DCMs in Practice

Session 3

Diagnosing Teachers' Multiplicative Reasoning

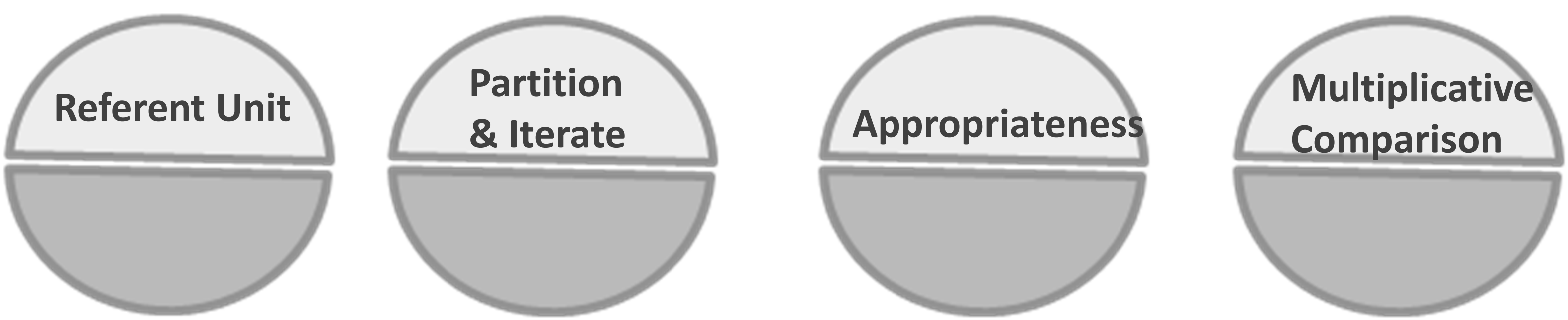
- *Diagnosing Teachers' Multiplicative Reasoning* (DTMR)*
  - NSF funded grant (DRL-0903411)
- Goal was to create a multidimensional test that will assess fine-grained components of teachers' reasoning multiplicatively with fractions
  - Understanding fractions as quantities as in Common Core State Standards
- The test would be used to
  - Tailor professional development to teachers' needs
  - Quantitatively study teachers' fine-grained reasoning abilities
    - Quantify findings based on extensive qualitative research base
    - Generalize to larger populations

Diagnosis from a Psychometric Model

- Diagnostic classification models would provide the type of diagnostic feedback we wanted to give teachers
  - We were trying to make **decisions** about teachers
- A diagnosis is a decision
  - Does a teacher need professional development on a given concept?

Attributes

- Instead of measuring an overall ability to reason multiplicatively with fractions, we broke down that continuous trait down into more fine-grained cognitive facilities or *attributes*:
  - Ability to identify appropriate **referent units** for numbers
  - Ability to **partition** quantities and **iterate** unit fractions
  - Ability to identify **appropriate** arithmetic operations
  - Ability to make **multiplicative comparisons**
- We were interested in two levels of mastery
  - Each attribute had two categories
  - Mastery of an attribute (= 1) or non-mastery of an attribute (= 0)



The DTMR Fractions Test

Designed to be Multidimensional using a DCM Framework

Steps in Test Construction

- 3-year test construction process
  - Interdisciplinary team from three (ish) universities
    - Mathematics Education Researchers & Psychometricians
- Key Steps
  - Operationalizing attributes
  - Writing Items
  - Validating Item/Attribute Alignment
  - Refining Items/Attributes

Defining Attributes

- Complex process
  - Not something that had been explicitly done at this level in mathematics education
- Relied on wealth of qualitative mathematics education research to define a set of workable attributes
  - Did not encompass all components of multiplicative reasoning
  - Focused on a set of components known to be difficult for students and teachers
    - Much less literature on teacher knowledge than student knowledge for this area
- Continued to refine for first few years of project

Designing Diagnostic Tests

- Items were written so that each item measures one or more of the attributes
  - Mapping is established by content experts
    - Confirmed by item response interviews
- Q-matrix for first several items on DTMR test:

	RU	PI	APP	MC
Item 1	1	0	0	0
Item 2	0	0	1	0
Item 3	1	0	0	0
Item 4	1	0	0	1

Example Item

- This item is analogous to Item 22 on the DTMR test
  - » Measures Referent Unit (Attribute 1) and Partitioning and Iterating (Attribute 2)

Ms. Roland gave her students the following problem to solve:

Candice has 4/5 of a meter of cloth. She uses 1/8 of a meter for a project.

How much cloth does she have left after the project?

She had students use the number line so that they could draw the lengths. Which of the following diagrams shows the solution? Assume all intervals are subdivided equally.

a)

b)

c)

d)

e)

Example Item Solution

- Option A is incorrect
  - » Takes 1/8 of 4/5 away from 4/5 instead of 1/8 of 1 whole
  - » Evidence of non-mastery of referent unit
- Option B is correct
  - » Correctly partitions the 1/5 segment so that the whole meter would be segmented into 40 pieces. Then, 5 of the 40 pieces or 1/8 of the whole are removed from 4/5 of the meter.

Ms. Roland gave her students the following problem to solve:

> *Candice has 4/5 of a meter of cloth. She uses 1/8 of a meter for a project. How much cloth does she have left after the project?*

She had students use the number line so that they could draw the lengths. Which of the following diagrams shows the solution? Assume all intervals are subdivided equally.

a)

b)

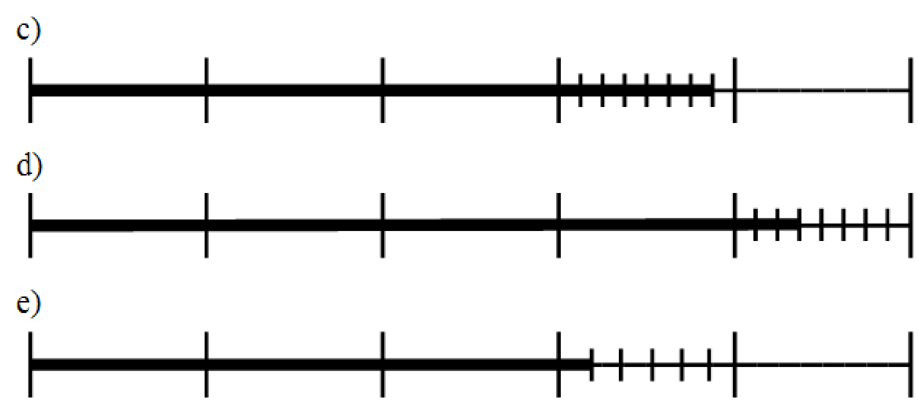
Example Item Solution

- Option C is incorrect
  - » Correctly partitions
  - » Takes 1/8 of 1/5 away from 4/5 instead of 1/8 of 1 whole
    - Evidence of non-mastery of referent unit
- Option D is incorrect
  - » Correctly partitions
  - » Removes 5/40 (or 1/8) from 1 whole instead of 4/5
    - Evidence of non-mastery of referent unit
- Option E is incorrect

Ms. Roland gave her students the following problem to solve:

Candice has 4/5 of a meter of cloth. She uses 1/8 of a meter for a project. How much cloth does she have left after the project?

She had students use the number line so that they could draw the lengths. Which of the following diagrams shows the solution? Assume all intervals are subdivided equally.



Attribute/Item Alignment

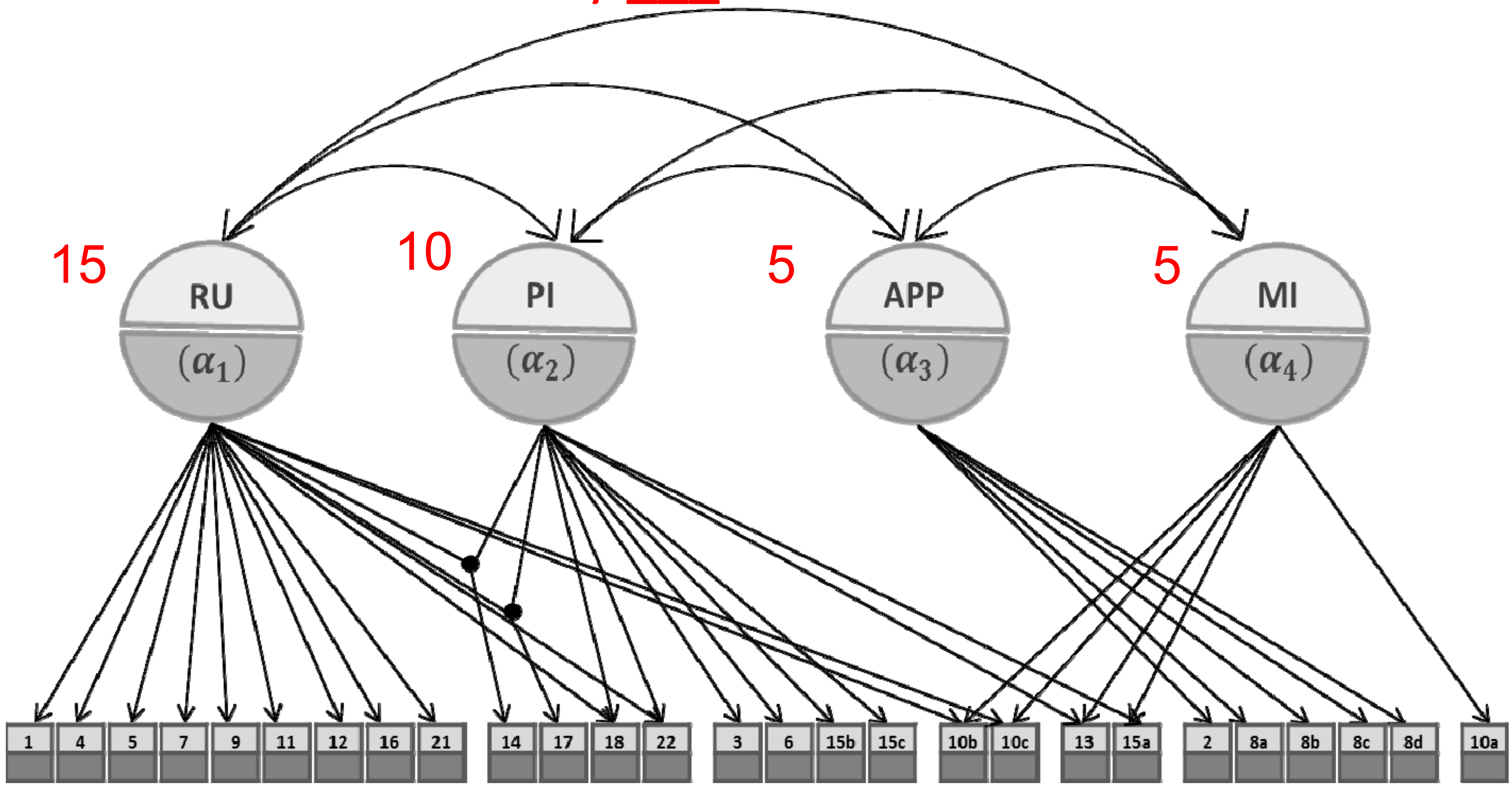
- Conducted three rounds of think aloud interviews with a total of 61 in-service teachers
  - These interviews are about an hour and a half long
    - Praise to our qualitative researchers!
  - This included 3 rounds of interviews (n= 14, 22, 25)
    - Each time revised test in light of interview data
    - After 3rd round felt we had a test that would yield valid classifications
      - Time to test it out!

Attribute/Item Alignment

- Analyzed data for
  - True Positive: answered item correctly; used intended attribute
  - False Positive: answered item correctly; did not use intended attribute
  - True Negative: missed item; did not use intended attribute
  - False Negative: missed item; used intended attribute
- Highlights
  - Common issue: Teachers found a way to circumvent attributes
    - Setting up and solving algebraic equations (instead of reasoning)
  - Open ended items were used in early pilots to more freely elicit reasoning
    - Helped write distracters

A Path Model of the DTMR Test

- Final Attribute/Item Alignment
- Attribute measured by ___ items



The University of Georgia

DCM Analysis of DTMR Data using Mplus

And a Little Help from our Friend, SAS

Mplus Files

- Input Files
  - DTMRdemo.inp
    - Mplus input file
  - DTMRdemo.dat
    - Data
    - Why “demo”?
      - Simulated from DTMR final model
- Output Files
  - DTMRdemo.out
    - Mplus output file
  - Resp_DTMRdemo.dat
    - Examinee classifications

SAS Macro for Estimating DTMR Data

- You can download this from the course website: http://wp.me/p3nkOf-nu

SAS Macro

- In SAS_MACRO folder:
- LCDM_Mplus2.sas
  - Creates Mplus syntax
  - Parses results into SAS output files
    - Useful for viewing/reporting results
    - Today Mplus output is shown, not SAS output
- DTMRdemo.sas
  - You actually run this file
  - Edit lines 23 -82 for your input information
  - You change this file when you want to change specifications for the model being run

Editing DTMRdemo.sas (1)

```
* Defining needed macro variables as global;
%GLOBAL macroloc filesave filename saslibname Qname dataname IDname
        itemstem itemlist numitem ordervar maxitemorder
        attstem attcat numatt numclass structon structorder loosen processors;

* Location of original data files - CHANGE ALL OF THEM;
* Permanent SAS library;                  LIBNAME folder      "FILE IS C:\Users\Laine\Dropbox\workshop\NCME 2013\Simulate\DTMRdemo"
* Path to SAS macro file;                 %LET macroloc=       C:\Users\Laine\Dropbox\workshop\NCME 2013\Simulate\DTMRdemo;
* Path to import/export files from;       %LET filesave=       C:\Users\Laine\Dropbox\workshop\NCME 2013\Simulate\DTMRdemo;

* Name prefix for files to be created;    %LET filename =     DTMR_demo1;
* Name of SAS library files are stored in; %LET saslibname=    work;
* Name of SAS dataset for Q matrix;       %LET Qname=         DTMR_q;
* Name of SAS dataset with original data;  %LET dataname=      DTMRdemo;

* Name of person ID variable (required);  %LET IDname=        ID;
* Item stem in Q matrix (cant be "item"); %LET itemstem=      x;
* List of items to be modeled;            %LET itemlist=      x1-x27;
* Total number of items;                  %LET numitem=       27;
* Variable for order of item model;       %LET ordervar=      itemorder;
* Max order of interaction in item model; %LET maxitemorder=  2;

* Attribute stem in Q matrix;             %LET attstem=       attribute;
* Number of categories for attributes;    %LET attcat =       2; * currently only set to 2;
* Total number of attributes;             %LET numatt=        4;
* Number of total classes (2^A);          %LET numclass=      16;
* Use structural model(0=N,1=Y);          %LET structon=      1;
* Order of interaction in structural model; %LET structorder=  3;

* Loosen congergence criteria (0=N,1=Y)?; %LET loosen=        0;
* Number of processors available for Mplus; %LET processors=   8;
********************************************************************************;
```

DTMRdemo.SAS

- Code is commented, making many input lines self-explanatory

For others:

- Itemorder
  - You set this in the next section for each item
- Attstem
  - Name for attributes (i.e., Attribute 1 vs. Att 1 vs. A1)
- Attcat = 2 for dichotomous attributes (e.g., mastery vs non-mastery)
- Structon
  - 1 produces code for the log-linear structural model
  - 0 runs the log-linear model in the background as default, but doesn't give you the structural parameters
- Convergence criteria
  - Zero leaves it at Mplus' default
- Processors available
  - Check the computer you're on

Editing DTMRdemo.SAS (2)

- Input Q-matrix
- Fill in the appropriate values in yellow under DATALINES;
  - Copy and paste from Excel
- The last column is the highest interaction term for the item

```
* Import Q-matrix into SAS;
DATA &saslibname..&Qname.;
INPUT &itemstem. attribute1-attribute4 &ordervar.;
DATALINES;
1    1    0    0    0    1
2    0    0    1    0    1
3    0    1    0    0    1

          ...

25   1    1    0    0    1
26   1    0    0    0    1
27   1    1    0    0    1
; RUN;
```

Editing DTMRdemo.SAS (2)

- From SAS macro, can specify LCDM or C-RUM using this input, but not DINA, DINO, or other sub-models
  - Open the input file SAS produces to make these changes
  - Needed for Item "19" and "24"

```
* Import Q-matrix into SAS;
DATA &saslibname..&Qname.;
INPUT &itemstem. attribute1-attribute4 &ordervar.;
DATALINES;
1    1    0    0    0    1
2    0    0    1    0    1
3    0    1    0    0    1



25   1    1    0    0    1
26   1    0    0    0    1
27   1    1    0    0    1
; RUN;
```

Original to Demo Item Mapping

- Macro renames items

Macro Item	DTMR Item
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8a
9	8b
10	8c
11	8d
12	9
13	10a
14	10b

Macro Item	DTMR Item
15	10c
16	11
17	12
18	13
19	14
20	15a
21	15b
22	15c
23	16
24	17
25	18
26	21
27	22

Editing DTMRexample.SAS (3)

- Change this to your data file name

```
* Import original data into SAS dataset;
DATA &saslibname..&dataname.;
    INFILE "&filesave.\DTMRdemo.dat" TRUNCOVER;
    INPUT &IDname. &itemlist.;
RUN;
```

Did it run?

- It will yell at you if it didn't
- Or not give you standard errors for your parameters

```
THE MODEL ESTIMATION TERMINATED NORMALLY

     THE CHI-SQUARE TEST CANNOT BE COMPUTED BECAUSE THE FREQUENCY TABLE FOR THE
     LATENT CLASS INDICATOR MODEL PART IS TOO LARGE.



MODEL FIT INFORMATION

Number of Free Parameters                         71

Loglikelihood

          H0 Value                         -14257.169
          H0 Scaling Correction Factor          1.063
            for MLR

Information Criteria

          Akaike (AIC)                      28656.338
          Bayesian (BIC)                    29004.075
          Sample-Size Adjusted BIC          28778.576
```

SAS Macro Tutorial

- For step-by-step how-to for using the SAS Macro and interpreting SAS macro output:

SAS Macro Tutorial

by Daniel Jurich at James Madison University

DTMR Test Results

Results Overview

- Data Collection
- Estimation
- Model Fit
- Items
  - How well did they function?
- Attribute Patterns
  - How many teachers are masters of each attribute?
  - What are the attribute mastery probabilities for a single teacher?
- Attribute Correlations
  - How highly correlated are the attributes?
  - Are any attributes dependent on another?

Data Collection

- National sample of 990 in-service middle grades mathematics teachers
- Sample stratified by
  - Region of the country (4 levels)
    - Northeast, Midwest, South, West
  - Urban-centric locale (12 levels)
    - City or suburb
      - Small, medium, large
    - Town or rural
      - Fringe, distant, remote
- Response rate: ≈20%
  - Received 990 of 5400 teachers
  - Demographics comparable to other math education national samples

Estimation

- LCDM as general modeling framework
  - Estimated with Mplus
  - Top-down approach
    - Estimate all higher order interactions and then remove non-significant interactions
- LCDM for Testlet Effects (Bifactor LCDM)
  - Estimated first
    - Three testlets on test
  - Did not converge
    - Testlet effects were negligible?
    - More research needed

Note about Model Fit

- Before we look at parameter estimates in the output, we would want to look at the model-data fit
  - See Rupp, Templin, Henson (2010) chapter
- This section will focus on
  - Interpreting output, the estimated parameters
    - Item parameter estimates
    - Examinee classifications
  - Evaluating the results
    - Were items highly “diagnostic”?

INTERPRETING ITEM PARAMETER RESULTS

Was this a “good” item?

- Example item seen previously
  - » Measures Referent Unit (Attribute 1) and Partitioning and Iterating (Attribute 2)

Ms. Roland gave her students the following problem to solve:

> *Candice has 4/5 of a meter of cloth. She uses 1/8 of a meter for a project. How much cloth does she have left after the project?*

She had students use the number line so that they could draw the lengths. Which of the following diagrams shows the solution? Assume all intervals are subdivided equally.

a)

b)

c)

d)

e)

LCDM Function for this Item

- Referent unit (α_1) and partitioning and iterating (α_2) are measured
  - Q-matrix entries:

	RU	PI	APP	MC
Item 22	1	1	0	0

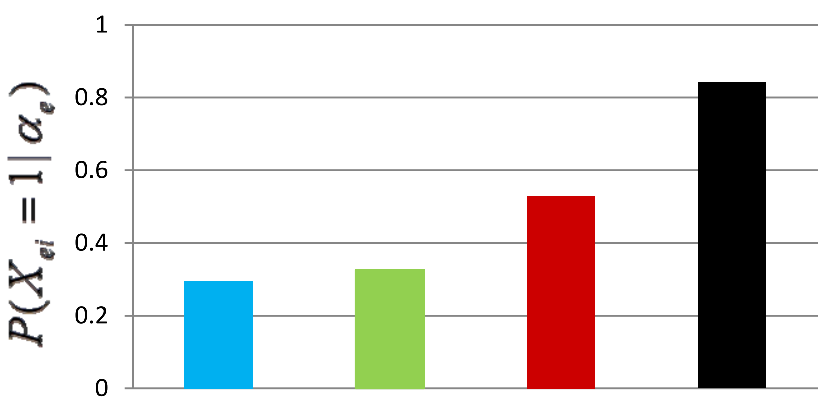
  - LCDM item response function:

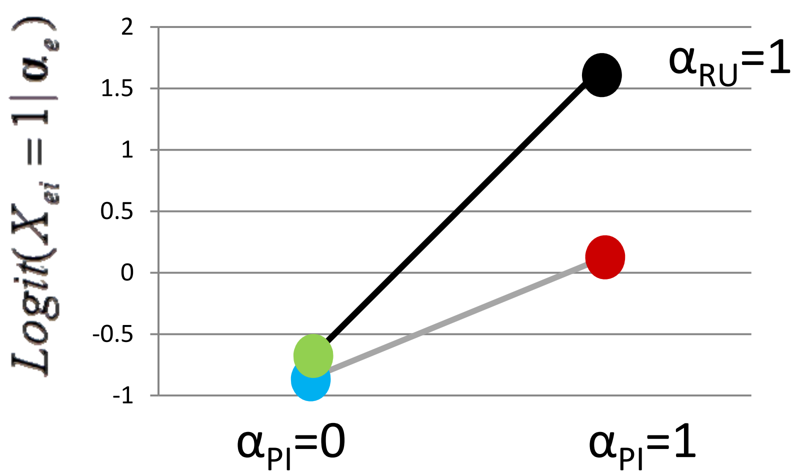
$$\log \frac{P(X_{ei} = 1 \mid \alpha_e)}{P(X_{ei} = 0 \mid \alpha_e)} = \underbrace{\boxed{\lambda_{i,0}}}_{\substack{\text{Intercept} \\ \text{(Guessing)}}} + \underbrace{\boxed{\lambda_{i,1(1)}}}_{\substack{\text{Main Effect} \\ \text{(RU)}}}(\alpha_{e1}) + \underbrace{\boxed{\lambda_{i,1(2)}}}_{\substack{\text{Main Effect} \\ \text{(PI)}}}(\alpha_{e2}) + \underbrace{\boxed{\lambda_{i,2(12)}}}_{\substack{\text{Interaction} \\ \text{(Between RU and PI)}}}(\alpha_{e1} \cdot \alpha_{e2})$$

Example Item Response Function

$$\log \frac{P(X_{ei}=1|\alpha_e)}{P(X_{ei}=0|\alpha_e)} = -.871 + .146(\alpha_{e1}) + .991(\alpha_{e2}) + 1.415(\alpha_{e1} \cdot \alpha_{e2})$$

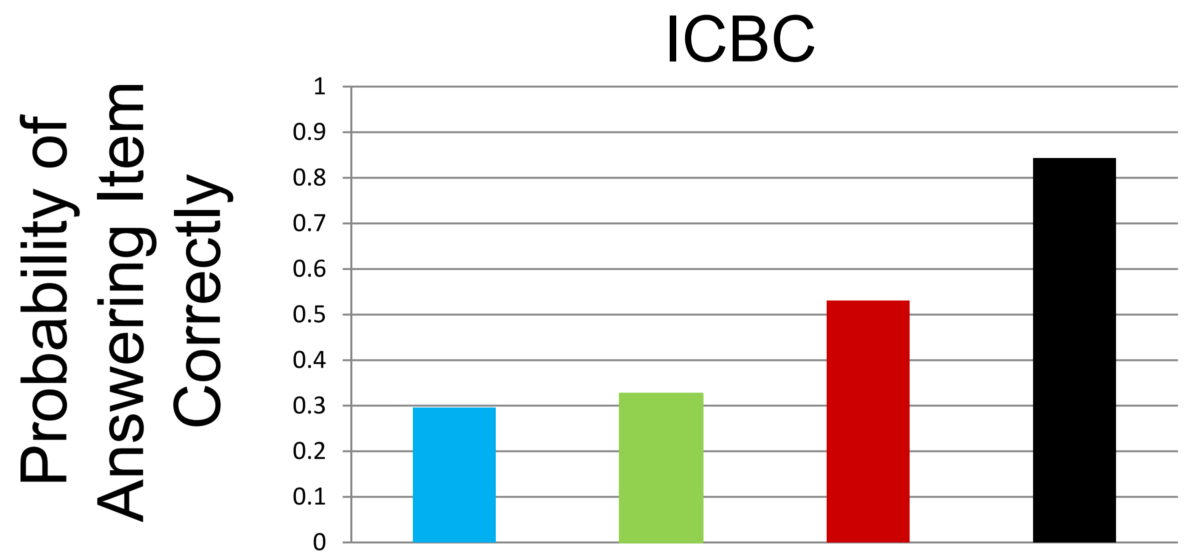
- On the logit scale, we can see the main effects are positive and the interaction is positive
- Item parameters provide construct validation
  - Is the item actually measuring the attribute?





Content Validity

- Results provide some evidence for content validity
  - Is the item actually measuring the attribute?



Key	
✓	Mastery
—	Non-mastery

Referent Unit	—	✓	—	✓
Partitioning & Iterating	—	—	✓	✓

Attribute Mastery Level

Mplus Output: Structural Parameter Estimates

- Structural parameter estimates
- More on these in the next section

New/Additional Parameters

Parameter	Estimate	SE	Estimate/SE	p
G_0	-0.183	0.142	-1.288	0.198
G_11	-4.542	0.758	-5.989	0.000
G_12	-1.721	0.319	-5.390	0.000
G_13	-1.338	0.358	-3.742	0.000
G_14	-1.112	0.241	-4.618	0.000
G_212	2.119	0.515	4.117	0.000
G_213	1.373	0.557	2.465	0.014
G_214	1.559	0.488	3.196	0.001
G_223	1.603	0.374	4.285	0.000
G_224	0.725	0.362	2.005	0.045
G_234	1.517	0.348	4.354	0.000
L1_0	-1.118	0.123	-9.106	0.000
L1_11	2.239	0.201	11.120	0.000
L2_0	0.585	0.130	4.492	0.000
L2_13	1.271	0.215	5.924	0.000
L3_0	-2.069	0.218	-9.497	0.000
L3_12	1.695	0.239	7.084	0.000
L4_0	-1.191	0.109	-10.906	0.000
L4_11	0.648	0.188	3.448	0.001
L5_0	-1.668	0.138	-12.047	0.000
L5_11	1.517	0.196	7.726	0.000

Mplus Output: Item Parameter Estimates

- Item parameter estimates
- For item i:
  - $\lambda_{i,0}$ =Li_0
  - $\lambda_{i,1(1)}$ =Li_11
  - $\lambda_{i,1(2)}$ =Li_12
  - $\lambda_{i,2(12)}$ =Li_212

Parameter	Estimate	SE	Estimate/SE	p
New/Additional Parameters				
G_0	-0.183	0.142	-1.288	0.198
G_11	-4.542	0.758	-5.989	0.000
G_12	-1.721	0.319	-5.390	0.000
G_13	-1.338	0.358	-3.742	0.000
G_14	-1.112	0.241	-4.618	0.000
G_212	2.119	0.515	4.117	0.000
G_213	1.373	0.557	2.465	0.014
G_214	1.559	0.488	3.196	0.001
G_223	1.603	0.374	4.285	0.000
G_224	0.725	0.362	2.005	0.045
G_234	1.517	0.348	4.354	0.000
L1_0	-1.118	0.123	-9.106	0.000
L1_11	2.239	0.201	11.120	0.000
L2_0	0.585	0.130	4.492	0.000
L2_13	1.271	0.215	5.924	0.000
L3_0	-2.069	0.218	-9.497	0.000
L3_12	1.695	0.239	7.084	0.000
L4_0	-1.191	0.109	-10.906	0.000
L4_11	0.648	0.188	3.448	0.001
L5_0	-1.668	0.138	-12.047	0.000
L5_11	1.517	0.196	7.726	0.000

Parameter Interpretation

- To demonstrate parameter interpretation, let's look at Item 18
  - Attributes measured:
    - Referent Unit (Attribute 1)
    - Partitioning and Iterating (Attribute 2)
- Parameter estimates:

Parameter	Estimate	SE	p-value
$\lambda_{18,0}$	-0.994	0.135	0.000
$\lambda_{18,1,(1)}$	1.132	0.260	0.000
$\lambda_{18,1,(2)}$	1.100	0.237	0.000

LCDM Intercepts

- Estimated Intercept: -0.994 (0.135)
- Indicates the logit of a correct response for a non-master of all attributes
  - Here, non-masters have an average probability of a correct response: $\exp(-0.994)/1 + \exp(-0.994) = 0.27$
- Hypothesis test is not important
  - Tests whether non-masters have a probability of a correct response of .50
- Problematic when very high
  - Difficult to identify other parameters
  - Indicates issues with test, Q-matrix, or attributes

Higher Order Model Parameters

- Interpretation of main effects and interactions proceeds sequentially:
- If interactions are present:
  - Examine highest level of interaction
    - If significantly different from zero, leave in model
    - If not, term can be omitted
  - Only 2 interaction terms found to be statistically significant on this test
    - Item 14 and 17
- When significant interactions are present, main effects cannot be easily interpreted
  - Sometimes called conditional main effects
  - Need to know *combination* of attributes mastered to fully describe item response function
- If interactions are not present:
  - Interpret main effects

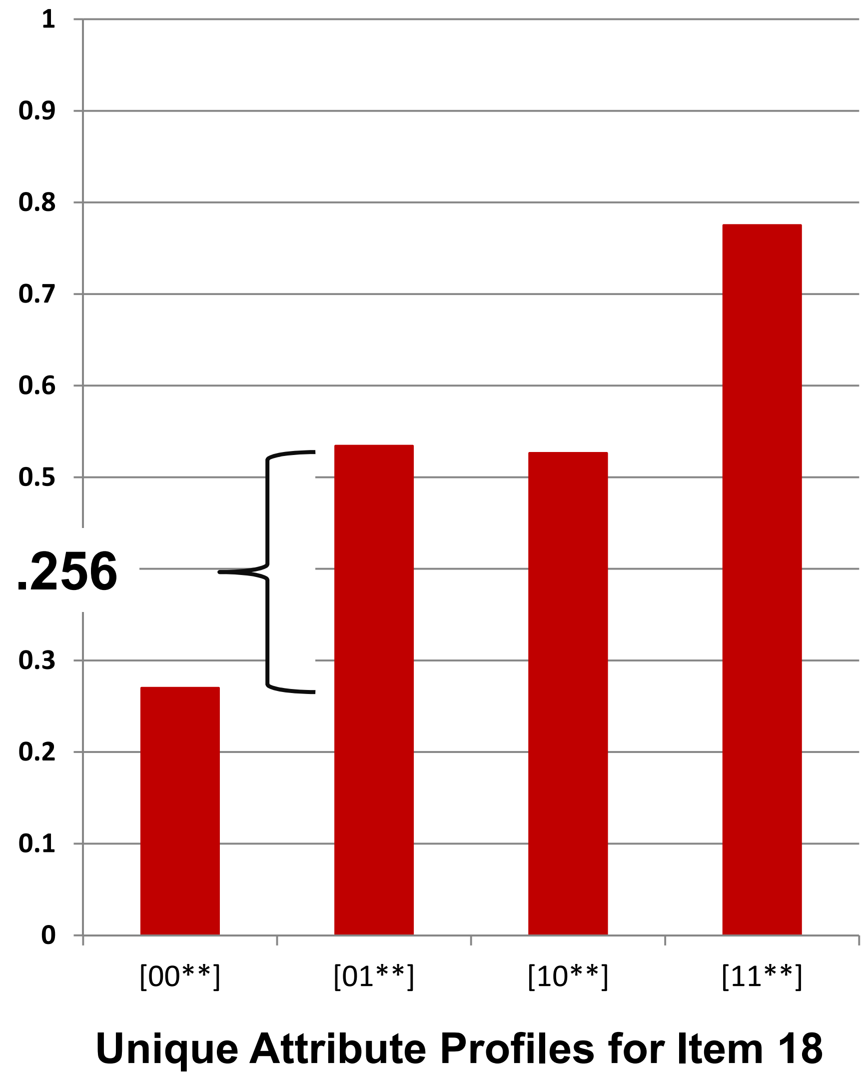
Interpreting Main Effects

- Main effects in LCDM cannot be tested for significance in the typical way
  - Lower bound is zero (for monotonicity)
  - *p*-values are inaccurate as they approach zero
  - Use practical significance
    - How much of an increase in probability for mastery of attribute is meaningful?
  - Or estimate the model with and without the main effect to compare fit
    - More in Model Fit section

Item 18 Partitioning and Iterating (P&I) Main Effect (Att 2)

Parameter	Estimate	SE	p-value
$\lambda_{18,0}$	-0.994	0.135	0.000
$\lambda_{18,1,(1)}$	1.132	0.260	0.000
$\lambda_{18,1,(2)}$	1.100	0.237	0.000

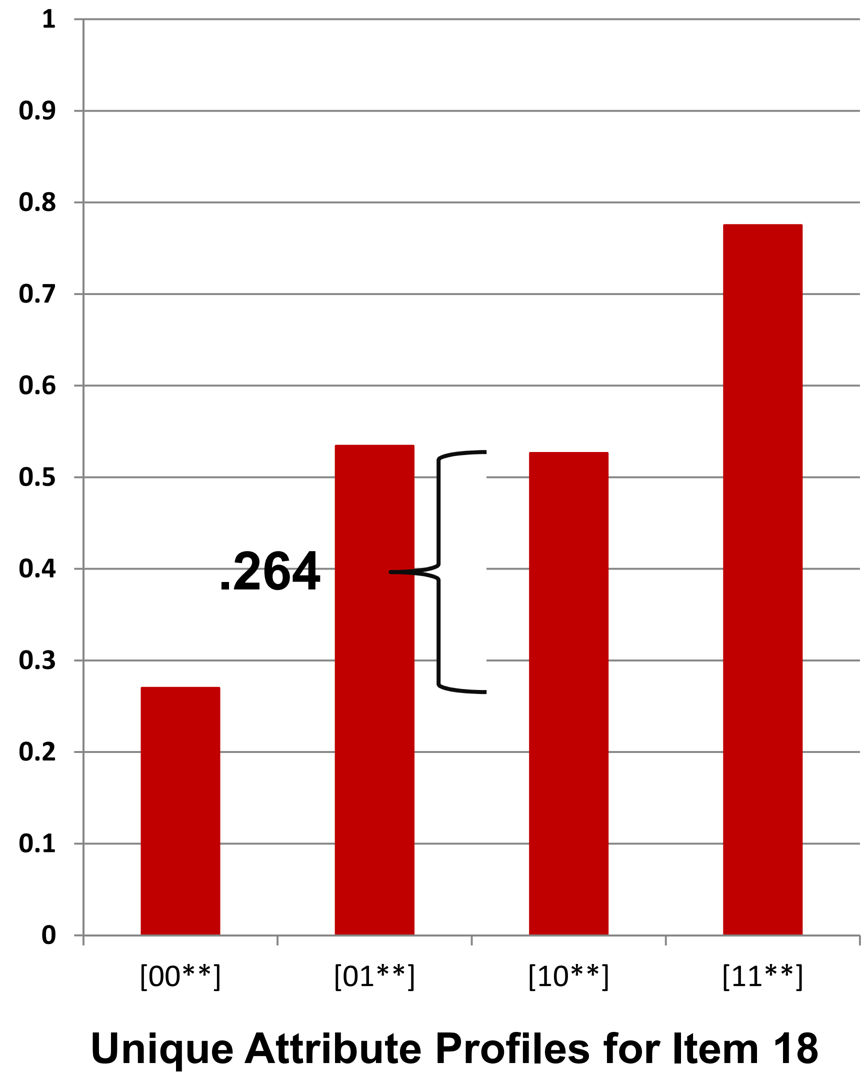
- When Referent Unit has not been mastered:
  - P&I main effect : $\lambda_{18,1,(2)}$ = 1.100
  - Respondents who have mastered P&I have an increase in logit of 1.1 over respondents who are non-masters
  - Respondents who have mastered P&I have an increase in probability of .256 over respondents who are non-masters



Item 18 Referent Unit (RU) Main Effect (Att 1)

Parameter	Estimate	SE	p-value
$\lambda_{18,0}$	-0.994	0.135	0.000
$\lambda_{18,1,(1)}$	1.132	0.260	0.000
$\lambda_{18,1,(2)}$	1.100	0.237	0.000

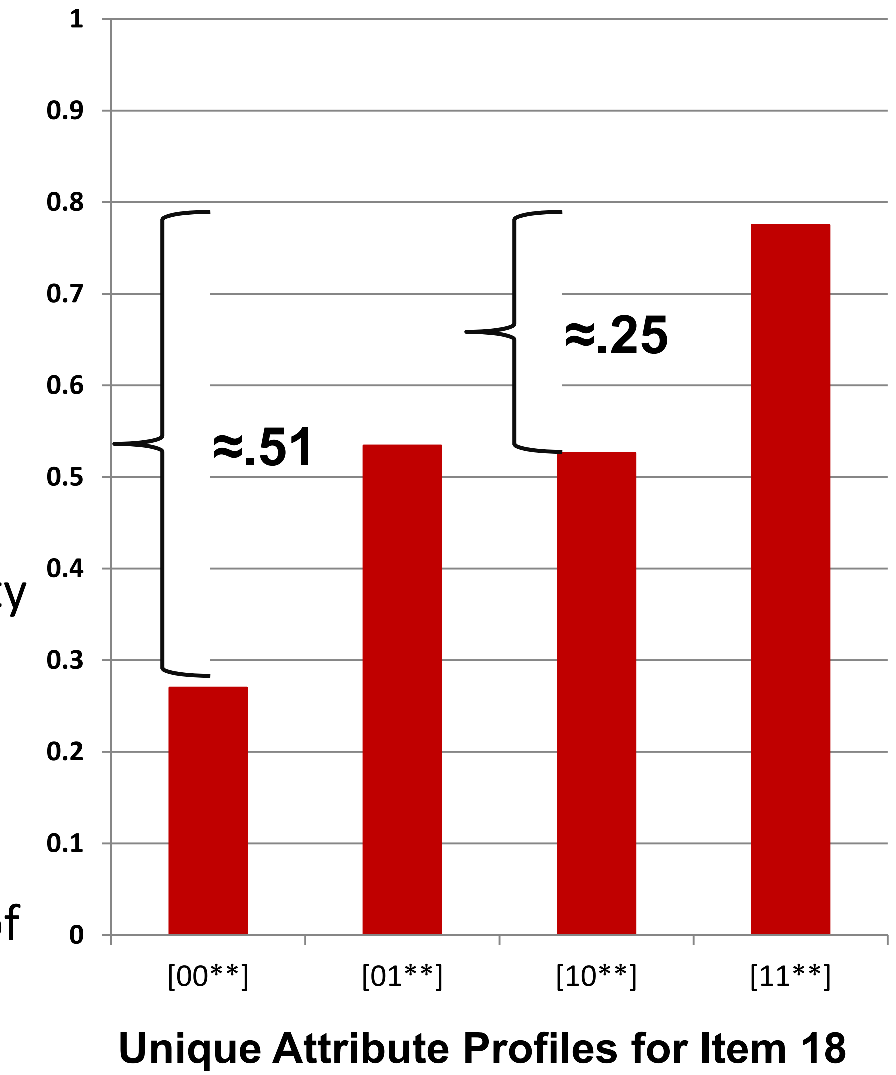
- When Partitioning & Iterating has not been mastered:
  - RU main effect : $\lambda_{18,1,(1)} = 1.132$
  - Respondents who have mastered Referent Unit have an increase in logit of 1.132 over respondents who are non-masters
  - Respondents who have mastered Referent Unit have an increase in probability of .264 over respondents who are non-masters of RU



Item 18 Referent Unit (RU) Main Effect (Att 1)

Parameter	Estimate	SE	p-value
$\lambda_{18,0}$	-0.994	0.135	0.000
$\lambda_{18,1,(1)}$	1.132	0.260	0.000
$\lambda_{18,1,(2)}$	1.100	0.237	0.000

- When both Referent Unit and Partitioning & Iterating has been mastered:
    - Respondents have a .775 probability of answering the item correctly
    - 51% increase of non-masters of both
    - About 25% increase over masters of only 1



Preview to Mplus Coding

- Mplus sets class-specific item response probabilities using the appropriate combinations of item parameters for each class
- Output looks like this:

```
    I21$1              1.502      0.128     11.733      0.000
    I22$1              1.247      0.161      7.742      0.000

Latent Class 5

 Thresholds
    I1$1               1.118      0.123      9.106      0.000
    I2$1              -0.585      0.130     -4.492      0.000
    I3$1               0.373      0.106      3.507      0.000
    I4$1               1.191      0.109     10.906      0.000
    I5$1               1.668      0.138     12.047      0.000
    I6$1               1.732      0.136     12.726      0.000
    I7$1               0.726      0.089      8.143      0.000
    I8A$1              0.615      0.249      2.468      0.014
    I8B$1              0.091      0.172      0.529      0.596
    I8C$1             -0.283      0.128     -2.215      0.027
    I8D$1              1.032      0.166      6.214      0.000
    I9$1               1.224      0.100     12.246      0.000
    I10A$1             0.503      0.184      2.738      0.006
    I10B$1             4.014      0.737      5.444      0.000
    I10C$1             4.888      0.868      5.629      0.000
    I11$1              0.875      0.097      9.031      0.000
    I12$1              1.288      0.111     11.646      0.000
    I13$1              0.287      0.206      1.389      0.165
    I14$1              2.143      0.140     15.354      0.000
    I15A$1            -0.243      0.292     -0.834      0.404
    I15B$1            -2.376      0.263     -9.023      0.000
    I15C$1            -2.605      0.302     -8.631      0.000
    I16$1              0.857      0.096      8.943      0.000
    I17$1              0.804      0.204      3.944      0.000
    I18$1             -0.106      0.194     -0.545      0.585
    I21$1              1.502      0.128     11.733      0.000
    I22$1             -0.183      0.207     -0.885      0.376

Latent Class 6

 Thresholds
    I1$1               1.118      0.123      9.106      0.000
```

Note about Thresholds

- Item parameters sum to give us (negative) thresholds
  - Negative because Mplus models P(X=0)/P(X=1) odds
- Example: Item 18 parameters

L18_0	-0.994	0.135	-7.366	0.000
L18_11	1.132	0.260	4.357	0.000
L18_12	1.100	0.237	4.632	0.000

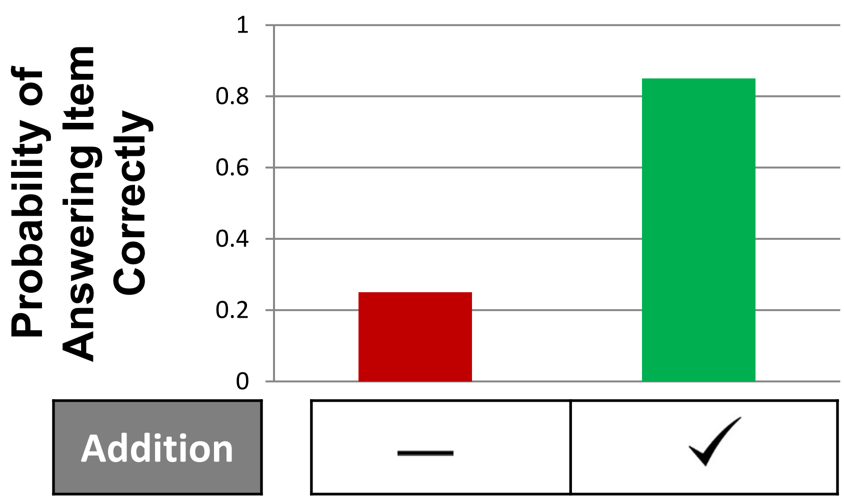
- For classes 1- 4: [00**], threshold = .994
- For classes 5-8: [01**], threshold = -.994 + 1.1 = 1.06
- For class 9-12: [10**], threshold = -.994 + 1.132 = 1.38
- For class 13-16: [11**], threshold = -.994+1.1+1.32 = 1.238

DIAGNOSTIC QUALITY OF ITEMS

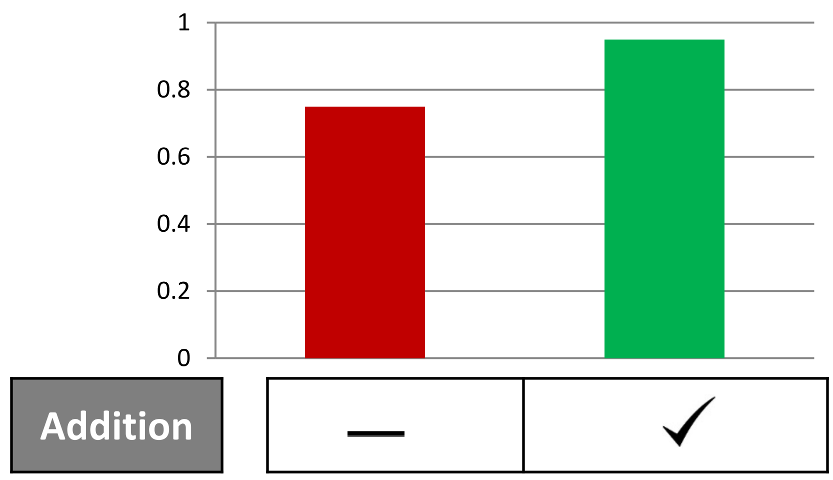
What makes a “good” diagnostic item?

- For “good” items, masters of the attribute(s) answer the item correctly and non-masters answer the item incorrectly
- For example, consider two hypothetical items:

Item A: 4+12=___



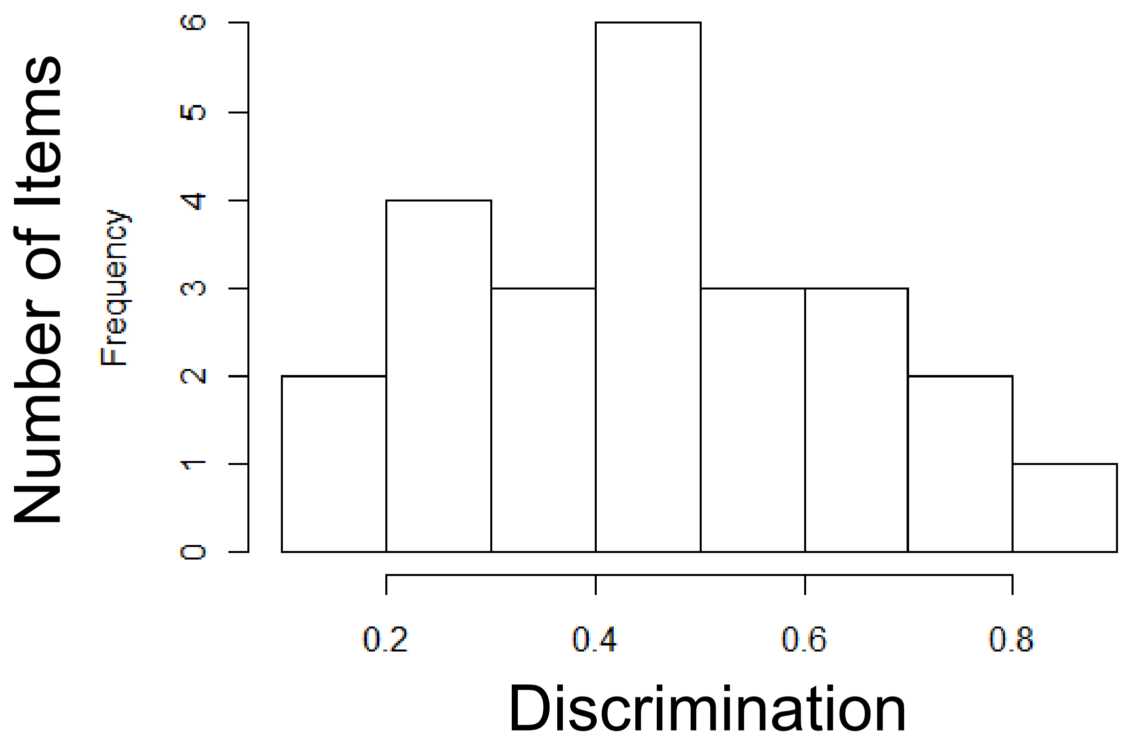
Item B: 1+12 = __



- Masters and non-masters should respond differently to the items
  - Evidence for the Q-matrix alignment
  - Is attribute required to answer the item correctly?

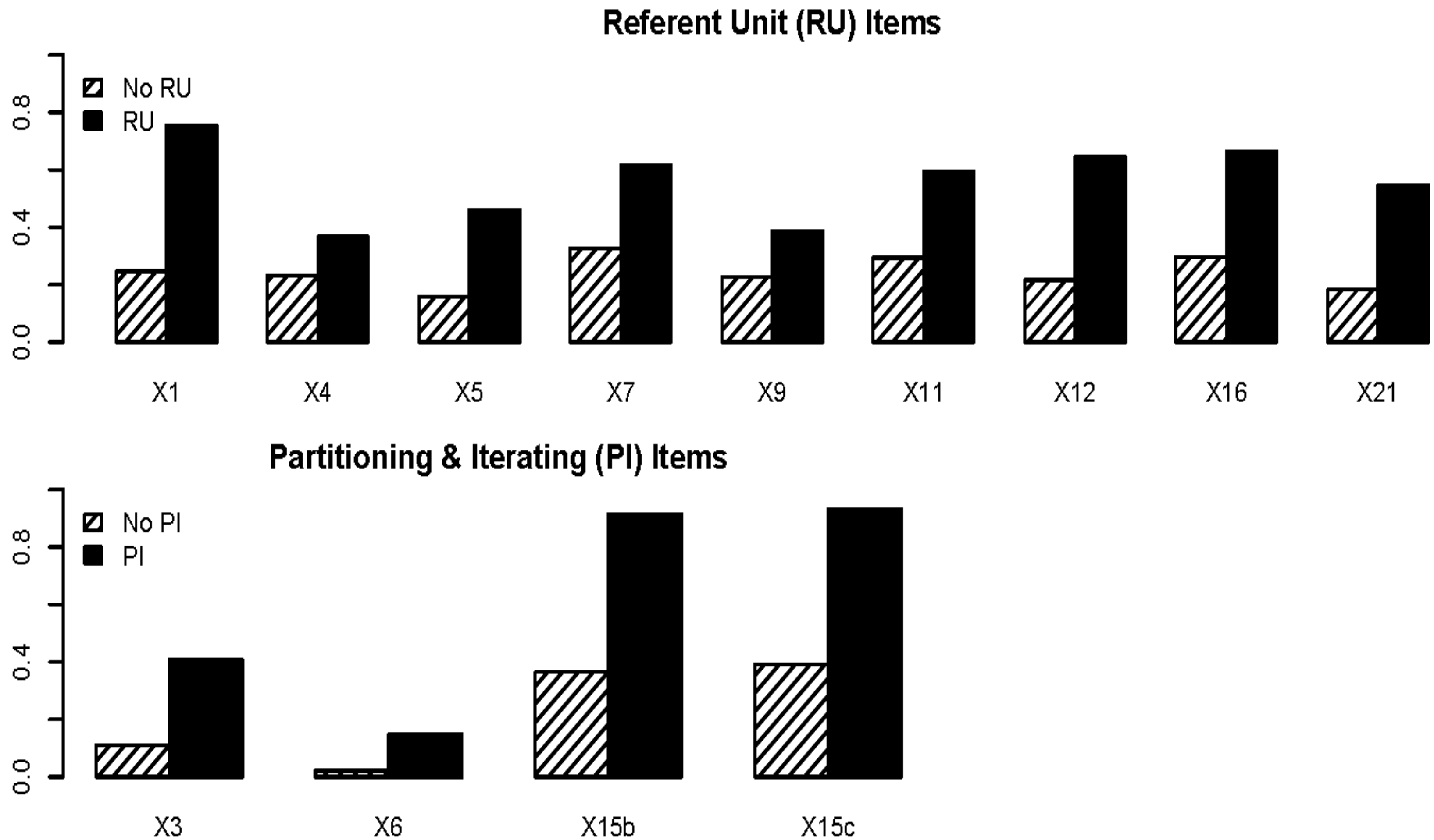
Item Discrimination

- This “goodness” quality is discrimination
  - ➢ Differences in the probability masters and non-masters answer the item correctly



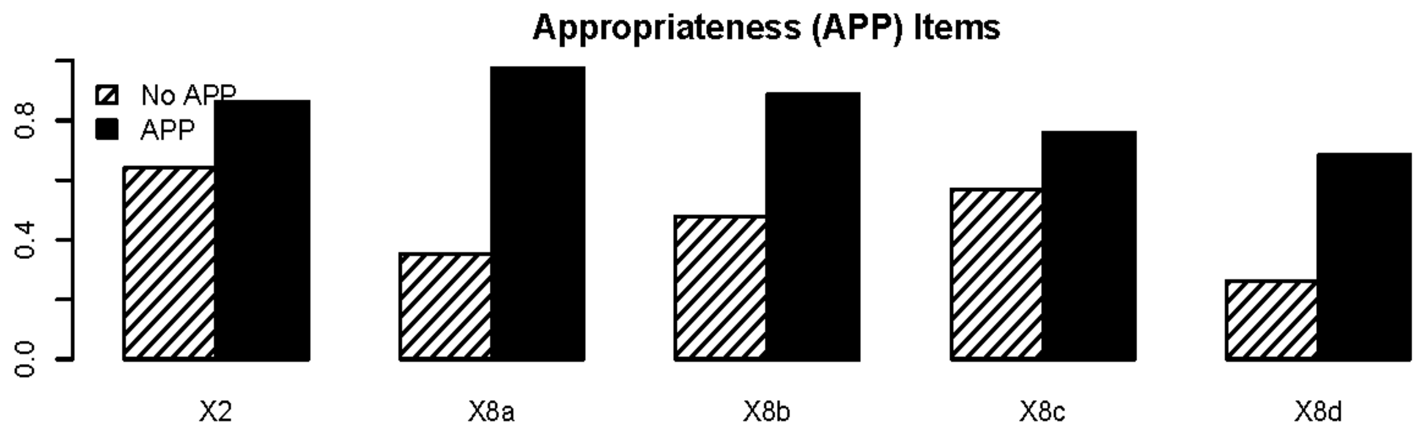
Simple Structure Items

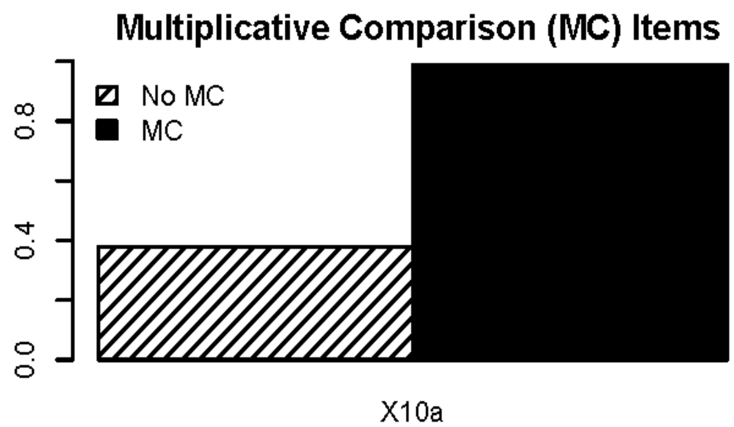
- Small intercepts and large main effects increase item discrimination



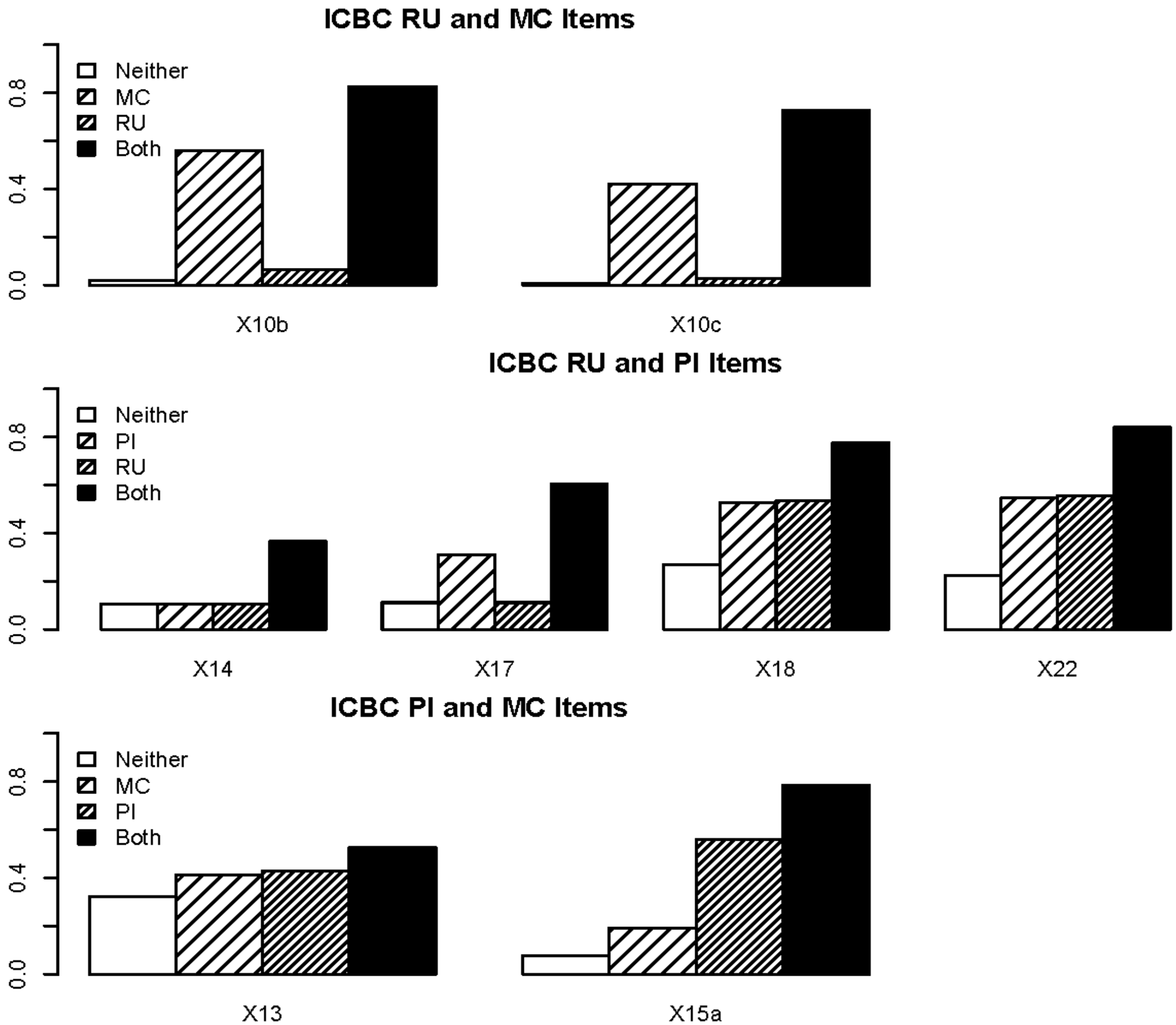
Simple Structure Items

- Odds ratios were calculated as an effect size
  - Almost all items had medium to large effects



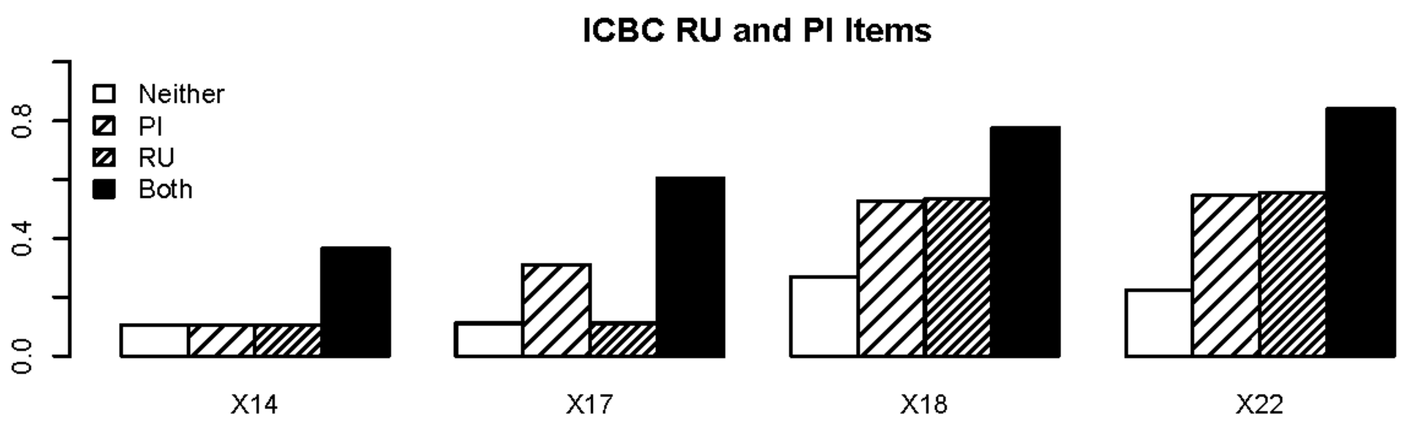


Complex Structure Items



Flexibility of LCDM Framework

- Notice the compensatory, non-compensatory nature of the items vary across the test
  - DINA model would force every item to look like Item 14
  - Empirically, we found only one item functioned like Item 14



  - Important not to assume these strict assumptions
    - Even if they are very popular in the literature
    - Don't do it!

INTERPRETING EXAMINEE CLASSIFICATIONS

Groups According to Attribute Mastery

Possible patterns or classes:

Pattern	RU	PI	APP	MC
1	0	0	0	0
2	0	0	0	1
3	0	0	1	0
4	0	0	1	1
5	0	1	0	0
6	0	1	0	1
7	0	1	1	0
8	0	1	1	1
9	1	0	0	0
10	1	0	0	1
11	1	0	1	0
12	1	0	1	1
13	1	1	0	0
14	1	1	0	1
15	1	1	1	0
16	1	1	1	1

Mplus Classification Output

- Proportion of examinees in each class

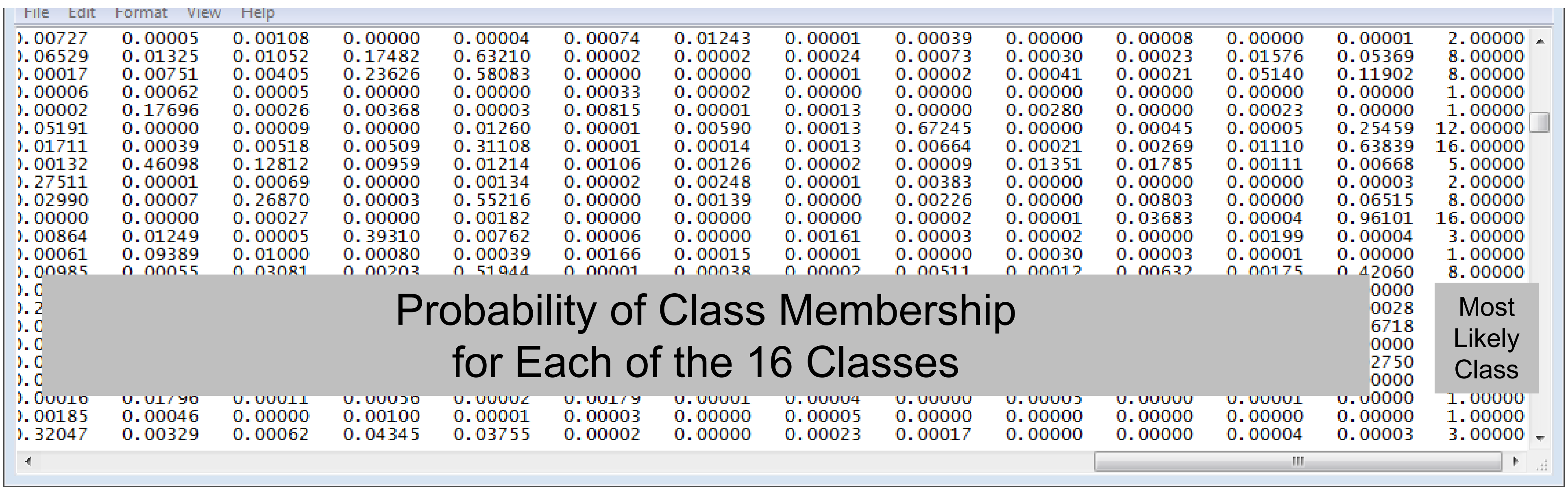
```
FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL

    Latent
   Classes

       1        210.34540          0.21247
       2         69.20994          0.06991
       3         55.19197          0.05575
       4         82.74175          0.08358
       5         37.63398          0.03801
       6         25.56661          0.02582
       7         49.04489          0.04954
       8        151.80987          0.15334
       9          2.24094          0.00226
      10          3.50645          0.00354
      11          2.32019          0.00234
      12         16.54147          0.01671
      13          3.33667          0.00337
      14         10.77973          0.01089
      15         17.15841          0.01733
      16        252.57174          0.25512
```

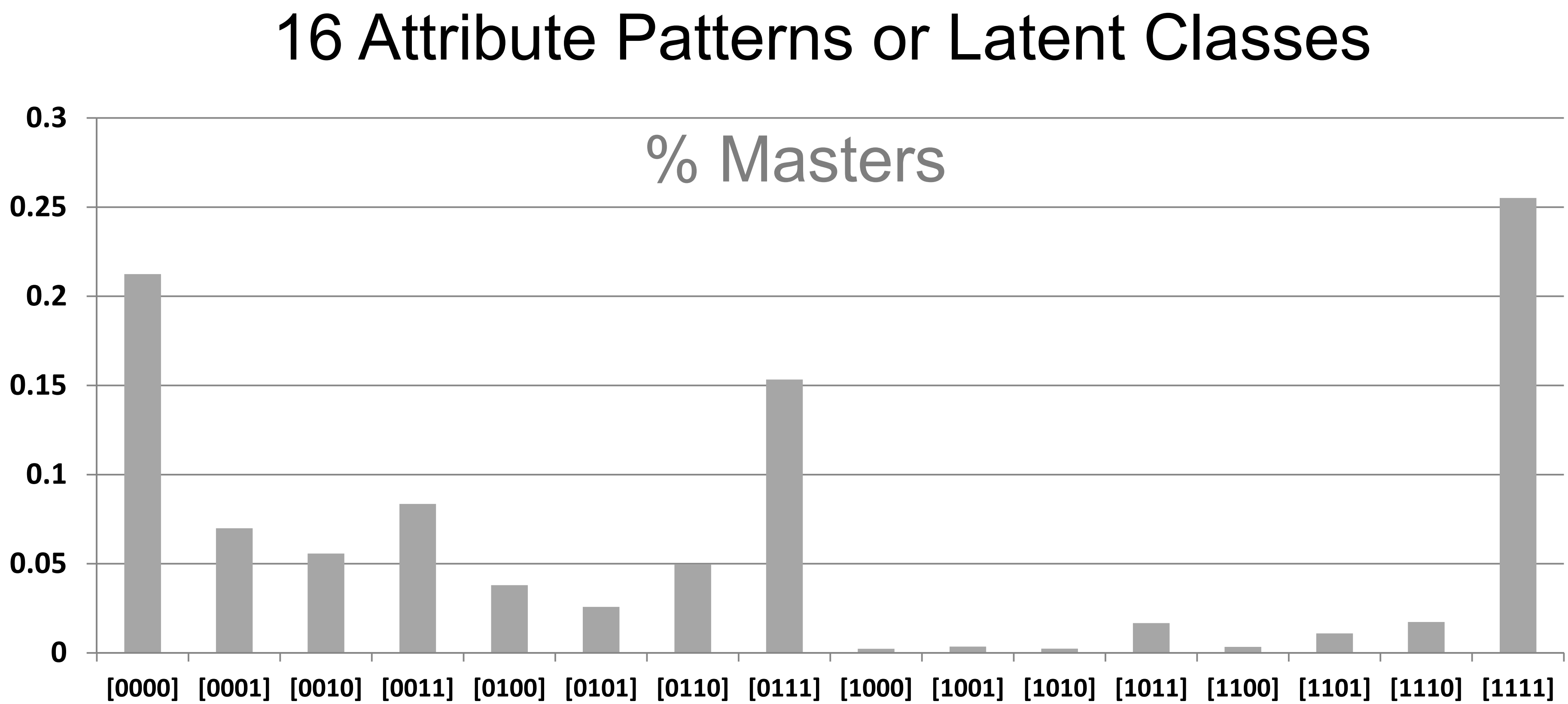
Examinee-level Classifications

- Resp_DTMRdemo.dat
  - Output file for examinee classifications
- First columns are any variables read into Mplus
  - scored responses to items, ID, etc.
- Last 17 Columns:

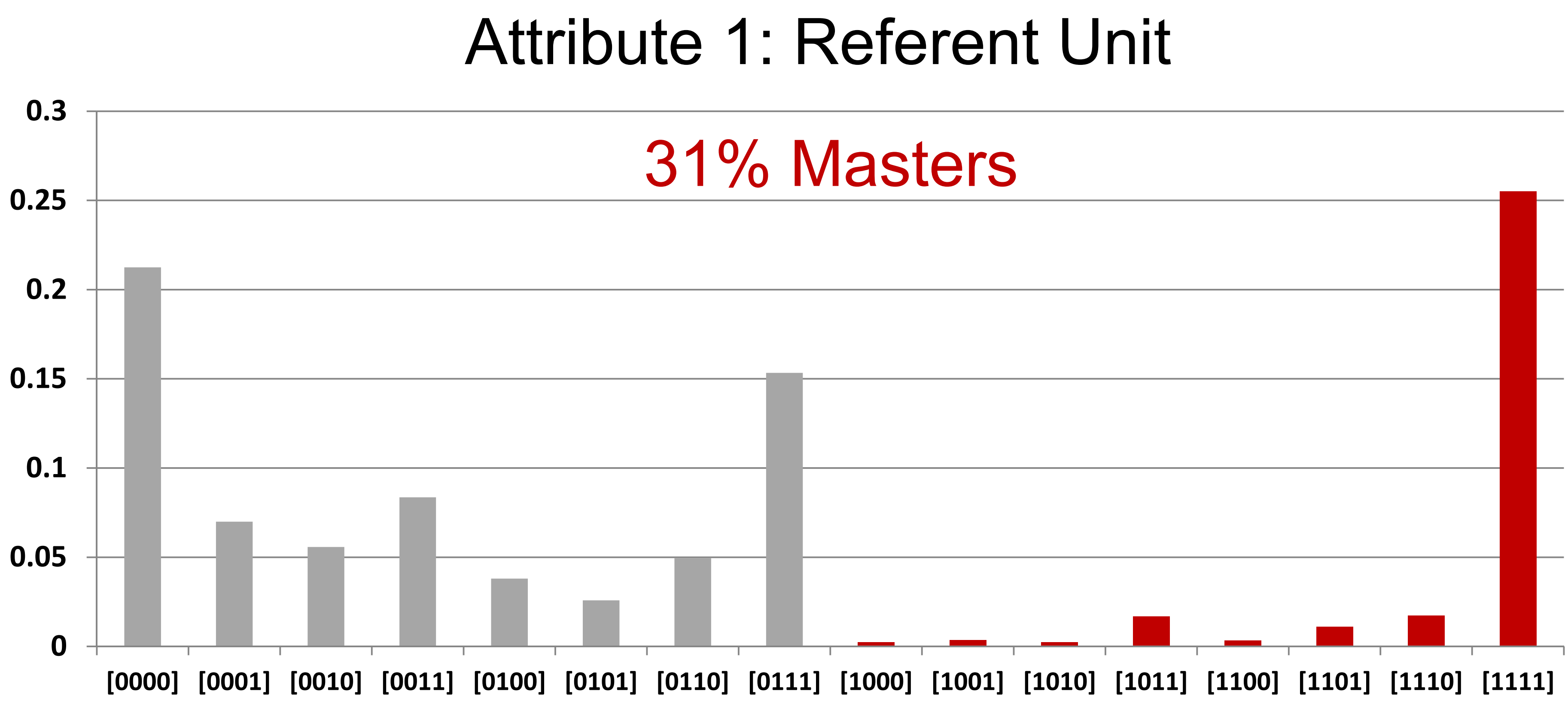


).00727	0.00005	0.00108	0.00000	0.00004	0.00074	0.01243	0.00001	0.00039	0.00000	0.00008	0.00000	0.00001	2.00000
).06529	0.01325	0.01052	0.17482	0.63210	0.00002	0.00002	0.00024	0.00073	0.00030	0.00023	0.01576	0.05369	8.00000
).00017	0.00751	0.00405	0.23626	0.58083	0.00000	0.00000	0.00001	0.00002	0.00041	0.00021	0.05140	0.11902	8.00000
).00006	0.00062	0.00005	0.00000	0.00000	0.00033	0.00002	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000
).00002	0.17696	0.00026	0.00368	0.00003	0.00815	0.00001	0.00013	0.00000	0.00280	0.00000	0.00023	0.00000	1.00000
).05191	0.00000	0.00009	0.00000	0.01260	0.00001	0.00590	0.00013	0.67245	0.00000	0.00045	0.00005	0.25459	12.00000
).01711	0.00039	0.00518	0.00509	0.31108	0.00001	0.00014	0.00013	0.00664	0.00021	0.00269	0.01110	0.63839	16.00000
).00132	0.46098	0.12812	0.00959	0.01214	0.00106	0.00126	0.00002	0.00009	0.01351	0.01785	0.00111	0.00668	5.00000
).27511	0.00001	0.00069	0.00000	0.00134	0.00002	0.00248	0.00001	0.00383	0.00000	0.00000	0.00000	0.00003	2.00000
).02990	0.00007	0.26870	0.00003	0.55216	0.00000	0.00139	0.00000	0.00226	0.00000	0.00803	0.00000	0.06515	8.00000
).00000	0.00000	0.00027	0.00000	0.00182	0.00000	0.00000	0.00000	0.00002	0.00001	0.03683	0.00004	0.96101	16.00000
).00864	0.01249	0.00005	0.39310	0.00762	0.00006	0.00000	0.00161	0.00003	0.00002	0.00000	0.00199	0.00004	3.00000
).00061	0.09389	0.01000	0.00080	0.00039	0.00166	0.00015	0.00001	0.00000	0.00030	0.00003	0.00001	0.00000	1.00000
).00985	0.00055	0.03081	0.00203	0.51944	0.00001	0.00038	0.00002	0.00511	0.00012	0.00632	0.00175	0.42060	8.00000
).0												0000	
).2												0028	
).0												6718	
).0												0000	
).0												2750	
).0												0000	
).00016	0.01796	0.00011	0.00036	0.00002	0.00179	0.00001	0.00004	0.00000	0.00005	0.00000	0.00001	0.00000	1.00000
).00185	0.00046	0.00000	0.00100	0.00001	0.00003	0.00000	0.00005	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000
).32047	0.00329	0.00062	0.04345	0.03755	0.00002	0.00000	0.00023	0.00017	0.00000	0.00000	0.00004	0.00003	3.00000

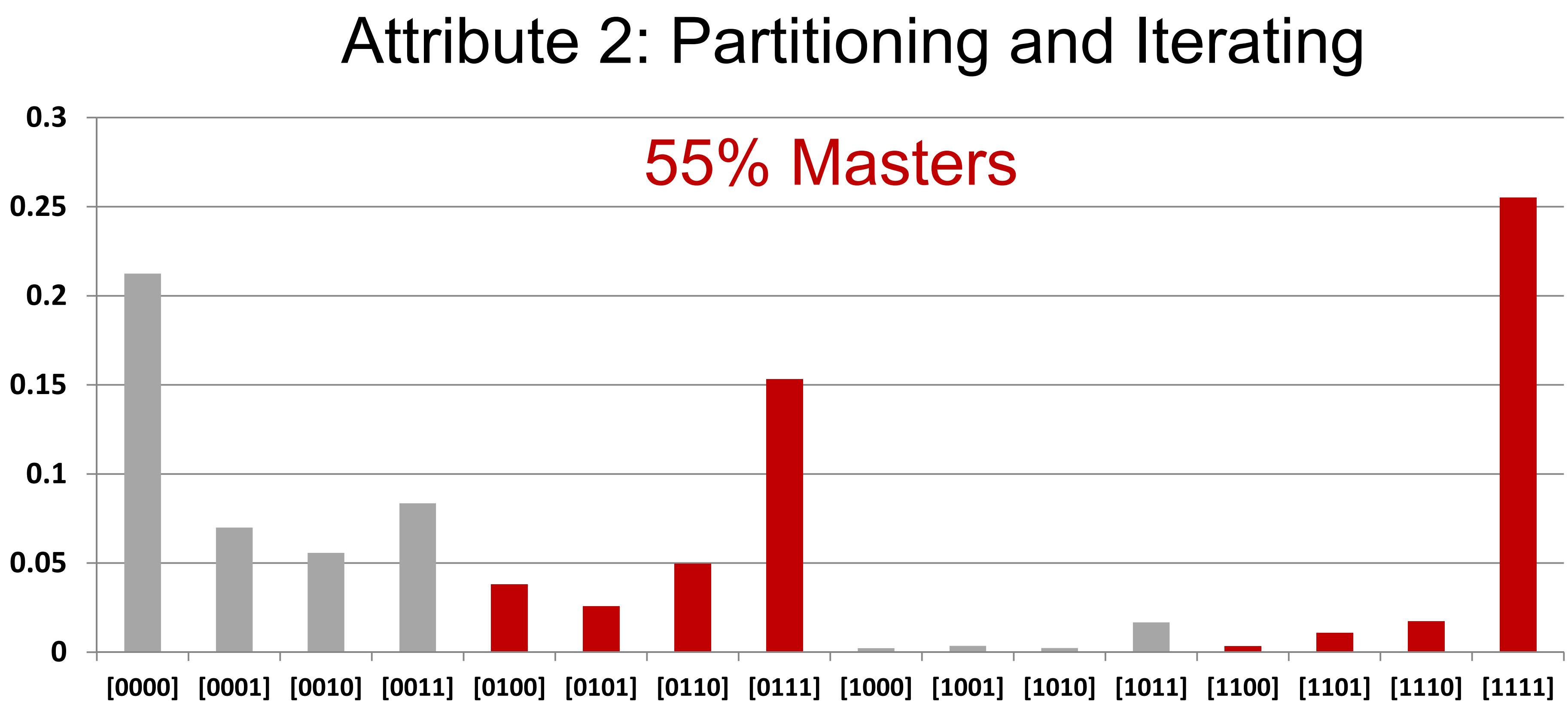
Attribute Patterns of Mastery



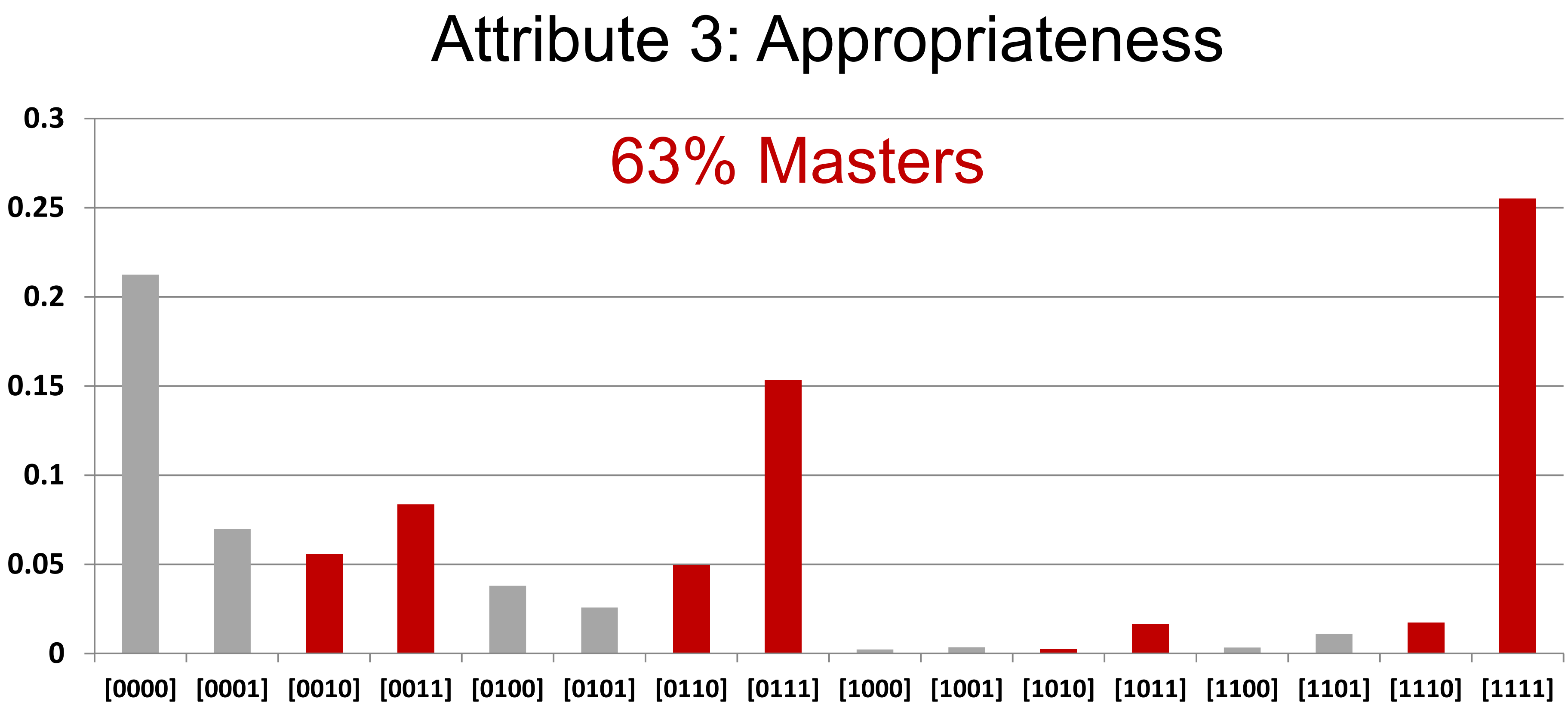
Attribute Patterns of Mastery



Attribute Patterns of Mastery

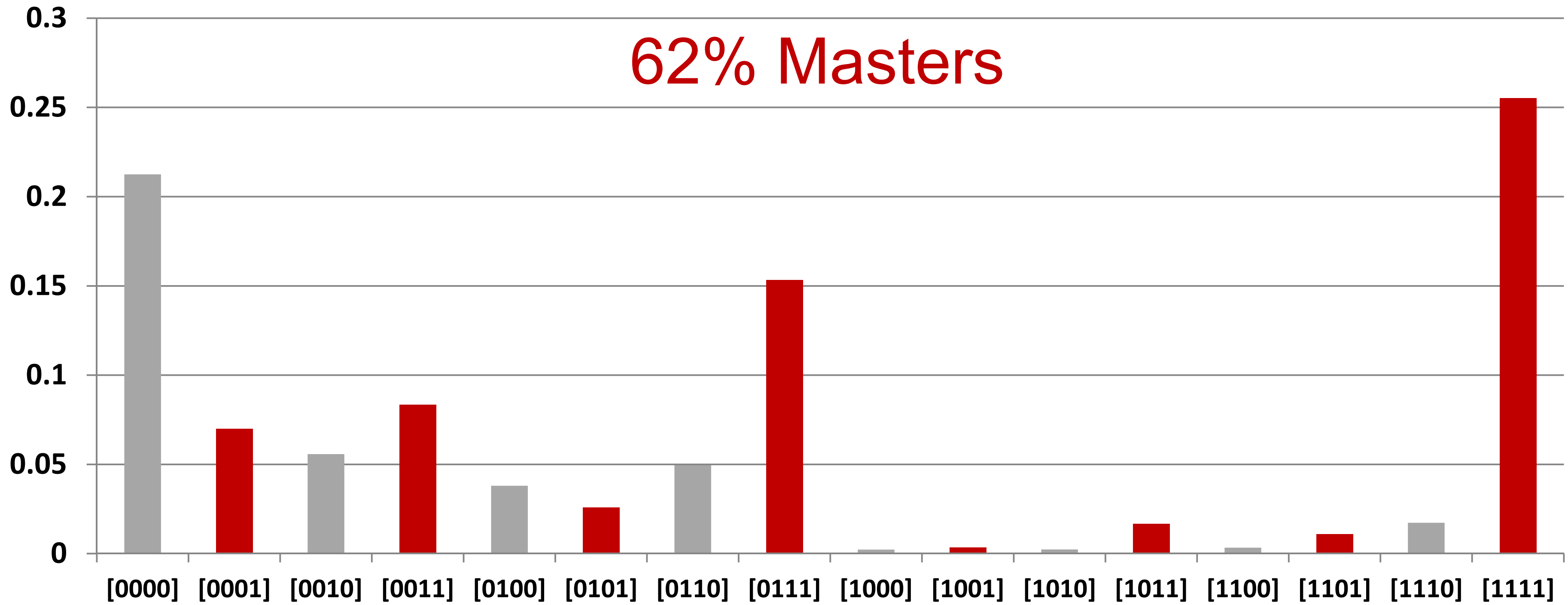


Attribute Patterns of Mastery

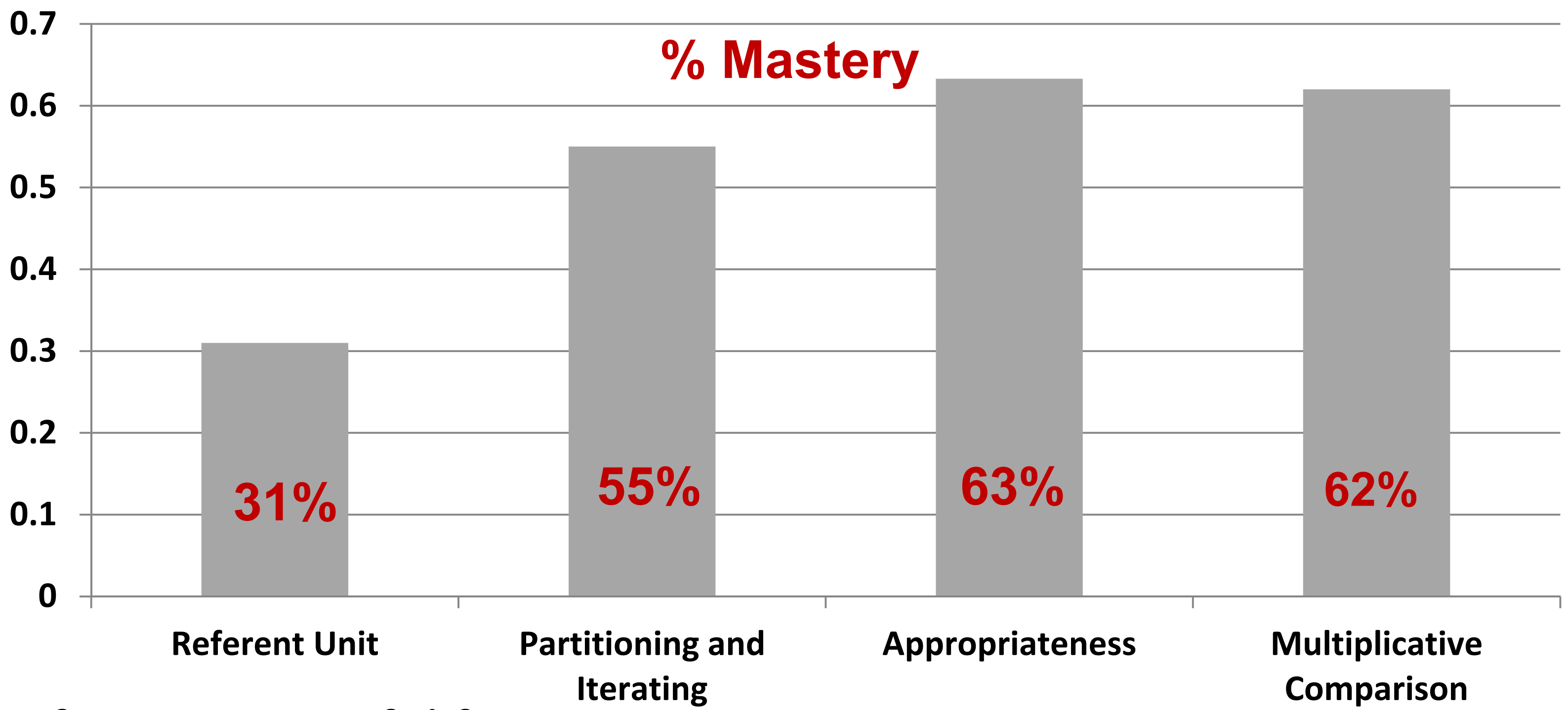


Attribute Patterns of Mastery

Attribute 4: Multiplicative Comparisons



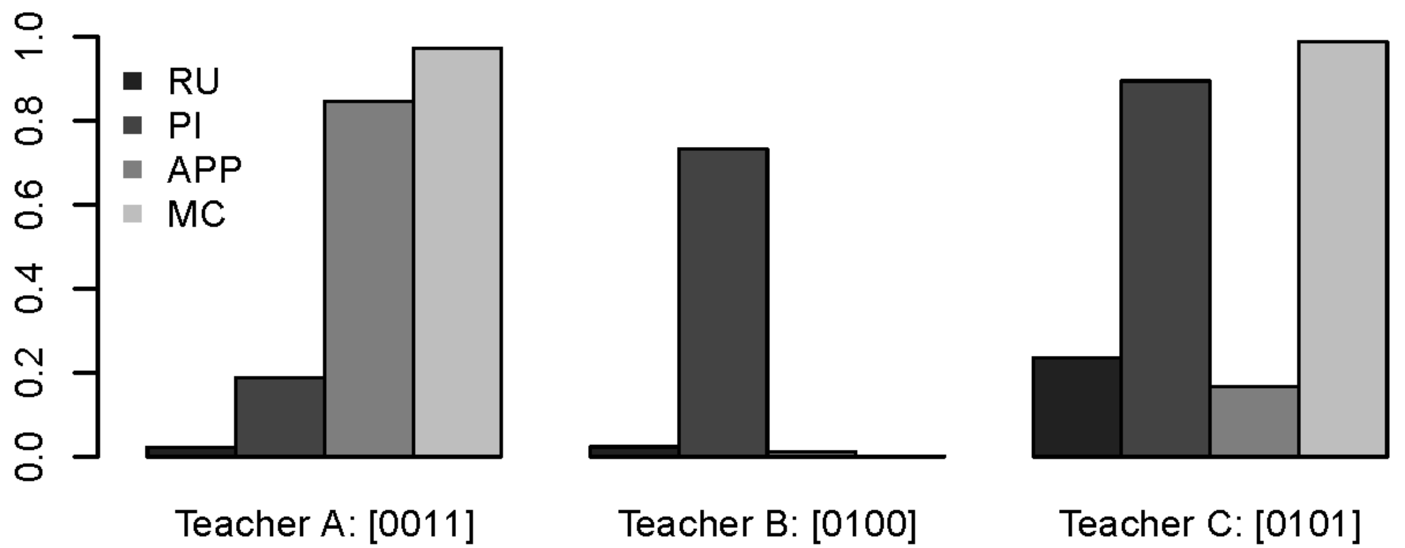
Individual Attribute Mastery



- Information useful for
  - Tailoring professional development
    - Many teachers may benefit from professional development on referent unit
  - Understanding base-rates of attribute mastery in the population of in-service teachers
    - Quantitative Research

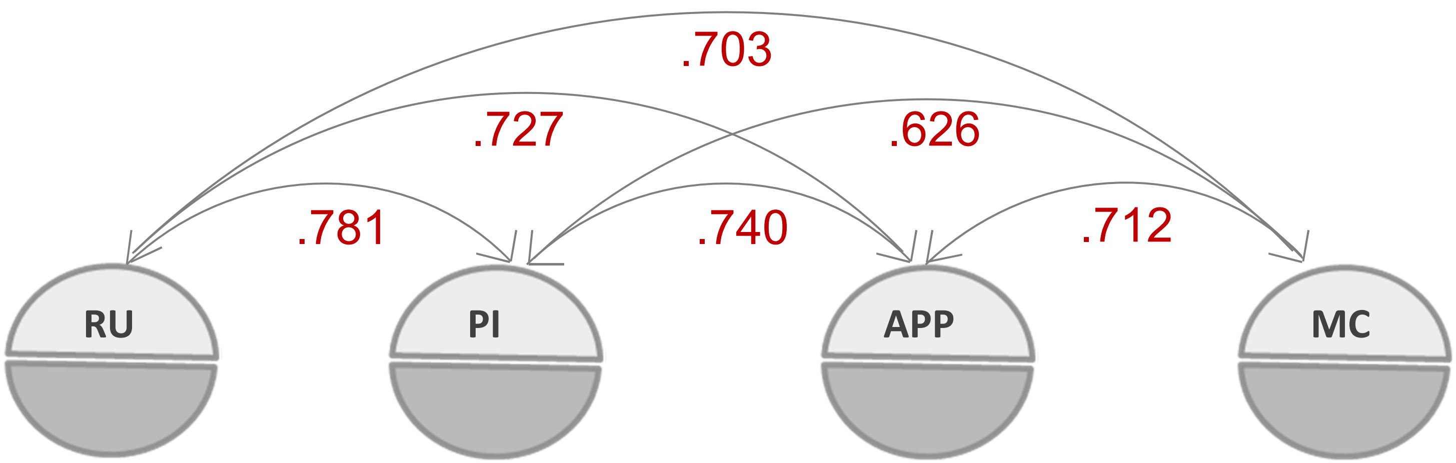
Teacher-level Individual Attribute Feedback

- Comparison of total scores and DCM diagnosis:
  - Teacher A, B, and C both answered 11 out of 27 items correctly
  - Teacher A has attribute pattern [0011]
  - Teacher B has attribute pattern [0100]
  - Teacher C has attribute pattern [0101]
    - Need different types of professional development

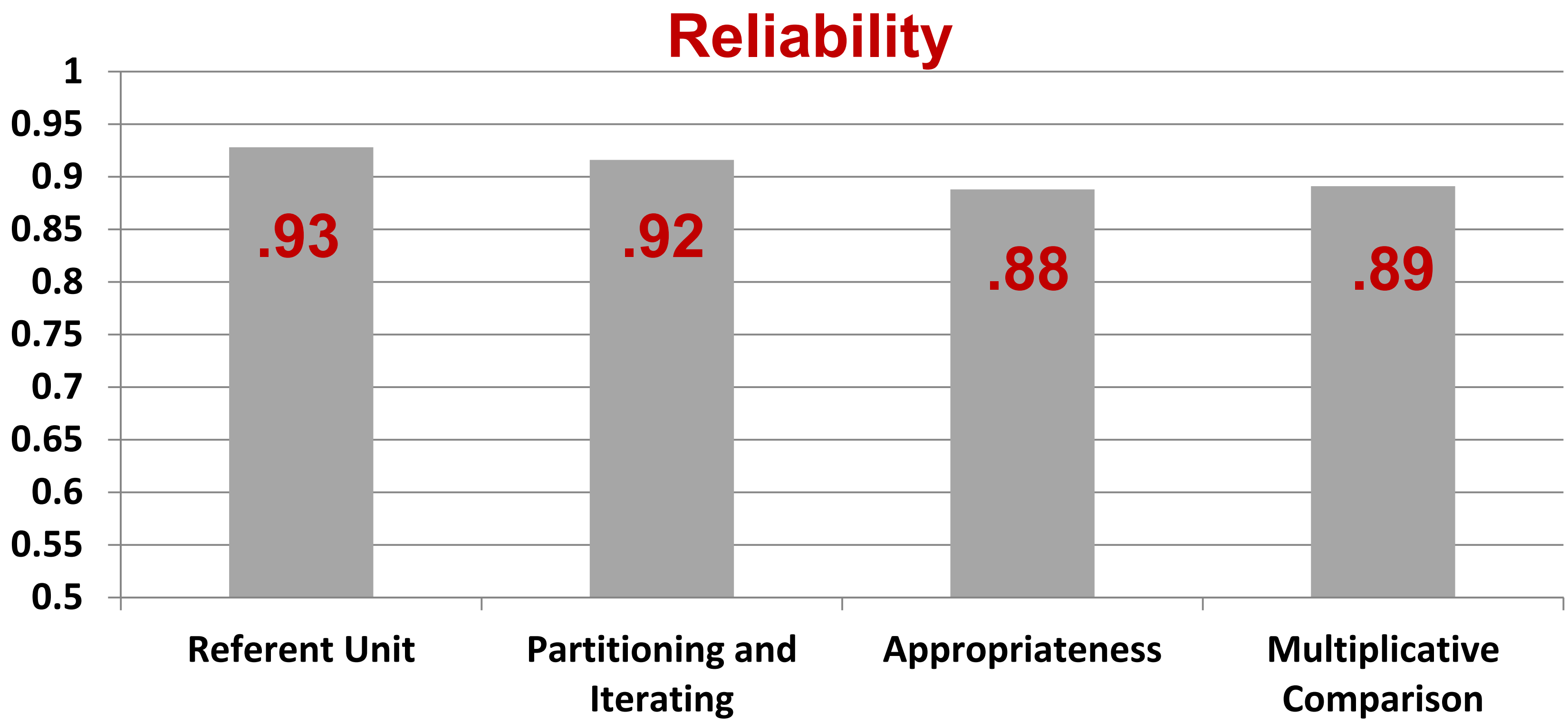


Attribute Correlations

- The attribute patterns are reflections of the correlations among the latent variables
  - Tetrachoric correlations (between categorical variables)
  - The relationships among the attributes are parameterized through a log-linear structural model
- Shows we have related but distinct dimensions
- More on attribute relationships in Section 4 about structural models



Individual Attribute Reliability



- Attributes measured by 15, 10, 5 and 5 items, respectively
- Gaining estimation precision by sacrificing latent trait precision

Concluding Remarks

Designing a Multidimensional Test

- Measuring multiple dimensions practically has been the argument for DCMs
  - However, many tests show traits too highly correlated
    - Typically retrofitting data
- This project is one of the first efforts to prospectively diagnose attributes
  - Further unique in that the attributes are cognitive in nature and very fine-grained
  - One possible model of how to do this in practice

Key Features of Successful DCM-based Test

1. Test purpose aligned with DCM purpose
2. Clearly-defined attributes
   - Theory delineating construct is strong
3. Strong items
   - High discrimination
   - High statistical information
4. Accurate Q-matrix
   - Strong understanding of how items relate to attributes
5. Accurate model parameterization
   - Model-data fit is critical to making valid inferences
   - Flexible model parameterization helps align theory and model
6. Sufficient data to estimate model
   - Examinees
   - Items
   - Items per attribute
     - Back to statistical information

Iterative Test Development

- Some grey area in confirmatory analyses
- Conjecture based approach
  - Begin with a hypothesis about the attributes, Q-matrix
  - Refine hypothesis with statistical feedback
    - LCDM parameters
- Both steps are important
  - Strong theory to inform model
  - Flexible model to inform theory
  - Cycles of test refinement yield stronger theory and more accurate inferences from test results

General Modeling Tips

- High-level interactions are difficult to estimate in most samples
  - More than 2-way interactions may not be possible
- Modeling strategy:
  - Try all item-level interactions
    - If model does not converge, limit to only 2-way interactions
  - Remove non-significant interactions from model
  - If all interactions and main effects for an attribute are close to zero:
    - Try removing entry for attribute in Q-matrix
    - More on this in Section 5 about model-data fit

Wrap-Up and Take-Home Points

- Session 3 demonstrated a potential use of DCMs

- Prospective applications of DCMs are rare
  - Tests aren’t designed to measure categorical attributes
    - Item information is different in DCMs
  - Users haven’t had access to software
    - Previously, most applications use software built by researchers
      - MCMC in Fortran or WinBugs
      - MML in Fortran
    - Now, researchers can use Mplus and more recently, FlexMIRT

Notes on Usefulness of DCMs

- Full utility of DCMs cannot be understood until applications become more frequent
  - Many papers to this point have used sub-optimal data and problems
- Funding opportunities exist and seem to review well
  - Educational Measurement: NSF (DR-K12); IES (Goals 2 and 5)
  - Psychological Measurement: NIH (NIMH; NIDA; NIA;…)
- Industry seems interested
  - ACT/College Board/ETS/Measurement Inc./Pearson