Multidimensional Item Response Theory Equating

Stella Y. Kim
University of North Carolina at Charlotte

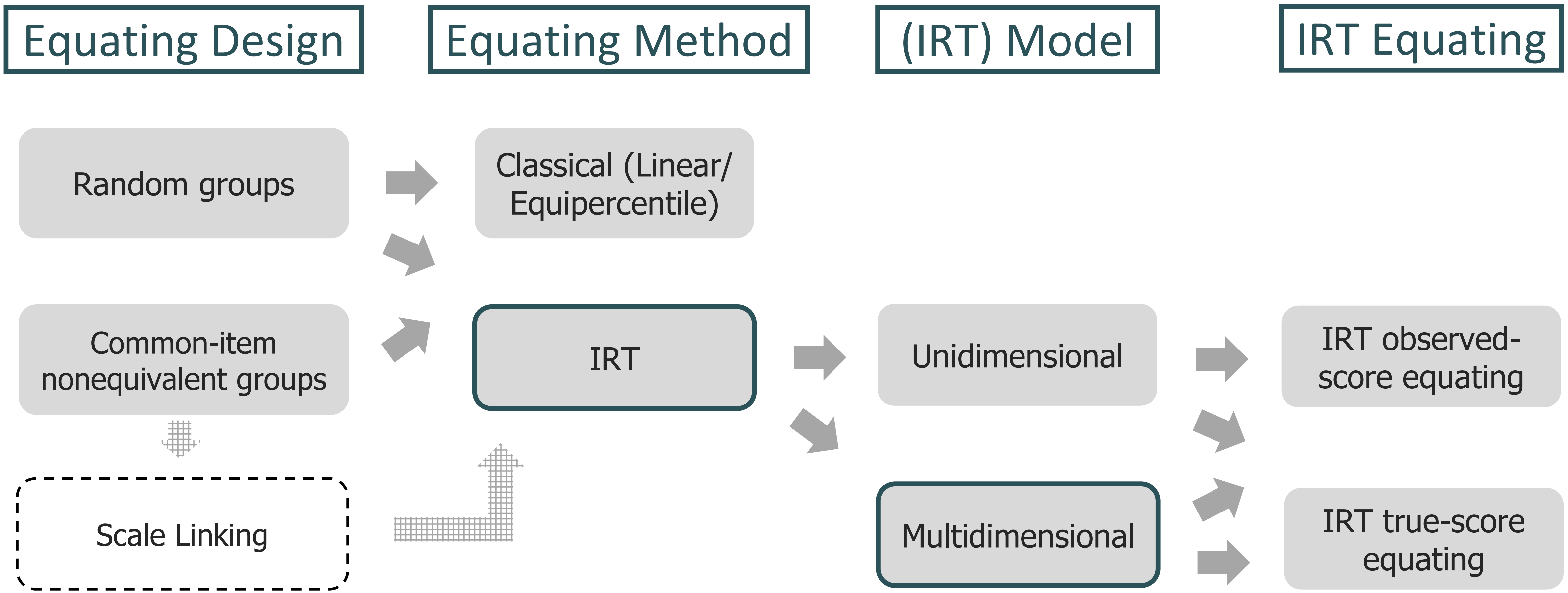
Learning Objectives

1. Explain benefits of using multidimensional IRT equating
2. List multidimensional IRT models that have been applied to equating in the literature
3. Describe the current literature on multidimensional IRT equating
4. Conduct multidimensional IRT observed-score equating
5. Examine potential research topics in multidimensional IRT equating

Prerequisite Knowledge

- Basic knowledge in IRT
  1. Unidimensionality assumption
  2. 3 parameter logistic model
  3. Lord-Wingersky formula
  4. IRT scale linking
- Basic knowledge in equating
  1. Equating designs
  2. IRT equating (observed-score equating and true-score equating)

Multidimensional IRT Equating is…



Useful Other Resources

- ITEMS Modules
    1. Module 6: Equating methods in classical test theory (Kolen)
    2. Module 10: Equating methods in item response theory (Cook & Eignor)
    3. Module 21: Multidimensional item response theory (Ackerman)
    4. Digital Module 25: Testlets models (Jiao & Liao)
- NCME Youtube Channel
    1. Introduction to Equating (by Dr. Robert Brennan)
    2. IRT Equating Methods (by Dr. Jaime Malatesta)

Introduction to MIRT Equating

1

Section Learning Objectives

1 Introduction to MIRT Equating

Describe the unidimensionality assumption required by IRT

List potential sources of multidimensionality of a test

Explain benefits of using multidimensional IRT in the context of equating

Unidimensionality Assumption

- Only a single ability is measured by items in a test
- Psychological and educational processes are very complex (Reckase, 2009)

To what extent is a test multidimensional ?

Sources of Multidimensionality

- Content domains (Reise, 2012)
- Multiple item formats (Bridgeman, 1992)
- Testlets (Jiao & Liao, 2021; Wainer et al., 2007)

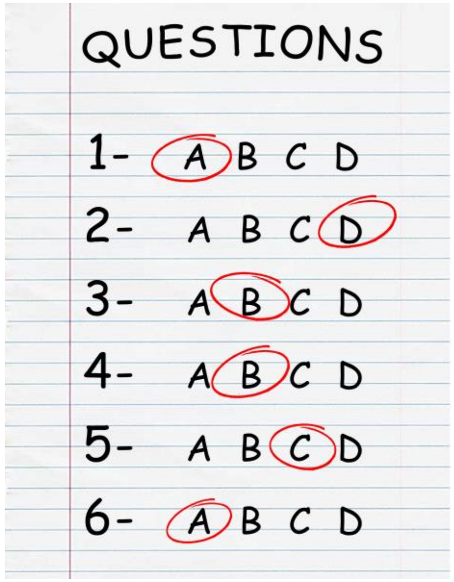
Sources of Multidimensionality

- Content domains (Reise, 2012)
- Multiple item formats (Bridgeman, 1992)
- Testlets (Jiao & Liao, 2021; Wainer et al., 2007)

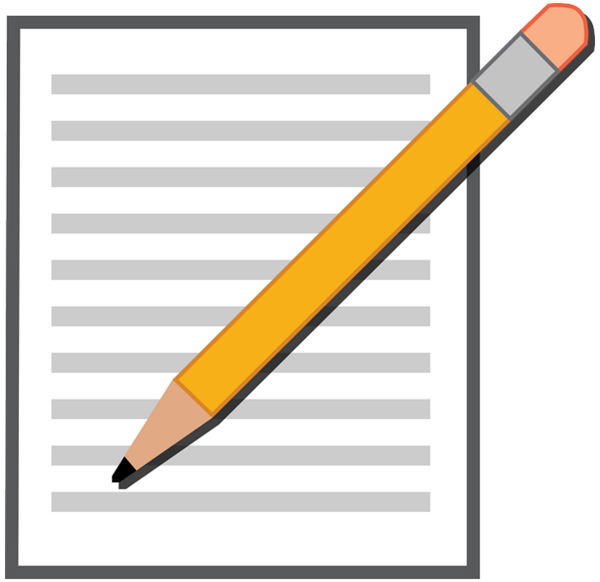
TIMSS Mathematics for Grade 8			
Number	Geometry	Data & Chance	Algebra

Sources of Multidimensionality

- Content domains (Reise, 2012)
- Multiple item formats (Bridgeman, 1992)
- Testlets (Jiao & Liao, 2021; Wainer et al., 2007)



Spanish Assessment	
Multiple-Choice items (Reading)	Essay (Writing)



Sources of Multidimensionality

- Content domains (Reise, 2012)
- Multiple item formats (Bridgeman, 1992)
- Testlets (Jiao & Liao, 2021; Wainer et al., 2007)

Reading Test			
A paragraph on social media	A paragraph on climate changes	A paragraph on the history of classic music	A paragraph on different kinds of fossils

Applications of MIRT

Application of MIRT

- Differential item functioning (Ackerman, 1992)
- Test construction (Reckase, et al., 1988)
- Computerized adaptive testing (Segall, 1996)
- Scale linking (Davey, et al., 1996)
- Equating (Lee & Brossman, 2012)

Impacts of Violation on Equating

- Inaccurate item and person parameter estimates (Ansley & Forsyth, 1985)

- MIRT equating methods provide fewer equating errors than UIRT equating methods (Lee & Brossman, 2012)

Multidimensional IRT Models

2

Section Learning Objectives

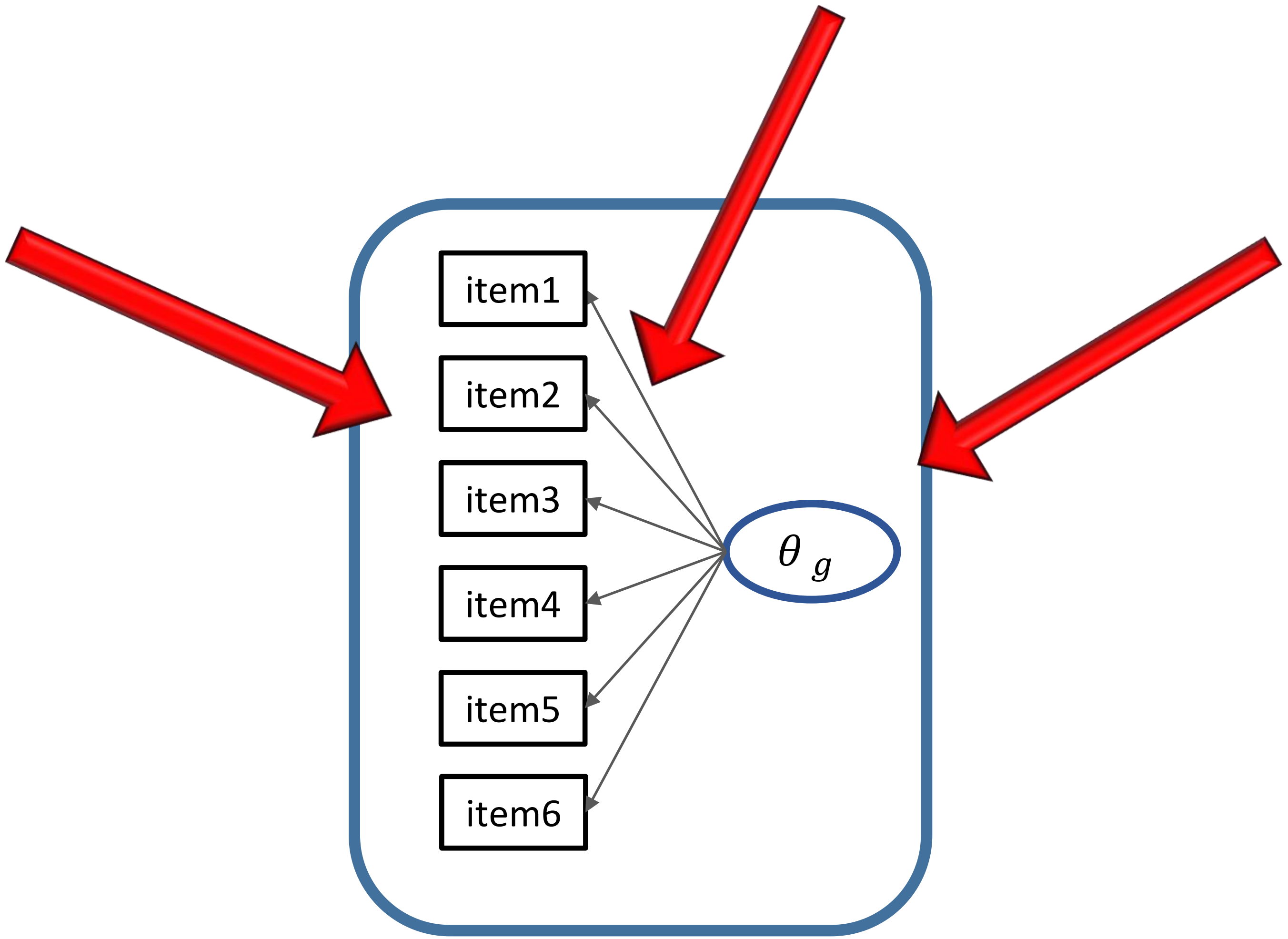
2 Multidimensional IRT Models

List multidimensional IRT models that have appeared in the equating context

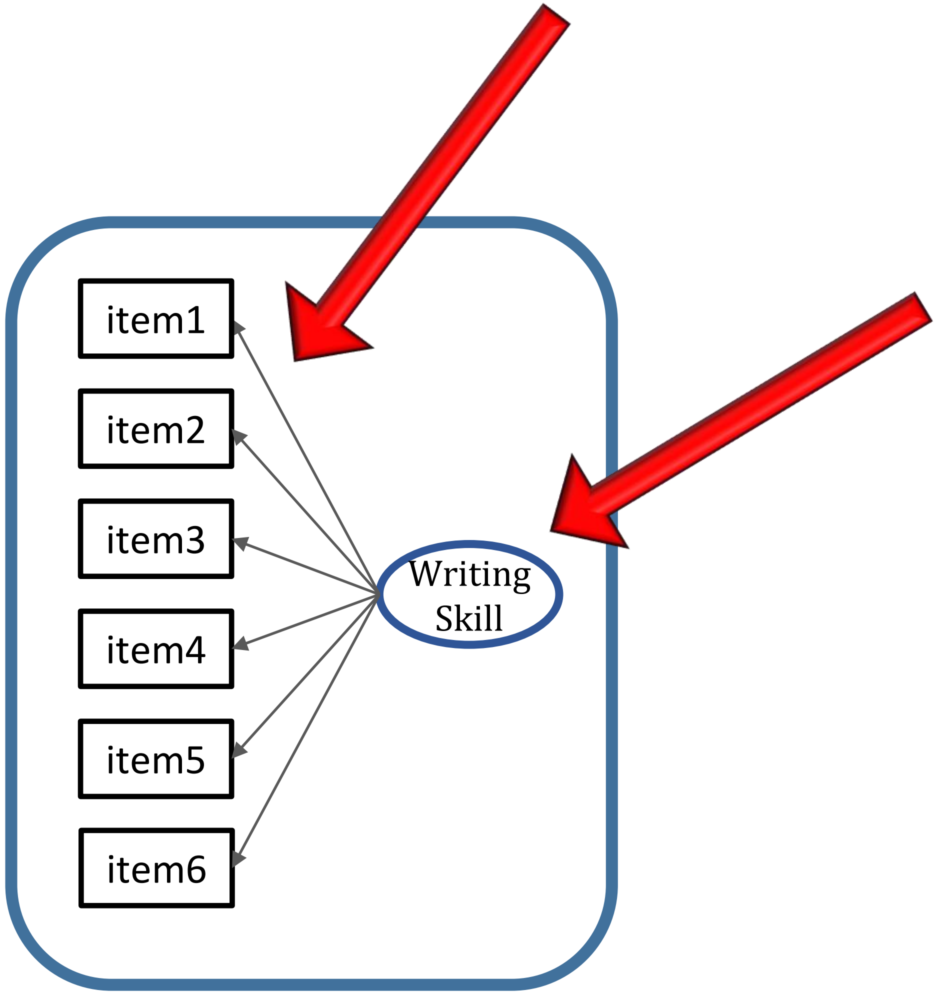
Describe characteristics of each MIRT model

Give examples in which each model is appropriate to be used

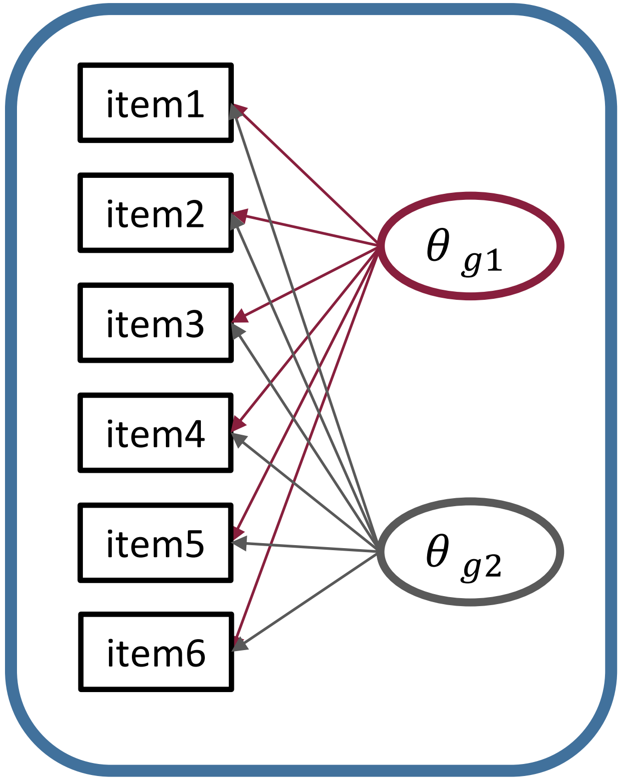
Unidimensional IRT



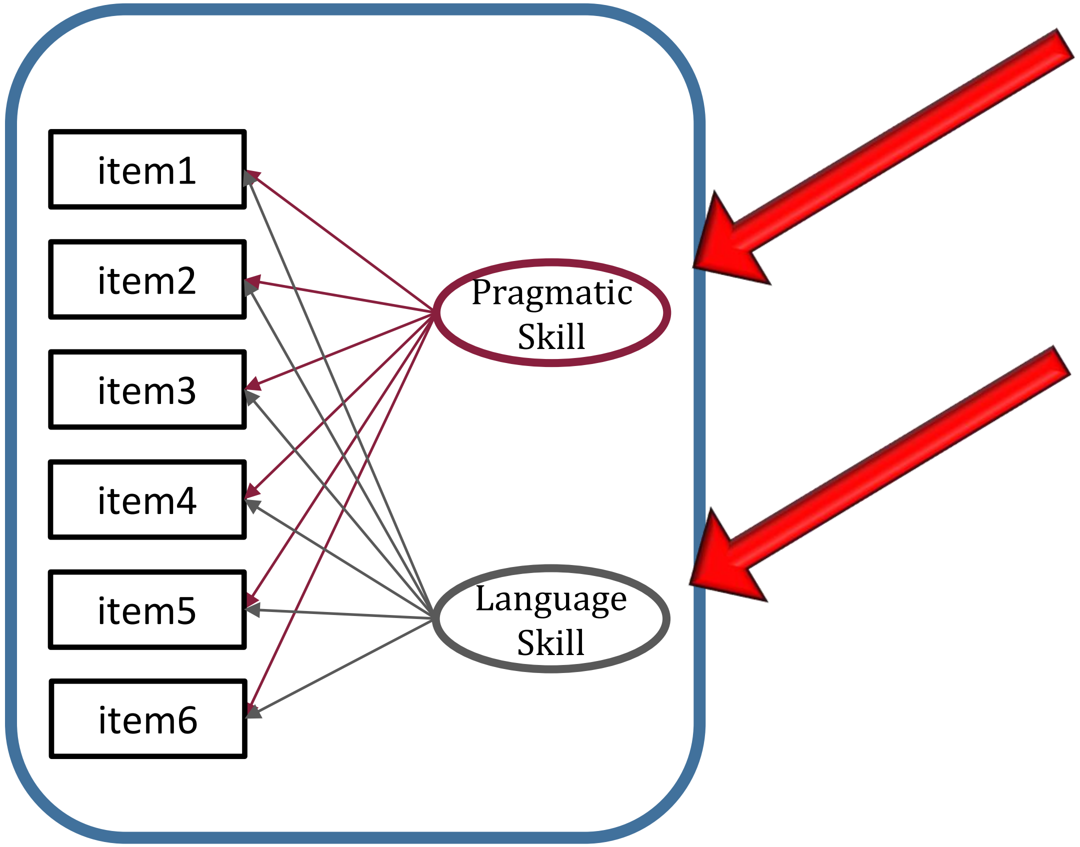
Unidimensional IRT



Full MIRT

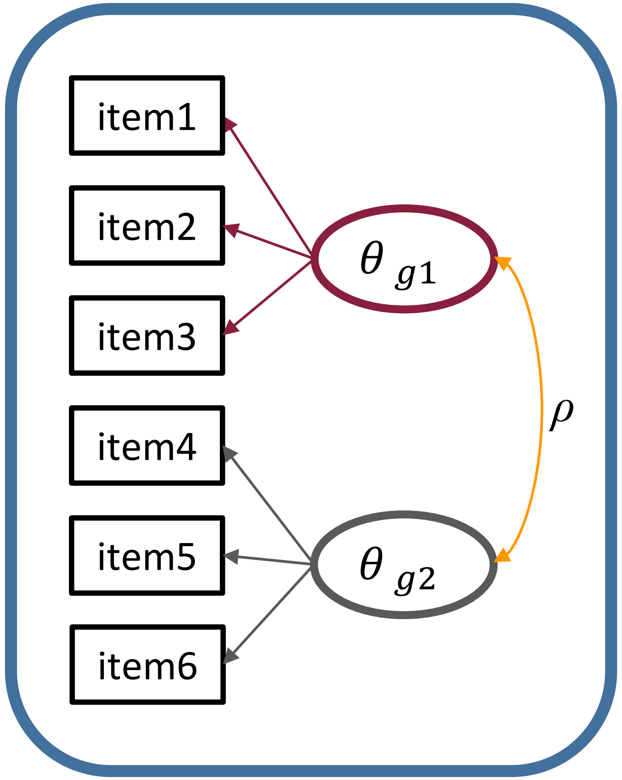


Full MIRT

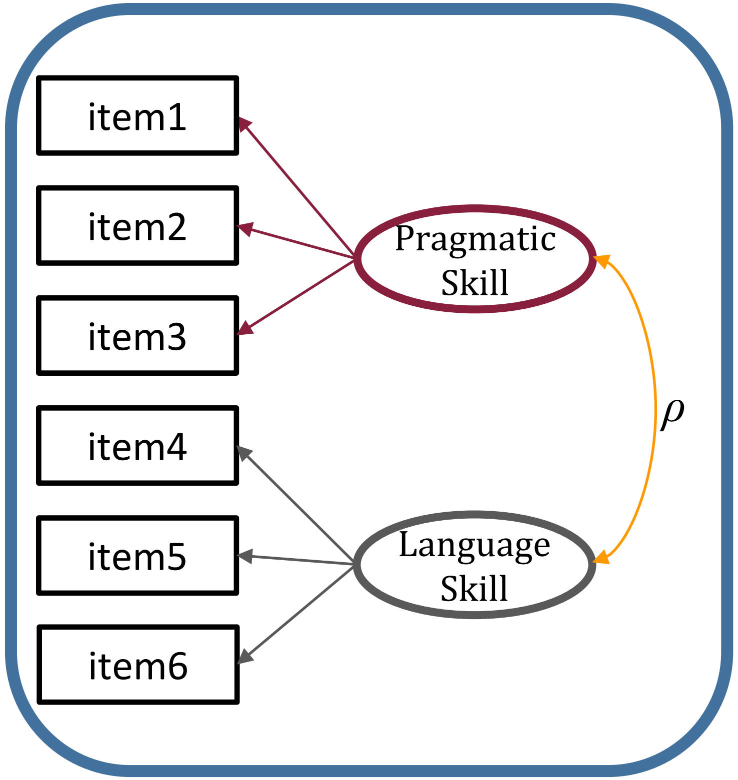


item1
item2
item3
item4
item5
item6
Pragmatic Skill
Language Skill

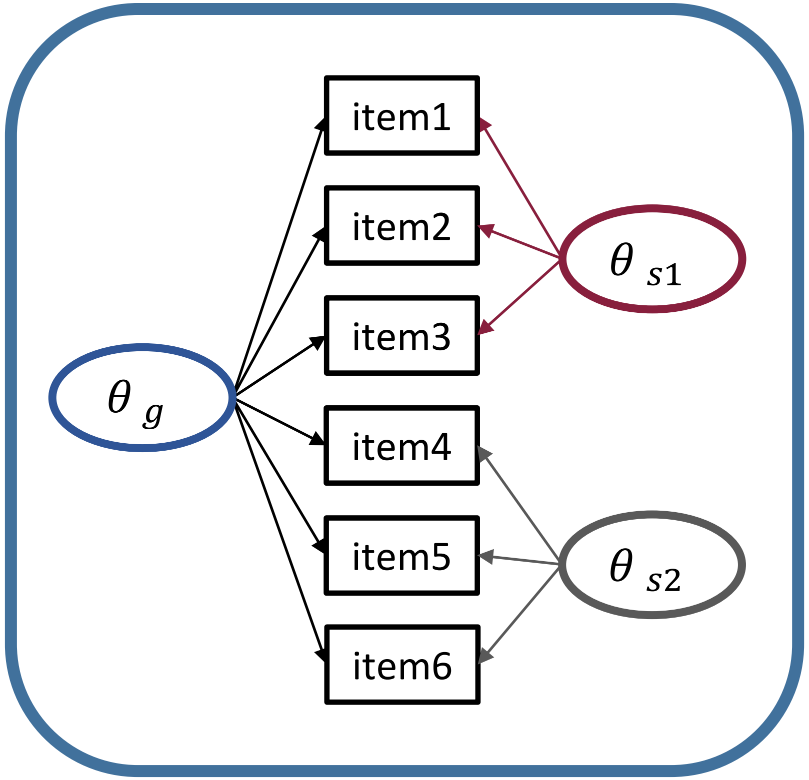
Simple-Structure MIRT



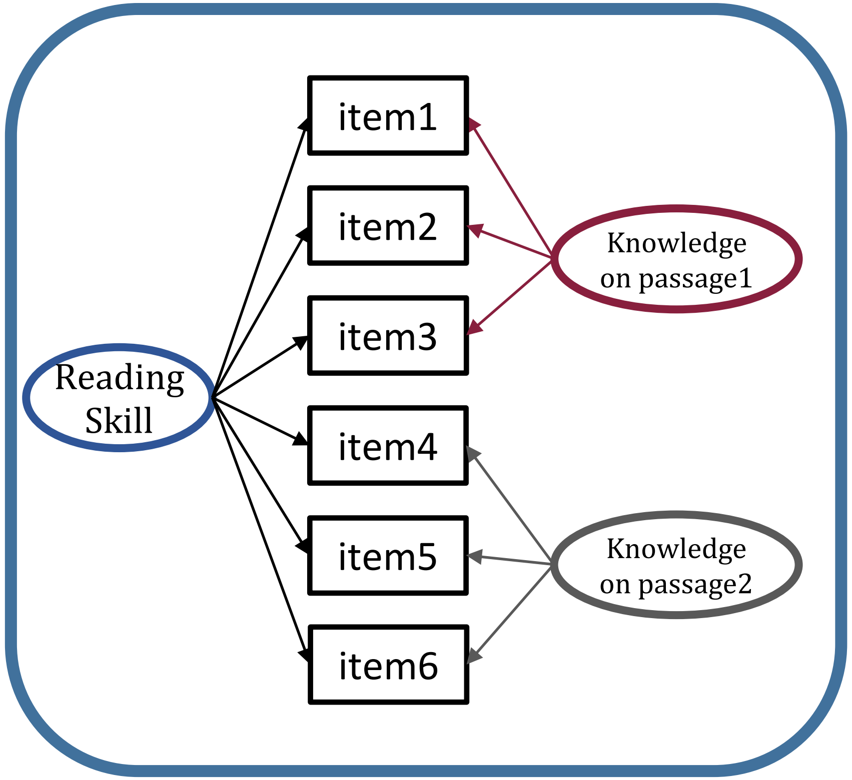
Simple-Structure MIRT



Bi-Factor MIRT



Bi-Factor MIRT



Literature on MIRT Equating

3

Section Learning Objectives

3 Literature on MIRT Equating

- Explain the history of MIRT equating research
- Summarize the major findings on MIRT equating reported by previous studies
- Articulate the gap in the MIRT equating literature
- Explain challenges associated with MIRT true-score equating

Literature Review: Development

Author(s) / Year	Purpose	Equating Design	Equating Methods	Data
Lee & Brossman (2012)	Development of MIRT equating methods	Random groups	Simple-structure MIRT observed	· Real
Brossman & Lee (2013)			Full MIRT observed Uni Approx. of MIRT observed Uni Approx. of MIRT true	· Real · Simulated
Lee & Lee (2016)			Bi-factor MIRT observed	· Pseudo-forms · Simulated
Tao & Cao (2016)			Testlet-response MIRT observed Testlet-response MIRT true	· Simulated
Lee et al. (2016)			Bi-factor MIRT true	· Real · Simulated
Kim & Cho (2020)			Projective bi-factor MIRT true	· Simulated
Kim, Lee, & Kolen (2020)		Random Groups CINEG	Simple-structure MIRT true	· Real · Pseudo-forms · Simulated

Literature Review: Comparison

Author(s) / Year	Purpose	Equating Design	Equating Methods	Data
Peterson & Lee (2014)	Comparison of equating methods	Random groups	Full MIRT observed Bi-factor MIRT observed	· Pseudo-forms · Equating back to itself
Lee, Lee, & Brennan (2014)		Random groups	Full MIRT observed Uni Approx. of MIRT observed Uni Approx. of MIRT true	· Simulated
Kim, Lim, & Lee (2019)		Random groups	Bi-factor MIRT observed Bi-factor MIRT true Testlet-response MIRT observed Testlet-response MIRT true	· Simulated
Choi (2019)		CINEG/NEAT	Simple-structure MIRT observed Bi-factor MIRT observed Full MIRT observed	· Real · Pseudo-forms
Panidvadtana et al.(2021)		CINEG/NEAT	Full MIRT observed	· Simulated

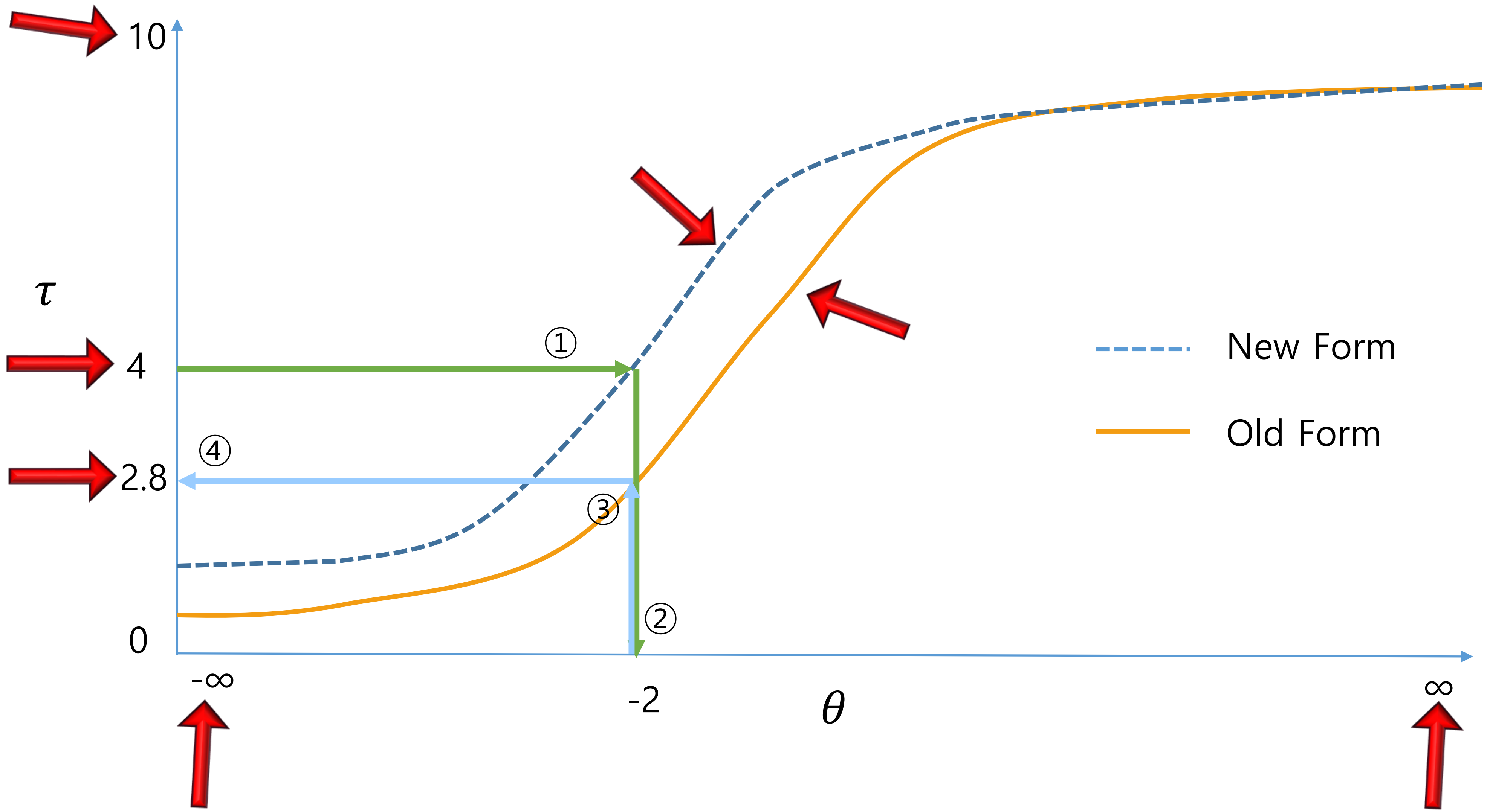
Summary of the Literature

- Multidimensional approaches provided **more accurate equating results** for multidimensional data (e.g., $r < .8$) (Kim et al., 2020; Peterson & Lee, 2014).
- A disattenuated correlation tells the relationship between two thetas with a higher value indicating less multidimensionality of data.
- For instance, under the simple-structure model, if the correlation is high (e.g., .95), then the test is essentially unidimensional as the two traits are not distinguishable and are almost the same or similar construct.

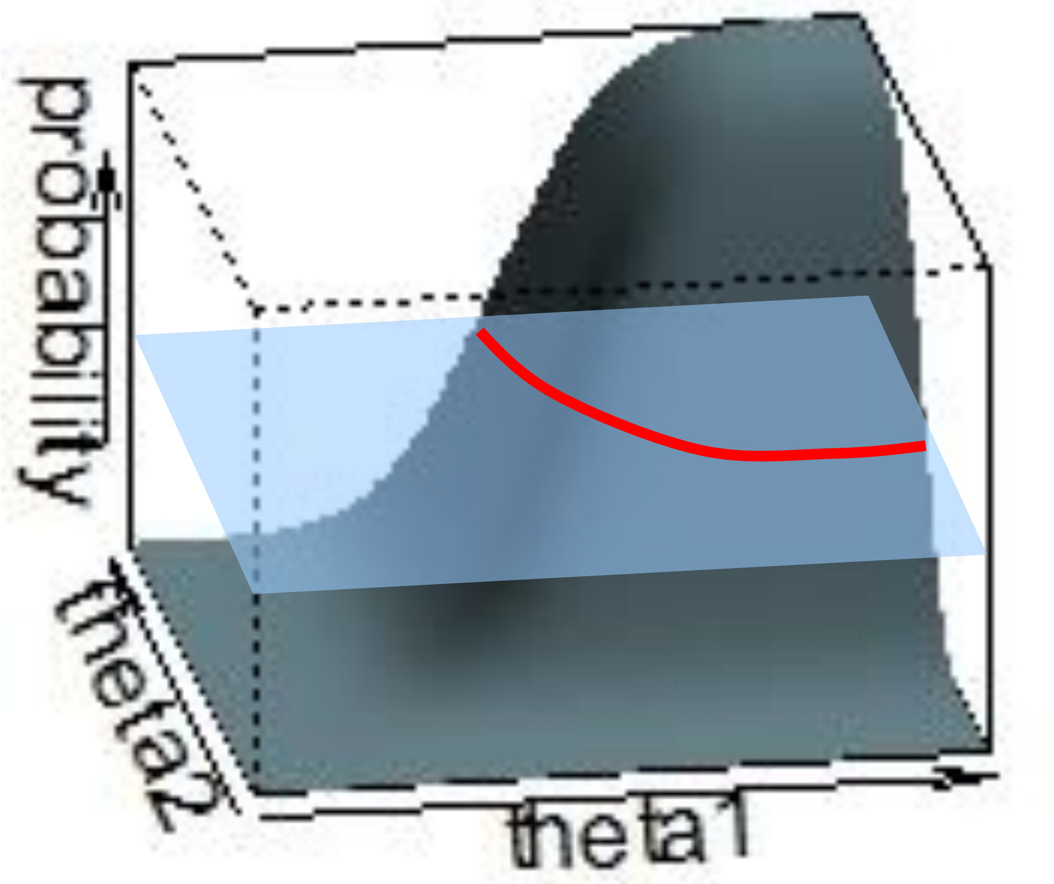
Summary of the Literature

- Study factors examined included a) degree of multidimensionality, b) test length, c) sample size, d) form difference, e) sources of dimensionality, and f) the number of dimensions.
- Most research was conducted under the random groups design, except for a few recent works.
- More attention has been paid to observed-score equating.

True Score Equating on UIRT



Challenges with MIRT True-Score Equating



Which pair of θ_1 & θ_2 to be used?

MIRT Observed-Score Equating

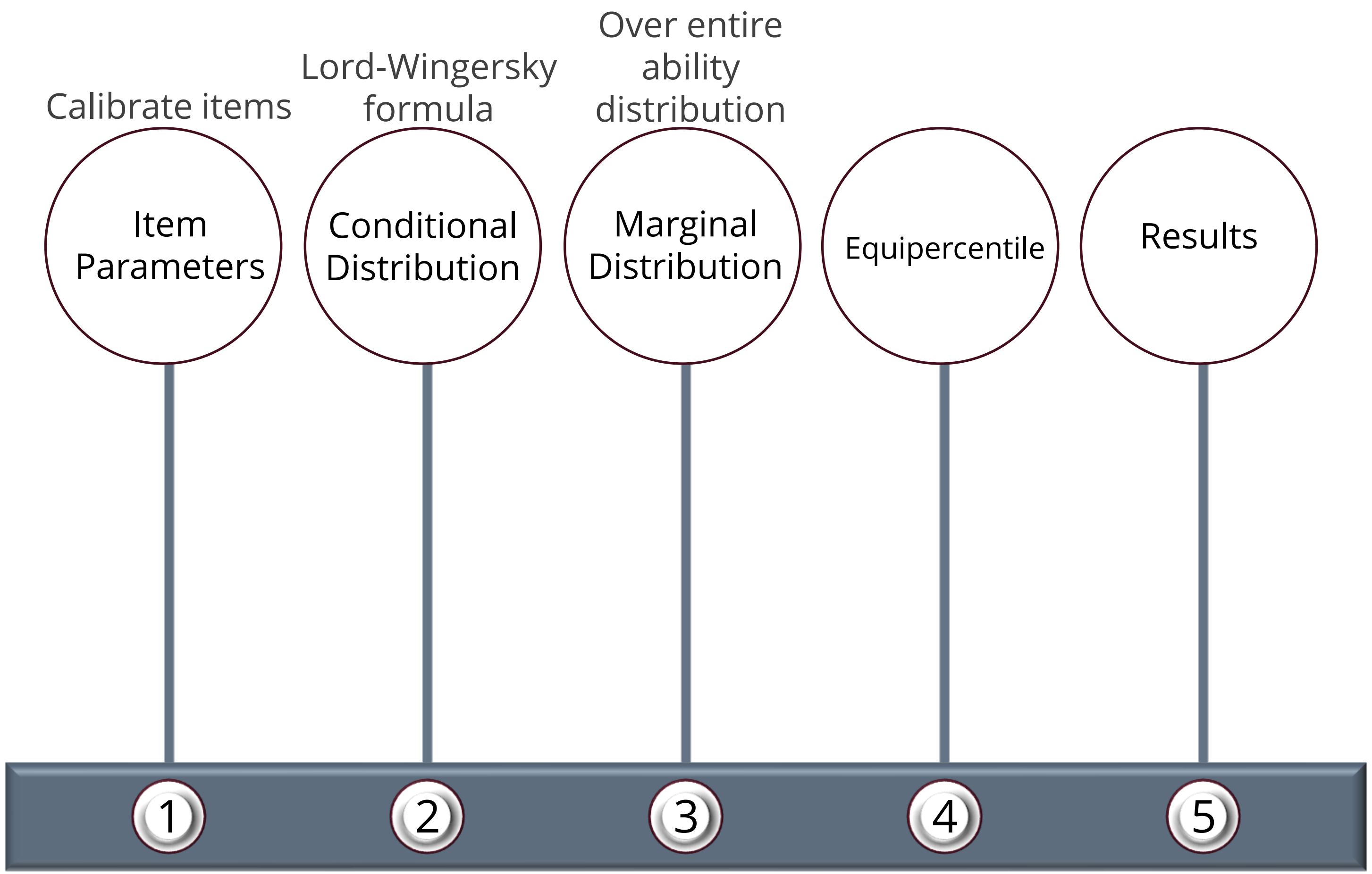
4

Section Learning Objectives

4 MIRT Observed-Score Equating

- Describe how observed-score equating is performed with MIRT models
- Identify differences between MIRT observed-score equating and UIRT observed-score equating

Procedure of MIRT Observed-Score Equating



Simple-Structure MIRT Observed-Score Equating

Under 3 parameter-logistic model
(*Note*. Item-specific and person-specific notations are omitted for simplicity)

Item Response Functions With Two Dimensions

Items tapping 1st dimension: $$P(\theta_1) = c + \frac{1-c}{1+e^{\{-1.7a[\theta_1+b]\}}}$$

Items tapping 2nd dimension: $$P(\theta_2) = c + \frac{1-c}{1+e^{\{-1.7a[\theta_2+b]\}}}$$

Simple-Structure MIRT Observed-Score Equating

Under 3 parameter-logistic model

Item Response Functions With Two Dimensions

Items tapping 1st dimension: $$P(\theta_1) = c + \frac{1-c}{1+e^{\{-1.7a[\theta_1+b]\}}}$$

Items tapping 2nd dimension: $$P(\theta_2) = c + \frac{1-c}{1+e^{\{-1.7a[\theta_2+b]\}}}$$

Lord-Wingersky formula with multiple abilities:

$$\begin{aligned} f_r(x|\boldsymbol{\theta}) &= f_{r-1}(x|\boldsymbol{\theta})(1-P_r), & x = 0 \\ &= f_{r-1}(x|\boldsymbol{\theta})(1-P_r)+f_{r-1}(x-1|\boldsymbol{\theta})P_r, & 0 < x < r \\ &= f_{r-1}(x-1|\boldsymbol{\theta})P_r & x = r \end{aligned}$$

Simple-Structure MIRT Observed-Score Equating

Under 3 parameter-logistic model

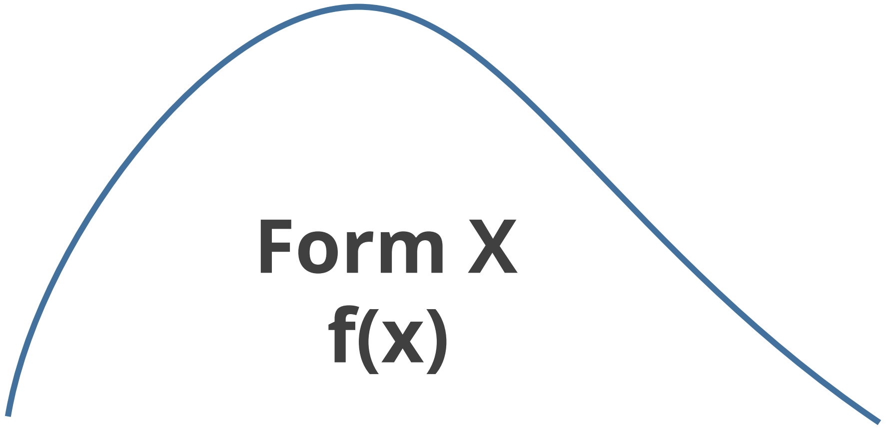
Marginal observed-score distribution:

$$f(x) = \iint_{-\infty}^{\infty} f(x|\theta_1, \theta_2) g(\theta_1, \theta_2) d(\theta_1) d(\theta_2)$$

From the previous slide

$$f(x) = \sum_{\theta_1} \sum_{\theta_2} f(x|\theta_1, \theta_2) g(\theta_1, \theta_2) .$$

Bivariate normal: MVN (0, I, ρ)



	A	B	C
1	theta1	theta2	p
2	-4	-4	3.08E-06
3	-4	-3.8	2.87E-06
4	-4	-3.6	1.47E-06
5	-4	-3.4	4.13E-07
6	-4	-3.2	6.36E-08
7	-4	-3	5.37E-09
8	-4	-2.8	2.48E-10
9	-4	-2.6	6.29E-12

Simple-Structure MIRT Observed-Score Equating

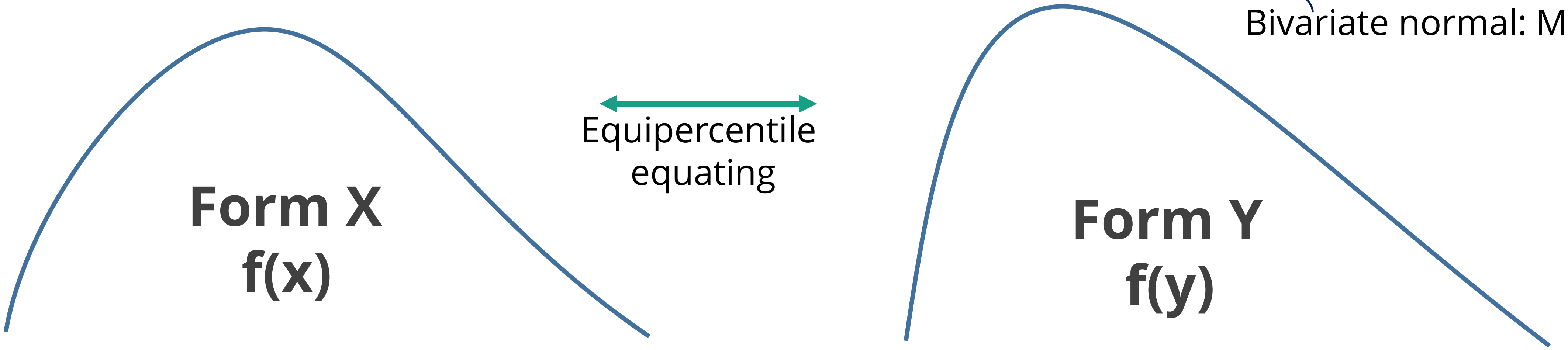
Under 3 parameter-logistic model

Marginal observed-score distribution:

$$f(x) = \iint_{-\infty}^{\infty} f(x|\theta_1, \theta_2) g(\theta_1, \theta_2) d(\theta_1) d(\theta_2)$$

$$f(x) = \sum_{\theta_1} \sum_{\theta_2} f(x|\theta_1, \theta_2)\, g(\theta_1, \theta_2)\,.$$

Bivariate normal: MVN (0, I, ρ)



Future Research

5

Section Learning Objectives

5 Future Research

- Recognize practical constraints or considerations in applying MIRT equating to practices
- Describe how to evaluate usefulness and appropriateness of MIRT equating in operational settings
- Examine potential research topics

Practical Considerations

- Dimensionality Assessment
  - Scree plot
  - Parallel analysis
  - Latent trait correlations or disattenuated correlations

- Multidimensionality benchmark
  - r < .8

```
    42                  FR2   90    2.90    0.07   91    2.37
    43                  FR3   99    3.01    0.07  100    2.42

Simple-structure MIRT item calibration
Example file for SS-MIRT

Group Latent Variable Means:
   Group               Label   P#    mu  1    s.e.   P#    mu  2
       1              Group1        0.00    ----         0.00

Latent Variable Variance-Covariance Matrix for Group  1: Group1
   P# Theta  1    s.e.   P# Theta  2    s.e.
          1.00    ----
  151     0.47    0.02          1.00    ----
```

Practical Considerations

- Number of Dimensions specified
  - Substantial reduction in estimation efficiency (with +4 dimensions)
  - The use of 41 quadrature points for 2 dimensions leads to 41^2 (=1,681) quadrature points to be evaluated
  - Metropolis-Hastings Robbins-Monro (MH-RM; Cai, 2010)

- Confirmatory nature of some MIRT models
  - Simple-structure, bi-factor, testlet-response, etc.
  - Model determined by the table of specifications

Potential Research Areas

- Common-item Nonequivalent Groups Design (CINEG)
  - For UIRT, several linking procedures exist including the Stocking-Lord method, Haebara method, the mean-mean and mean-sigma methods
  - For MIRT, the rotation and the orientation of the coordinate axis need to be adjusted
  - The rotation issue can be resolved by setting correlations to zero while adjusting the orientation of the axis is more complex to achieve
  - Most research conducted using concurrent calibration
  - MIRT linking with separate calibration (Oshima, Davey, & Lee, 2000)

Potential Research Areas

- Number of Dimensions
  - Mostly 2-3 dimensions in the current literature
  - Quadrature points and weights – dramatically increases as the number of dimensions increases
  - Parameter estimation process needs to be enhanced

- Differential multidimensionality
  - More than 2 dimensions (unequal level of multidimensionality across dimensions)
  - Between groups (very multidimensional for one group and essentially unidimensional for the other group)

Potential Research Areas

- True-score equating
  - How to approximate the relationship between a set of thetas on two forms?
- Practical implications
  - Multidimensionality due to group characteristics or form (test) characteristics?
  - Impact on reported scores, classification decisions, etc.

You have reached the end of this section

Kim, S. Y. (2022). Multidimensional Item Response Theory Equating [Digital ITEMS Module 29]. *Educational Measurement: Issues and Practice, 41*(3), 85-86.