MEASURING INTERVENTION EFFECTIVENESS:

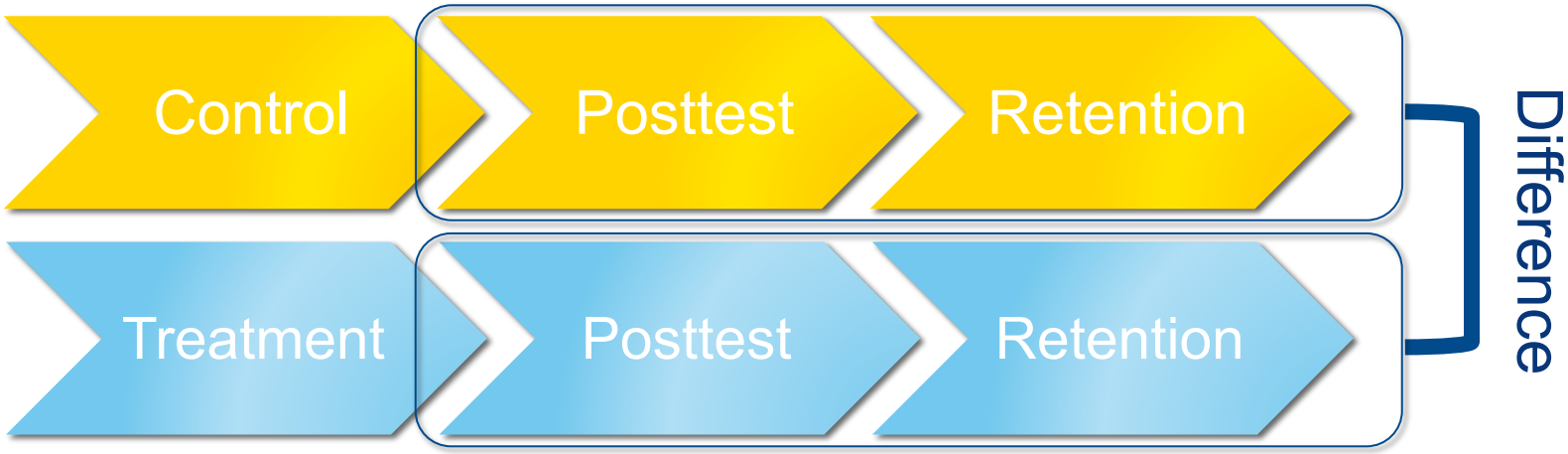
THE BENEFITS OF AN ITEM RESPONSE THEORY APPROACH

Society for Research in Educational Effectiveness
September 6, 2012

Katherine McEldoon, Sun-Joo Cho,
& Bethany Rittle-Johnson

Department of Psychology & Human Development
Peabody College, Vanderbilt University

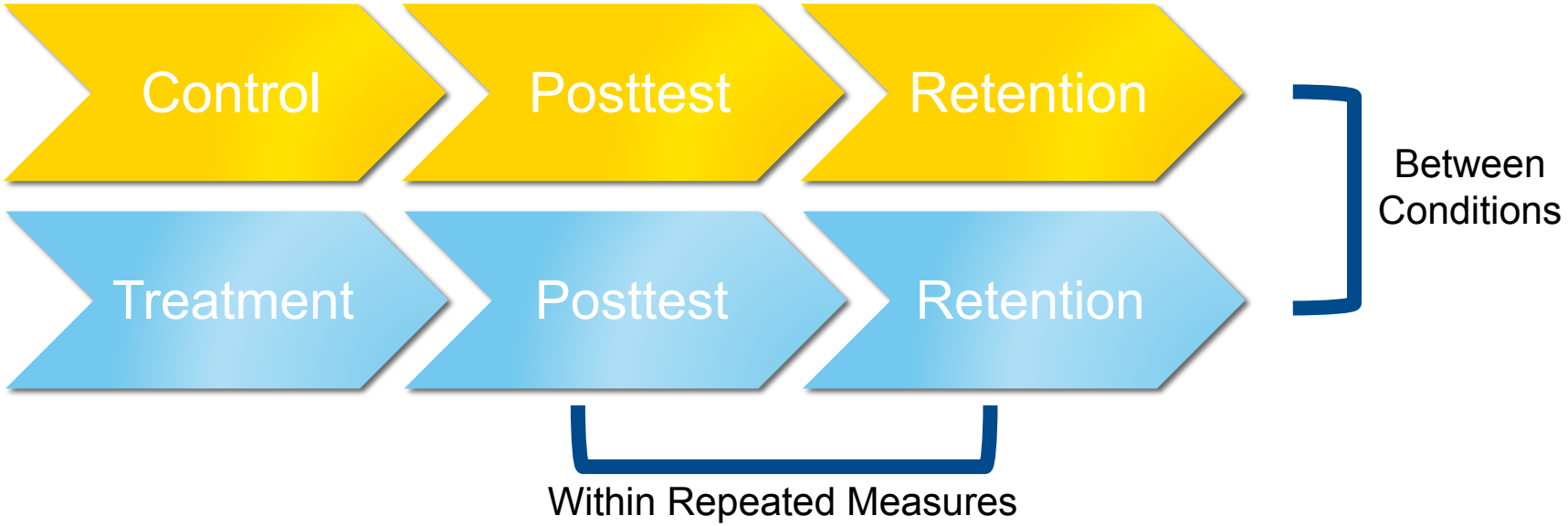
Intervention Research Design



Roadmap

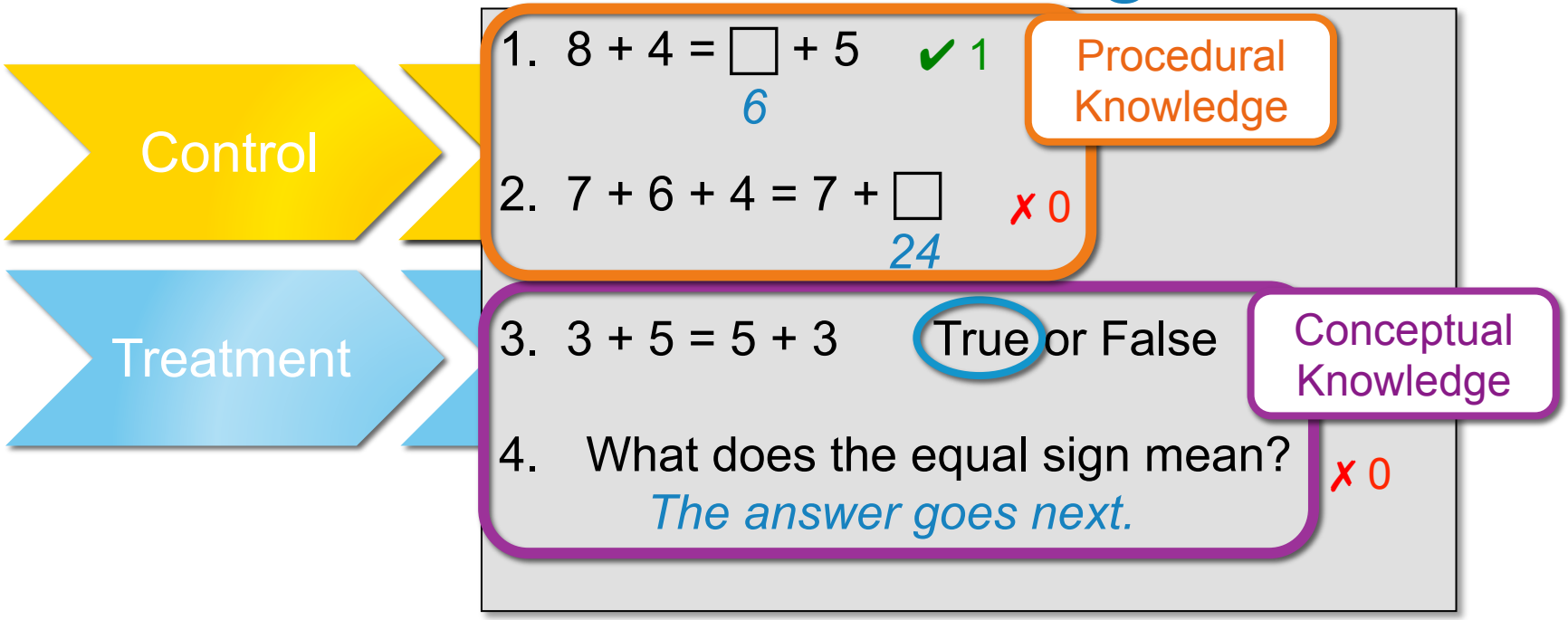
- Issues with RM ANOVA
- Item Response Theory (IRT) Models
  - Overcome limitations of RM ANOVA
  - Offer additional benefits
- GEL MIRT Model as alternative to RM ANOVA

Intervention Research Design



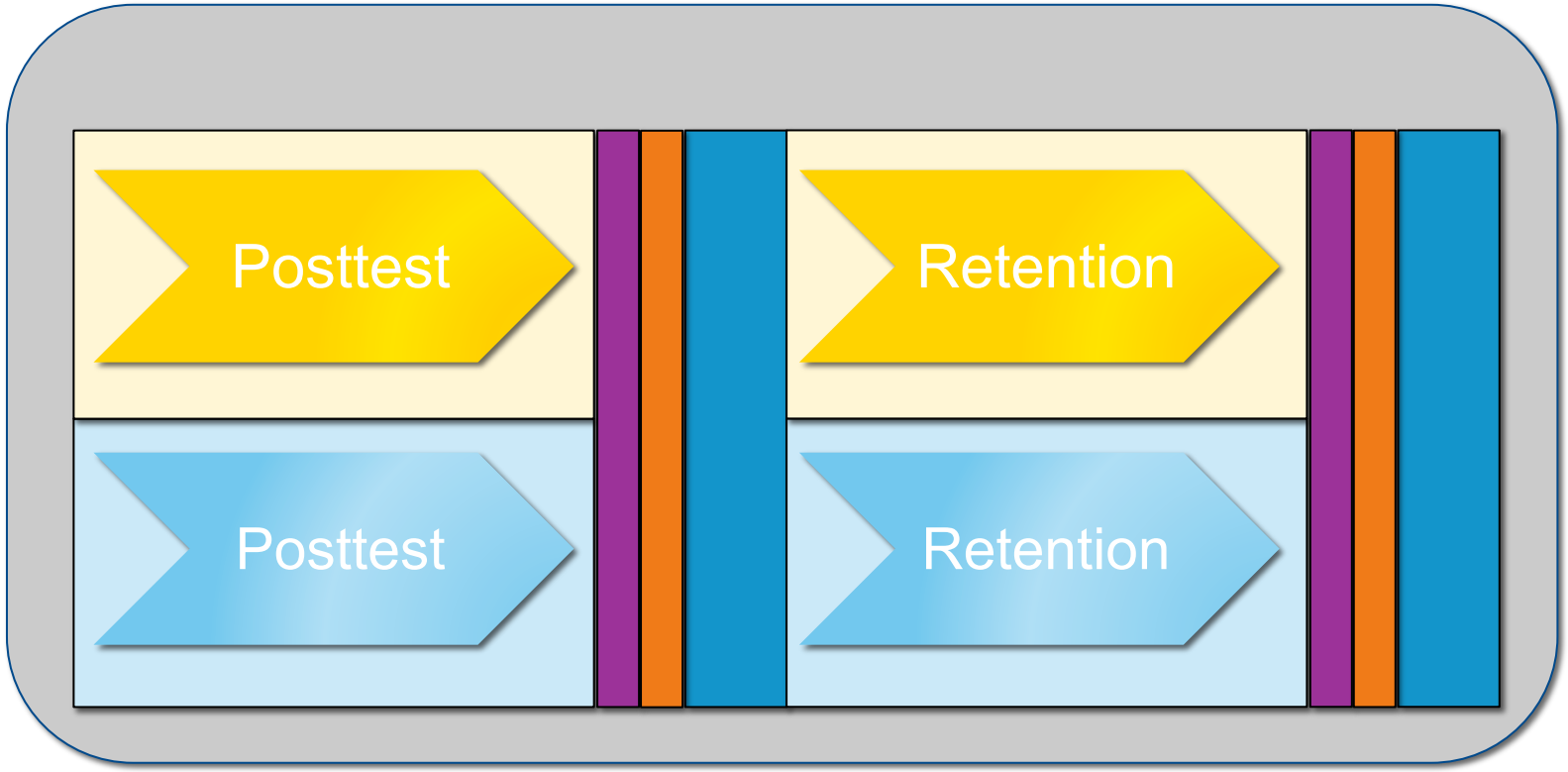
- Between-Within (split-plot) Design

Intervention Research Design

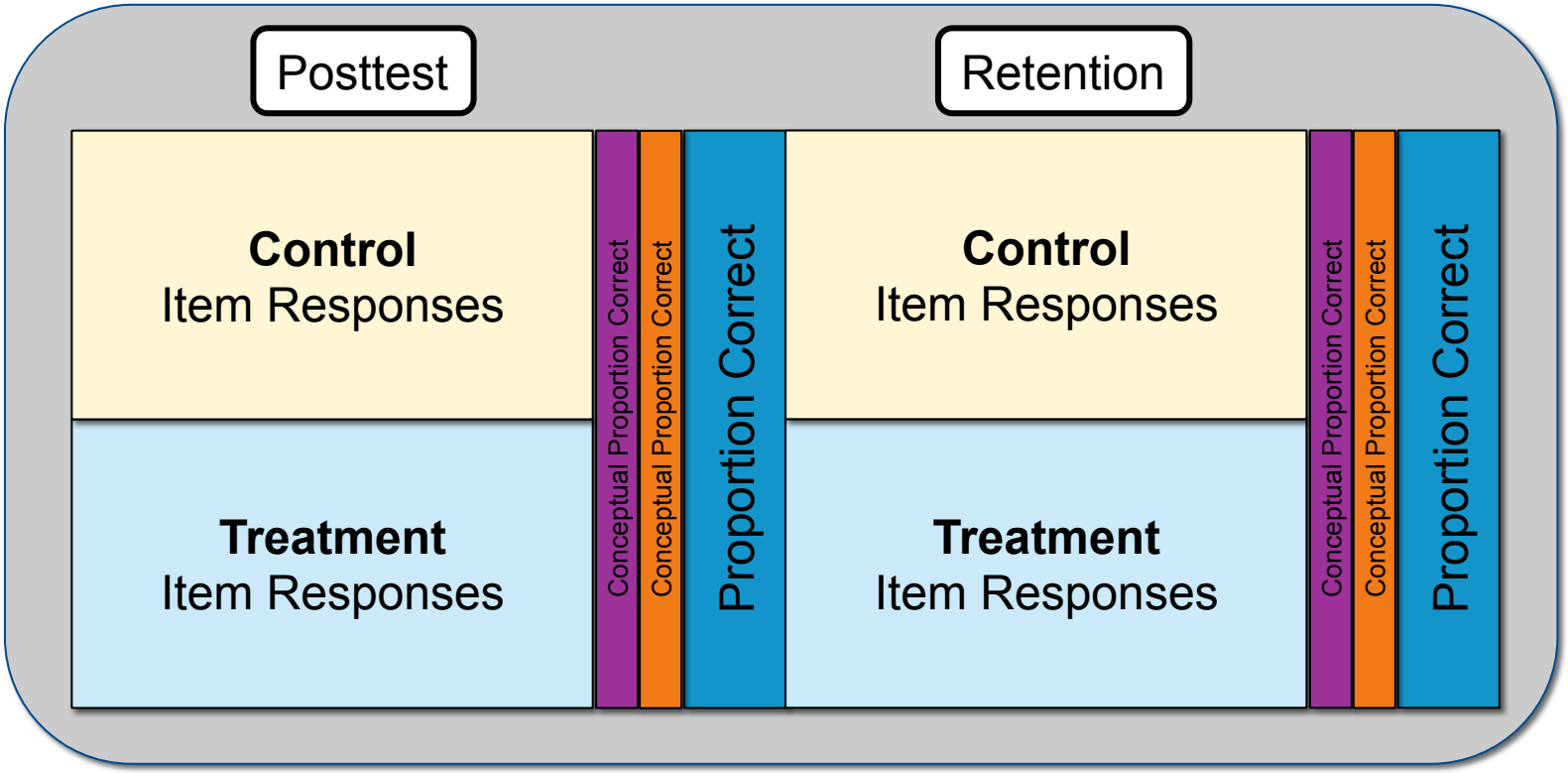


- Binary Scores
  - Proportion Correct
- Assessment Subcomponents
  - Multidimensionality

Data Structure for Analysis



Data Structure for Analysis



Data Structure for Analysis

Condition	Subject	Conceptual Item Responses			Procedural Item Responses			Concept. P. Correct	Proced. P. Correct	**Total Proportion Correct**
Control	1	1	0	0	0	1	0	.33	.33	**.33**
Control	2	1	1	0	1	1	1	.66	1.0	**.83**
Control	3	0	1	0	1	1	0	.33	.66	**.50**
Control	4	1	1	0	1	0	0	.66	.33	**.50**
Treatment	5	1	1	0	1	0	1	.66	.66	**.66**
Treatment	6	1	1	1	1	1	1	1.0	1.0	**1.0**
Treatment	7	1	0	0	1	1	0	.33	.66	**.50**
Treatment	8	0	1	1	1	1	0	.66	.66	**.66**

Binary Responses to Items

ANOVA Dependent Variables

Issues with RM ANOVA Analysis

1. Samples used in intervention research often violate RM ANOVA assumptions

2. Proportion correct scale not an interval scale, and does not allow for meaningful comparisons

Issues with RM ANOVA - Assumptions

1. Samples used in intervention research often violate RM ANOVA assumptions

RM ANOVA Assumptions:

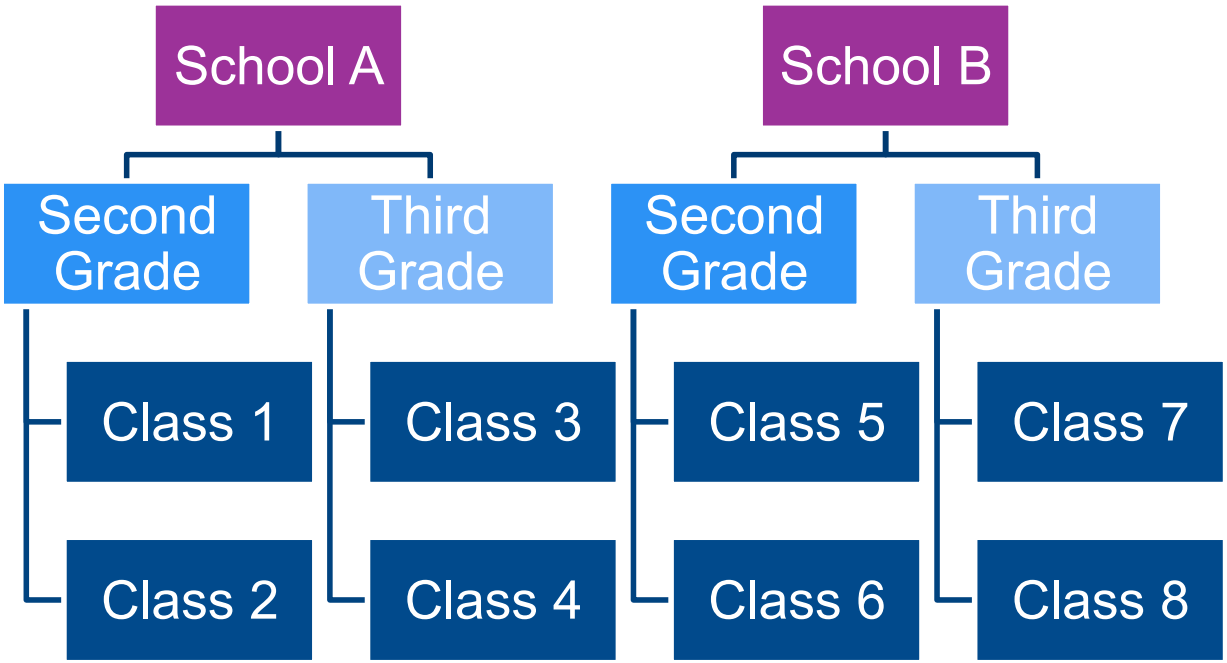
A. Independence
B. Normality
C. Equality of Variances

Although relatively robust, when violated, conclusions can be biased (Dixon, 2008; Embretson, 1991; Jaeger, 2008)

Issues with RM ANOVA - Assumptions

A. Independence

- Violated by hierarchical and nested structure of educational settings

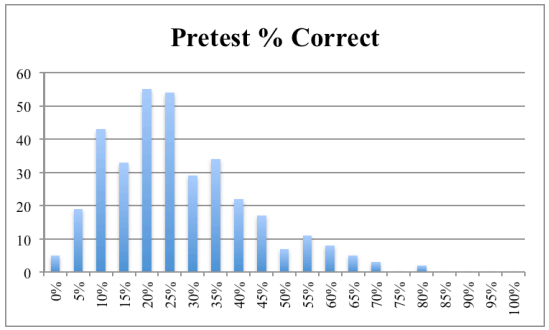


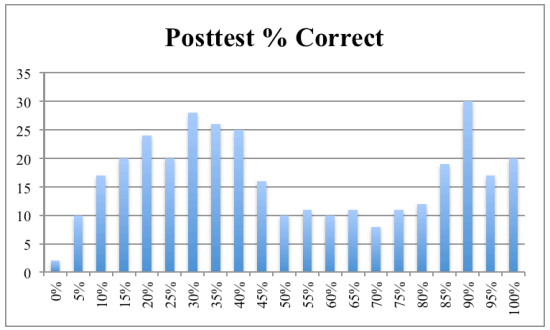
Issues with RM ANOVA - Assumptions

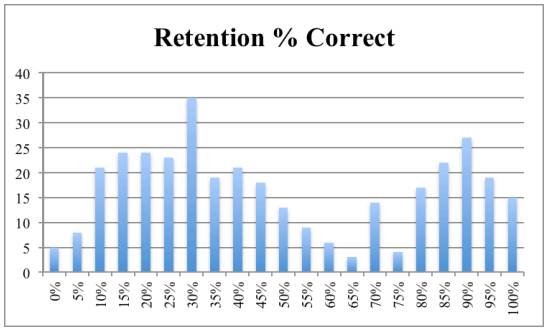
B. Normality

- Between: Conditions
- Within: Timepoints

- Distributions often not normal
  - Bimodal distributions due to differential intervention effects
- When group sizes are unequal, results biased (Wilcox, 2005)





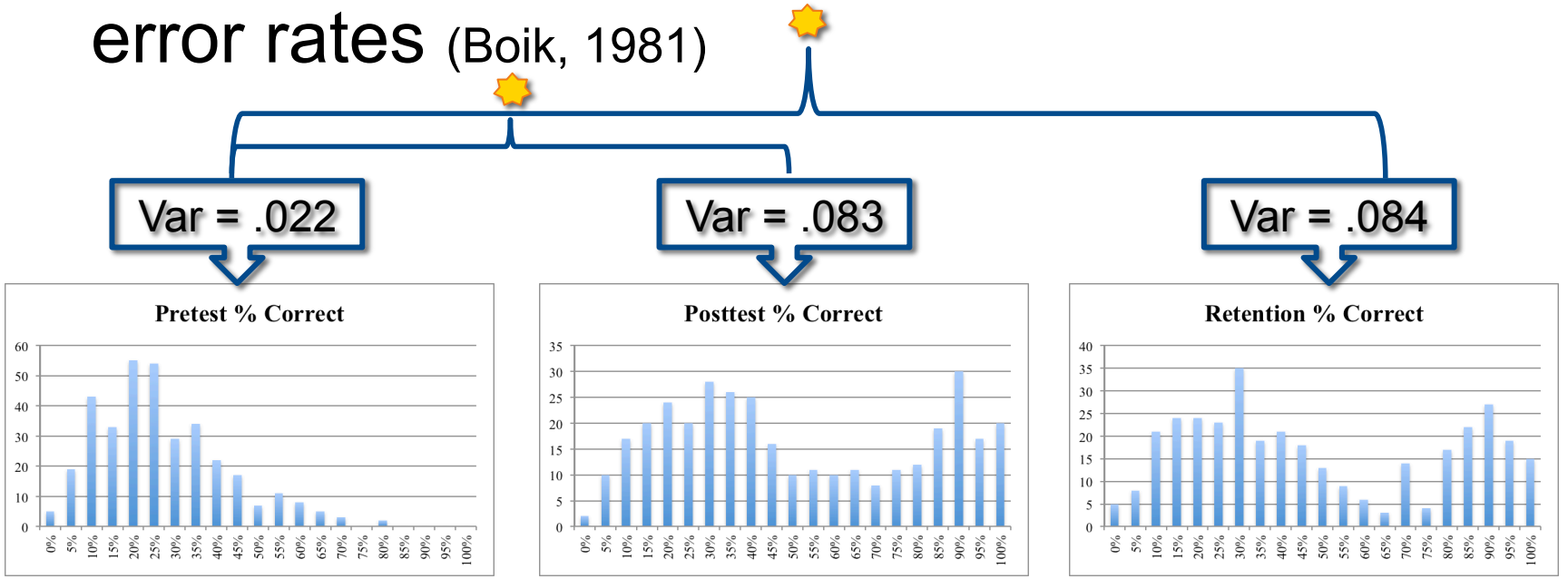


Issues with RM ANOVA - Assumptions

C. Equal Group Variances Between

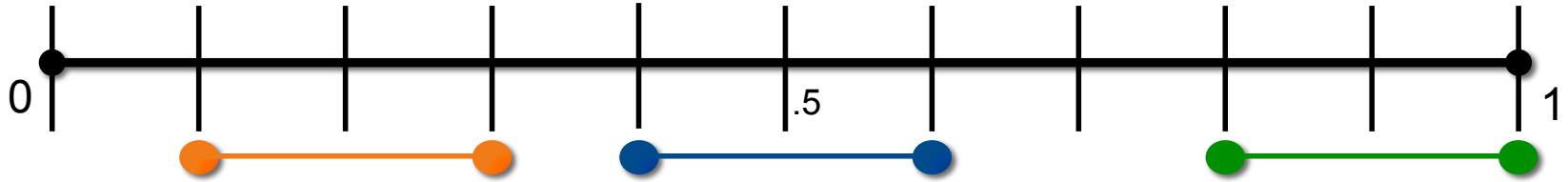
- Between: Conditions
- Within: Timepoints

- Small violations can lead to inflated Type I error rates (Boik, 1981)

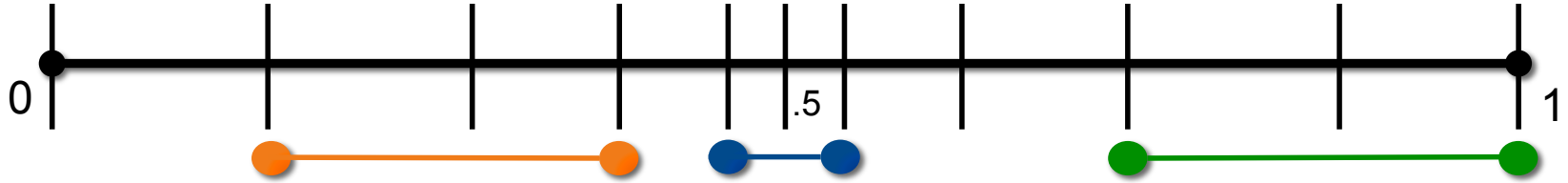


Issues with RM ANOVA – Interval Scale

~~Interval Scale~~



Proportion Correct Scale



- Proportion correct is not empirically an interval scale
- Meaningful comparisons can only be made from the same initial value
- Can lead to biased results (Agresti, 2002; Dixon, 2008)

An Alternative: IRT Models

- Item Response Theory Models
  - Overcomes limitations of RM ANOVA models
    - No assumptions of independence, normality, or equal variance
      - However, does have assumptions of dimensionality and local independence
    - Interval scale, allowing for meaningful comparisons
  - Has additional benefits
    - Incorporates more informative metrics
    - Accounts for measurement error

Advances in IRT Models

- IRT Models in the past have been used for:
  - Individual Differences
  - Large-scales tests

- Recent Advances:
  - Individual & Group Differences
  - Smaller sample sizes typical of education research
    - E.g. 100 subjects and 20 assessment items
    - New estimation methods
      (Random item approach, Cho & Rabe-Hesketh, 2011, 2012)
  - Simultaneously Handle:
    - Longitudinal Designs
    - Multidimensional Constructs

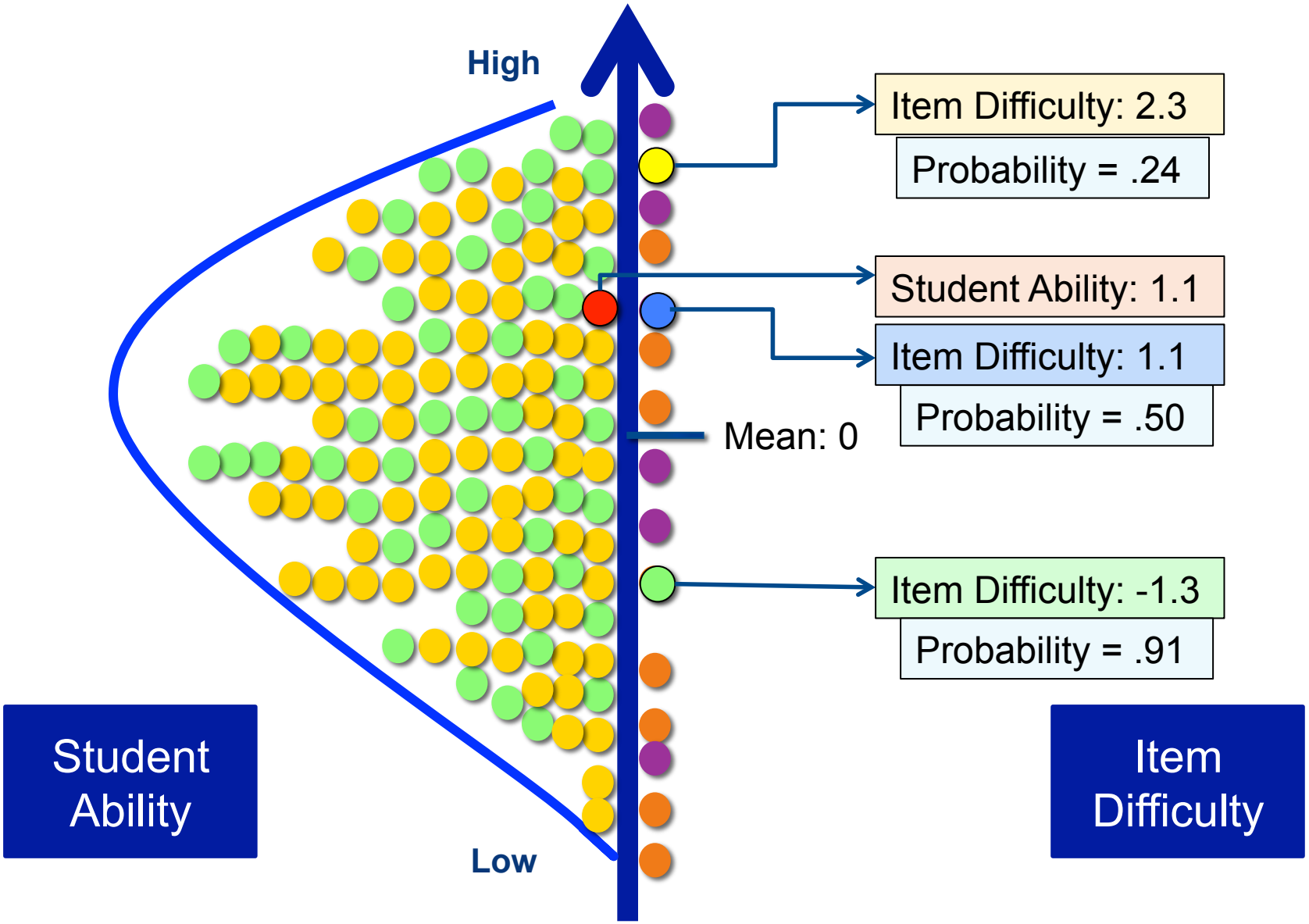
Data Structure for Analysis

Condition	Subject	Conceptual Item Responses			Procedural Item Responses			Concept. P. Correct	Proced. P. Correct	**Total Proportion Correct**
Control	1	1	0	0	0	1	0	.33	.33	**.33**
Control	2	1	1	0	1	1	1	.66	1.0	**.83**
Control	3	0	1	0	1	1	0	.33	.66	**.50**
Control	4	1	1	0	1	0	0	.66	.33	**.50**
Treatment	5	1	1	0	1	0	1	.66	.66	**.66**
Treatment	6	1	1	1	1	1	1	1.0	1.0	**1.0**
Treatment	7	1	0	0	1	1	0	.33	.66	**.50**
Treatment	8	0	1	1	1	1	0	.66	.66	**.66**

IRT Dependent Variables

ANOVA Dependent Variables

IRT Model

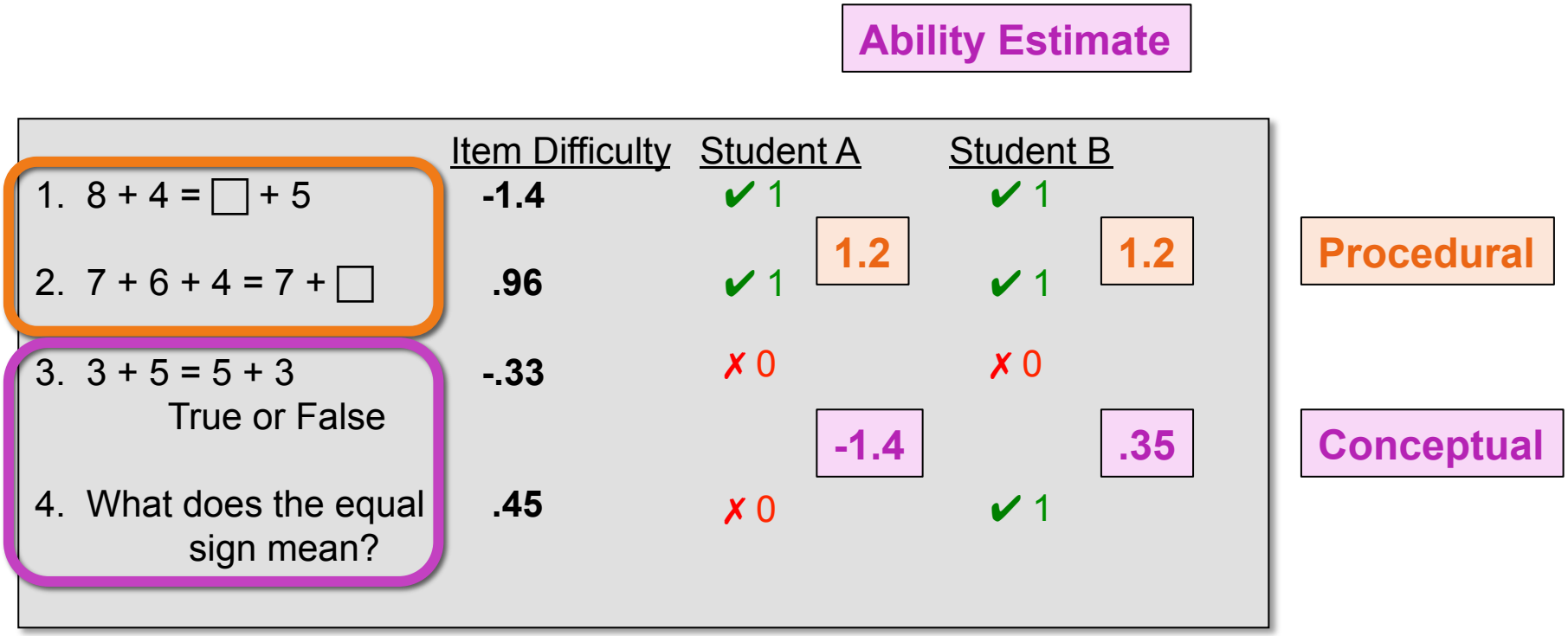

High
Low
Item Difficulty: 2.3
Probability = .24
Student Ability: 1.1
Item Difficulty: 1.1
Probability = .50
Mean: 0
Item Difficulty: -1.3
Probability = .91
Student Ability
Item Difficulty

Additional Benefits of IRT Models

1. Less sensitive to violations of RM ANOVA assumptions
2. Interval scale allows for meaningful comparisons
3. More informative metrics of student ability and item difficulty
4. Latent variable separates true group difference from measurement error

Additional Benefits of IRT Models

3. More informative metrics of student ability and item difficulty
   - Ability estimates for each subscale
   - Correlation structure between subscales



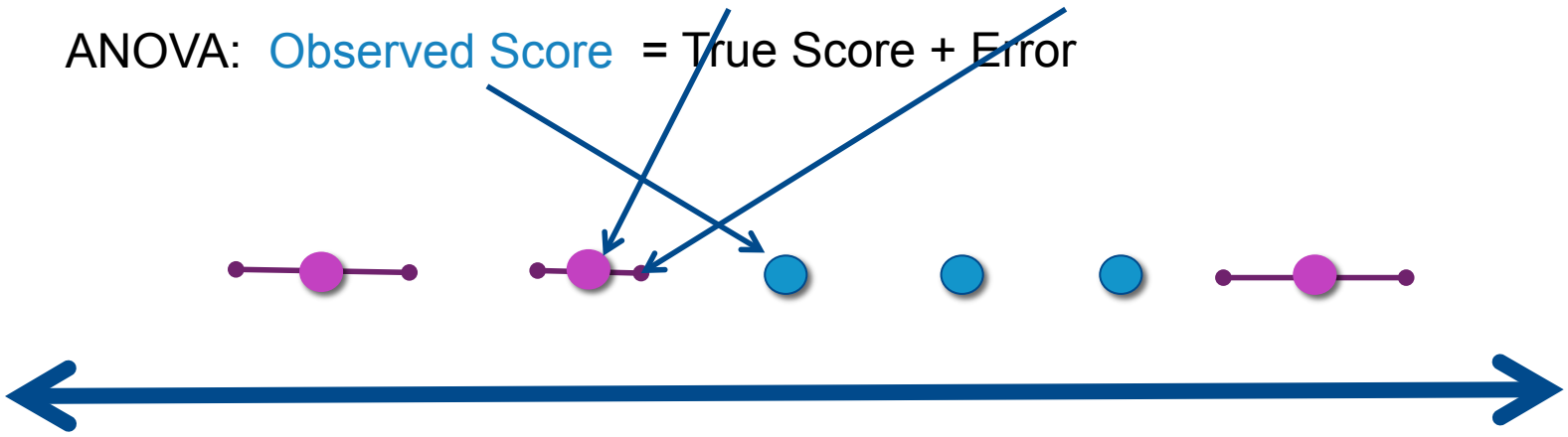
Additional Benefits of IRT Models

4. Latent variable separates true group difference from measurement error

Latent Variable: not directly observed, but inferred from other variables that are observed

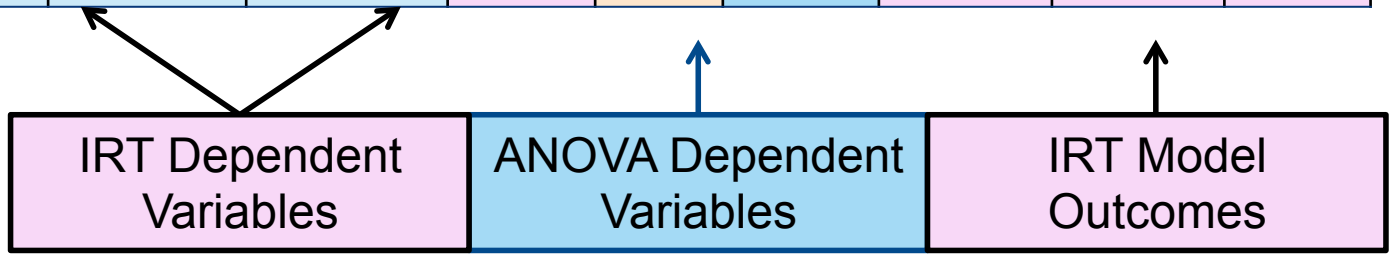
IRT: Latent Ability = Latent Ability Estimate + Standard Error

ANOVA: Observed Score = True Score + Error

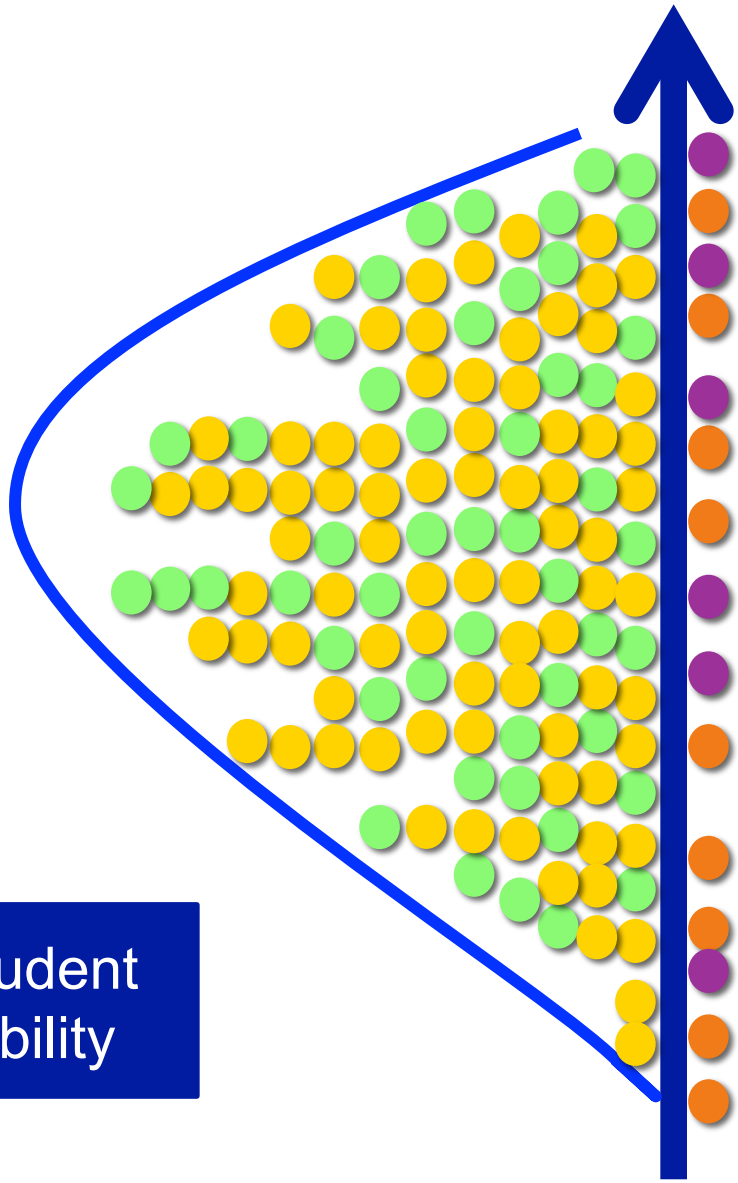


Data Structure for Analysis

Condition	Subject	Conceptual Item Responses	Procedural Item Responses	Concept. P. Correct	Proced. P. Correct	Total Proportion Correct	Concept Ability Est (se)	Proced. Ability Est (se)	Total Ability Est (se)
Control	1	1 0 0	0 1 0	.33	.33	**.33**	**-.42** (.12)	**-1.3** (.11)	**-.86** (.14)
Control	2	1 1 0	1 1 1	.66	1.0	**.83**	**.15** (.13)	**1.1** (.08)	**.63** (.11)
Control	3	0 1 0	1 1 0	.33	.66	**.50**	**-.42** (.09)	**.58** (.10)	**.08** (.09)
Control	4	1 1 0	1 0 0	.66	.33	**.50**	**.15** (.08)	**-1.3** (.13)	**-.58** (.12)
Treatment	5	1 1 0	1 0 1	.66	.66	**.66**	**.15** (.08)	**.58** (.06)	**.37** (.11)
Treatment	6	1 1 1	1 1 1	1.0	1.0	**1.0**	**1.3** (.11)	**1.1** (.08)	**1.2** (.18)
Treatment	7	1 0 0	1 1 0	.33	.66	**.50**	**-.42** (.13)	**.58** (.07)	**.08** (.07)
Treatment	8	0 1 1	1 1 0	.66	.66	**.66**	**.15** (.07)	**.58** (.07)	**.37** (.10)



IRT Model

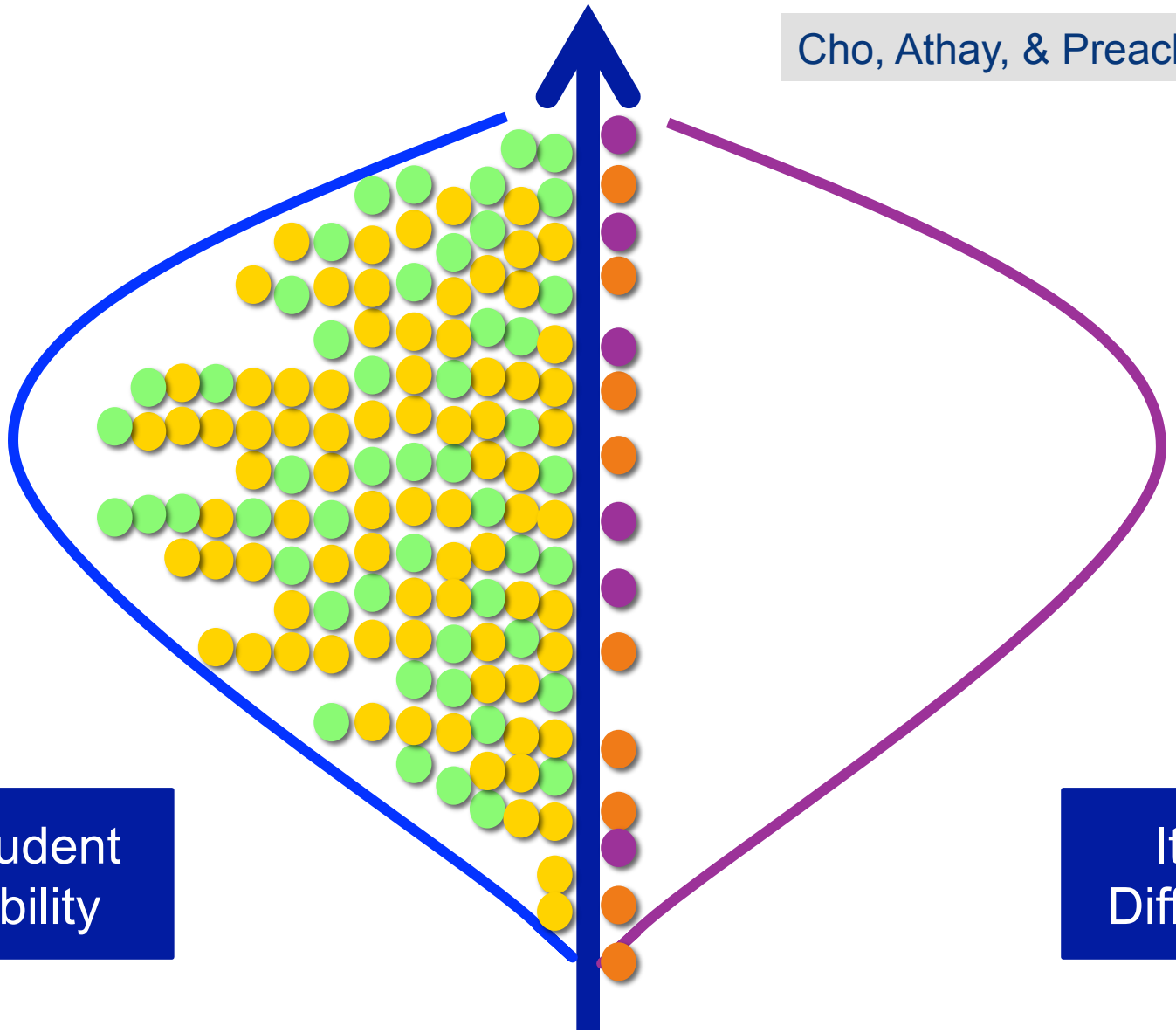


Student Ability

Item Difficulty

GEL MIRT Alternative to RM ANOVA

Cho, Athay, & Preacher (2012)

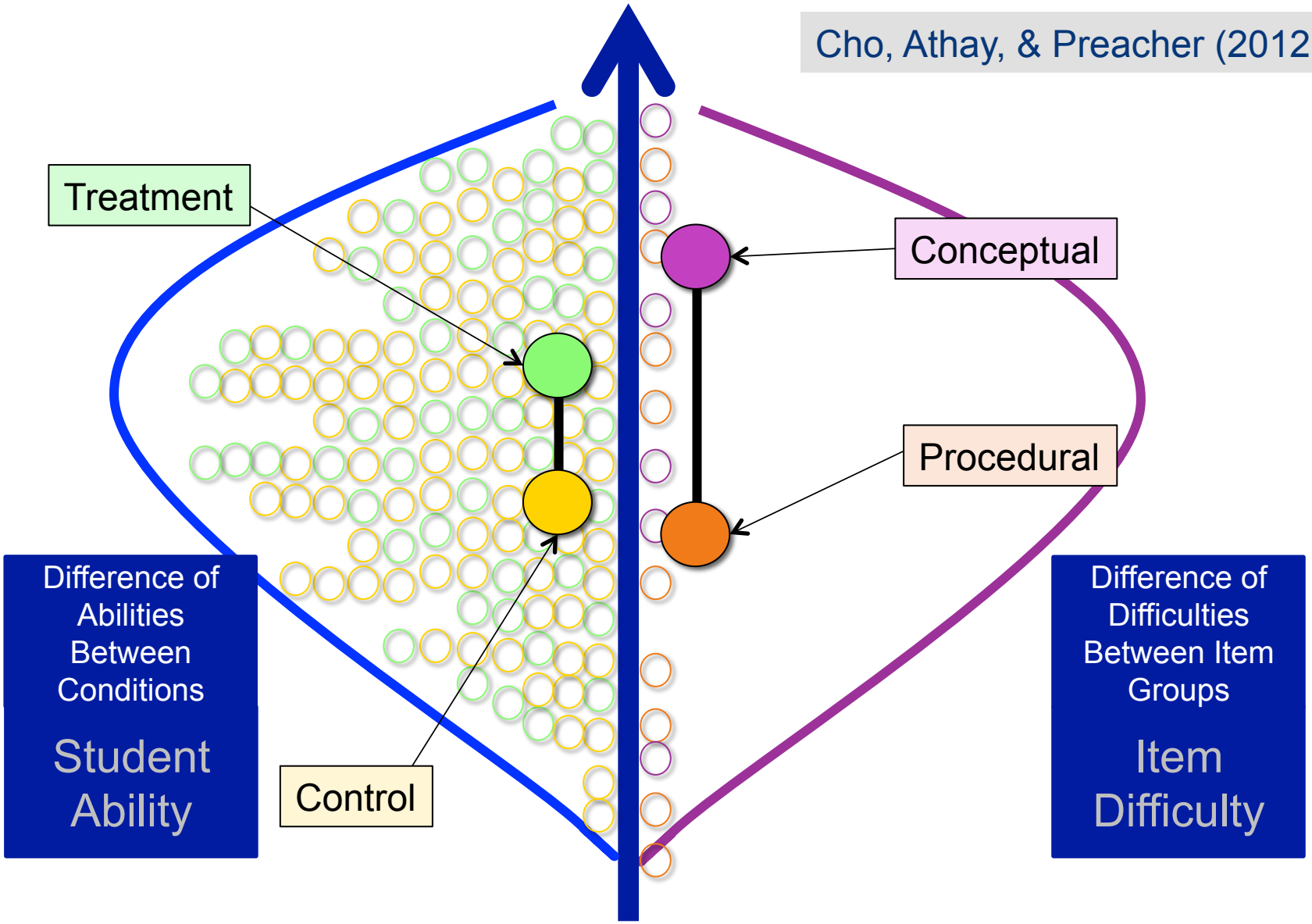


Student Ability

Item Difficulty

GEL MIRT Alternative to RM ANOVA

Cho, Athay, & Preacher (2012)



The Dataset

156 2nd & 3rd Grade Students

~ 20 minutes

Control N = 79	Conceptual Instruction	Problem Solving with accuracy feedback
Treatment N = 77	Problem Solving with accuracy feedback	Conceptual Instruction

Problem Solving: Math Equivalence

2 + 5 + 8 = ___ + 8

3 + 4 + 6 = ___ + 4

Immediate Posttest

2 Week Retention

Matthews, Rittle-Johnson, McEldoon & Taylor, 2012; Rittle-Johnson, Matthews, Taylor & McEldoon, 2011

Conceptual

3 + 5 = 5 + 3 T/F

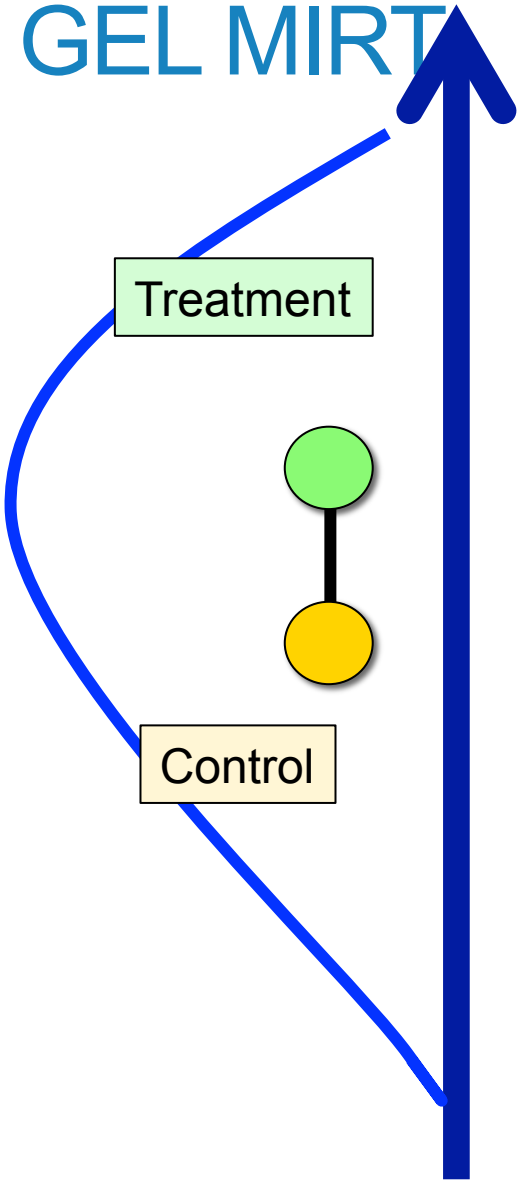
What does the equal sign mean?

Procedural

8 + 4 = □ + 5

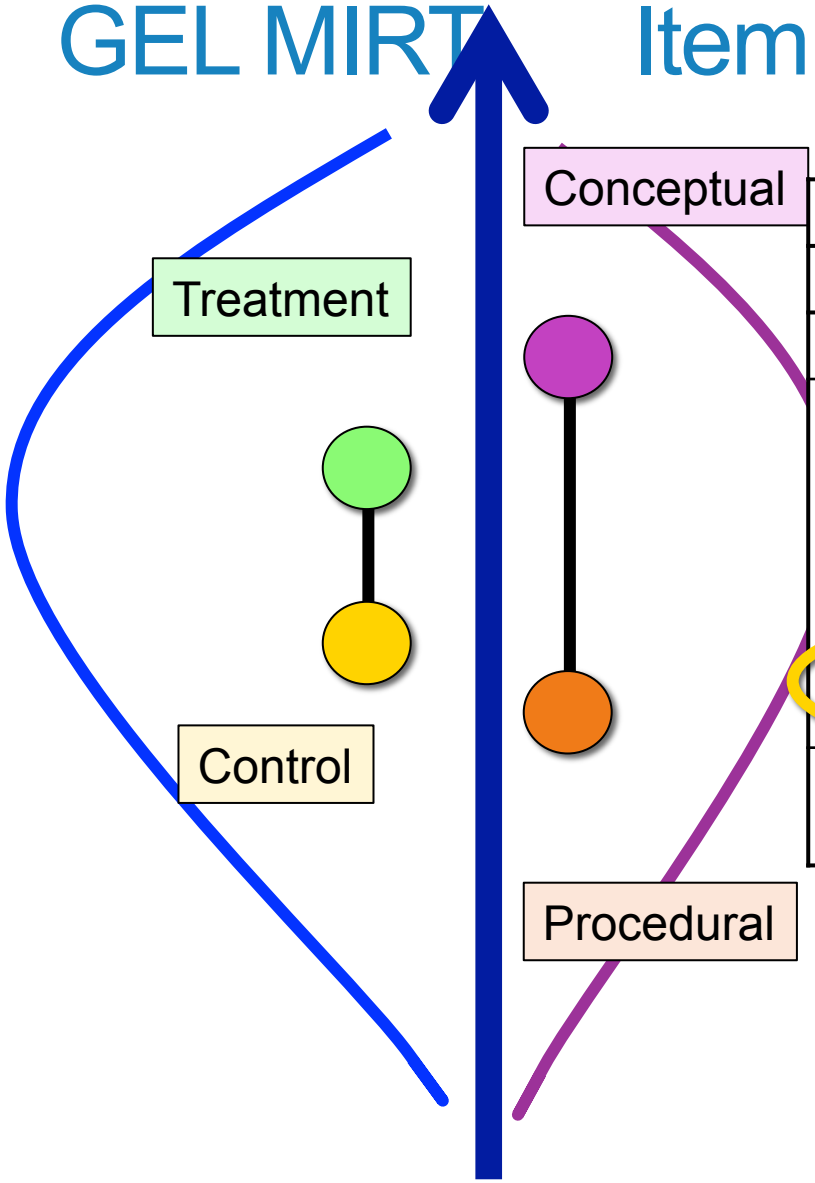
7 + 6 + 4 = 7 + □

Comparing Output – RM ANOVA and GEL MIRT



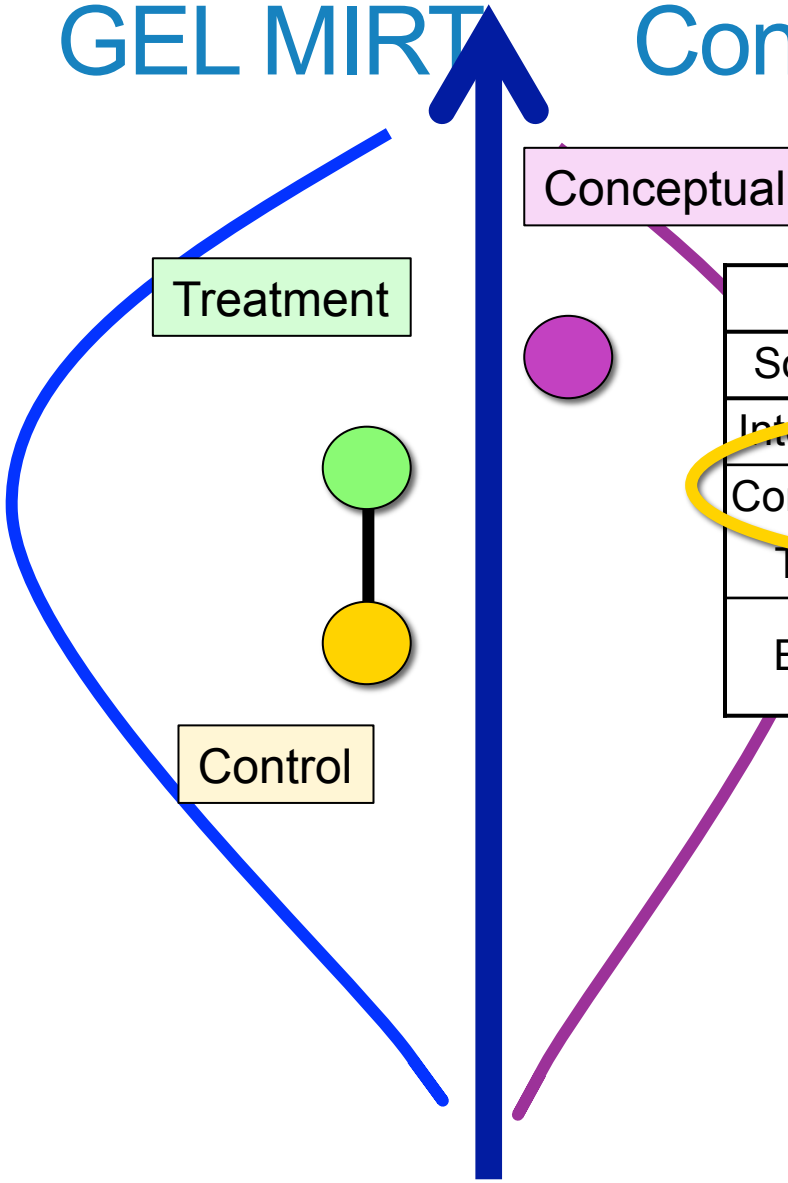
	RM ANOVA		GEL MIRT		
Source	F	Sig.(p)	Estimate	Z Score	Sig. (p)
Intercept	1298.2	<.01	1.14	3.18	<.01
Condition	0.592	0.443	0.113	0.55	0.58
Time	15.02	<.01*	0.207	2.09	.036*
Time X Condition	0.54	0.464	0.188	1.35	0.176
Error	(SS) 16.2		Individual Student Ability Estimates		

Comparing Output – RM ANOVA and GEL MIRT Item Group Differences



	RM ANOVA		GEL MIRT		
Source	F	Sig.(p)	Estimate	Z Score	Sig. (p)
Intercept	1298.2	<.01	1.826	3.87	<.01
Condition	0.592	0.443	0.113	0.54	0.584
Time	15.02	<.01*	0.208	2.1	0.036*
Time X Condition	0.54	0.464	0.188	1.35	0.176
Item Group	na	na	-1.302	-2.07	.038*
Error	(SS) 11.1		Individual Student Ability Estimates		

Comparing Output – RM ANOVA and GEL MIRT Conceptual Subscale



	RM ANOVA		GEL MIRT		
Source	F	Sig.(p)	Estimate	Z Score	Sig. (p)
Intercept	1231.9	<.01	0.488	1.28	0.2
Condition	3.169	0.077	0.32	1.49	0.137
Time	1.134	0.289	0.188	2.05	0.04*
Error	(SS) 3.123		Individual Student Ability Estimates		

Support from Simulation Studies

- Simulation studies support that IRT models are more accurate at detecting true group differences than RM ANOVA Models
- When:
  - True group differences on latent variable
  - RM ANOVA assumptions are violated
- Detection Rates of Group Differences:
  - RM ANOVA: 44%
  - GEL MIRT: 99%

Cho, Suh, & Nelson (2012)

Conclusions

- Researchers should consider the advantages of an IRT approach for evaluating intervention effectiveness
  - GEL MIRT model (Cho, Athay, & Preacher, 2012)
- Pro:
  - More informative metrics
  - Less prone to biased results
  - Can be performed using the open-source and free program R
    - Details of the model, as well as information how to run these analyses can be found in Cho, Athay, & Preacher (2012)
- Con:
  - Requires more technical proficiency on the part of the data analyst
  - Challenging to understand

Thank You

Sun-Joo Cho & Bethany Rittle-Johnson
Michael Nelson & Marci DeCaro
Children's Learning Lab

Vanderbilt Children's Learning Lab

http://peabody.vanderbilt.edu/departments/psych/research/research_labs/childrens_learning_lab/index.php

GEL MIRT Model Paper and Details

http://quantpsy.org/pubs.htm

The first author is supported by a predoctoral training grant provided by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B040110 to Vanderbilt University. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.