Diagnostic Test Design with Blueprint Specifications

Matthew J. Madison & Nancy Alila
University of Georgia

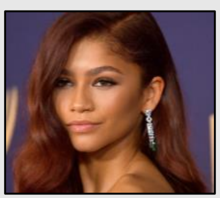
2025 Meeting of the National Council on Measurement in Education

Talk Overview

- Background
  - » Introduction to DCMs
  - » Item influence
  - » DCM empirical blueprints
- Simulated scenarios
  - » Two models and two assessment scenarios
- Results and conclusions

Diagnostic Classification Models

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

Student	Addition	Subtraction	Multiplication	Division
	✓	—	✓	—
	✓	✓	✓	✓

- Defining features: multidimensionality and diagnostic interpretations

Item Influence

- DCM applications use fewer items

- ***Item influence*** – one item (or subset) can have a disproportionate impact on classifications (Jurich & Madison, 2023)
  - » Problematic for construct and content validity

- Item influence metrics:
  - » Item override: how many classifications change if an item is omitted?
  - » **Proportion of attribute information**

Attribute Information

- Item response theory information
  - » Item information function → test information function

- Analogous concept for DCMs
  - » Cognitive diagnostic index (CDI; Henson & Douglas, 2005)
  - » All the items measuring an attribute contribute to the overall information

Item	CDI	Compute	Proportion of Att Info
1	.24		
2	.44		
3	.36		
Total	1.04		

Attribute Information

- Item response theory information
  - » Item information function → test information function

- Analogous concept for DCMs
  - » Cognitive diagnostic index (CDI; Henson & Douglas, 2005)
  - » All the items measuring an attribute contribute to the overall information

Item	CDI	Compute	Proportion of Att Info
1	.24	.24 ÷ 1.04	
2	.44		
3	.36		
Total	1.04		

Attribute Information

- Item response theory information
  - Item information function → test information function
- Analogous concept for DCMs
  - Cognitive diagnostic index (CDI; Henson & Douglas, 2005)
  - All the items measuring an attribute contribute to the overall information

Item	CDI	Compute	Proportion of Att Info
1	.24	.24 ÷ 1.04	23%
2	.44		
3	.36		
Total	1.04		

Attribute Information

- Item response theory information
  - Item information function → test information function
- Analogous concept for DCMs
  - Cognitive diagnostic index (CDI; Henson & Douglas, 2005)
  - All the items measuring an attribute contribute to the overall information

Item	CDI	Compute	Proportion of Att Info
1	.24	.24 ÷ 1.04	23%
2	.44	.44 ÷ 1.04	42%
3	.36	.36 ÷ 1.04	35%
Total	1.04	1.04 ÷ 1.04	100%

Blueprint Example

- Test blueprints guide test development efforts
- Common Core Mathematics Standards
  - » Grade 4 Measurement and Data

Attribute

Two Subattributes

Geometric measurement: understand concepts of angle and measure angles.

5. Recognize angles as geometric shapes that are formed wherever two rays share a common endpoint, and understand concepts of angle measurement:
   a. An angle is measured with reference to a circle with its center at the common endpoint of the rays, by considering the fraction of the circular arc between the points where the two rays intersect the circle.
   b. An angle that turns through n one-degree angles is said to have an angle measure of n degrees.

Blueprint Example

- Suppose in this case, that test developers had prespecified blueprint proportions of 25% and 75%
  - » Then they might allocate 2 and 6 items to the two subattributes

Item	Proportion of Att Information	Empirical Blueprint	Prespecified Blueprint
1			25%
2			
3			75%
4			
5			
6			
7			
8			

Blueprint Example

- Suppose in this case, that test developers had prespecified blueprint proportions of 25% and 75%
  - » Then they might allocate 2 and 6 items to the two subattributes

Item	Proportion of Att Information	Empirical Blueprint	Prespecified Blueprint
1	12		25%
2	28		
3	10		75%
4	8		
5	8		
6	9		
7	11		
8	14		

Blueprint Example

- Suppose in this case, that test developers had prespecified blueprint proportions of 25% and 75%
  - » Then they might allocate 2 and 6 items to the two subattributes

Item	Proportion of Att Information	Empirical Blueprint	Prespecified Blueprint
1	12	40%	25%
2	28		
3	10	60%	75%
4	8		
5	8		
6	9		
7	11		
8	14		

Blueprint Example

- Suppose in this case, that test developers had prespecified blueprint proportions of 25% and 75%
  - » Then they might allocate 2 and 6 items to the two subattributes
  - » This allocation does not guarantee a match

- **The purpose of this study is to examine the ability of DCMs to adhere to prespecified blueprints**

General and Constrained DCMs

- General model: log-linear cognitive diagnosis model (LCDM)
  - Subsumes many other DCMs
  - Allows for top-down approach to model building
  - Blueprint matching not guaranteed

- Constrained model: one-parameter LCDM (1-PLCDM)
  - Special case of LCDM where attribute main effects are constrained
  - Analogous to 1-PL IRT model
  - Nice measurement properties (sufficiency, invariant item ordering)
  - Assumptions: simple structure Q-matrix and independent attributes

Scenario #1: Summative

- Guiding example: North Carolina end-of-course mathematics tests
  - Use unidimensional IRT with cutscores to classify examinees into four levels

Table 1. EOC Mathematics domain weight distributions.

Domain	NC Math 1
Number and Quantity and Algebra	36 - 40%
Functions	32 - 36%
Geometry	8 - 12%
Statistics and Probability	18 - 20%
Total	100%

Scenario #1: Summative

- One polytomous attribute with four subattributes
  - » Prespecified blueprint: 40 / 28 / 12 / 20
  - » Item allocation (25 items): 10 / 7 / 3 / 5

- N = 2000

- Item parameters:
  - » Level 1: probability correct uniform on (0,.25)
  - » Each subsequent level increased by (.05,.30)

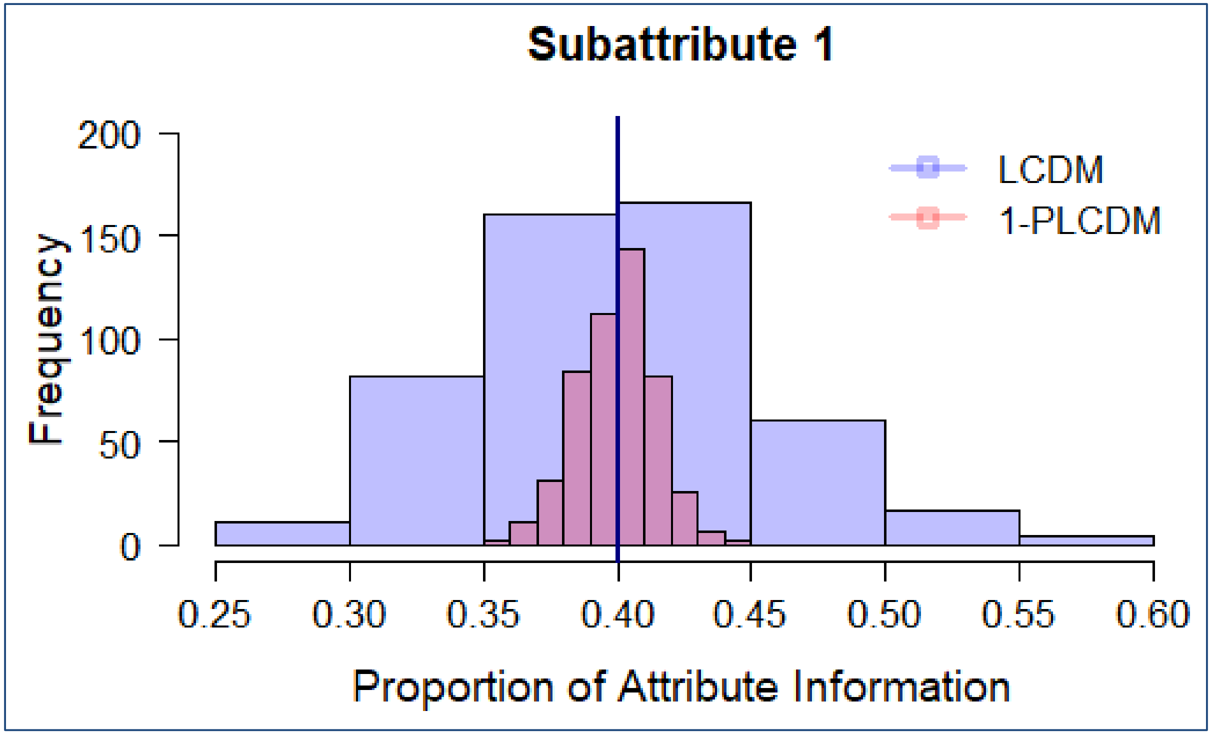
Scenario #1: Summative

- Estimated the polytomous LCDM and 1-PLCDM
  - » *mirt* package (Chalmers, 2012)

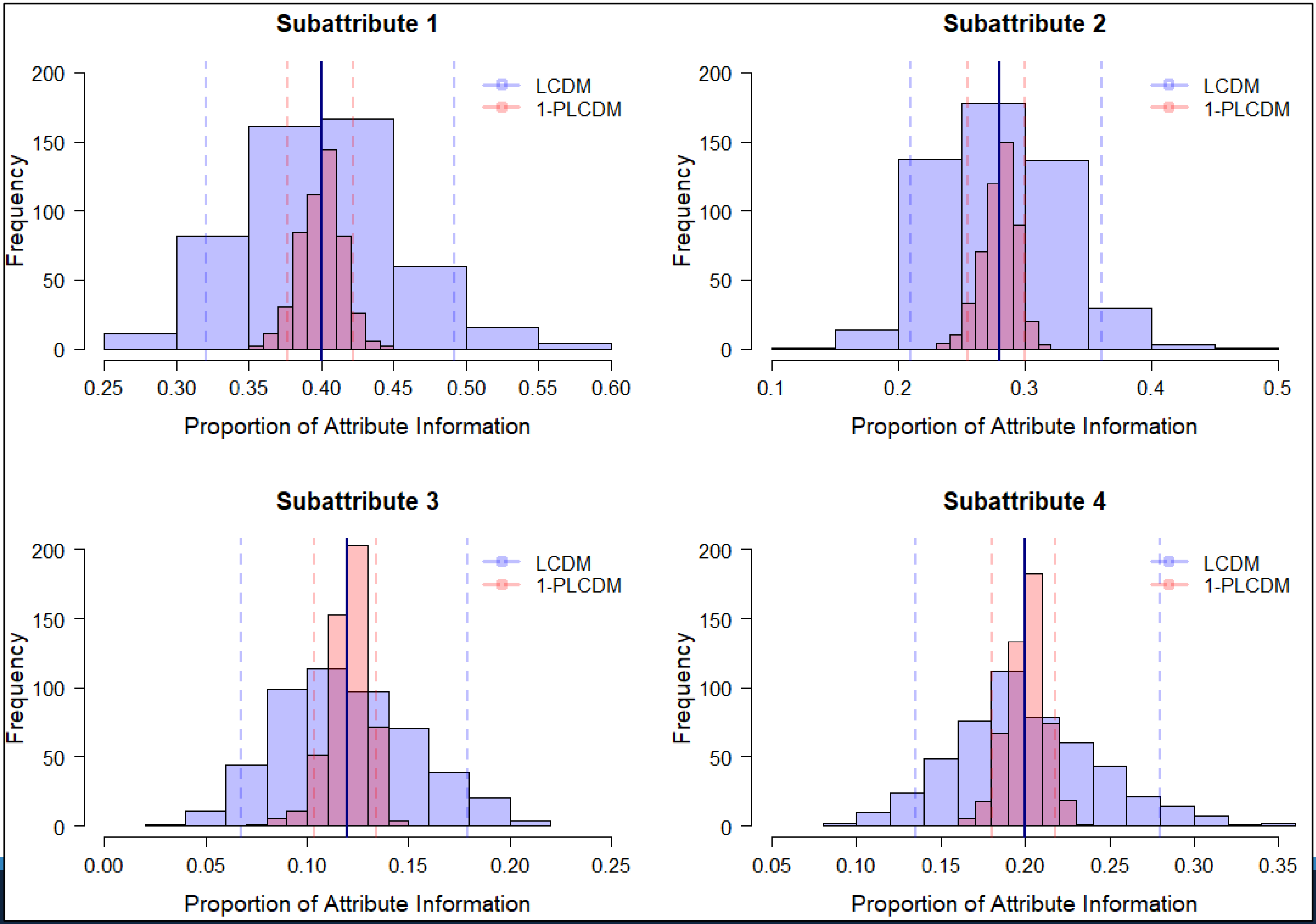
- Compared the empirical and prespecified blueprints
  - » 500 replications

Scenario #1 Results

- On *average*, both models approximated the prespecified blueprint
  - » Precision was much better for 1-PLCDM



Scenario #1 Results



Scenario #1 Results

- On *average*, both models approximated the prespecified blueprint
  - » Precision was much better for 1-PLCDM
  - » LCDM ranged from (25, 57)
  - » 1-PLCDM ranged from (35,45)

- Combined absolute error much lower for 1-PLCDM
  - » LCDM had mean blueprint error of 14%, max of 37%
  - » 1-PLCDM had mean blueprint error of 4%, max of 10%

Scenario #2: Intermediate

- Guiding example: Common Core State Standards

Cluster

Attribute #1

Attribute #2

Build fractions from unit fractions by applying and extending previous understandings of operations on whole numbers.

3. Understand a fraction *a/b* with *a* > 1 as a sum of fractions 1/*b*.
   a. Understand addition and subtraction of fractions as joining and separating parts referring to the same whole.
   b. Decompose a fraction into a sum of fractions with the same denominator in more than one way, recording each decomposition by an equation. Justify decompositions, e.g., by using a visual fraction model. *Examples: 3/8 = 1/8 + 1/8 + 1/8 ; 3/8 = 1/8 + 2/8 ; 2 1/8 = 1 + 1 + 1/8 = 8/8 + 8/8 + 1/8.*
   c. Add and subtract mixed numbers with like denominators, e.g., by replacing each mixed number with an equivalent fraction, and/or by using properties of operations and the relationship between addition and subtraction.
   d. Solve word problems involving addition and subtraction of fractions referring to the same whole and having like denominators, e.g., by using visual fraction models and equations to represent the problem.
4. Apply and extend previous understandings of multiplication to multiply a fraction by a whole number.
   a. Understand a fraction *a/b* as a multiple of 1/*b*. *For example, use a visual fraction model to represent 5/4 as the product 5 × (1/4), recording the conclusion by the equation 5/4 = 5 × (1/4).*
   b. Understand a multiple of *a/b* as a multiple of 1/*b*, and use this understanding to multiply a fraction by a whole number. *For example, use a visual fraction model to express 3 × (2/5) as 6 × (1/5), recognizing this product as 6/5. (In general, n × (a/b) = (n × a)/b.)*
   c. Solve word problems involving multiplication of a fraction by a whole number, e.g., by using visual fraction models and equations to represent the problem. *For example, if each person at a party will eat 3/8 of a pound of roast beef, and there will be 5 people at the party, how many pounds of roast beef will be needed? Between what two whole numbers does your answer lie?*

Scenario #2: Intermediate

- Three dichotomous attributes with subattributes
  - » Attribute 1 blueprint: 50 / 50 (3 / 3)
  - » Attribute 2 blueprint: 40 / 40 / 20 (4 / 4 / 2)
  - » Attribute 3 blueprint: 60 / 40 (6 / 4)

- Similar design to Scenario #1
  - » N = 2000
  - » Simple structure Q-matrix
  - » Attribute correlations uniform (.25, .75)

Scenario #2 Results

- Similar trend to Scenario #1
  - » LCDM error exacerbated in the multiattribute settings
- On *average*, both models approximated the prespecified blueprint
  - » Precision was much better for 1-PLCDM
  - » Attribute #1 (50/50)
    - LCDM 90% interval was (41, 59); worst replication was 0% / 100%
    - 1-PLCDM 90% interval was (49, 51); worst replication was 48% / 52%
- 1-PLCDM had much less total error
  - » LCDM mean error of 16%, max of 255%
  - » 1-PLCDM mean error of 1%, max of 5%

Conclusions

- Examined DCMs' ability to adhere to prespecified blueprints
  - » Critical for classification validity and interpretation
- Developed a framework for estimating empirical blueprints
  - » Proportion of attribute information
- Two simulated scenarios (summative and intermediate)
  - » 1-PLCDM was able to approximate prespecified blueprints
  - » Confirmed that general models struggle to match blueprints

Conclusions

- Not a criticism of the LCDM or general models
  - » Need general models to support use of constrained models
  - » 1-PLCDM limited to simple structure and makes strong assumptions

- DCMs can be used in contexts where blueprints are applied
  - » Increases validity and interpretability of classifications
  - » Expands the settings in which they can be applied

Conclusions

- Encourage prospective development of diagnostic tests
  - » Attributes are predefined and operationalized
  - » Items are written to reflect the definitions
- DCM applications should examine empirical blueprints
  - » Better understanding of how attributes are defined
  - » Evaluate the congruence of model-based attribute definitions and practitioner-based definitions
  - » Use item influence metrics to revise tests, if needed
- We hope that this work contributes to wider application of DCMs and better understanding of classifications

If you have questions or comments, feel free to contact us:

mjmadison@uga.edu

nancy.alila@uga.edu