Logit Response Function Graph]

### Probability Response Function

![Probability Response Function Graph]
No Latent Variable Interaction

- **No interaction**: parallel lines for the logit
  - Compensatory RUM (Hartz, 2002)

\[
\text{Logit}(X_{ri} = 1 | \alpha_r) = \lambda_{i}^{T} h(q_i, \alpha_r) = \lambda_{i,0} + \lambda_{i,1,(1)} \alpha_{r1} + \lambda_{i,1,(2)} \alpha_{r2}
\]

Logit Response Function

Probability Response Function

![Graphs showing logit and probability response functions with different parameter values.](image)


**Strong Positive Interactions**

- **Positive interaction**: over-additive logit model
  - Conjunctive model (i.e., all-or-none)
  - DINA model (Haertel, 1989; Junker & Sijtsma, 1999)

---

### Logit Response Function

- $\alpha_1 = 0$
- $\alpha_1 = 1$

### Probability Response Function

- $P(X=1|\alpha)$

---

Possible Attribute Patterns:

- $\alpha_1 = 0; \alpha_2 = 0$
- $\alpha_1 = 0; \alpha_2 = 1$
- $\alpha_1 = 1; \alpha_2 = 0$
- $\alpha_1 = 1; \alpha_2 = 1$
DINA: Only Highest Interaction

- DINA Model (Haertel, 1989; Junker & Sijtsma, 1999):

  \[ P(X_{ri} = 1 \mid \alpha_r) = (1 - s_i) \xi_{ri} g_i^{1 - \xi_{ri}} \]

  \[ \xi_{ri} = \prod_{a=1}^{A} \alpha_{ra}^{q_{ia}} \]

- Under LCDM:

  \[ \text{Logit}(X_{ri} = 1 \mid \alpha_r) = \lambda_{i,0} + \lambda_{i,2,(1,2)} \alpha_{r1} \alpha_{r2} \]

  - Highest order interaction positive
  - No lower order effects present

University of South Carolina Talk
• **Negative interaction**: under-additive logit model
  - Disjunctive model (i.e., one-or-more)
  - DINO model (Templin & Henson, 2006)
DINO: Interaction Cancels Out Main Effects

- **DINO Model** (Templin & Henson, 2006):

\[
P(X_{ri} = 1 | \alpha_r) = \left(1 - s_i \right)^{\omega_{ri}} g_i^{1-\omega_{ri}}
\]

\[
\omega_{ri} = 1 - \prod_{a=1}^{A} (1 - \alpha_{ra})^{q_{ri}}
\]

- **Under LCDM:**

\[
\text{Logit}(X_{ri} = 1 | \alpha_r) = \lambda_{i,0} + \lambda_{i,1} \alpha_{r1} + \lambda_{i,2} \alpha_{r2} - \lambda_{i,1} \alpha_{r1} \alpha_{r2}
\]

- Highest order cancels out gains from lower order main effects
Less Extreme Interactions

- Extreme interactions are unlikely in practice
- Below: positive interaction with positive main effects

\[
\text{Logit}(P(X_{ri} = 1|\alpha_r)) = \lambda_{i,0} + \lambda_{i,1,1}(\alpha_{r1} + \lambda_{i,1,2}\alpha_{r2} + \lambda_{i,2,1,2}\alpha_{r1}\alpha_{r2})
\]

Logit Response Function

Probability Response Function

Possible Attribute Patterns
Linking DCMs with MIRT

- DCMs and Multidimensional IRT models are similar
  - Different distributional assumptions for the latent variables (i.e., Templin & Rupp, in press)
    \[
    \text{Logit}(P(X_{ri} = 1 | \gamma_r)) = h(q_i, \gamma_r)
    \]

- DCMs typically have:
  - Confirmatory loading patterns (set by \(q_i\))
  - Latent variable interactions (defined in \(h(\cdot)\))

- Under MIRT (\(\gamma_r\) typically denoted \(\theta_r\)): \(\gamma_r \sim N_A(\mu_\gamma, \Sigma_\gamma)\)
  - Assumed multivariate normal distribution for latent variables

- Under DCMs (\(\gamma_r\) typically denoted \(\alpha_r\)): \(\gamma_r \sim MVBA(\eta)\)
  - Assumed multivariate Bernoulli distribution for latent variables
  - \(\eta\) is vector of LCA structural model parameters (class probabilities)
Estimation Methods

- Bayesian methods:
  - Software such as Arpeggio
    (DiBello, Stout, et al.; Assessment Systems Corporation)
  - Psychometrician created code
    (e.g., R, WinBUGS, FORTRAN, MATLAB)

- Marginal Maximum Likelihood:
  - Can be done in any package with LCA and constraints:
    - LCDM in Mplus (Templin & Hoffman, under review)
    - GDM in MLDTM (general latent variable package by von Davier)
      - Lacks ability to estimate interactions
    - GDINA in Ox (De la Torre, 2011)
      - Interaction terms are not directly testable
MOTIVATION FOR DCM USE
Why Use DCMs?

- Other psychometric approaches have been developed for measuring multiple dimensions
  - Classical Test Theory - Scale Subscores
  - Multidimensional Item Response Theory (MIRT)

- Yet, issues in application have remained:
  - Reliability
  - Large samples are needed
  - Dimensions are often very highly correlated

- Many tests are used for classification: “proficiency”

- DCMs can provide direct “proficiency” classifications
Nature of DCM Attributes: Dichotomous?

• Perhaps the biggest question about DCMS: are latent attributes truly dichotomous?
  ➢ Some theories of learning say yes
  ➢ More realistically, the answer is no…but…

• DCMs can also be thought of as approximations to MIRT models (Haberman, von Davier, & Lee, 2008)
  ➢ Approximations allow for:
    ▶ Quick non-parametric (M)IRT estimation
      – Loosening linearity constraints in 3+ category attribute DCMs
  ➢ Explorations of latent variable interactions
    ▶ LCDM can provide parameters/hypothesis tests
  ➢ Attribute hierarchies can be tested
    ▶ Using structural parameters in DCMs
ATTRIBUTE RELIABILITY IN DCMS
Attribute Reliability in DCMs

- Until recently, reliability of attributes in DCMs was not thought much about
  - Is needed for use in practice – need quantification of how accurate attributes are measured

- Studying reliability is difficult
  - Classical notions of true score variance and total variance are different due to categorical nature of attributes
  - Comparing reliability of dichotomous (categorical) attributes to continuous latent variables in IRT is very difficult

- Difficulty doesn’t stop us...only makes it harder to tell
• Consider reliability as a measure of consistency of an estimated latent trait

• We can quantify the reliability of a latent trait by considering how stable successive draws from a posterior distribution are
  ➢ Making some assumptions about posterior distribution

• This method would allow for rough comparisons of either categorical or continuous traits
Reliability for Continuous Traits

- For continuous traits: \( \theta_r \sim N(\hat{\theta}_r, SE(\hat{\theta}_r)) \)

- Across all respondents \( r \), if we drew multiple values \((\theta_1^*, \theta_2^*)\) from these distributions, we could quantify reliability as:
  \[
  \rho_\theta = CORR(\theta_1^*, \theta_2^*)
  \]

- If \( SE(\hat{\theta}_r) \) was constant \( \rho_\theta \) would be an intraclass correlation
  - In IRT, \( SE(\hat{\theta}_r) \) depends on \( \hat{\theta} \) - so index of reliability describes accuracy of measurement across range of \( \theta \)

- If \( SE(\hat{\theta}_r) = 0, \rho_\theta = 1; \) as \( SE(\hat{\theta}_r) \to \infty, \rho_\theta \to 0 \)
• For dichotomous attributes: $\hat{\alpha} \sim B(\hat{p}_r); \alpha \in \{0,1\}$

• Across all respondents $r$, if we drew multiple values $(\alpha_1^*, \alpha_2^*)$ from these distributions, we could quantify reliability as:
  
  $$\rho_\alpha = TCORR(\alpha_1^*, \alpha_2^*)$$

  ➢ Tetrachoric correlation used as $\rho_\alpha$ is bounded below usual range (-1,1) for marginal $\hat{p}_\alpha \neq .5$

• If for all respondents:
  
  ➢ $\hat{p}_r = 1$ or $\hat{p}_r = 0$ then $\rho_\alpha = 1$: perfect consistency
  ➢ $\hat{p}_r = .5$ then $\rho_\alpha = 0$: assignment is random

• Attribute reliability is closely related to classification rate
### Theoretical Reliability Comparison

#### Reliability vs. Number of Items

<table>
<thead>
<tr>
<th>Reliability Level</th>
<th>DCM</th>
<th>IRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80</td>
<td>8 Items</td>
<td>34 Items</td>
</tr>
<tr>
<td>0.85</td>
<td>10 Items</td>
<td>48 Items</td>
</tr>
<tr>
<td>0.90</td>
<td>13 Items</td>
<td>77 Items</td>
</tr>
</tbody>
</table>

From: Templin and Bradshaw (in press)

---

**Diagram:**
- **DCM** and **Rasch IRT** reliability comparisons.
- Graph shows the increase in reliability as the number of items increases.

---

**Table:**
- Reliability levels and corresponding number of items for DCM and IRT.
Uni- and Multidimensional Comparison

DCM

IRT

1.0

0.9

0.8

0.7

0.6

0.5

1-Dimension

2-Dimension

BiFactor

DCM

IRT

DCM

IRT

BiFactor

Dimensional Model
DCMs for an EOC Test

2PL $\rho_{\theta} = .87$

Number of Items

Reliability

- 2 Category: 24 Items
- 3 Category: 42 Items
- 4 Category: 50 Items
- 5 Category: 54 Items
Ramifications for Use of DCMs

- Such characteristics allow DCMs to potentially change how large scale testing is conducted (Templin & Henson, 2009)
  - Investigated reliability in End-of-Grade tests in Midwestern State (USA)

- Reliable measurement of multiple dimensions
  - Two-attribute DCM application to empirical data:
    - Reliabilities of 0.95 and 0.90 (compared to 0.72 and 0.70 for IRT)

- Multidimensional proficiency standards
  - Respondents must demonstrate proficiency on multiple areas to be considered proficient for an overall content domain
  - “Teaching to the test” would therefore represent covering more curricular content to best prepare respondents

- Shorter unidimensional tests
  - Two-category unidimensional DCM application to empirical data:
    - Test needed only 24 items to have same reliability as IRT with 73 items
EVALUATING ATTRIBUTE HIERARCHIES
• Often attribute hierarchies exist in education
  ➢ Part of cognitive theory

• Diagnostic modeling methods exist for hierarchies
  ➢ Attribute Hierarchy Method
  ➢ Rule Space Method

• Neither approach allows for:
  ➢ Statistical hypothesis test for attribute hierarchies
  ➢ Analysis of attribute hierarchies with latent class-based DCMs
The Hierarchical Diagnostic Classification Model

- Whereas the LCDM represented a crossed-factors ANOVA model, the HDCM uses nested factors
  - Profiles not possible are not estimated (no longer $2^A$)
    • Reduces number of structural model parameters (probabilities)
    • Changes nature of item parameters (nested interactions)
  - HDCM is nested within LCDM
    • Allows for hypothesis test for attribute hierarchy

- Under LCDM and item measuring two crossed attributes:
  \[
  \logit(X_{e7} = 1 | \alpha_e) = \lambda_{7,0} + \lambda_{7,1,(1)} \alpha_{e1} + \lambda_{7,1,(3)} \alpha_{e3} + \lambda_{7,2,(1,3)} \alpha_{e1} \alpha_{e3}
  \]

- Under HDCM with Attribute 1 nested within attribute 3:
  \[
  \logit(X_{e7} = 1 | \alpha_e) = \lambda_{7,0} + \lambda_{7,1,(3)} \alpha_{e3} + \lambda_{7,2,(1(3))} \alpha_{e1(3)} \alpha_{e3}
  \]
• The HDCM with an attribute hierarchy can be phrased as a model nested within the LCDM

• A nested-model comparison test can be constructed
  - Deviance test: -2 difference in model log-likelihood values

• Deviance test does not follow typical Chi-Square distribution
  - HDCM fixes LCDM model parameters at boundaries

• Test is mixture of Chi-Squares
  - Cannot easily be derived analytically
  - Simulation can approximate p-value
  - If naïve test used, likely result is conservative Type-I error rates
• The ECPE is a test developed and scored by the English Language Institute of the University of Michigan
  • Measures advanced English ability in respondents for which English is not their first language

• LCDM analysis of grammar section of the ECPE
  • 28 multiple choice items
  • 3 purported attributes: morphosyntactic, cohesive, and lexical rules
    – 19 items measure one attribute
    – 9 items measure two attributes
    – 0 items measure three attributes
95% of base-rates of profiles indicated a linear attribute hierarchy

Is remaining 5% meaningful or simply statistical noise?
Analysis of the ECPE with the HDCM

• The suspected attribute hierarchy in the ECPE was evaluated with the HDCM
  ➢ Morphosyntactic nested within Cohesive nested within Lexical

• Classification of examinees for HDCM and LCDM had high overlap: 93.6% agreement (.909 kappa)

• Results presented:
  ➢ Structural model
  ➢ Example item
  ➢ Hypothesis test results
Most of 5% shifted into Classes 1 and 2

Structural Model Parameter Estimates for the ECPE Data

<table>
<thead>
<tr>
<th>Morphosyntactic</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohesive</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Lexical</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Attribute Profile c</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>$\hat{\eta}_c$ - LCDM (solid)</td>
<td>.301</td>
<td>.129</td>
<td>.012</td>
<td>.175</td>
<td>.009</td>
<td>.018</td>
<td>.011</td>
<td>.346</td>
</tr>
<tr>
<td>$\hat{\eta}_c$ - HDCM (dashed)</td>
<td>.320</td>
<td>.144</td>
<td>-</td>
<td>.184</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.351</td>
</tr>
</tbody>
</table>
Item 7 HCDM and LCDM Estimates

- **Item 7:**
  \[
  \text{logit}(X_{e7} = 1 | \alpha_e) = \lambda_{7,0} + \lambda_{7,1,(3)} \alpha_{e3} + \lambda_{7,2,(1(3))} \alpha_{e1(3)} \alpha_{e3}
  \]

  - Morphosyntactic rules (Attribute 1)
  - Lexical rules (Attribute 3)

- **Parameter estimates:**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>HDCM Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{7,0}$</td>
<td>-0.040</td>
<td>0.088</td>
</tr>
<tr>
<td>$\lambda_{7,1,(3)}$</td>
<td>0.924</td>
<td>0.140</td>
</tr>
<tr>
<td>$\lambda_{7,2,(1(3),3)}$</td>
<td>1.949</td>
<td>0.229</td>
</tr>
</tbody>
</table>

LCDM Response Function

HDCM Response Function

HDCM does not model [1,,0] due to hierarchy
HDCM Ramifications

- Attribute hierarchies are present in many data sets
  - Can be indicative of less dimensionality
  - More reasons *not* to use DINA (or DINO)

- Until the HDCM, latent class-based DCMs have not been able to be adapted for hierarchies
  - No model was able to test for existence of hierarchy

- The HDCM fills these gaps and can help researchers test theories about nature of hypotheses
LONGITUDINAL DCMS
Longitudinal Modeling of Attributes

- Field of educational measurement is rapidly looking to longitudinal models to chart student progress
  - Learning progressions
  - Value added models

- Effective feedback from tests is multidimensional
  - Can provide information to help tailor learning plans

- Multidimensional feedback is difficult to attain in practice
  - Motivation for DCM use
Longitudinal DCMs

- If longitudinal DCMs can be developed and estimated, more informative feedback can be given throughout the academic year.

- As DCMs are types of latent class models, we can start by envisioning longitudinal versions of such models:
  - Latent Markov models/latent transition models

\[
P(X_p = x_p) = \sum_{s_1=1}^{S_1} \sum_{s_2=1}^{S_2} \cdots \sum_{s_M=1}^{S_M} \eta_{s_1,s_2,\ldots,s_M} \prod_{m=1}^{M} \prod_{i=1}^{I} \pi_{i,s_m}^{x_{p,m,i}} (1 - \pi_{i,s_m})^{1-x_{p,m,i}}
\]
Statistical Issues

• Direct application of LMM methods are inappropriate:
  
  ➢ With \(2^A\) attributes measured at a given occasion, number of transition probability parameters becomes huge
    ✷ For test with 10 attributes and 4 occasions: over 1 trillion

  ➢ General methods do not allow for unbalanced time
    ✷ Students will not all take tests at same time in academic year

  ➢ Tests may not all have the same items

• NSF MMS funding to develop this idea further
Possible Solution?

• As a possible solution, we could consider the higher-order modeling of each attribute
  
  ➢ Provide a multivariate growth model using a generalized linear mixed model (but for latent attributes)

• Probability of mastery for responant \( r \), attribute \( a \), at time \( t_{rm} \) as a function of fixed and random effects:

\[
\varphi_{rat_{rm}} = P(\alpha_{rat_{rm}} = 1|t_{rm}, u_r) = \frac{\exp(\beta_{0a} + \beta_{1a}t_{rm} + u_{0ra} + u_{1ra}t_{rm})}{1 + \exp(\beta_{0a} + \beta_{1a}t_{pm} + u_{0ra} + u_{1ra}t_{rm})}
\]
Distribution of Random Effects

- For the random intercept/slope:

\[
\mathbf{u_p} = \begin{bmatrix}
    u_{0p1} \\
    u_{1p1} \\
    \vdots \\
    u_{0pA} \\
    u_{1pA}
\end{bmatrix} \sim N(\mathbf{0}_{(2A \times 1)}, \Sigma_{(2A \times 2A)})
\]

\[
\Sigma = \begin{bmatrix}
    \sigma_{01}^2 & \sigma_{01,11} & \cdots & \sigma_{01,0A} & \sigma_{11,0A} \\
    \sigma_{01,11} & \sigma_{11}^2 & \cdots & \sigma_{01,1A} & \sigma_{11,1A} \\
    \vdots & \vdots & \ddots & \vdots & \vdots \\
    \sigma_{01,0A} & \sigma_{11,0A} & \cdots & \sigma_{0A}^2 & \sigma_{0A,AA} \\
    \sigma_{11,0A} & \sigma_{11,1A} & \cdots & \sigma_{0A,AA} & \sigma_{1A}^2
\end{bmatrix}
\]
Transition Probabilities Function

- With such a model, the overall transition probability is expressed as the marginal product of attribute probabilities:

\[ \eta_{s_1, s_2, \ldots, s_M} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{m=1}^{M} \prod_{a=1}^{A} \varphi_{pat_{pm}} (1 - \varphi_{pat_{pm}})^{1-\alpha_{sa}} f(u_p; 0, \Xi) du_{0p1} \cdots du_{1pA} \]

- Model reduces number of needed parameters while allowing for additional modeling of covariates (such as time)

- Other issues: measurement invariance of attributes across covariates (especially time)
CONCLUDING REMARKS
Concluding Remarks

• DCMs provide direct link between diagnosis and behavior
  - Provide diagnostic classifications directly
  - Diagnoses set by psychometric model parameters

• DCMs can be used in many contexts
  - Can be used to create highly informative tests
  - Can be used to measure multiple dimensions

• Applications of DCMs are in their infancy
  - Not many tests have been built for use with DCMs
  - Time will tell their effectiveness
The Paradox of DCMs

• DCMs are often pitched as models that allow for measurement of “fine-grained” skills (e.g., Rupp & Templin, 2008)

• Paradox of DCMs:
  - Coarse measurement of a latent trait for only several categories
  - Increased capacity to measure psychological traits multidimensionality
• Questions? Comments? References?
  ➢ Email: jtemplin@uga.edu

• Thank you!