

ITEMS Digital Module 08: Foundations of Operational Item Analysis

This document contains all core content slides from sections 1-3 with the exception of slides that show video screens. In the digital module all slides can be accessed individually.

Module Organization

The module starts with an introductory section that leads to the main menu from which learners can select individual content and activity sections:

The image shows a digital module's main menu. At the top left are icons for Home and Navigation. The title "Main Menu" is centered above two main categories: "Theory" and "Practice".

Theory section (left, green background):

- 01 Test Development [20 Minutes]
- 02 Item Analysis with CTT [30 Minutes]
- 03 Item Analysis with IRT [30 Minutes]

Practice section (right, blue background):

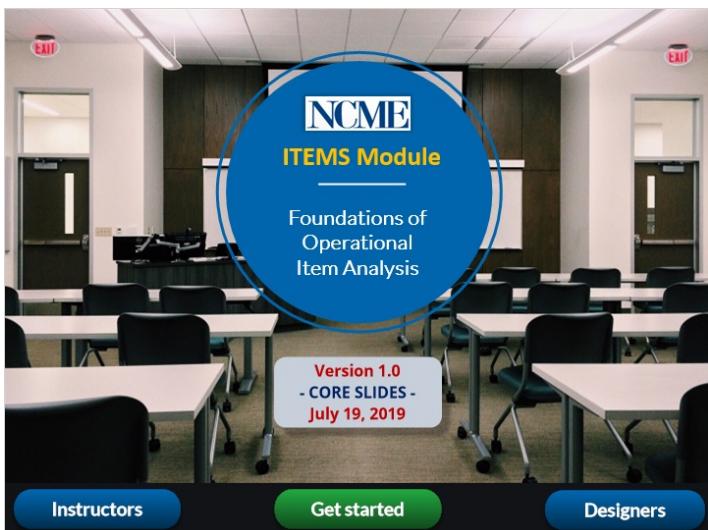
- 04 Example & Activity: Tachiny [15 Minutes] (circled with a large red "no")
- 05 Example & Activity: TAP [45 Minutes] (partially visible)
- Quizzes [15 Minutes] (partially visible)

A purple callout at the bottom states: "This slide deck contains the first three content sections only."

DM08 (Item Analysis, Version 1.0)

1. Module Overview

1.1 Module Cover (START)



1.2 Instructors

A screenshot of a slide titled "Meet the content development team:". It features two photographs of men: Hanwook [Henry] Yoo on the left and Ronald K. Hambleton on the right. Below each photo is a blue callout box with their names and affiliations: "Hanwook [Henry] Yoo Educational Testing Service (ETS)" and "Ronald K. Hambleton University of Massachusetts Amherst". At the bottom, a button reads "Click on the images to get to know them a bit!".

1.3 Designers

Meet the instructional design team:



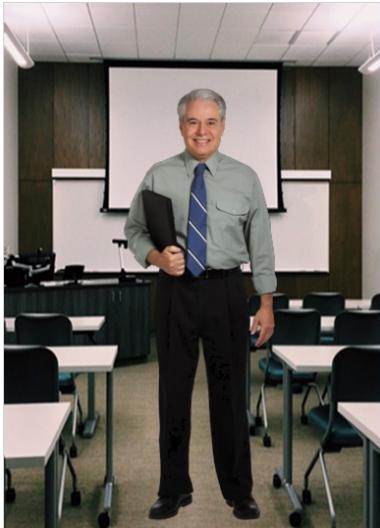
André A. Rupp
Educational
Testing Service

Xi Lu
Florida State
University

Special Thanks: ETS

NCME national council on measurement in education

1.4 Welcome

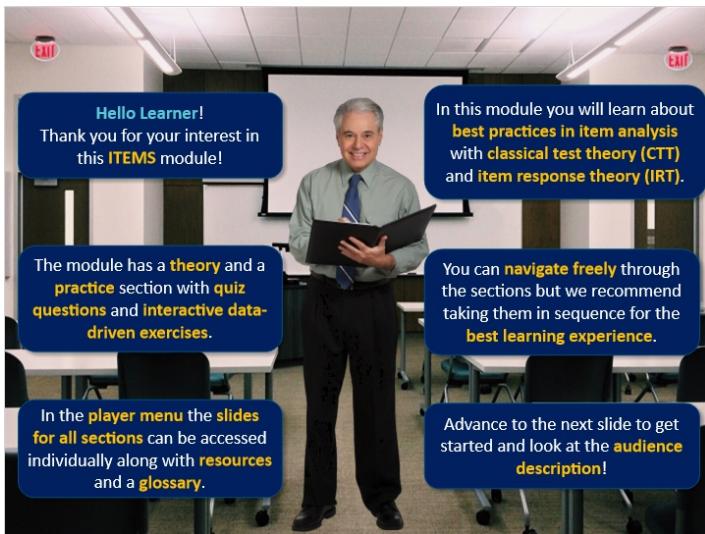


Welcome to the
ITEMS Module!

The man to the left is Jonathan!

Along with the instructors he will
be guiding you through the
module content.

1.5 Overview



1.6 Target Audience

Target Audience

Anyone who would like a gentle statistical introduction to this topic:

- graduate students and faculty in Master's, Ph.D., or certificate programs
- psychometricians and other measurement professionals
- data scientists / analysts
- research assistants or research scientists
- technical project directors
- assessment developers

However, we hope that you find the information in this module useful no matter what your official title or role in an organization is!



1.7 Expectations (I)

Let's discuss expectations....

1.8 Expectations (II)

ITEMS Modules in Context

1.9 Learning Objective

Learning Objectives



The image shows several wooden blocks with letters on them, spelling out the word "PLAN". The blocks are arranged on a surface with some red stars in the background.

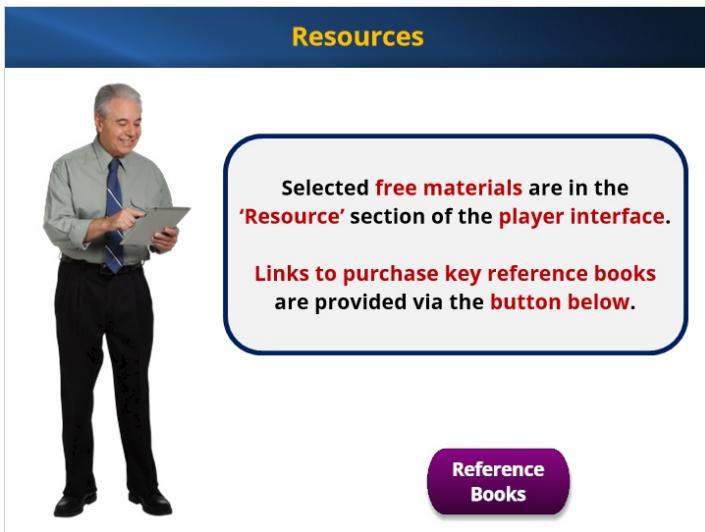
1. Describe key phases in test development and essential roles of item analysis	2. Understand issues associated with item analysis (e.g., sample size, missing data)
3. Interpret and use CTT item statistics in item review and test development	4. Interpret and use IRT item statistics in item review and test development
5. Identify the pros and cons of CTT- and IRT-based item statistics	6. Use software packages (TAP, R Shiny) for conducting item analysis

1.10 Prerequisites

Prerequisites

- Working knowledge of foundational measurement concepts:
 - ✓ Construct definitions / latent variables
 - ✓ Assessment formats
 - ✓ Item / task types
 - ✓ Scales and scale scores
 - ✓ Basic aspects of assessment development
- Working knowledge of foundational statistical concepts:
 - ✓ Descriptive statistics for distributions
 - ✓ Working with two-dimensional graphs
 - ✓ Making quantitative judgments with statistics

1.11 Resources



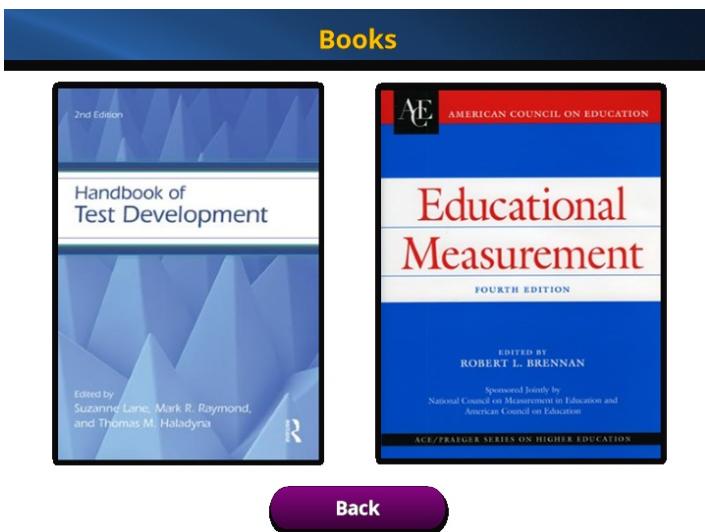
Resources

Selected **free materials** are in the '**Resource**' section of the **player interface**.

Links to purchase key reference books are provided via the **button below**.

Reference Books

Books (Slide Layer)



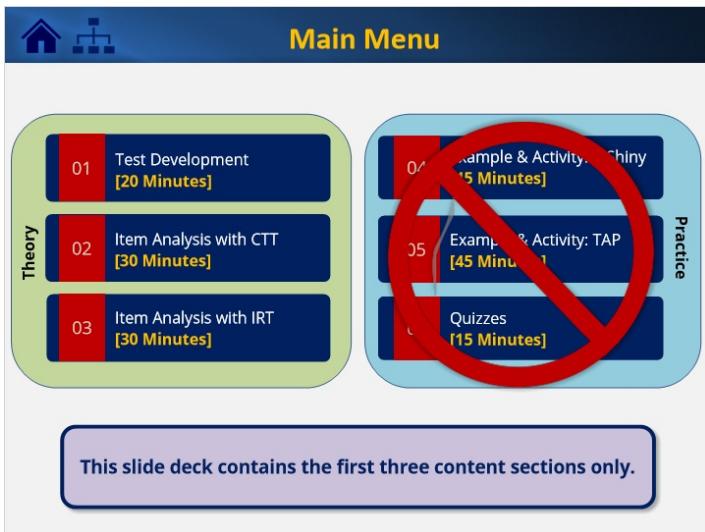
Books

Handbook of Test Development
2nd Edition
Edited by Suzanne Lane, Mark R. Raymond, and Thomas M. Haladyna

Educational Measurement
FOURTH EDITION
EDITED BY ROBERT L. BRENNAN
Sponsored Jointly by National Council on Measurement in Education and American Council on Education
ACR/PRAGER SERIES ON HIGHER EDUCATION

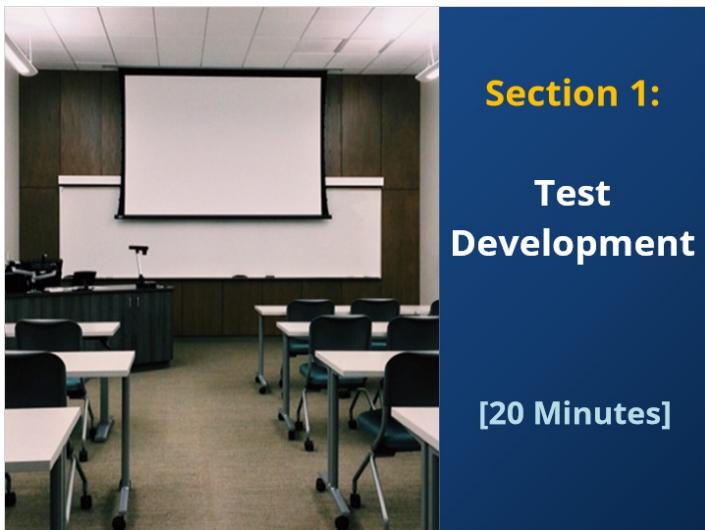
Back

1.12 Main Menu



2. Section 1: Test Development

2.1 Cover: Section 1



2.2 Objectives: Section 1

Learning Objectives



1. Describe the key steps in test construction, especially during pre-testing
2. Provide a definition and description of the main purposes of item analysis
3. Describe the key differences between CTT and IRT as they relate to item analysis
4. Describe the influence of sample size, missing data, and other key topics on item analysis

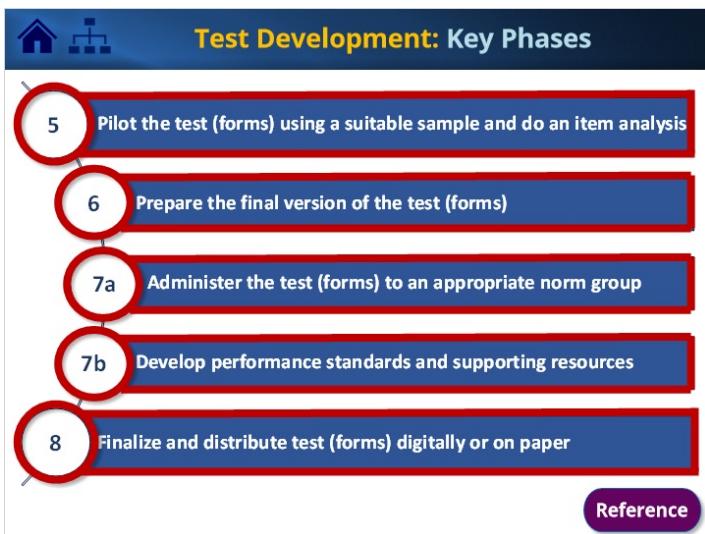
2.3 Test Development Phases (I)

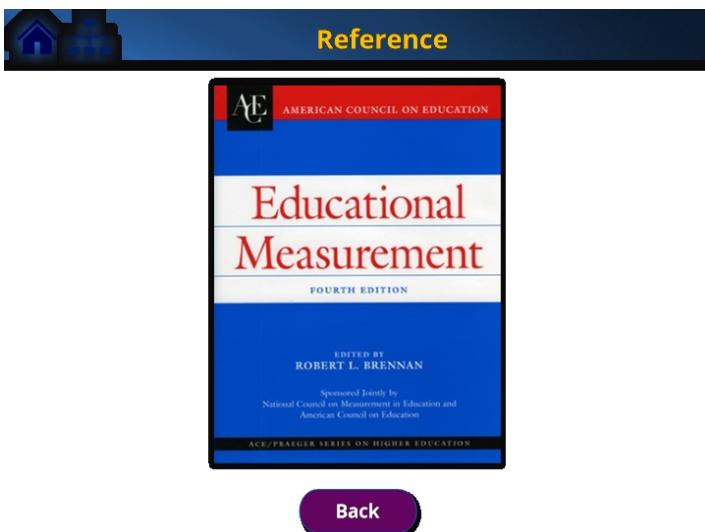
Test Development: Key Phases

- 1 Develop test specifications / blueprints
- 2 Draft and review initial versions of test items
- 3 Pre-test / field-test items and do an item analysis
- 4 Create one or multiple test forms

2.4 Test Development Phases (II)



Reference (Slide Layer)



2.5 Pre-testing: Design Options



Pre-testing: Design Options

- Design 1:** Embed pretest items in operational forms
- Design 2:** Create a separate test section
- Design 3:** Create a separate administration



2.6 Pre-testing: Considerations

Pre-testing: Considerations

- Calibration sample should be representative of the population of interest
- Pretest items should not influence examinee performance on the operational items
- Item context should be controlled
- Motivation for examinees should be high
- Pretest exposure should be minimized

2.7 Item Analysis: Definition

The slide has a dark blue header bar with a house icon and a tree icon on the left, and the title 'Item Analysis: Definition' in yellow text on the right.

"Item analysis is a term broadly used to define the computation and examination of any statistical property of examinees' responses to an individual test item." (Crocker & Algina, 1986, p. 311)

"Item analysis procedures refer to a set of statistical measures used by testing experts to review and revise items, to estimate the characteristics of potential test forms, and to make judgments about the quality of items and assembled test forms." (Moses, 2017, p. 19)

2.8 Item Analysis: Purposes

The slide has a dark blue header bar with a house icon and a tree icon on the left, and the title 'Item Analysis: Purposes' in yellow text on the right.

- **Estimate key item characteristics**
 - ✓ Create item pools with known properties
 - ✓ Assemble new test forms
 - ✓ Select anchor items for equating across forms
- **Detect flawed items**

(e.g., poor distractors, too easy / too difficult, poor discrimination)
- **Identify items with differential functioning**

(e.g., potential bias against gender, language, or ethnicity)

2.9 Analysis Framework

 **Analysis Framework**



Option 1
Classical Test Theory
(CTT)

Option 2
Item Response Theory
(IRT)

2.10 Advantages of CTT

 **Advantages of CTT**

- Basic principles are **easy** to communicate to stakeholders
- Smaller sample sizes are **sufficient** for descriptive analyses
- Estimation of statistics is **conceptually straightforward**
- Reliability information is **easy** to compute
- Distractor analyses are **relatively straightforward**



2.11 Item Analysis: CTT

Item Analysis: CTT Details

- **Item analysis consists of:**
 - ✓ Determining sample-specific item parameters
 - ✓ Deleting items based on statistical criteria
- **Item analysis involves evaluating:**
 - ✓ Item difficulty indices
 - ✓ Item discrimination indices
 - ✓ Answer choices (for selected-response items)
- **Problematic items show:**
 - ✓ Item difficulty values that are very high or very low
 - ✓ Item discrimination values that are very low
 - ✓ Item choices / distractors that are selected rarely or oddly

2.12 Advantages of IRT

Advantages of IRT

- Models are expressed at the **item-level** rather than at the **test-level**
- Item statistics are **not sample dependent**
- Examinee proficiency scores are **not item dependent**
- Examinees and items can be placed on a **common reporting scale**
- Parameter estimates can be used for **item banking** and **adaptive testing**
- Models can be easily extended to **multiple dimensions, levels, and so on**



2.13 Item Analysis: IRT

- **Item analysis consists of:**
 - ✓ Determining sample-invariant item parameters
 - ✓ Utilizing goodness-of-fit criteria to detect misfit item
- **Item analysis involves evaluating:**
 - ✓ Item parameter estimates and associated errors
 - ✓ Shape of item characteristic curves
 - ✓ Shape of item information functions
- **Problematic items show:**
 - ✓ Poor model fit via a statistical test or residual analysis
 - ✓ Item difficulty parameters that are very high or very low
 - ✓ Item discrimination parameters that are very low
 - ✓ Guessing parameters that are very high

2.14 Parameter Relationships: Discrimination

$$a_i \cong \frac{r_{bis}}{\sqrt{1-(r_{bis})^2}}$$
$$r_{bis} \cong \frac{a_i}{\sqrt{1+(a_i)^2}}$$

- a_i is **discrimination parameter** of item response model
- r_{bis} is the **biserial correlation**

Example:
if the **biserial correlation is .37**, then the approximate estimate of the **discrimination parameter is .40** (and vice versa)

Technical
Caveats

Excel
Computation

Caveats (Slide Layer)

Parameter Relationships: Difficulty

Relationship between CTT and IRT estimates is valid only when examinee proficiency is normally distributed (Mean=0 and SD=1) and there is no guessing on item responses

The relationship is approximate rather than accurate as a consequence of the different distributions and assigned scores of the two models:

- ✓ The number-correct score of CTT and the ability score of IRT have distributions with different shapes
- ✓ The mathematical relationship between number-correct score and ability score is nonlinear
- ✓ The number-correct score is subject to errors of measurement whereas the ability score is not

Back

Excel (Slide Layer)

Approximate Parameter Relationships

	A	B	C	D	E
1					
2		Model	Statistic	Value	Equation
3		IRT	Discrimination parameter	0.40	= (D5)/SQRT(1-(D5*D5))
4			Difficulty parameter	0.48	=D7/D5
5		CTT	Biserial correlation	0.37	=D7/D4
6			p-value	0.43	
7			Normal deviate of p-value 0.43	0.18	=-(NORM.S.INV(D6))
8					

Back

2.15 Parameter Relationships: Difficulty

Parameter Relationships: Difficulty

$b_i \cong \frac{\gamma_i}{r_{bis}}$

$r_{bis} \cong \frac{\gamma_i}{b_i}$

• b_i is the **difficulty parameter** of item response model
 • γ_i is the **normal curve deviate** corresponding to the **p-value**
 • r_{bis} is the **biserial correlation**

Example:

If the **biserial correlation is .37** and the **p-value is .43** then the approximate estimate of the **difficulty parameter is .48** (and vice versa)

[Excel Computation](#)

Excel: parameter relationship between CTT & IRT (Slide Layer)

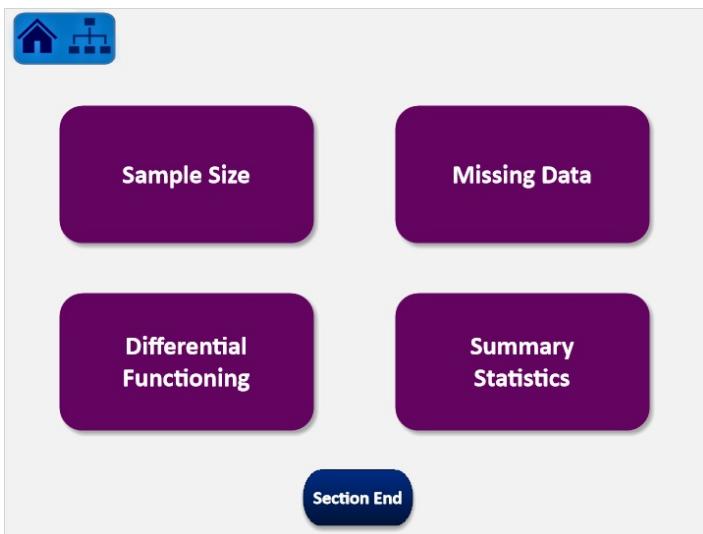
 

Approximate Parameter Relationships

A	B	C	D	E
1				
2	Model	Statistic	Value	Equation
3	IRT	Discrimination parameter	0.40	= (D5)/SQRT(1-(D5*D5))
4		Difficulty parameter	0.48	=D7/D5
5	CTT	Biserial correlation	0.37	=D7/D4
6		p -value	0.43	
7		Normal deviate of p-value 0.43	0.18	=-(NORM.S.INV(D6))
8				

[Back](#)

2.16 Topic Selection



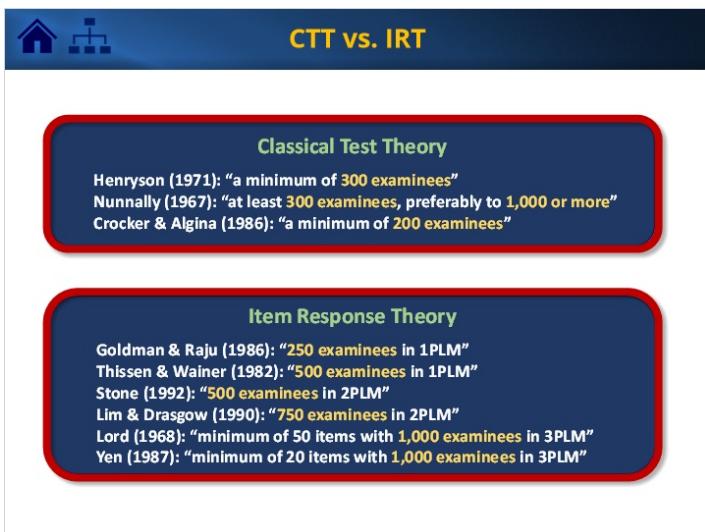
2.17 Bookmark: Sample Size



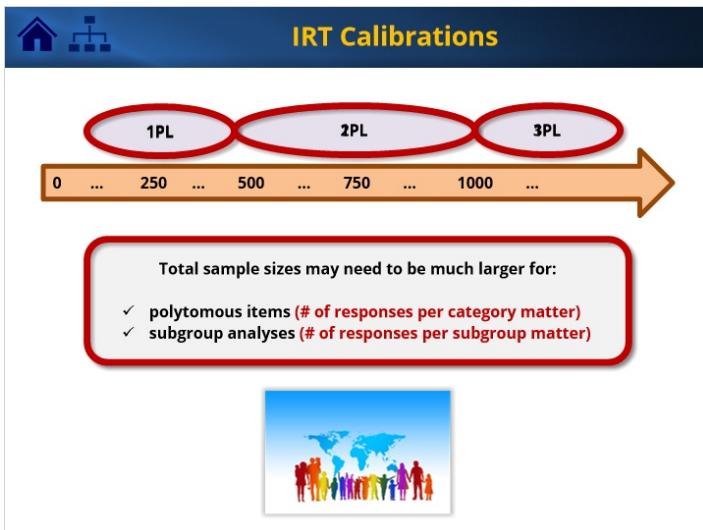
2.18 Sample Size (I)



2.19 Sample Size (II)



2.20 Sample Size (III)



2.21 Bookend: Sample Size



2.22 Bookmark: Missing Data



2.23 Missing Data (I)

The slide has a dark blue header bar with a house icon and a tree icon on the left, and the text "General Principles" in yellow on the right. Below the header, there are three numbered items in white circles with red outlines:

- 1** **Omitted:**
the examinee saw the item but didn't respond
- 2** **Not-reached:**
the examinee didn't see the item at the end of the test and didn't respond
- 3** **Not-presented (or not-administered):**
the examinee didn't see the item by design

2.24 Missing Data (II)



Statistical Effects

A	B	Treated as			F
		Missing	Incorrect	Random	
1	1	1	1	1	
2	0	0	0	0	
3	1	1	1	1	
4	1		0	0	
5	1		0	0	
6	1		1	1	
7	1		0	0	
8	1	1	1	1	
9	0	0	0	0	
10	0	0	0	0	
11	1		0	0	
12	1	1	1	1	
13	1	1	1	1	
14	0	0	0	0	
15	0	0	0	0	
16	1	1	1	1	
17	0	0	0	0	
18	1	1	1	1	
19	0		0	0	
20	1	1	1	1	
21	1		0	1	
22	0	0	0	0	
23	1	1	1	1	
24	p -value	0.60	0.56	0.45	0.50
25					
26					

2.25 Bookend: Missing Data



2.26 Bookmark: DIF



2.27 Differential Functioning (II)

Subgroups

Item analysis statistics, especially item difficulty indices, can be used to compare reference group and focal groups such as:

- 1 Gender
(Male vs. Female)
- 2 Race/Ethnicity
(White American vs. Asian American, African American, American Indian, or Hispanic/Latino American)
- 3 English Language Status
(English language learner vs. Non-English language learners; for language assessments)
- 4 Countries
(United States vs. other countries; for international assessments)

Reference

Reference (Slide Layer)

The slide is titled "Reference" at the top. Below it is a white rectangular box containing the following text:

ITEMS • Instructional Topics in Educational Measurement

An NCME Instructional Module on

Using Statistical Procedures to Identify Differentially Functioning Test Items

Brian E. Clauzer, National Board of Medical Examiners
Kathleen M. Mazor, University of Massachusetts School of Medicine

This module is intended to prepare the reader for an in-depth review of differential item functioning (DIF). It is distributed free and is intended for use by test developers, test item writers, and test item reviewers. **DEFINITION:** The Standardized procedure and rationale for identifying differential item functioning procedures and rationale are described. **DEFINITION:** The Standardized procedure and rationale for identifying differential item functioning procedures and rationale are described. The theoretical framework is presented, the relative strengths and weaknesses of various statistical procedures are discussed, and practical guidelines for the practitioner in applying statistical methods to identify DIF are provided. The reader is encouraged to apply the concepts and procedures described in this module to his/her own work, as are the policy and practice statements implemented at the test developer's organization.

NOTES: These results are routinely used as the basis for decisions regarding placement, achievement, and licensure. These decisions are often made on the basis of test scores, which must be valid for valid interpretation. One potential threat to validity is item bias.

REFERENCES:

Brian E. Clauzer & Kathleen M. Mazor, "Performance of the National Board of Medical Examiners' 2008 Market Street Examination," *Journal of Educational Measurement*, Vol. 46, No. 1, Spring 2009, pp. 1-16.

Kathleen M. Mazor is a Professor of Research and Roderick A. Smith is a Professor of Psychology at the University of Massachusetts School of Medicine 55 Lake Ave. N., Worcester, MA 01655. © 2009 National Board of Medical Examiners. All rights reserved.

Author Information:

ITEMS is a series of units designed to facilitate learning content in measurement and assessment. This module may be used in conjunction with the NCME's *Instructional Topics in Educational Measurement*. It is intended for use by test developers, test item writers, and test item reviewers. The module may also be used as a resource for test item writers and test item reviewers. The module should be submitted to Dr. Michael Eddy, Executive Director, NCME, 1801 Alexander Bell Drive, Suite 500, Reston, VA 20190-5102, USA. Tel: +1 703 243 9600; Fax: +1 703 243 9601; Email: eddy@ncme.org.

381

Back

2.28 Differential Functioning (I)

The slide is titled "General Principles" at the top. It features several sections of text in boxes and an illustration of a cartoon character holding a red book.

Detecting potentially biased test items against different subgroups is an important part of comprehensive fairness evaluations.

DIF analysis is usually conducted

a) after pre-testing or b) at the preliminary item analysis stage.

2.29 Bookend: DIF



This is the end of this part.

2.30 Bookmark: Summary Statistics



2.31 Summary Statistics (I)

CTT: Summary Statistics

$p_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$ <ul style="list-style-type: none"> ▪ p_j is the item p-value ▪ x_{ij} is the item score $i = 1, \dots, N$ [examinees] $j = 1, \dots, J$ [items] 	$X_i = \sum_{j=1}^J x_{ij}$ <ul style="list-style-type: none"> ▪ X_i is the total test score ▪ x_{ij} is the item score $i = 1, \dots, N$ [examinees] $j = 1, \dots, J$ [items] 	$\bar{X} = \sum_{j=1}^J \bar{x}_j$ <ul style="list-style-type: none"> ▪ \bar{X} is the average test score ▪ \bar{x}_j is the average item score $i = 1, \dots, N$ [examinees] $j = 1, \dots, J$ [items]
--	---	---

A screenshot of a Microsoft Excel spreadsheet titled "CTT: Summary Statistics". The spreadsheet contains three tables of data and their corresponding formulas.

Data Tables:

- Table 1 (Examinee Data):** A grid where rows represent examinees (1-10) and columns represent items (1-7). Cell C15 contains the formula: `=AVERAGE(C1:C10)`.
- Table 2 (Test Score Data):** A grid where rows represent items (1-7) and columns represent examinees (1-10). Cell K15 contains the formula: `=AVERAGE(K1:K10)`.
- Table 3 (Mean Test Score Data):** A grid where rows represent items (1-7) and columns represent examinees (1-10). Cell L15 contains the formula: `=SUM(K1:K10)/10`.

Formulas:

- Cell C15 (p-value):** `=AVERAGE(C1:C10)`
- Cell K15 (Mean):** `=AVERAGE(K1:K10)`
- Cell L15 (test score):** `=SUM(K1:K10)/10`

2.32 Summary Statistics (II)

IRT: Functional Relationships

Item Characteristic Curve (ICC) and Test Characteristic Curve (TCC)

$$TCC(\theta) = \sum_{i=1}^N P_i(\theta)$$

where $P_i(\theta)$ is the item characteristic function of item i at level θ

Item Information Function (IIF) and Test Information Function (TIF)

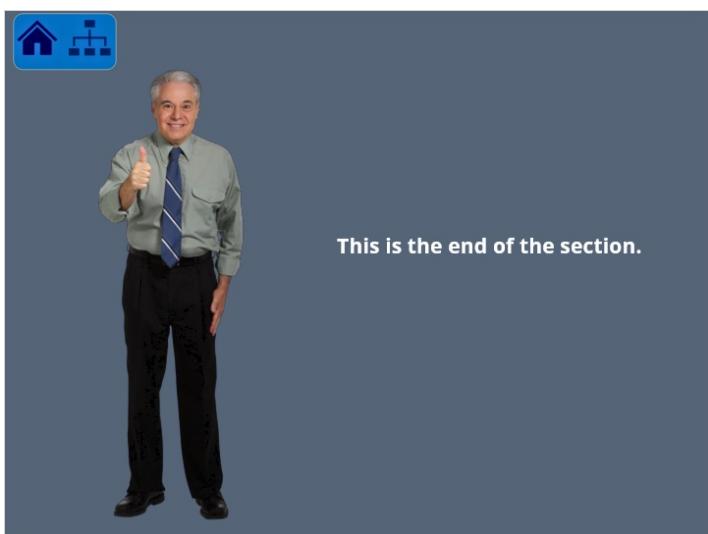
$$TIF(\theta) = \sum_{i=1}^N I_i(\theta)$$

where $I_i(\theta)$ is the item information function of item i at level θ

2.33 Bookend: Summary Statistics

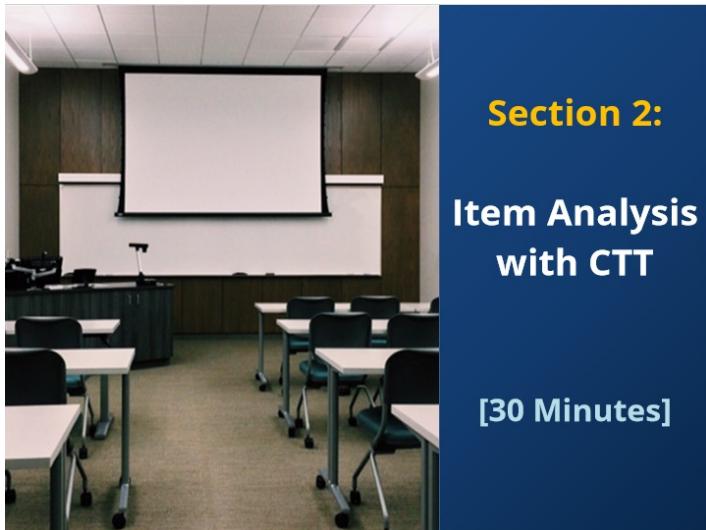


2.34 Bookend: Section 1



3. Section 2: Item Analysis with CTT

3.1 Cover: Section 2



3.2 Objectives: Section 2

A slide titled "Learning Objectives" featuring a wooden block arrangement spelling "PLAN". Below the title are four numbered objectives in boxes.

Learning Objectives



- 1.** Compute and interpret item difficulty and discrimination statistics
- 2.** Understand common flagging thresholds and their rationales
- 3.** Conduct item distractor analyses using numerical and graphical summaries
- 4.** Make recommendations about item revision and elimination

3.3 Definition of Item Analysis



“Item analysis consists of statistical analyses of the data produced when test takers respond to test items—analyses conducted for the purpose of providing information about the items, rather than the test takers.” (Livingston, 2006, p. 421)

“Item analysis procedures refer to a set of statistical measures used by testing experts to review and revise items, to estimate the characteristics of potential test forms, and to make judgments about the quality of items and assembled test forms.” (Moses, 2017, p. 19)

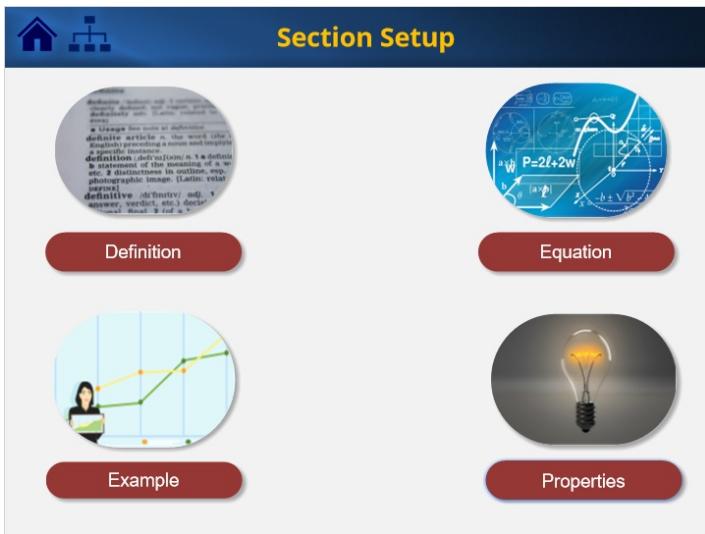
3.4 Purposes of Item Analysis



Purposes of Item Analysis

- **Describe items statistically for use in test development**
 - To assemble a new test form
 - To select anchor items for equating
 - Other
- **Detect flawed items**
 - Multiple correct answers
 - Non-functioning distractors
 - Items that are too easy or to difficult
 - Miskeyed items
- **Identify items that show differential functioning across**
 - Subgroups defined for legal reasons (e.g., gender, ethnicity)
 - Subgroups defined for research reasons (e.g., instructional context)

3.5 Section Setup



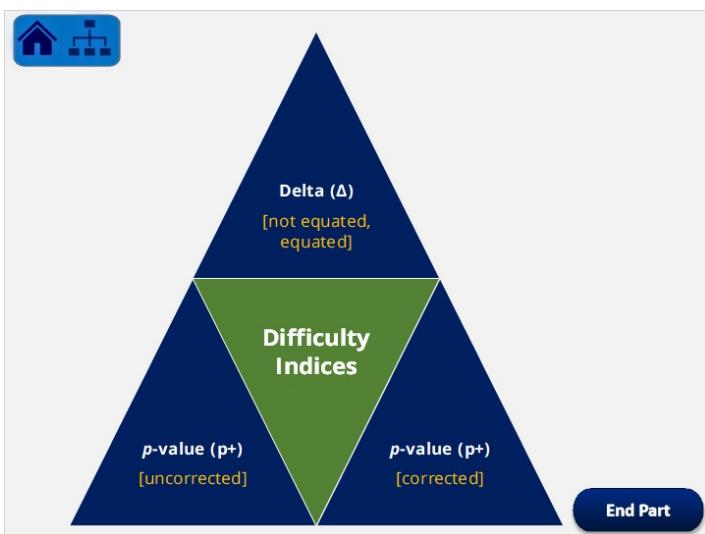
3.6 Topic Selection: Approaches



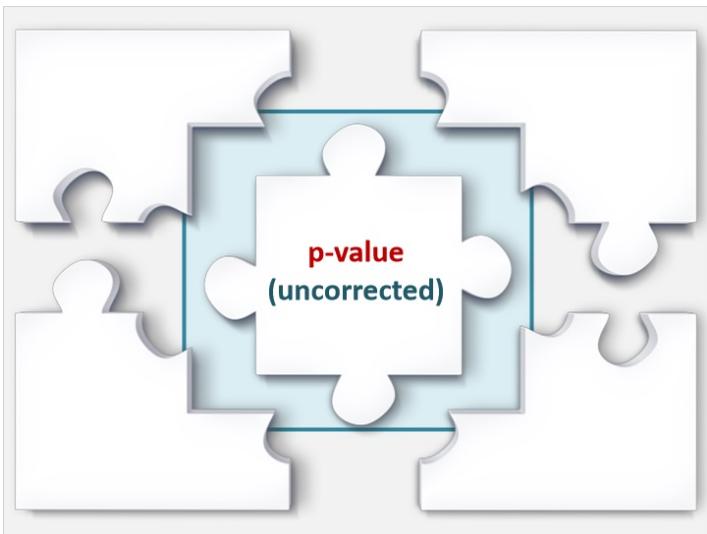
3.7 Bookmark: Item Difficulty



3.8 Topic Selection: Difficulty



3.9 Bookmark: p-value

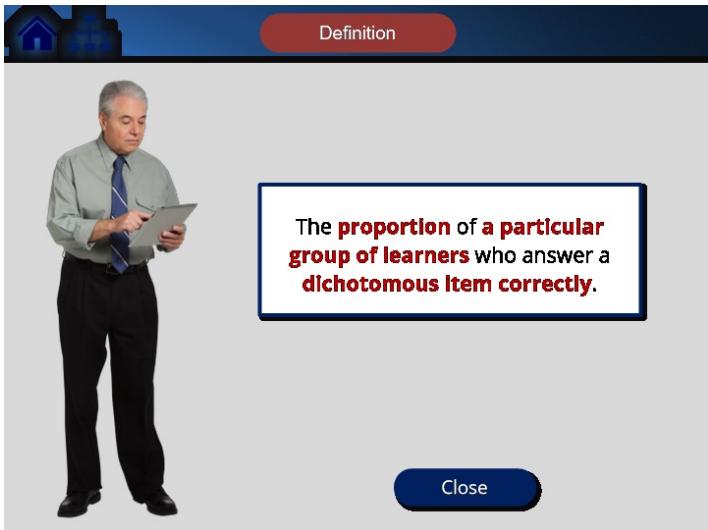


3.10 Proportion Correct / p-value (p+)

The figure consists of four circular icons arranged in a 2x2 grid, each containing a different element from the slide:

- Definition:** A circular icon containing a bulleted list of definitions for words like 'definite', 'definite article', 'distinctness', 'definition', 'definitive', and 'demonstrative'.
- Equation:** A circular icon containing a graph of a parabola opening upwards, with a vertex at $(-b/2a, \text{min})$. It includes the quadratic formula $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ and the equation $P=2L+2W$.
- Example:** A circular icon featuring a cartoon character of a woman holding a book, with a yellow arrow pointing to a green line graph on a grid.
- Properties:** A circular icon showing a glowing yellow lightbulb.

Definition (Slide Layer)

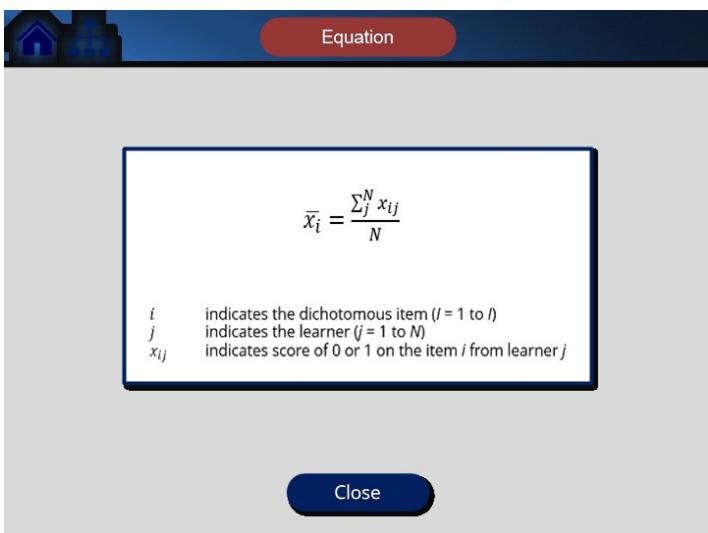


Definition

The proportion of a particular group of learners who answer a dichotomous item correctly.

Close

Equation (Slide Layer)



Equation

$$\bar{x}_i = \frac{\sum_j^N x_{ij}}{N}$$

i indicates the dichotomous item ($i = 1$ to I)
 j indicates the learner ($j = 1$ to N)
 x_{ij} indicates score of 0 or 1 on the item i from learner j

Close

Example (Slide Layer)

The screenshot shows a slide layer with a dark blue header bar. On the left of the header is a small icon of a house. In the center of the header is a red rounded rectangle containing the word "Example". Below the header is a light gray content area. In the top-left corner of this area is a white rectangular box with a black border. Inside the box, the text "If 43 learners out of 100 learners answered an item correctly:" is displayed in bold black font. Below this text is a mathematical equation:
$$\bar{x}_i = \frac{\sum_j^N x_{ij}}{N} = \frac{43}{100} = 0.43$$
. At the bottom of the slide layer is a dark blue "Close" button.

Properties (Slide Layer)

The screenshot shows a slide layer with a dark blue header bar. On the left of the header is a small icon of a house. In the center of the header is a red rounded rectangle containing the word "Properties". Below the header is a light gray content area. In the top-left corner of this area is a white rectangular box with a black border. Inside the box, the text "Range is from 0.0 to 1.0. A low p-value (e.g., below 0.1) describes a difficult item and a high p-value (e.g., above 0.9) describes an easy item." is displayed in bold black font. To the right of the text box is a full-body illustration of a man with gray hair, wearing a light green button-down shirt, a blue striped tie, and black trousers. He is standing with his hands at his sides, looking slightly upwards and to the left with a thoughtful or confused expression. At the bottom of the slide layer is a dark blue "Close" button.

3.11 Average Item Score (AIS) (I)

The slide has a dark blue header with the title "Average Item Score (AIS)" in yellow. Below the header are four circular icons, each with a red button below it:

- Definition:** A circular icon containing a snippet of text from a dictionary definition of the word "definite".
- Equation:** A circular icon showing a geometric diagram with a right-angled triangle and the equation $P=2l+2w$.
- Example:** A circular icon showing a person holding a tablet displaying a graph of a piecewise linear function.
- Properties:** A circular icon showing a glowing lightbulb.

Definition (Slide Layer)

The slide layer has a dark blue header with the word "Definition" in white. Below the header is a photograph of a man in a suit holding a tablet. To his right is a white text box with a black border containing the following text:

The average score of a particular group of learners who respond to a **polytomous** item.

At the bottom right of the slide layer is a blue "Close" button.

Equation (Slide Layer)

The slide layer has a dark blue header with a house icon and the word "Equation". The main content area contains the following text and equation:

$$\bar{x}_i = \frac{\sum_0^K N_k x_{i_k}}{N}$$

x_{i_k} indicates the k^{th} category score of item i ($k = 0$ to K)
 N_k indicates the number of examinees who answered x_{i_k}
 N implies the sum of N_k

A "Close" button is located at the bottom of the slide layer.

Example (Slide Layer)

The slide layer has a dark blue header with a house icon and the word "Example". The main content area contains the following text and equation:

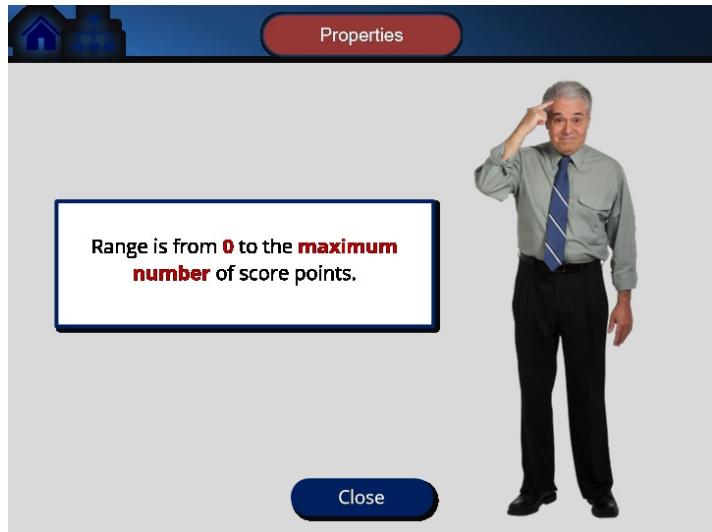
100 learners respond to a 4-point item:

- 10 learners get a score of '0'
- 25 learners get a score of '1'
- 35 learners get a score of '2'
- 20 learners get a score of '3'
- 10 learners get a score of '4'

$$\bar{x}_i = \frac{\sum_0^K N_k x_{i_k}}{N} = \frac{(0 * 10) + (1 * 25) + (2 * 35) + (3 * 20) + (4 * 10)}{100} = 1.95$$

A "Close" button is located at the bottom of the slide layer.

Properties (Slide Layer)



3.12 Average Item Score (AIS) (II)

Transformed AIS

- Definition**: A circular icon containing a detailed definition of a definite article.
- Equation**: A circular icon showing a geometric diagram of a rectangle with width 'w' and height 'l', and the formula $P=2l+2w$.
- Example**: A circular icon showing a person sitting at a desk with a graph on the screen.
- Properties**: A circular icon showing a glowing lightbulb.

Definition (Slide Layer)



Definition

Transformation

The **average item score** can be divided by the **maximum possible score** to have an **equivalent range** as the **p-value** (i.e., 0.0 - 1.0)

Close

Equation (Slide Layer)



Equation

$$p\text{-value} = \frac{AIS - AIS_{min}}{AIS_{max} - AIS_{min}}$$

Close

Example (Slide Layer)

The screenshot shows a slide layer with a dark blue header containing a house icon and the word "Example". The main content area contains a text box with the following text:
For example, the average item score of **1.95** can be transformed into a *p*-value of **0.49**:
$$p\text{-value} = \frac{AIS - AIS_{min}}{AIS_{max} - AIS_{min}} = \frac{1.95 - 0}{4 - 0} = 0.49$$

A "Close" button is located at the bottom of the slide layer.

Properties (Slide Layer)

The screenshot shows a slide layer with a dark blue header containing a house icon and the word "Properties". The main content area contains a text box with the following text:
Useful for **assembling parallel test forms** since the item difficulty of both **dichotomously scored** and **polytomously scored** items are **comparable / on the same scale**.

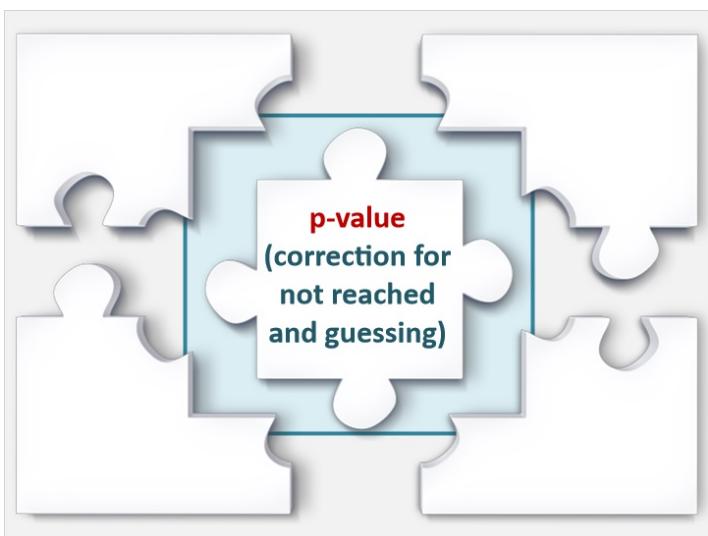
To the right of the text box, there is a photograph of a man in a light green shirt, dark trousers, and a striped tie, standing with his hand to his forehead in a thoughtful pose. A "Close" button is located at the bottom of the slide layer.

3.13 Bookend: p-value



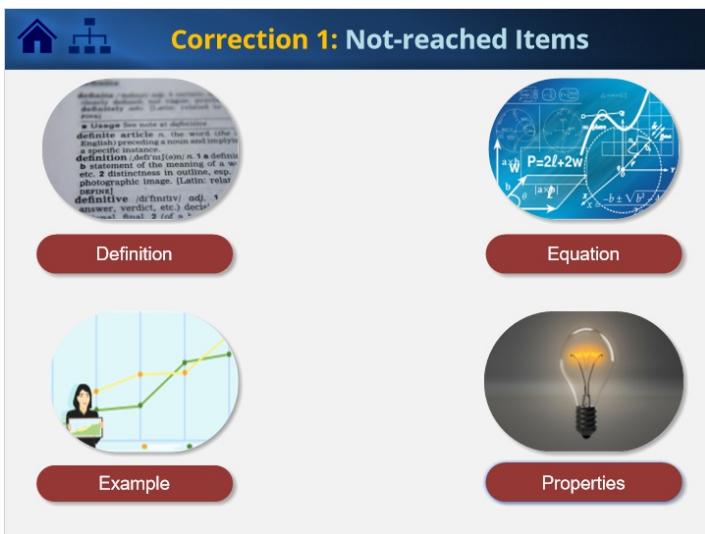
This is the end of this part.

3.14 Bookmark: Corrections



3.15 Correction 1: Not-reached

Correction 1: Not-reached Items



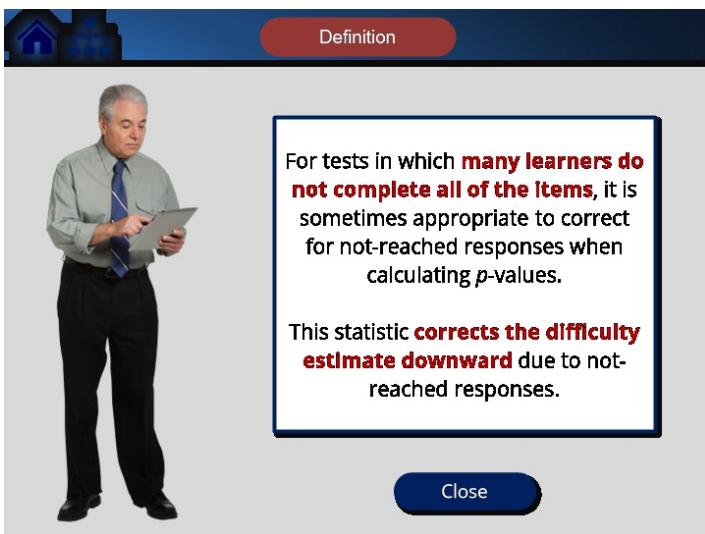
Definition

Equation

Example

Properties

Definition (Slide Layer)



Definition

For tests in which **many learners do not complete all of the items**, it is sometimes appropriate to correct for not-reached responses when calculating p -values.

This statistic **corrects the difficulty estimate downward** due to not-reached responses.

Close

Equation (Slide Layer)

The screenshot shows a slide layer with a dark blue header bar. On the left of the header is a small icon of a house with a blue roof. To the right of the icon is a red rounded rectangle containing the word "Equation". Below the header is a large white rectangular area with a thin black border. Inside this area, there is a mathematical formula:
$$\bar{x}_i = \frac{\sum_j^N x_{ij}}{N - N_{NR}}$$
 Below the formula, there is explanatory text: "N Indicates the number of total learners" and "N_{NR} indicates the number of not-reached responses". At the bottom of the slide layer is a dark blue rounded rectangle containing the word "Close".

Example (Slide Layer)

The screenshot shows a slide layer with a dark blue header bar. On the left of the header is a small icon of a house with a blue roof. To the right of the icon is a red rounded rectangle containing the word "Example". Below the header is a large white rectangular area with a thin black border. Inside this area, there is text: "43 learners out of 100 answer an item correctly" and "20 learners do not provide a response". Below this text is a mathematical formula:
$$\bar{x}_i = \frac{\sum_j^N x_{ij}}{N - N_{NR}} = \frac{43}{100 - 20} = 0.54$$
 Below the formula, there is text in red: "(uncorrected p-value = 0.43)". At the bottom of the slide layer is a dark blue rounded rectangle containing the word "Close".

Properties (Slide Layer)

This procedure will prevent **not-reached items** at the end of the test from **looking artificially difficult**.

It is applicable to field-test statistics when **learner motivation may be low**.

Close

A man in a green shirt and tie stands with his hand to his forehead, looking thoughtful.

3.16 Correction 2: Guessing

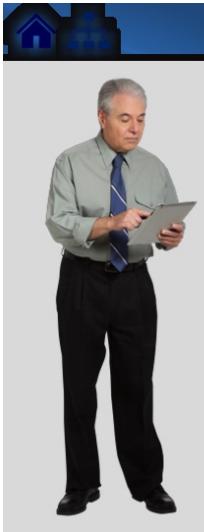
Definition

Equation

Example

Properties

Definition (Slide Layer)



Definition

Sometimes, it is desirable to determine **what proportion of learners** was able to answer an item correctly without guessing. In that case, the **p-value** should be corrected **downward**.

Close

Equation (Slide Layer)



Equation

When it is reasonable to assume that all distractors are approximately equally attractive:

$$\bar{x}_i = \frac{R - \frac{W}{n-1}}{N}$$

R Indicates the number of learners who choose the correct answer
W Indicates the number of examinees who choose distractors
n Indicates the number of options on the item
N Indicates the number of total learners

Close

Example (Slide Layer)

The screenshot shows a slide layer with a dark blue header bar. On the left of the header is a small icon of a house and trees. To the right of the icon is a red rounded rectangle containing the word "Example". Below the header is a white rectangular content area with a thin black border. Inside this content area, there is text in red and black. The text reads: "43 learners out of 100 choose the correct answer" and "4 response options for the item". Below this text is a mathematical equation:
$$\bar{x}_i = \frac{R - \frac{W}{n-1}}{N} = \frac{43 - \frac{57}{4-1}}{100} = 0.24$$
 followed by the note "(uncorrected p-value = 0.43)". At the bottom of the content area is a dark blue "Close" button.

Properties (Slide Layer)

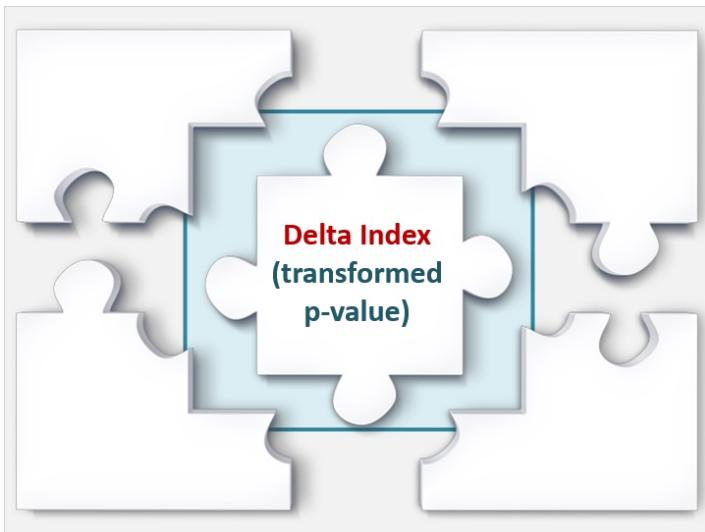
The screenshot shows a slide layer with a dark blue header bar. On the left of the header is a small icon of a house and trees. To the right of the icon is a red rounded rectangle containing the word "Properties". Below the header is a white rectangular content area with a thin black border. Inside this content area, there is text in red and black. The text reads: "This correction is based on a **strong assumption** that learners **either know the correct answer or guess randomly.**". Below this text is a red triangular warning sign with a white exclamation mark. To the right of the content area is a full-body illustration of a man in a light green shirt, a blue striped tie, and black trousers. He is standing with his left hand resting against his head, looking thoughtful. At the bottom of the content area is a dark blue "Close" button.

3.17 Bookend: Corrections



This is the end of this part.

3.18 Bookmark: Delta Index



3.19 Delta Index

The slide has a dark blue header with the title "Delta Index". Below the header are four circular icons, each with a red button at the bottom:

- Definition:** Shows a circular diagram with text about the usage of the definite article "the".
- Equation:** Shows a circular diagram with a geometric proof involving a rectangle and a diagonal, with the formula $P=2l+2w$.
- Example:** Shows a circular diagram with a graph of a piecewise function and a person holding a flag.
- Properties:** Shows a circular diagram with a glowing lightbulb.

Definition (Slide Layer)

The slide has a dark blue header with a house icon and the word "Definition". Below the header is a large image of a man in a suit reading a tablet. To his right is a white text box with a black border containing the following text:

A nonlinear transformation of the
average item score (AIS)
to a **normal distribution** with
mean = 13 and **standard deviation = 4**.

Below the text box is a circular icon with a blue bell curve. At the bottom right is a blue button labeled "Close".

Equation (Slide Layer)

Equation

$$\Delta = 13 - 4\Phi^{-1}(p)$$

Φ is the normal cumulated distribution function
 p is p-value
13 mean of distribution
4 standard deviation of distribution

Close

Example (Slide Layer)

Example

If p-value = 0.60 then the corresponding delta value = 12.0

$$\Delta = 13 - 4\Phi^{-1}(p)$$

	A	B	C	D	E	F	G
1							
2	p-value	0.6			Equation: p-value to Delta	=NORM.INV(1-C2,13,4)	
3	Delta value	12.0			Equation: Delta to p-value	=NORM.DIST(13-C3,0,4,TRUE)	
4							

Close

Properties (Slide Layer)

Properties

- Range is from **1 to 25** with **± 3 standard deviations** (positive numbers only).
- Unlike the *p*-value, a **high** delta value implies a **difficult** item and a **low** delta value implies an **easy** item.
- Delta serves to express **Item difficulty** on an **equal interval scale** whereas *p*-values are measures on an **ordinal scale**.
- Delta **magnifies small differences** between items at the **extreme ends of the scale**.

Close

3.20 Equated Delta Index

Equated Delta Index

Definition

Equation

Example

Properties

Definition (Slide Layer)



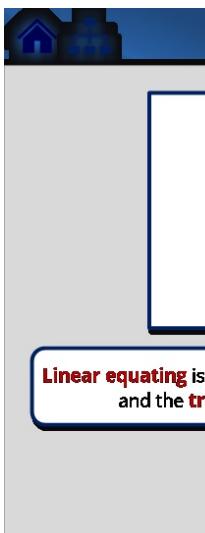
Definition

To make delta values **comparable over different examinee samples**, delta values can be "**adjusted**" using **linear equating**:

- ✓ **two administrations**
- ✓ **one reference sample**
- ✓ **one "anchor set" of items**
- ✓ **one resulting transformation**

Close

Equation (Slide Layer)



Equation

$$\frac{\Delta_R - \bar{\Delta}_R}{SD(\Delta_R)} = \frac{\Delta_N - \bar{\Delta}_N}{SD(\Delta_N)}$$

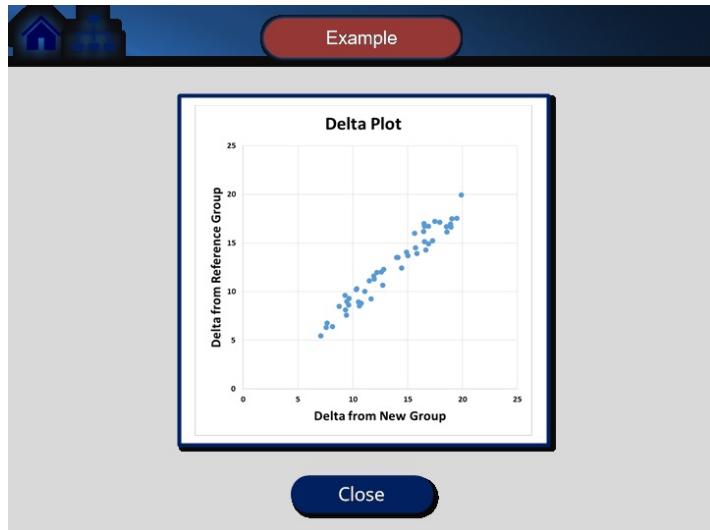
$$\bar{\Delta}_R = \frac{SD(\Delta_R)}{SD(\Delta_S)} (\Delta_N - \bar{\Delta}_N) + \bar{\Delta}_S$$

R indicates reference group
N indicates target group

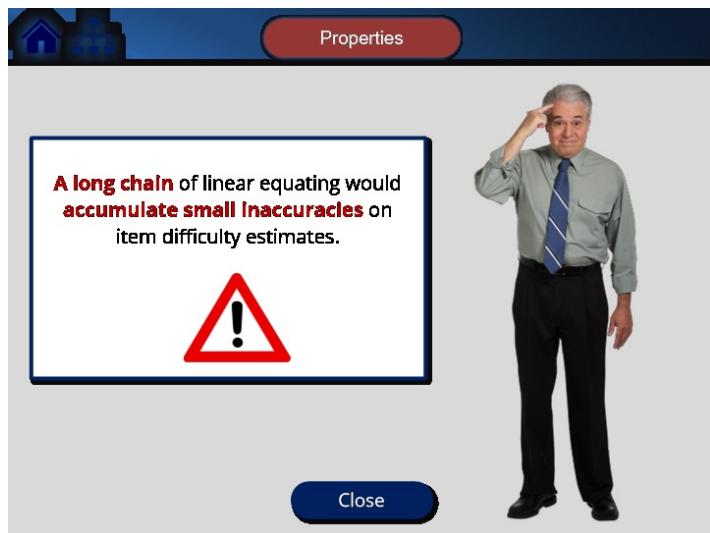
Linear equating is carried out with **anchor items** for **two administrations** and the **transformation** is then applied to **all new items**.

Close

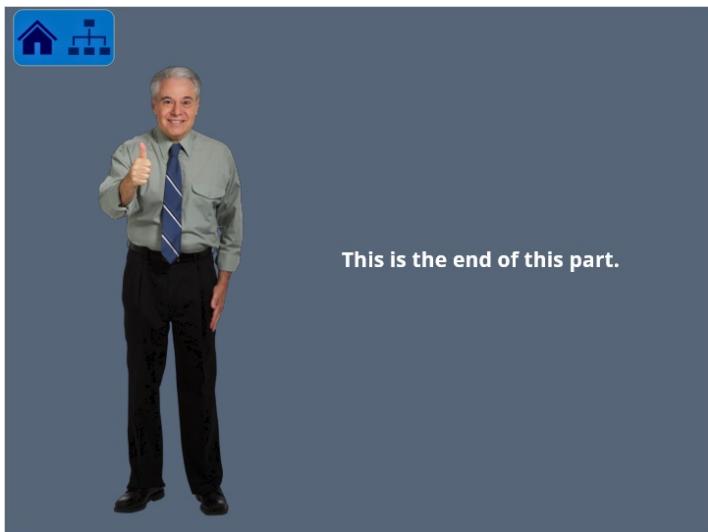
Example (Slide Layer)



Properties (Slide Layer)



3.21 Bookend: Delta Index



This is the end of this part.

3.22 Bookmark: Item Discrimination



3.23 Correlation Coefficients

Foundations: Scales & Assumptions

Indices in the two green boxes only are the focus of this module.

Item Variable	Criterion Variable		
	Dichotomous	Polytomous	Continuous
Dichotomous	Phi Tetrachoric (*)	Other Polychoric (*)	Point-biserial Biserial (*)
Polytomous		Other Polychoric (*)	Point-polyserial Polyserial (*)
Continuous			Pearson

* Assumptions about underlying unobserved (latent) variable are made (beyond scope of module).

3.24 Foundations: Pearson Correlation

Item-Total Score Correlations

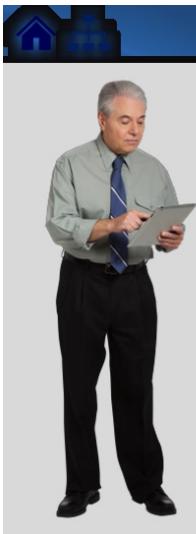
Definition

Equation

Example

Properties

Definition (Slide Layer)



Definition

A **Pearson correlation** coefficient captures the degree of **linear association** between **two continuous variables**.

Item discrimination is defined as the **correlation** between **item scores** and **some criterion scores of interest** (e.g., total test scores, pass-fail).

Many **discrimination indices** are based on **Pearson correlation** and associated variants.

Close

Equation (Slide Layer)



Equation

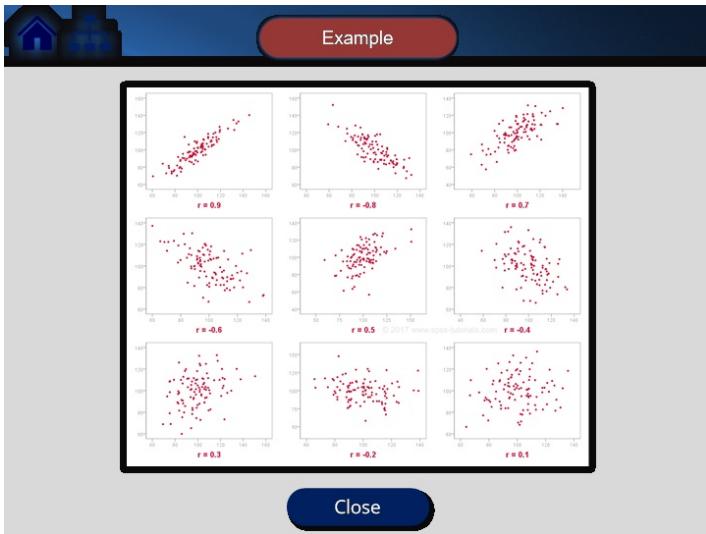
Most Indices are measures of association based on the **Pearson product-moment correlation**, which is a correlation between **two continuous variables** (e.g., height and weight):

$$r_{(x,y)} = \frac{\sigma_{(x,y)}}{\sigma_{(x)}\sigma_{(y)}}$$

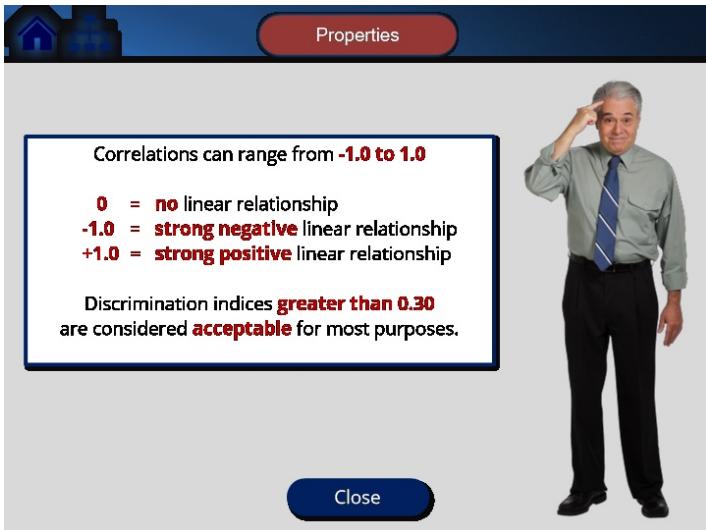
$\sigma_{(x,y)}$ indicates the covariance of the two variables X and Y
 $\sigma_{(x)}$ indicates the standard deviation of variable x
 $\sigma_{(y)}$ indicates the standard deviation of variable y

Close

Example (Slide Layer)



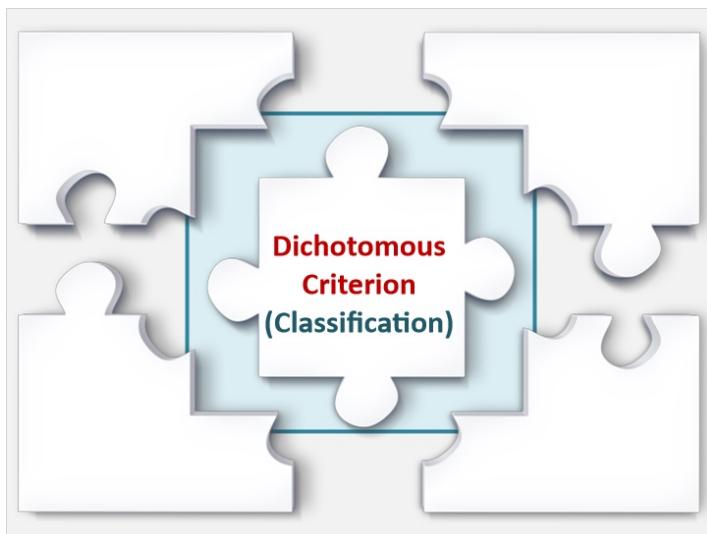
Properties (Slide Layer)



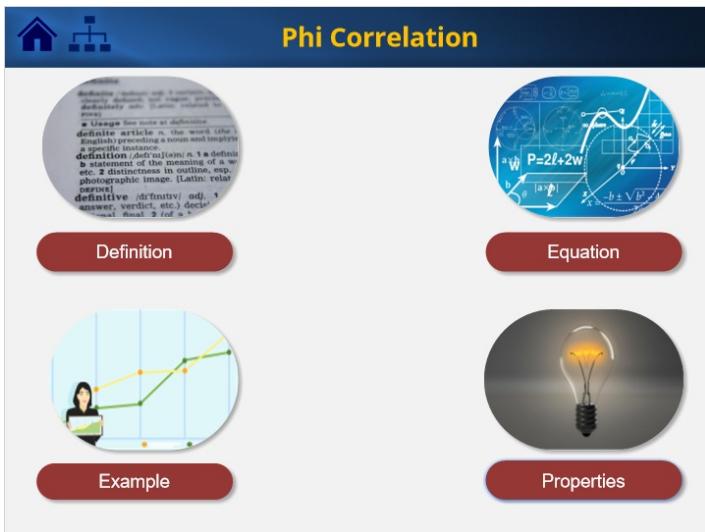
3.25 Topic Selection: Discrimination



3.26 Bookmark: Dichotomous Criterion



3.27 Phi Correlation



Phi Correlation

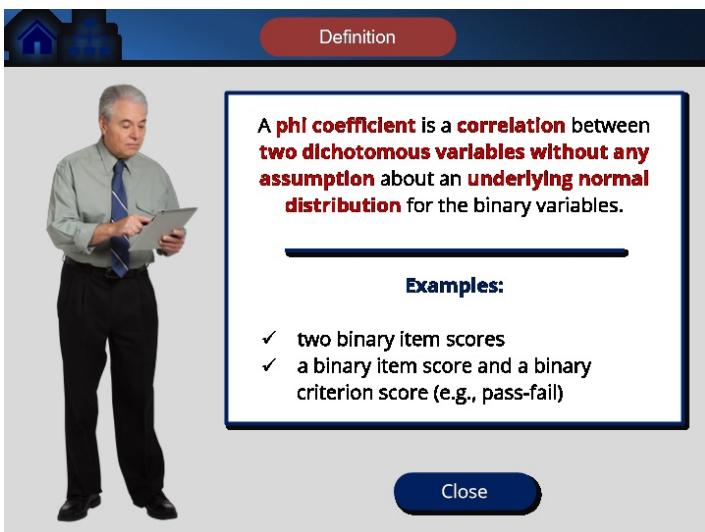
Definition: *a* Usage like *the* at definite article *a*, the word *the* is used to refer to a specific instance.
b Statement of the meaning of a word or words; definition: *a* & *b* distinctness in outcome, especially in meaning, resulting from parallel processes (definitive, definitive, verdict, etc.) decisive answer, final 2 inf n

Equation: $P=2\sqrt{W}$

Example: A graph showing a path from a person to a lightbulb.

Properties: A glowing lightbulb.

Definition (Slide Layer)



Definition

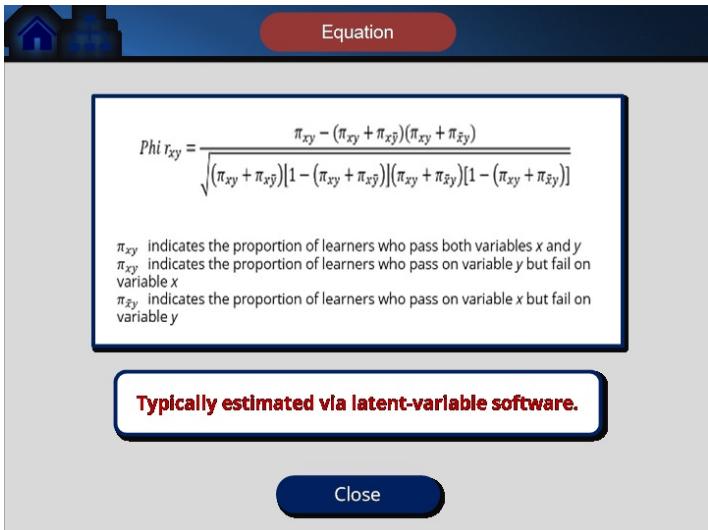
A **phi coefficient** is a **correlation** between **two dichotomous variables without any assumption** about an **underlying normal distribution** for the binary variables.

Examples:

- ✓ two binary item scores
- ✓ a binary item score and a binary criterion score (e.g., pass-fail)

Close

Equation (Slide Layer)



The slide layer has a dark blue header with a house icon and the word "Equation". The main content area contains a mathematical formula for Phi (Φ) and its interpretation:

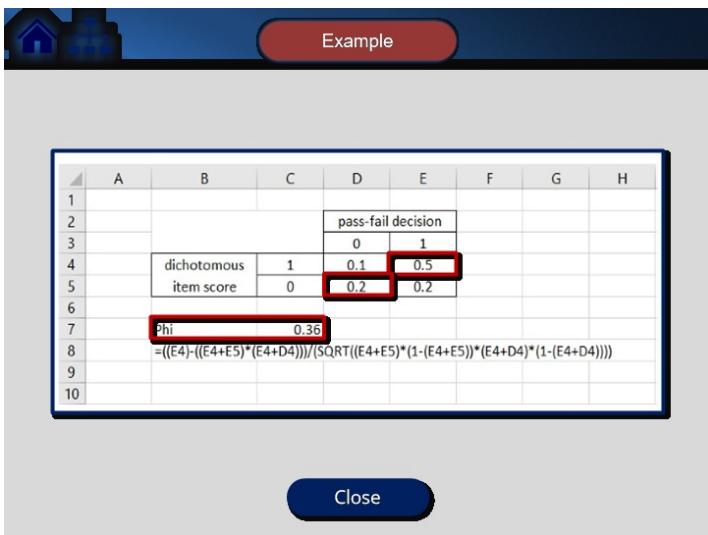
$$\Phi_{xy} = \frac{\pi_{xy} - (\pi_{xy} + \pi_{x\bar{y}})(\pi_{xy} + \pi_{\bar{x}y})}{\sqrt{(\pi_{xy} + \pi_{x\bar{y}})[1 - (\pi_{xy} + \pi_{x\bar{y}})][(\pi_{xy} + \pi_{\bar{x}y})[1 - (\pi_{xy} + \pi_{\bar{x}y})]}}$$

π_{xy} indicates the proportion of learners who pass both variables x and y
 $\pi_{x\bar{y}}$ indicates the proportion of learners who pass on variable y but fail on variable x
 $\pi_{\bar{x}y}$ indicates the proportion of learners who pass on variable x but fail on variable y

Typically estimated via latent-variable software.

Close

Example (Slide Layer)



The slide layer has a dark blue header with a house icon and the word "Example". The main content area shows a Microsoft Excel spreadsheet with data related to a pass-fail decision:

		pass-fail decision	
		0	1
dichotomous	1	0.1	0.5
item score	0	0.2	0.2
Phi	0.36		

=((E4)-(E4+E5)*(E4+D4))/(SQRT((E4+E5)*(1-(E4+E5))*(E4+D4)*(1-(E4+D4))))

Close

Properties (Slide Layer)

The slide has a dark blue header bar with a house icon and the word 'Properties'. The main content area contains a white box with a black border. Inside the box, the text reads:
Range is from **-1 to +1**
-1 implies a perfect **negative relationship**
+1 implies a perfect **positive relationship**
Both **extreme** cases (i.e., -1 or +1) would have **no values** in the **off-diagonals** of the **2 x 2 cross-classification table.**

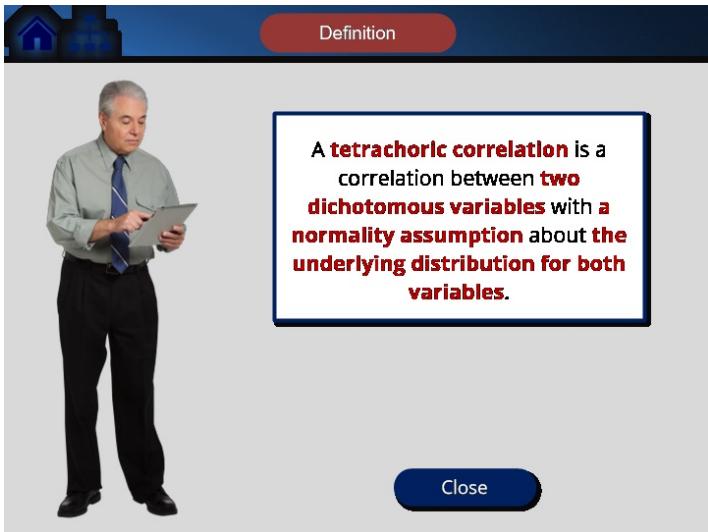
To the right of the text box is a photograph of a man in a grey shirt and striped tie, standing with his hand to his forehead in a thinking pose. At the bottom of the slide is a blue 'Close' button.

3.28 Tetrachoric Correlation

The slide has a dark blue header bar with a house icon and the title 'Tetrachoric Correlation'. It features four circular content boxes arranged in a 2x2 grid:

- Definition:** A circular icon containing a snippet of text about the usage of the definite article 'the'.
- Equation:** A circular icon showing a geometric diagram of a rectangle with width ℓ and height w , with the formula $P=2\ell+2w$ displayed.
- Example:** A circular icon showing a scatter plot with a fitted curve.
- Properties:** A circular icon showing a glowing lightbulb.

Definition (Slide Layer)

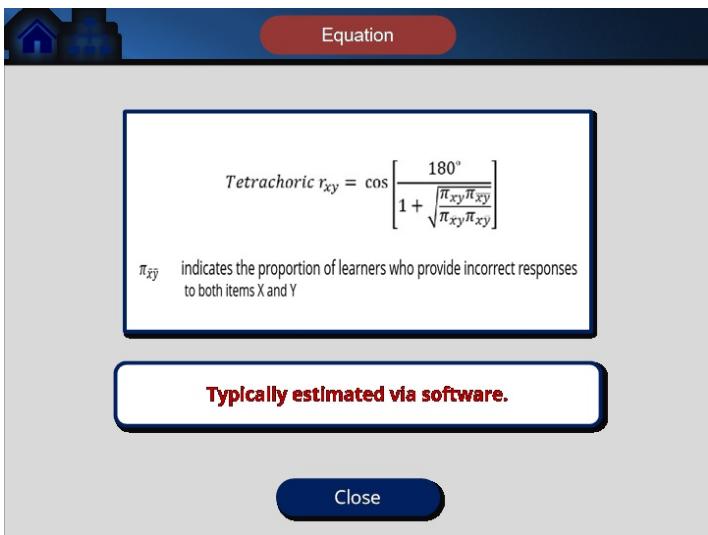


Definition

A **tetrachoric correlation** is a correlation between **two dichotomous variables** with a **normality assumption** about the **underlying distribution for both variables**.

Close

Equation (Slide Layer)



Equation

$$\text{Tetrachoric } r_{xy} = \cos \left[\frac{180^\circ}{1 + \sqrt{\pi_{xy}\pi_{\bar{x}\bar{y}}}} \right]$$

$\pi_{\bar{x}\bar{y}}$ indicates the proportion of learners who provide incorrect responses to both items X and Y

Typically estimated via software.

Close

Example (Slide Layer)

Example

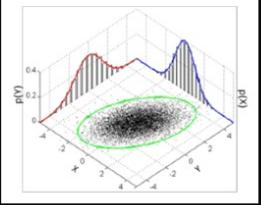
	A	B	C	D	E	F
18						
19					pass-fail decision	
20				0	1	
21	dichotomous	1	0.1	0.5		
22	item score	0	0.2	0.2		
23						
24	tetrachoric	0.60				
25	$=\text{COS}(180/(1+(\text{SQRT}((E21*D22)/(D21*E22))))))$					
26						

Close

Properties (Slide Layer)

Properties

Unlike a phi correlation, a **tetrachoric correlation** assumes that a **bivariate normal distribution** underlies the **dichotomously scored variables**.



Close



3.29 Bookend: Dichotomous Criterion

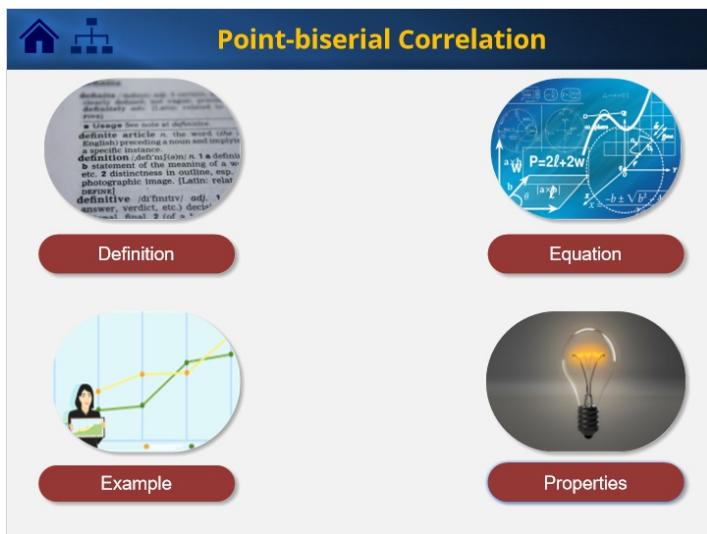


This is the end of this part.

3.30 Bookmark: Continuous Criterion



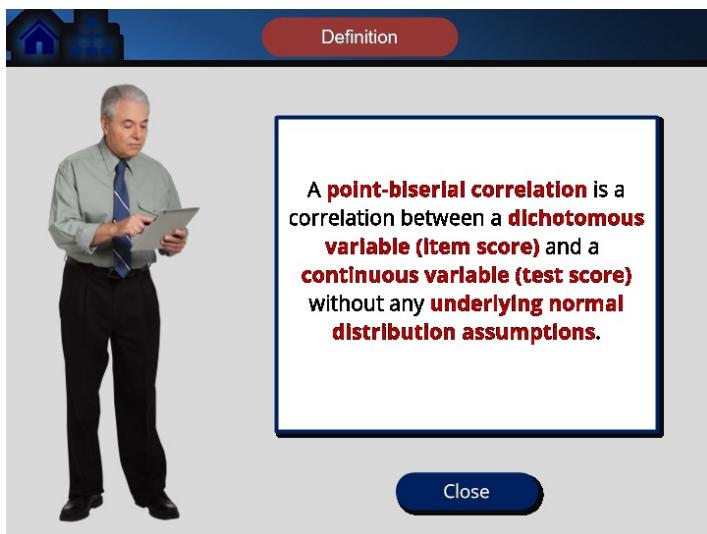
3.31 Point-biserial Correlation



The slide has a dark blue header with a house icon and a navigation tree. The title "Point-biserial Correlation" is in yellow. Below the title are four circular sections with labels: "Definition", "Equation", "Example", and "Properties".

- Definition:** A circular icon containing a snippet from a grammar book about the usage of the definite article "the".
- Equation:** A circular icon showing a graph of a normal distribution curve with a point marked at $P=2t+2\sigma$.
- Example:** A circular icon showing a scatter plot with a regression line and a person holding a chart.
- Properties:** A circular icon showing a glowing lightbulb.

Definition (Slide Layer)

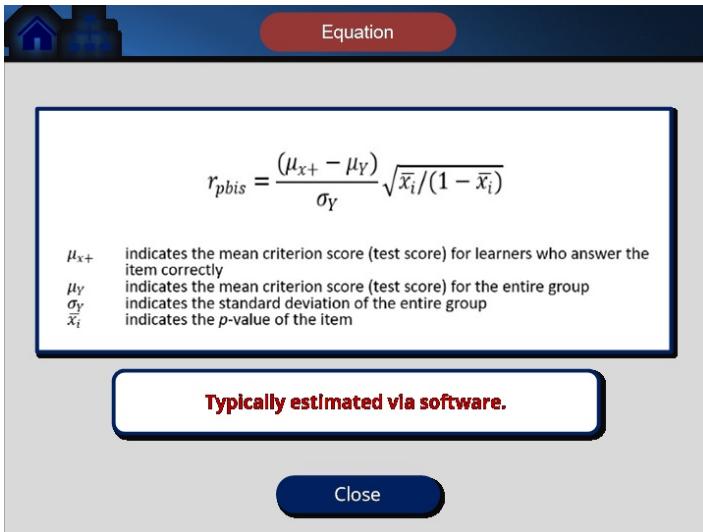


The slide layer has a dark blue header with a house icon and a "Definition" button. It features a man in a suit holding a tablet on the left and a text box on the right.

A point-biserial correlation is a correlation between a **dichotomous variable (item score)** and a **continuous variable (test score)** without any **underlying normal distribution assumptions**.

Close

Equation (Slide Layer)



The slide layer has a dark blue header bar with a house icon and the word "Equation". The main content area contains the formula for r_{pbis} and its components with their meanings. A note states it's typically estimated via software, and a "Close" button is at the bottom.

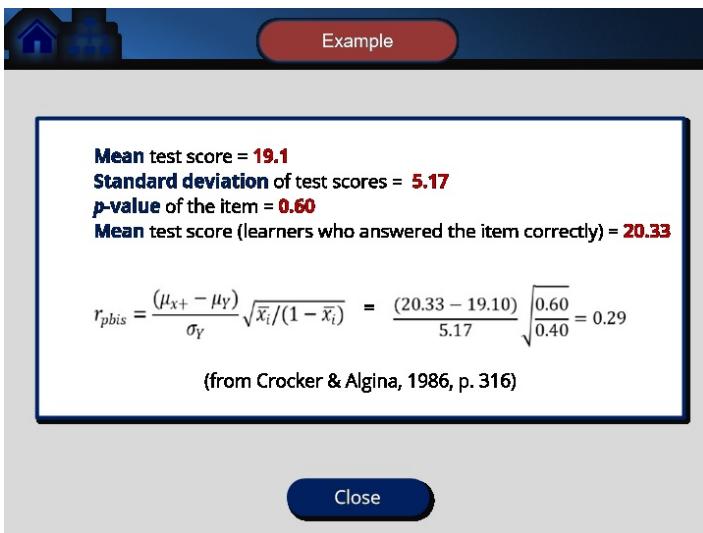
$$r_{pbis} = \frac{(\mu_{x+} - \mu_Y)}{\sigma_Y} \sqrt{\bar{x}_i / (1 - \bar{x}_i)}$$

μ_{x+} indicates the mean criterion score (test score) for learners who answer the item correctly
 μ_Y indicates the mean criterion score (test score) for the entire group
 σ_Y indicates the standard deviation of the entire group
 \bar{x}_i indicates the *p*-value of the item

Typically estimated via software.

Close

Example (Slide Layer)



The slide layer has a dark blue header bar with a house icon and the word "Example". It displays calculated statistics and the formula for r_{pbis} . A note credits Crocker & Algina (1986). A "Close" button is at the bottom.

Mean test score = **19.1**
Standard deviation of test scores = **5.17**
p-value of the item = **0.60**
Mean test score (learners who answered the item correctly) = **20.33**

$$r_{pbis} = \frac{(\mu_{x+} - \mu_Y)}{\sigma_Y} \sqrt{\bar{x}_i / (1 - \bar{x}_i)} = \frac{(20.33 - 19.10)}{5.17} \sqrt{\frac{0.60}{0.40}} = 0.29$$

(from Crocker & Algina, 1986, p. 316)

Close

Properties (Slide Layer)

The slide layer features a man in a grey shirt and blue striped tie, standing with his hand to his forehead in a thinking pose. To his left is a white box containing text about Pearson's bivariate correlation. At the bottom right is a 'Close' button.

Range is from -1 to +1

-1 implies a perfect **negative relationship**
+1 implies a perfect **positive relationship**

Computing **Pearson's bivariate correlation** between each item's score and the total test score provides the **same results**

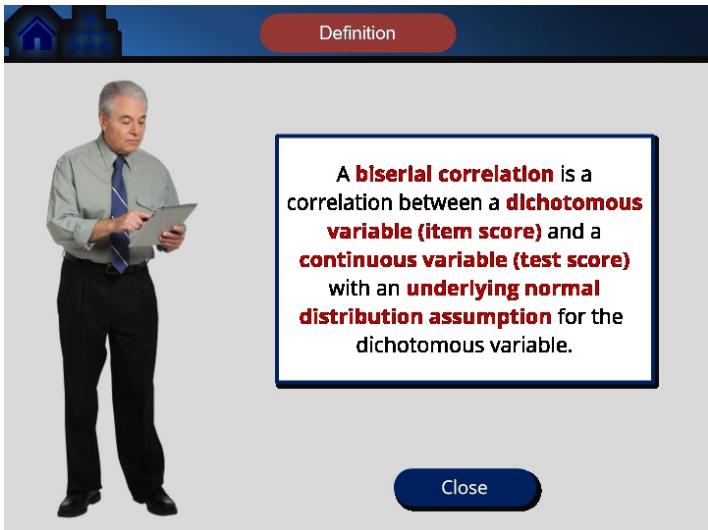
Close

3.32 Biserial Correlation

The slide layer for Biserial Correlation contains four main sections:

- Definition:** A circular icon showing a definition of the term.
- Equation:** A circular icon showing the formula $P=2\ell+2w$.
- Example:** A circular icon showing a scatter plot with a regression line.
- Properties:** A circular icon showing a glowing lightbulb.

Definition (Slide Layer)

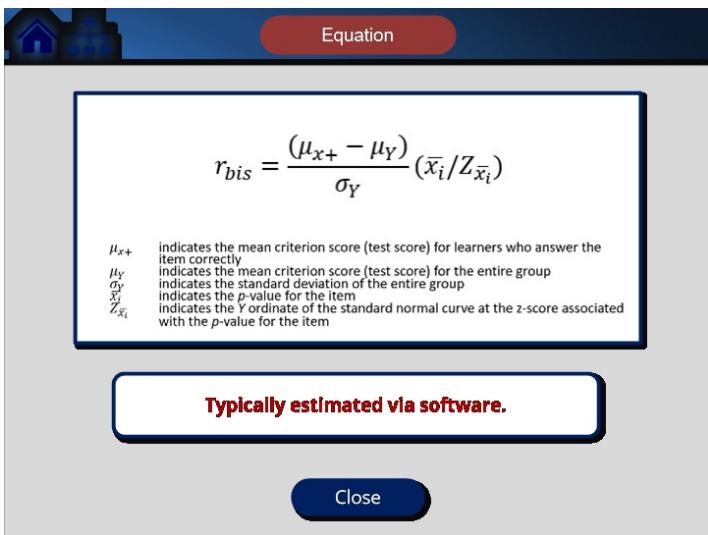


Definition

A **biserial correlation** is a correlation between a **dichotomous variable (item score)** and a **continuous variable (test score)** with an **underlying normal distribution assumption** for the dichotomous variable.

Close

Equation (Slide Layer)



Equation

$$r_{bis} = \frac{(\mu_{x+} - \mu_Y)}{\sigma_Y} (\bar{x}_i / Z_{\bar{x}_i})$$

μ_{x+} indicates the mean criterion score (test score) for learners who answer the item correctly
 μ_Y indicates the mean criterion score (test score) for the entire group
 σ_Y indicates the standard deviation of the entire group
 \bar{x}_i indicates the *p*-value for the item
 $Z_{\bar{x}_i}$ indicates the *Y* ordinate of the standard normal curve at the *z*-score associated with the *p*-value for the item

Typically estimated via software.

Close

Example (Slide Layer)

The slide layer has a dark blue header with a house icon and the word 'Example'. The main content area contains the following text:

Mean test score = 19.1
Standard deviation of test scores = 5.17
p-value of the item in question = 0.60 (normal ordinate = 0.39)
Mean test score (learners who answered the item correctly) = 20.33

Below this is a mathematical formula:

$$r_{bis} = \frac{(\mu_{x+} - \mu_Y)}{\sigma_Y} (\bar{x}_i / Z_{\bar{x}_i}) = \frac{(20.33 - 19.10)}{5.17} \left(\frac{0.60}{0.39} \right) = 0.37$$

At the bottom right is a 'Close' button.

Properties (Slide Layer)

The slide layer has a dark blue header with a house icon and the word 'Properties'. The main content area contains the following text:

Unlike a point-biserial correlation, a **biserial correlation** assumes that a **normal distribution** underlies the **dichotomously variable** and that there is a **linear relationship** between the **item score** and the **total test score**.

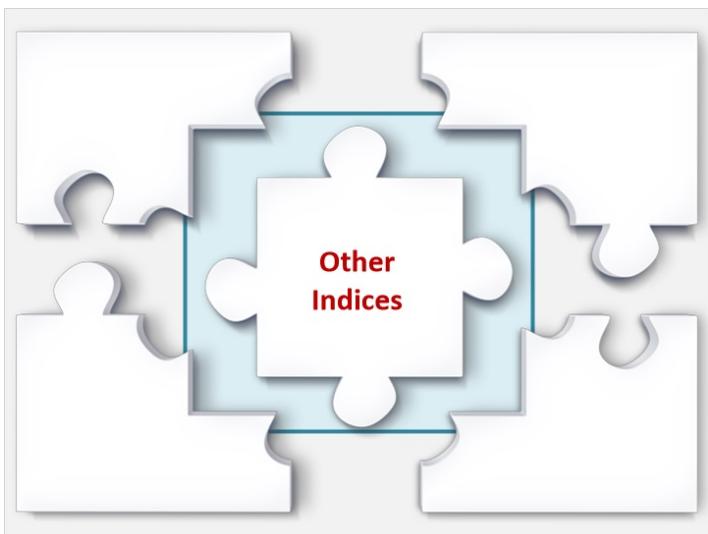
Below the text is a circular icon containing a blue normal distribution curve. At the bottom right is a 'Close' button.

3.33 Bookend: Continuous Criterion

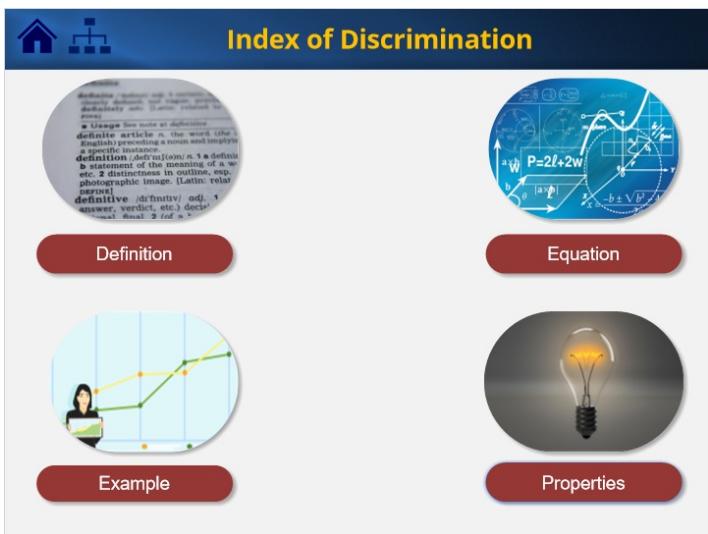


This is the end of this part.

3.34 Bookmark: Other Indices



3.35 Index of Discrimination



The slide is titled "Index of Discrimination". It features four circular icons with labels: "Definition" (with a text box), "Equation" (with a mathematical diagram), "Example" (with a graph), and "Properties" (with a lightbulb).

Definition:
discriminate: *verb* verb 1. to consider or treat differently from others in a way that is unfair or prejudiced against them. 2. to consider or treat (something) as being different from others in a way that is unfair or prejudiced against them.
a Usage: See note at definition.
definite article: *n.* the word "the" used before a noun to refer to a specific person, place, thing, etc.
definition: *n.* 1. a formal statement of the meaning of a word, etc. 2. distinctness in outcome, especially in a trial or competition, which permits a definite (definitive) adj., answer, verdict, etc.) decision.

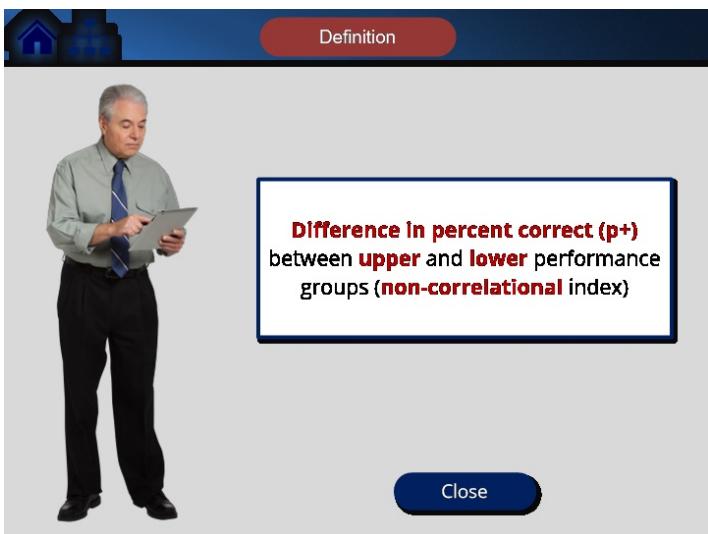
Equation:
$$P=2l+2w$$

$$b = \sqrt{a^2 + b^2}$$

Example:


Properties:


Definition (Slide Layer)



A man in a suit is standing and looking at a tablet. A callout box contains the following text:

Difference in percent correct ($p+$) between upper and lower performance groups (non-correlational index)

Close

Equation (Slide Layer)

The screenshot shows a slide layer with a dark blue header bar. On the left of the header is a small icon of a house. In the center of the header is a red button-like shape containing the word "Equation". Below the header is a light gray content area. In the center of this area is a white rectangular box with a black border. Inside the box, the equation $D = P_u - P_l$ is displayed. Below the equation, there is explanatory text: " P_u indicates the proportion in the upper group who answered the item correctly" and " P_l indicates the proportion in the lower group who answered the item correctly". At the bottom of the slide layer is a dark blue "Close" button.

Example (Slide Layer)

The screenshot shows a slide layer with a dark blue header bar. On the left of the header is a small icon of a house. In the center of the header is a red button-like shape containing the word "Example". Below the header is a light gray content area. In the center of this area is a white rectangular box with a black border. Inside the box, there is explanatory text: "Upper proportion = total score of at least 73%", "Lower proportion = total score of no more than 27%", "p-value in upper proportion = 86%", and "p-value in lower proportion = 12%". Below this text is the equation $D = P_u - P_l = .86 - .12 = .74$. At the bottom of the slide layer is a dark blue "Close" button.

Properties (Slide Layer)

Range is from **-1.0** to **+1.0**

Positive values indicate that an item discriminates in favor of the **upper group** while **negative values** indicate that an item discriminates in favor of the **lower group**.

D ≥ 0.40	the item is functioning fine
0.39 ≥ D ≥ 0.20	the item may need revision
D ≤ 0.19	the item should be removed

Close

A thinking man icon is standing on the right.

3.36 Correction for Item Influence

Definition

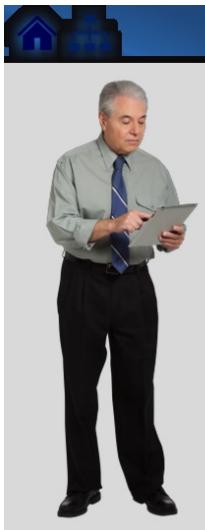
Equation

$$P = 2\ell + 2w$$

Example

Properties

Definition (Slide Layer)



Definition

- Researchers worry about **item bias** since an **item's score was included** in the calculation of the **total test score**.
- A **correlation** between an item score and the total score **with that item removed**.
- The **computation of any item statistic** can be **modified** in this way.

Close

Equation (Slide Layer)



Equation

$$\text{Corrected } r_{pbis} = \frac{\sigma_Y(r_{pbis}) - \sigma_x}{\sqrt{\sigma_x^2 + \sigma_y^2 - 2(r_{pbis})\sigma_x\sigma_y}}$$

σ_x indicates the standard deviation of x (e.g., the item score)
 σ_y indicates the standard deviation of y (e.g., the total test score)
 r_{pbis} indicates the uncorrected point-biserial correlation

Typically estimated via software.

Close

Example (Slide Layer)

Example

Uncorrected point-biserial correlation = 0.29
Standard deviation of the total test scores = 5.17
Standard deviation of the item scores = 0.30

$$\begin{aligned} \text{Corrected } r_{pbis} &= \frac{\sigma_Y(r_{pbis}) - \sigma_x}{\sqrt{\sigma_x^2 + \sigma_y^2 - 2(r_{pbis})\sigma_x\sigma_y}} \\ &= \frac{5.17 * 0.29 - 0.30}{\sqrt{0.30^2 + 5.17^2 - 2 * 0.29 * 0.30 * 5.17}} = 0.24 \end{aligned}$$

Close

Properties (Slide Layer)

Properties

- As the **standard deviation** of the item score increases, the **corrected correlation** will be **smaller** than **original correlation**.
- The correction will work when the **test length is short**.
- If test length is **longer** (e.g., more than 25) items, the effect of correction will be **minimal**.

Close



3.37 Bookend: Other Indices



This is the end of this part.

3.38 Bookmark: Flagging Rules



3.39 Flagging Rules from Item Analysis (1)

Flagging Rules

Many operational programs are using similar flagging rules to remove problematic items during item analysis:

- ✓ **High difficulty** (e.g., % correct less than .20)
- ✓ **Low difficulty** (e.g., % correct more than .80)
- ✓ **Low discrimination** (e.g., point-biserial less than .20)
- ✓ **High percentage of missing responses** (e.g., % greater than 50%)
- ✓ **Some distractors chosen often** by high-ability test takers
- ✓ **Other customized rules** (e.g., using process data)

3.40 Conceptual Foundations

General Principles

Using a set of response curves for each item, reviewers can immediately view multiple aspects of item performance such as:

- **Overall difficulty** of the item
- **How item difficulty varies with the total score** of learners
- **Which answers are being chosen** by learners at different performance levels **(for MC items)**
- **Which scores are attained** by learner at different performance levels **(for CR items)**

Reference

References (Slide Layer)

The screenshot shows a slide with a dark blue header containing a house icon and the word "Reference". Below the header is a white rectangular area representing a document cover. The cover has the "ETS" logo at the top left, the words "Research Report" at the top right, and the title "A Graphical Approach to Item Analysis" in the center. At the bottom of the cover, it says "Samuel A. Livingston" and "Neil J. Dorans". In the bottom right corner of the slide, there is a purple button with the word "Back".

3.41 Example 1: PARCC

The screenshot shows a slide with a dark blue header containing a house icon and the title "Example 1: PARCC". Below the header is a white rectangular area with a blue rounded rectangle containing a list of 8 items. The items are numbered 1 through 8 and describe various statistical and performance metrics for PARCC items. At the bottom of the slide, there is a purple button with the text "Access Report in Resource Tab".

1. p -value above .95 for dichotomous items and above .80 for polytomous items
2. p -value below .25 for dichotomous items, and below .30 for polytomous items
3. Item-total correlation below .15
4. Any distractor-total correlation above .00
5. Greater number of high-performing students (top 20 percent) choosing a distractor rather than the keyed response
6. High percentage of omits: above 5 percent for selected-response items and above 15 percent for constructed-response items
7. High percentage that did not reach the item: above 5 percent for selected-response items and above 15 percent for constructed-response items
8. CR items with a score value obtained by less than 3 percent of responses

3.42 Example 2: SmarterBalanced

Example 2: SmarterBalanced

Flag	Definition
A	High difficulty (p -value less than 0.10)
B	Polytomous items with percentage obtaining any score category less than three percent of total N
C	Polytomous items with higher criterion score mean for students in a lower score-point category
D	Selected response items with proportionally more high-proficient students selecting a distractor over the key
F	Selected response items with higher criterion score mean for students choosing a distractor than the mean for those choosing the key
H	Low difficulty (p -value greater than 0.95)
P	Selected response items with positive distractor point-biserial correlation
R	Low item-total correlation (p -value less than 0.30)
V	Item more difficult at the higher-grade level for vertical linking items
Z	Item needs content review (judgmental decision)

**Access Report
in Resource Tab**

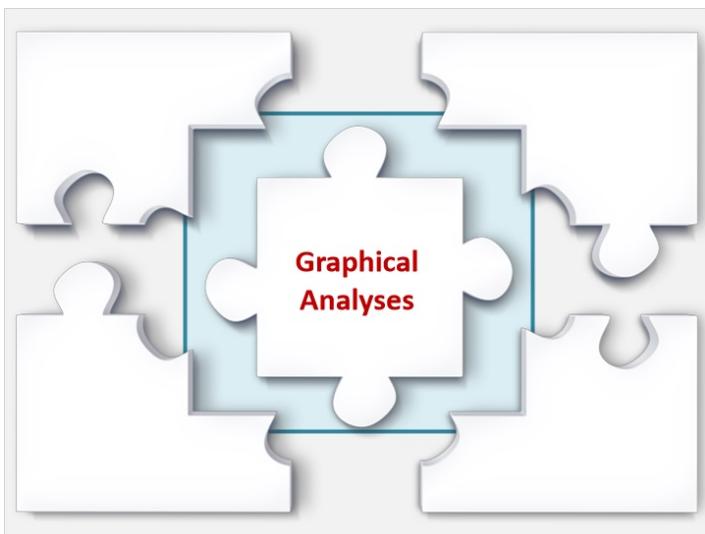
3.43 Bookend: Graphical Analyses



3.44 Bookmark: Distractors



3.45 Bookmark: Graphical Analyses



3.46 Conceptual Foundations

Graphical Layout

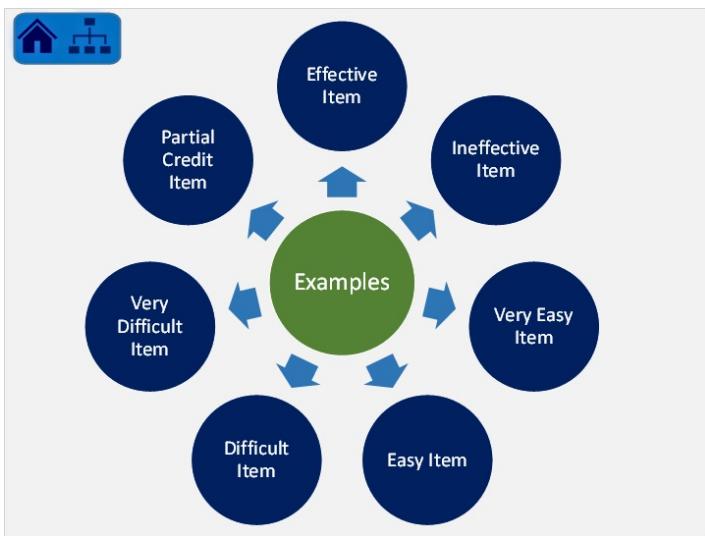
- Each **graph** per item consists of one or more **response curves**
- The **horizontal axis** represents the score scale for the criterion
- The **vertical axis** represents the probability of a correct response
- The **response curve** shows a learner's probability of a particular response to the item as a function of the learner's total score

3.47 Conceptual Foundations

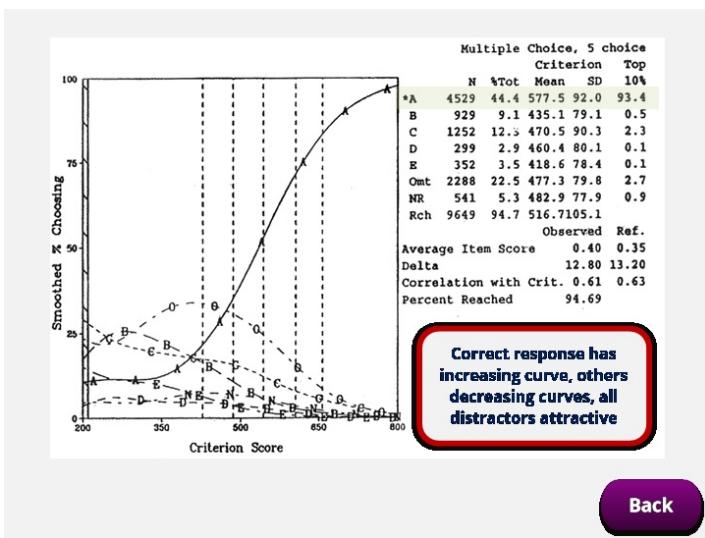
Response Curve Interpretation

- ✓ The **higher** the response curve (in any given range of the total score), the **easier** the item (**and vice versa**)
- ✓ The **more steeply** the height of the curve **increases** from left (low performers) to right (high performers), the **greater** the discriminating power of the item (**and vice versa**)

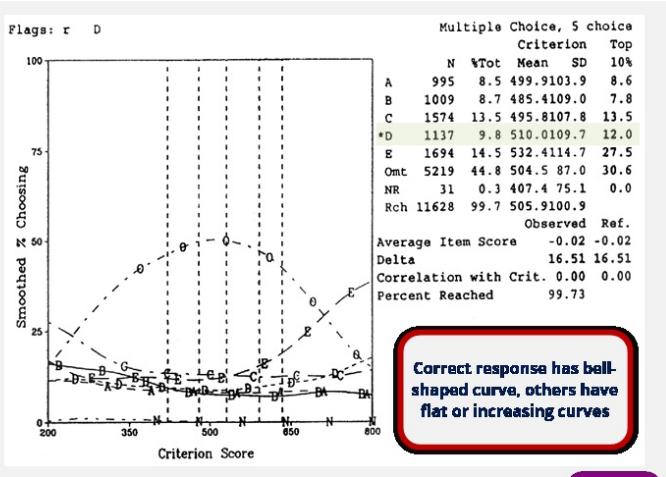
3.48 Item Selection



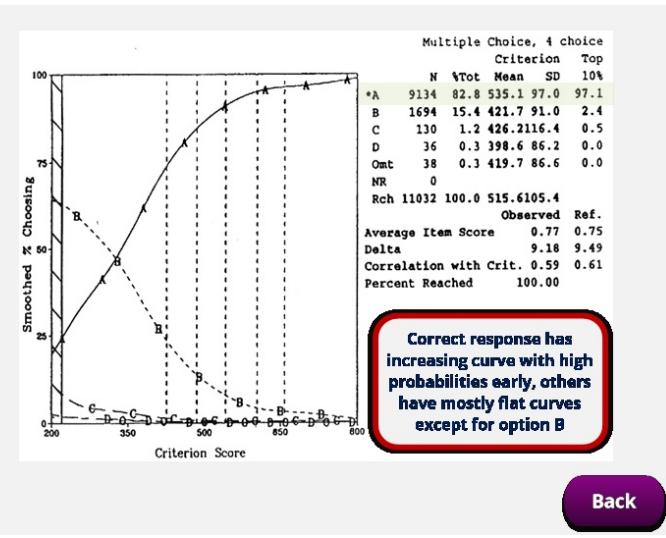
Effective Item (Slide Layer)



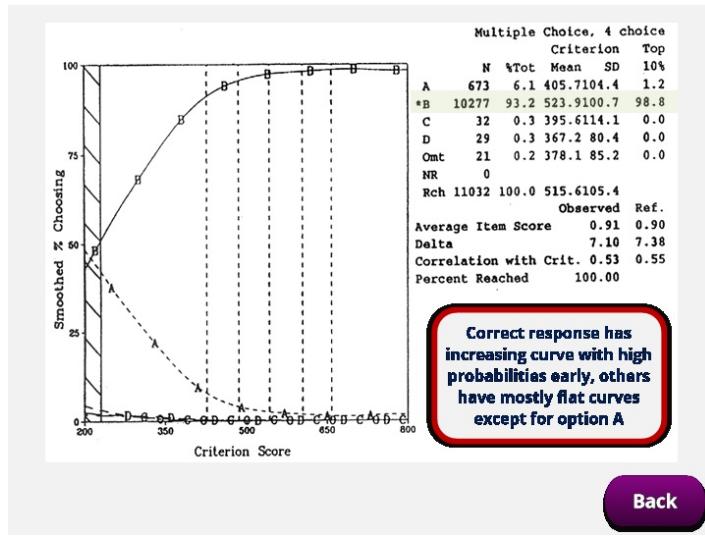
Ineffective Item (Slide Layer)



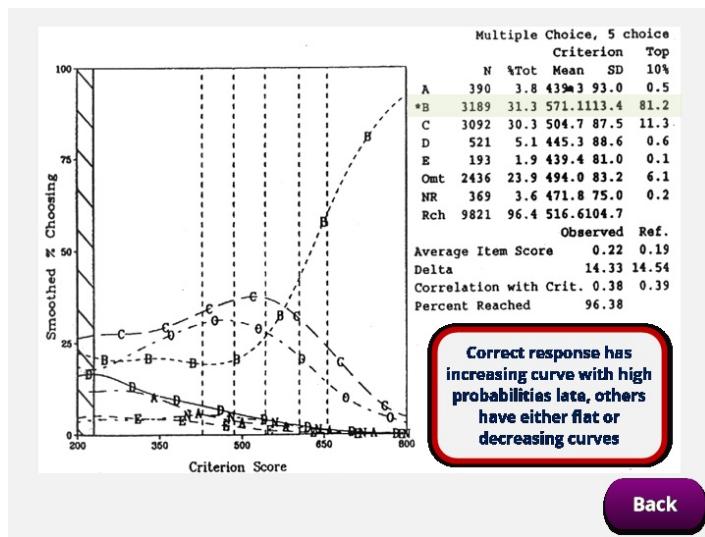
Easy Item (Slide Layer)



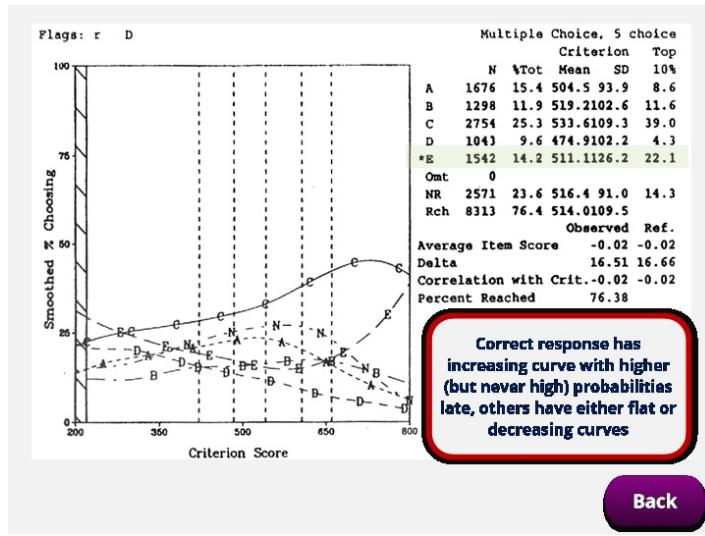
Very Easy Item (Slide Layer)



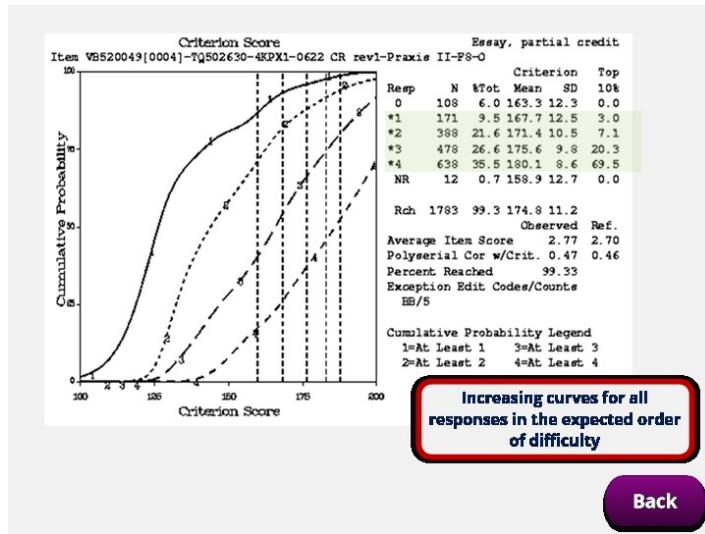
Difficult Item (Slide Layer)



Very Difficult Item (Slide Layer)



Partial Credit Item (Slide Layer)

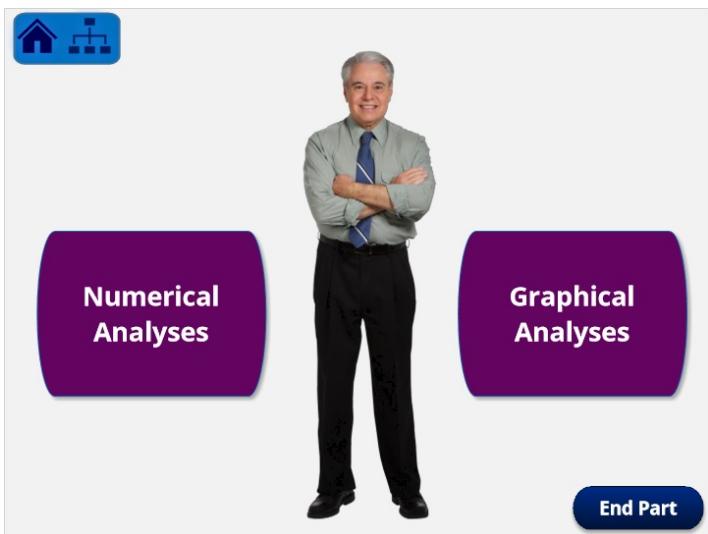


3.49 Bookend: Graphical Analyses

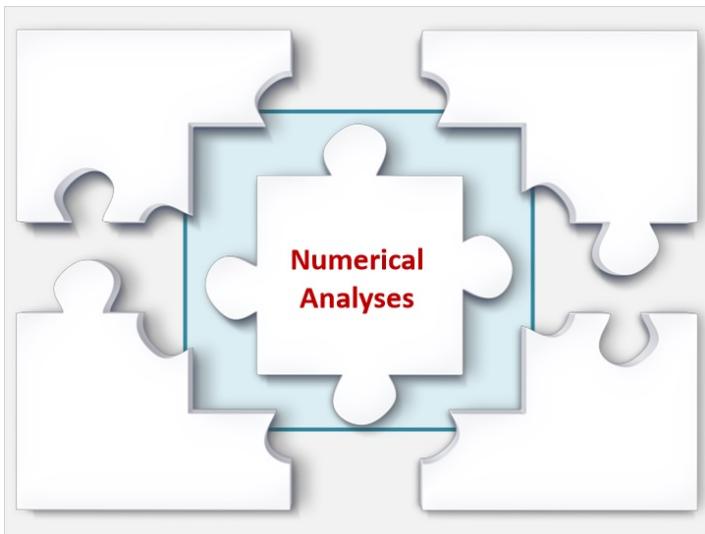


This is the end of this part.

3.50 Topic Selection: Distractors



3.51 Bookmark: Numerical Analyses



3.52 Conceptual Foundations

General Principles

A multiple-choice item may not be serving its intended purpose if:

- The item contains **two correct answers**
- The item contains **no correct answer**
- One or more distractors are **not attracting any learner**
- The item stem provides **clues to the correct answer**
- other reasons

3.53 Conceptual Foundations

Approaches

Comparing learners who choose different options can be done in various ways via the:

- **Mean score on the criterion** (e.g., the total test score) for the group of learners who chose each option (this cannot differentiate the answers by guessing at random)
- **Performance of top 20% and bottom 20% of learners** who chose each option (those should be notably different)
- **Item-total correlation of each option** (only correct answer should have positive value)
- **Response curve for each option** (visual inspection)

3.54 Example 1: Miskeyed Options

Example 1: Miskeyed Options

• **Statistics:**

- Number of examinees = 100 (top 20% and bottom 20%)
- Correct answer = C (supposedly)
- p -value = 0.11 (hard item)
- D index = -0.55 and biserial correlation = -0.24 (negative discrimination)

Group	N	A	B	C*	D
Total	100	15	64	11	10
Top 20%	20	10%	85%	5%	0%
Bottom 20%	20	15%	10%	60%	15%

• **Observations:**

- Only 11 learners chose the correct answer, which translates to 60% of the low-performing group and 5% of the high-performing group
- More than half of examinees - 64 learners - and 85% of the top 20% learners chose distractor B. Thus, distractor B may possibly be the correct answer instead of option C
- Options C and D are not chosen by many learners overall and may not be effective distractors

 This item should either be revised or simply recoded by the design team

3.55 Example 2: Unattractive Distractors

Example 2: Unattractive Distractors

• Statistics:

- Number of learners = 100 (top 20% and bottom 20%)
- Correct answer = B
- p-value = 0.83 (easy item)
- D index = 0.35 and biserial correlation = 0.11 (low discrimination power)

Group	N	A	B*	C	D
Total	100	4	83	13	0
Top 20%	20	0%	95%	5%	0%
Bottom 20%	20	15%	60%	25%	0%

• Observations:

- A total of 83 learners chose the correct answer, which translates into 95% of the high-performing group and 60% of the low-performing group
- Only the low-performing group chose distractor A (not attractive for high performers)
- None of learners chose distractor D (not attractive at all)

 Distractors A and D need to be revised to be more attractive to learners

3.56 Example 3: No Correct Answer

Example 3: No Correct Answer

• Statistics:

- Number of learners = 100 (top 20% and bottom 20%)
- Correct answer = D
- p-value = 0.29 (difficult item)
- D index = 0.05 and biserial correlation = -0.01 (essentially no discrimination)

Group	N	A	B	C	D*
Total	100	21	28	22	29
Top 20%	20	15%	35%	15%	35%
Bottom 20%	20	20%	25%	15%	30%

• Observations:

- A total of 29 learners chose the correct answer, which translates into 35% of the high-performing group and 30% of the low-performing group
- There is no clear performance difference between the high- and low-performing groups across all distractors
- The answer key is not clearly attractive to either the high- or low-performing groups

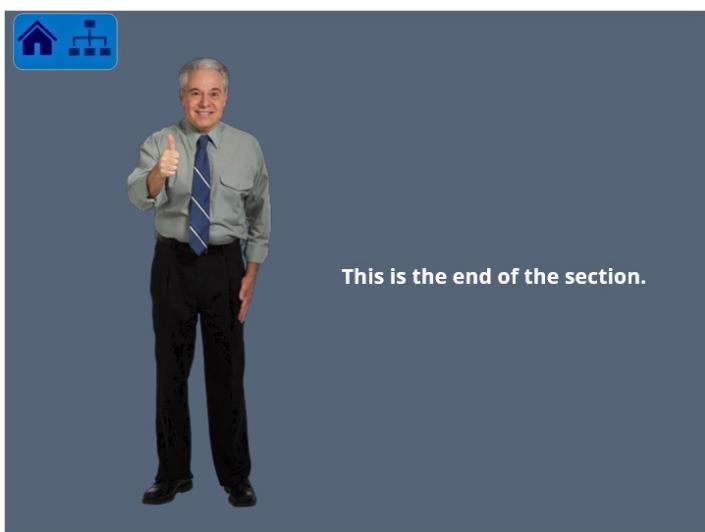
 This item should be reviewed with regards to whether the answer key is correct. The key should also be revised to be more attractive to the high-performing group and less attractive to the low-performing group

3.57 Bookend: Numerical Analyses



This is the end of this part.

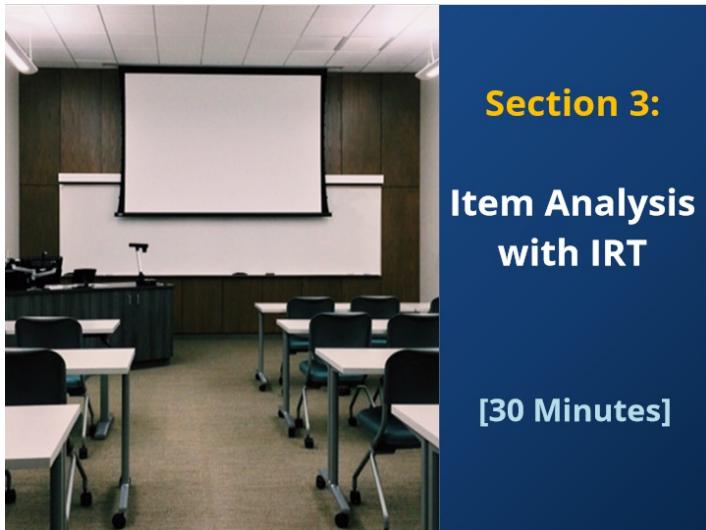
3.58 Bookend: Section 2



This is the end of the section.

4. Section 3: Item Analysis with IRT

4.1 Cover: Section 3



4.2 Objectives: Section 3

This is a screenshot of a learning objectives slide. At the top, there's a dark blue header bar with a house icon, a tree icon, and the text "Learning Objectives". Below the header is a photograph of several wooden blocks with letters on them, spelling out "P", "L", "A", and "N". The slide is divided into four quadrants, each containing a numbered objective. The objectives are:

1. Interpret key item parameters from dichotomous IRT models to judge item characteristics
2. Understand the impact of IRT model fit on the precision and standard error of estimates
3. Describe the key properties of item characteristic curves and item information functions
4. Make decisions about item quality using item characteristic curves and item information functions

4.3 Overview (I)

Overview: IRT Properties

- Item response theory (IRT) is a **statistical / psychometric framework** linking examinee scores and observed responses to items **at the item and test levels**
- Links between proficiencies and item responses are made through **non-linear probability models** with two **strong assumptions**:
 - ✓ **Appropriate dimensionality** (unidimensional vs. multidimensional)
 - ✓ **Local independence** (model accounts for all important relationships)

4.4 Overview (II)

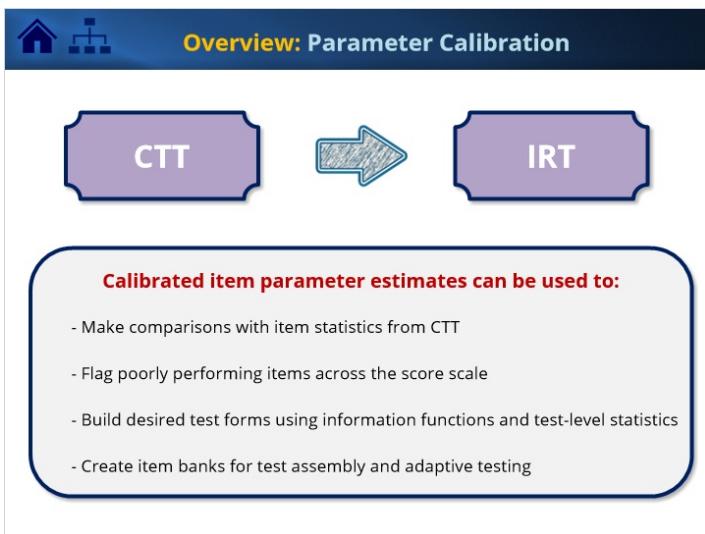
Overview: IRT Properties

Under acceptable fit:

- Item and examinee proficiency parameters are invariant
- Probabilistic relationships between examinees and items exists
- Examinee and item parameters are on a common reporting scale



4.5 Overview (III)



4.6 Topic Selection



4.7 Bookmark: IRT Models



4.8 IRT Models (I)

Three-parameter Model

$$P(u_i = 1|\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

- $P(u_i = 1|\theta)$ is the probability of a correct response to item i as a function of examinee score θ
- a_i is item i 's discrimination parameter
- b_i is item i 's difficulty parameter
- c_i is item i 's pseudo-guessing parameter
- D is a scaling factor (1.702 for the logistic model)

Example

Excel Computation: 3PL (Slide Layer)

IRT Models: Three-parameter Model

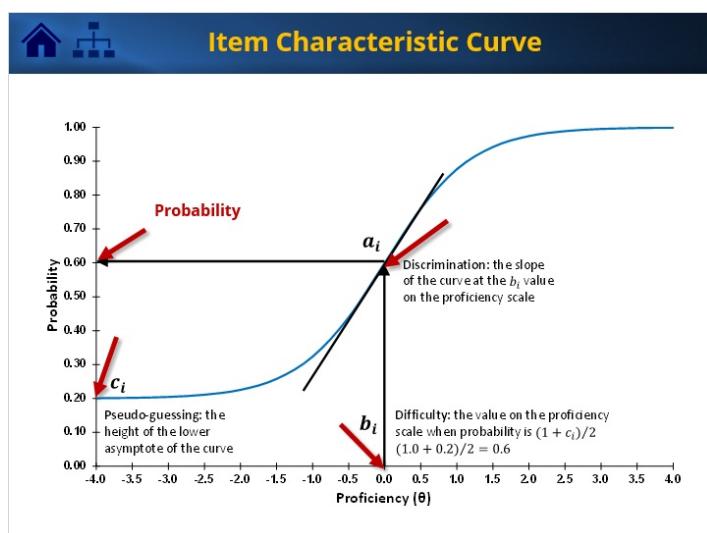
Let's assume $a_i=1.00$, $b_i=0.00$, $c_i=0.20$, and $\theta = 0.00$.

Then the probability to answer this item correct is 0.60:

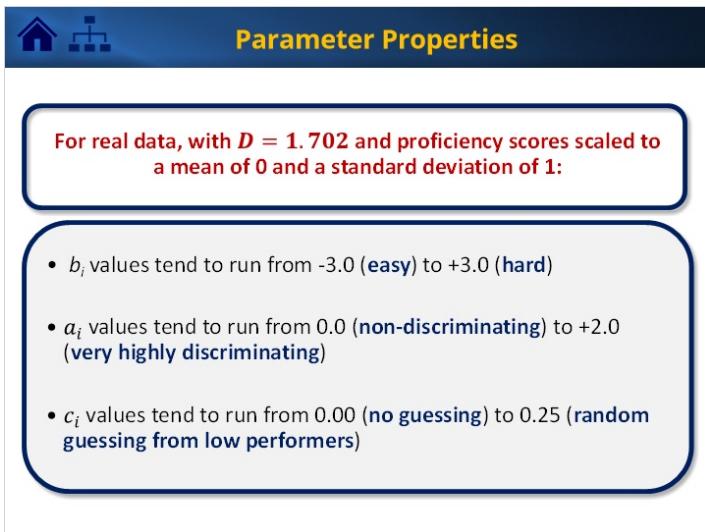
Parameters				
A	B	C	Proficiency	
1.00	0.00	0.20	0.00	
Probability				
0.60	$=D4+((1-D4)*(1/(1+EXP(-1.702*B4*(E4-C4))))$)			

Back

4.9 IRT Models (II)



4.10 IRT Models (III)

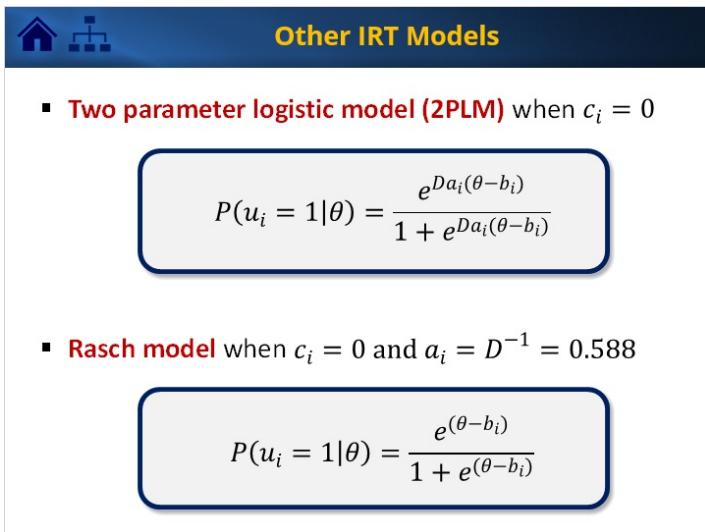


Parameter Properties

For real data, with $D = 1.702$ and proficiency scores scaled to a mean of 0 and a standard deviation of 1:

- b_i values tend to run from -3.0 (**easy**) to +3.0 (**hard**)
- a_i values tend to run from 0.0 (**non-discriminating**) to +2.0 (**very highly discriminating**)
- c_i values tend to run from 0.00 (**no guessing**) to 0.25 (**random guessing from low performers**)

4.11 IRT Models (IV)



Other IRT Models

- **Two parameter logistic model (2PLM)** when $c_i = 0$
$$P(u_i = 1|\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}$$
- **Rasch model** when $c_i = 0$ and $a_i = D^{-1} = 0.588$
$$P(u_i = 1|\theta) = \frac{e^{(\theta-b_i)}}{1 + e^{(\theta-b_i)}}$$

4.12 IRT Models (V)

Polytomous Models

Commonly used models for polytomous data (e.g., 1-2-3-4 from Likert scales or other graded responses):

Graded Response Model (Samejima, 1969)
Generalized Partial Credit Model (Muraki, 1992)
Partial Credit Model (Masters, 1982)
Nominal Response Model (Bock, 1972)
Rating Scale Model (Andrich, 1978)

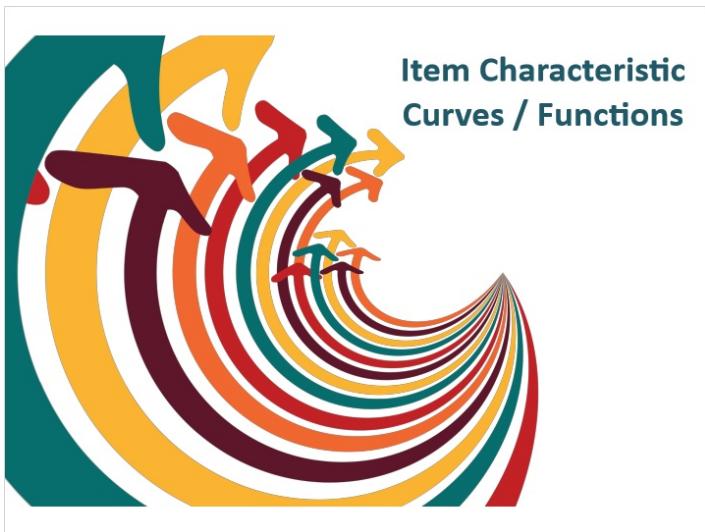
Click below to access the free Wikipedia page and ITEMS module or explore a common reference book on this topic

[Wikipedia](#) [ITEMS Module](#) [Handbook of Modern IRT](#)

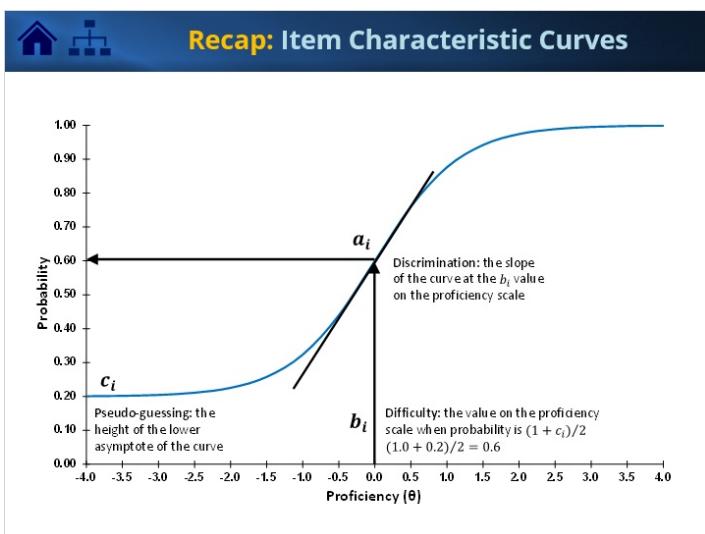
4.13 Bookend: IRT Models



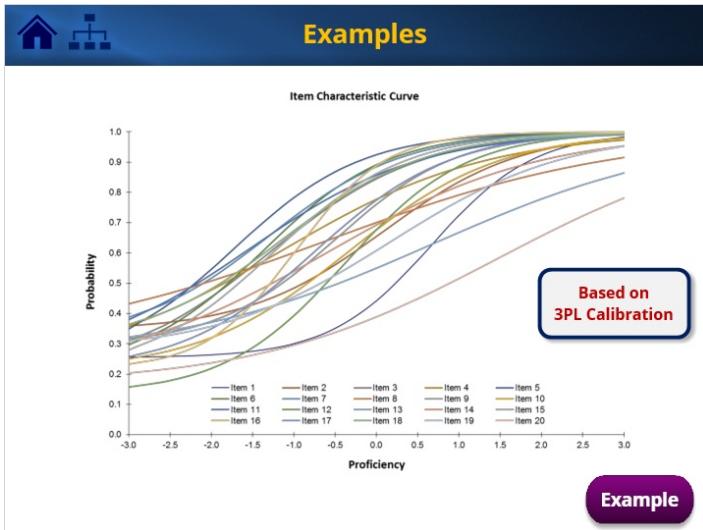
4.14 Bookmark: ICCs



4.15 ICCs (I)



4.16 ICCs (II)

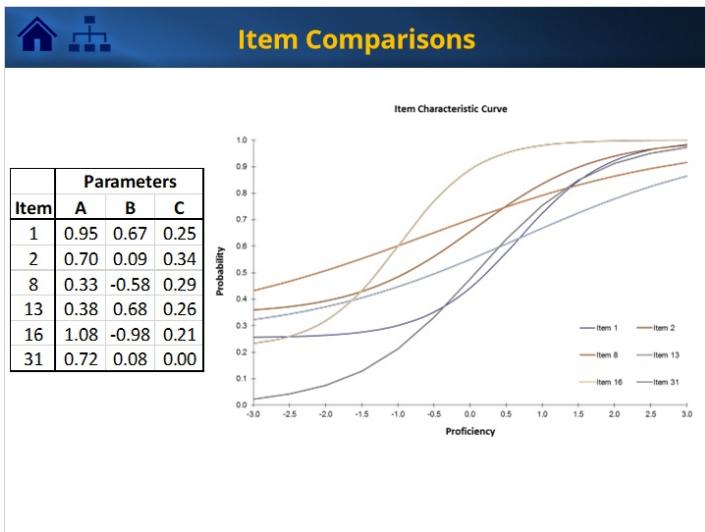


Example (Slide Layer)

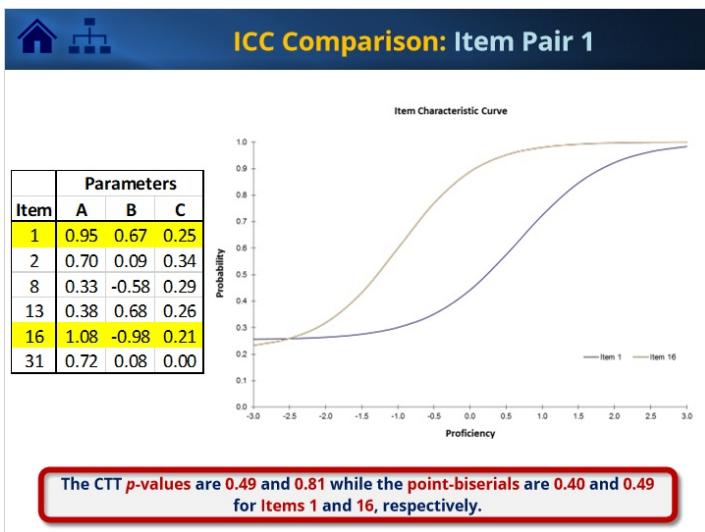
The screenshot shows a Microsoft Excel spreadsheet titled "IRT Models: ICC Examples". The table has columns A-S and rows 1-26. Row 1 contains column headers A through S. Row 2 contains sub-headers for columns A-C and then "Parameters" followed by a grid of 10x10 cells representing "Probability Function". Row 3 contains sub-headers for columns D-S. Rows 4-23 contain data for 20 items, each with columns A-C and then a 10x10 grid of probability values. Row 24 contains a formula "TCC = SUM(A4:S4)" and row 25 contains numerical values for the TCC calculation. A purple button at the bottom right is labeled "Back".

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1																			
2																			
3																			
4	Item	A	B	C	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0	2.5	3.0		
5	1	0.954	0.671	0.254	0.26	0.26	0.26	0.28	0.30	0.35	0.44	0.58	0.72	0.85	0.92	0.96	0.98		
6	2	0.703	0.086	0.343	0.36	0.37	0.39	0.43	0.48	0.56	0.65	0.75	0.83	0.90	0.94	0.97	0.98		
7	3	0.645	-1.296	0.203	0.31	0.37	0.45	0.56	0.67	0.77	0.84	0.90	0.94	0.96	0.98	0.99	0.99		
8	4	0.473	-1.155	0.213	0.36	0.41	0.48	0.55	0.63	0.71	0.78	0.84	0.88	0.92	0.94	0.96	0.97		
9	5	0.753	-1.744	0.217	0.35	0.43	0.55	0.67	0.78	0.87	0.92	0.96	0.98	0.99	1.00	1.00			
10	6	0.784	-1.368	0.217	0.30	0.38	0.45	0.57	0.70	0.81	0.89	0.94	0.97	0.98	0.99	1.00	1.00		
11	7	0.724	-1.303	0.317	0.39	0.46	0.52	0.61	0.72	0.81	0.88	0.93	0.96	0.98	0.99	0.99	1.00		
12	8	0.130	-0.578	0.280	0.43	0.47	0.51	0.55	0.60	0.65	0.70	0.75	0.79	0.83	0.86	0.89	0.92		
13	9	0.803	-0.487	0.295	0.32	0.38	0.37	0.44	0.53	0.64	0.76	0.85	0.92	0.96	0.98	0.99	0.99		
14	10	0.641	-0.338	0.209	0.25	0.28	0.32	0.38	0.47	0.57	0.68	0.77	0.85	0.91	0.94	0.97	0.98		
15	11	0.569	-1.530	0.237	0.38	0.45	0.53	0.62	0.71	0.79	0.86	0.91	0.94	0.96	0.98	0.98	0.99		
16	12	0.806	-0.378	0.134	0.16	0.18	0.22	0.29	0.39	0.53	0.68	0.80	0.89	0.94	0.97	0.98	0.99		
17	13	0.379	0.679	0.259	0.32	0.34	0.37	0.41	0.45	0.50	0.55	0.61	0.67	0.73	0.78	0.82	0.86		
18	14	0.455	-0.616	0.197	0.31	0.35	0.40	0.47	0.54	0.62	0.69	0.76	0.82	0.87	0.91	0.93	0.95		
19	15	0.732	-1.263	0.197	0.26	0.34	0.43	0.54	0.66	0.78	0.86	0.92	0.95	0.98	0.99	0.99	1.00		
20	16	1.081	-0.979	0.214	0.23	0.26	0.32	0.43	0.60	0.77	0.89	0.95	0.98	0.99	1.00	1.00	1.00		
21	17	0.724	-0.745	0.211	0.26	0.29	0.35	0.43	0.54	0.66	0.77	0.86	0.92	0.95	0.97	0.99	0.99		
22	18	0.680	-1.137	0.290	0.36	0.41	0.48	0.57	0.67	0.77	0.85	0.91	0.94	0.97	0.98	0.99	0.99		
23	19	0.554	0.174	0.275	0.31	0.33	0.36	0.40	0.46	0.53	0.61	0.69	0.77	0.84	0.89	0.93	0.95		
24	20	0.401	1.480	0.166	0.20	0.22	0.24	0.26	0.30	0.34	0.39	0.45	0.52	0.59	0.66	0.72	0.78		
25	TCC	6.13	6.90	8.00	9.46	11.20	13.02	14.70	16.13	17.25	18.07	18.65	19.05	19.33					

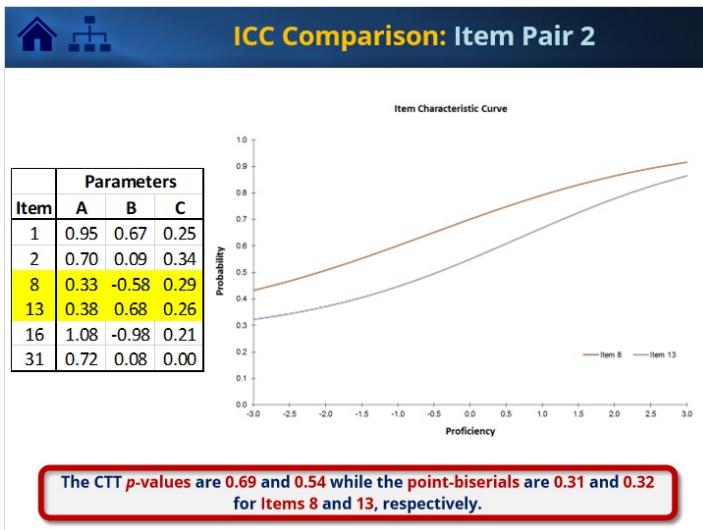
4.17 ICCs (III)



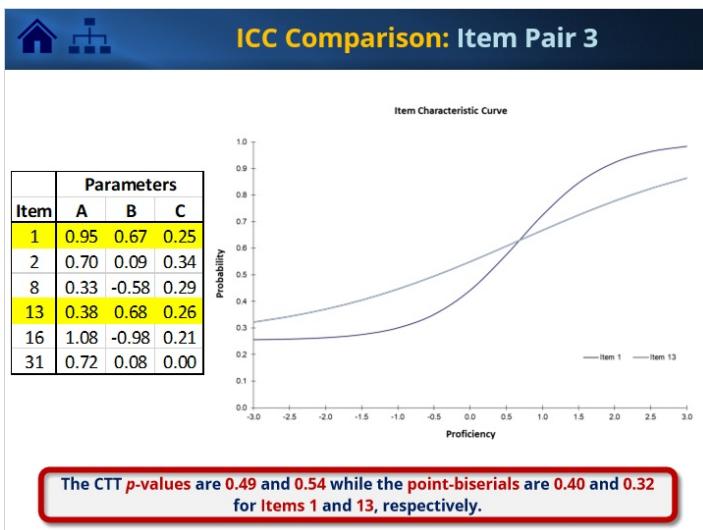
4.18 ICCs (IV)



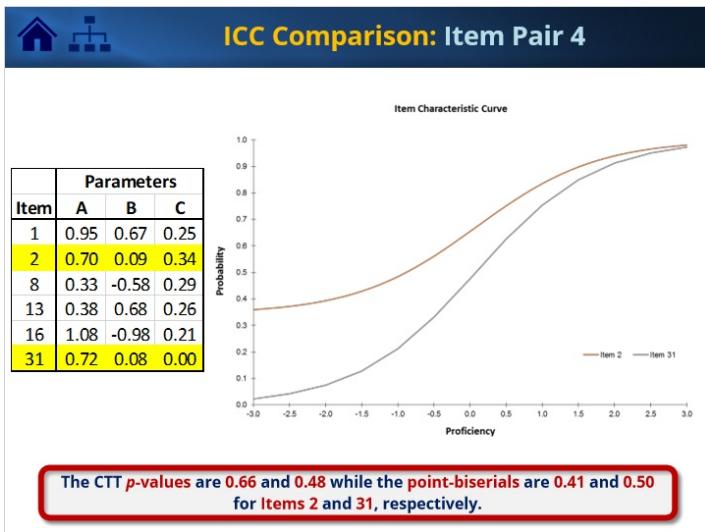
4.19 ICCs (V)



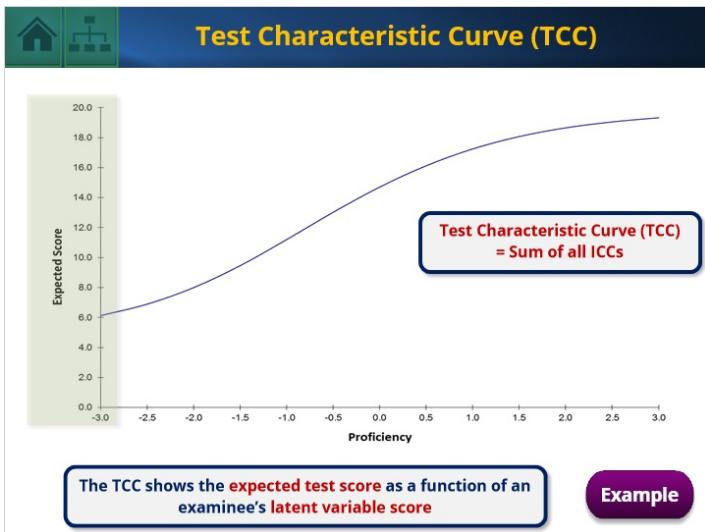
4.20 ICCs (V)



4.21 ICCs (VI)



4.22 ICCs (VII)



Example (Slide Layer)

 Example

Item	Parameters			Probability Function												
	A	B	C	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0	2.5	3.0
1	0.954	0.671	0.254	0.26	0.26	0.26	0.28	0.30	0.35	0.44	0.58	0.72	0.85	0.92	0.96	0.98
2	0.703	0.086	0.343	0.36	0.37	0.39	0.43	0.48	0.56	0.65	0.75	0.83	0.90	0.94	0.97	0.98
3	0.645	-1.296	0.203	0.31	0.37	0.45	0.56	0.67	0.77	0.84	0.90	0.94	0.96	0.98	0.99	0.99
4	0.473	-1.155	0.213	0.36	0.41	0.48	0.55	0.63	0.71	0.78	0.84	0.88	0.92	0.94	0.96	0.97
5	0.753	-1.744	0.219	0.35	0.43	0.55	0.67	0.78	0.87	0.92	0.96	0.98	0.99	0.99	1.00	1.00
6	0.784	-1.368	0.217	0.36	0.38	0.45	0.57	0.70	0.81	0.89	0.94	0.97	0.98	0.99	1.00	1.00
7	0.724	-1.303	0.312	0.39	0.44	0.52	0.61	0.72	0.81	0.88	0.93	0.96	0.98	0.99	0.99	1.00
8	0.330	-0.578	0.286	0.43	0.47	0.51	0.55	0.60	0.65	0.70	0.75	0.79	0.83	0.86	0.89	0.92
9	0.803	-0.487	0.295	0.32	0.34	0.37	0.44	0.53	0.64	0.76	0.85	0.92	0.96	0.98	0.99	0.99
10	0.641	-0.338	0.209	0.25	0.28	0.32	0.38	0.47	0.57	0.68	0.77	0.85	0.91	0.94	0.97	0.98
11	0.569	-1.530	0.230	0.38	0.45	0.53	0.62	0.71	0.79	0.86	0.91	0.94	0.96	0.98	0.98	0.99
12	0.806	-0.378	0.134	0.16	0.18	0.22	0.29	0.39	0.53	0.68	0.80	0.89	0.94	0.97	0.98	0.99
13	0.379	0.679	0.259	0.32	0.34	0.37	0.41	0.45	0.50	0.55	0.61	0.67	0.73	0.78	0.82	0.86
14	0.455	-0.616	0.197	0.31	0.35	0.40	0.47	0.54	0.62	0.69	0.76	0.82	0.87	0.91	0.93	0.95
15	0.732	-1.263	0.197	0.26	0.34	0.43	0.54	0.66	0.78	0.86	0.92	0.95	0.98	0.99	1.00	1.00
16	1.081	-0.979	0.214	0.23	0.26	0.32	0.43	0.60	0.77	0.89	0.95	0.98	0.99	1.00	1.00	1.00
17	0.724	-0.745	0.211	0.26	0.29	0.35	0.43	0.54	0.66	0.77	0.86	0.92	0.95	0.97	0.99	0.99
18	0.680	-1.137	0.290	0.36	0.41	0.48	0.57	0.67	0.77	0.85	0.91	0.94	0.97	0.98	0.99	0.99
19	0.554	0.174	0.275	0.31	0.33	0.36	0.40	0.46	0.53	0.61	0.69	0.77	0.84	0.89	0.93	0.93
20	0.401	1.480	0.166	0.20	0.22	0.24	0.26	0.30	0.34	0.39	0.45	0.52	0.59	0.66	0.72	0.78
25	ICC															
26																

Back

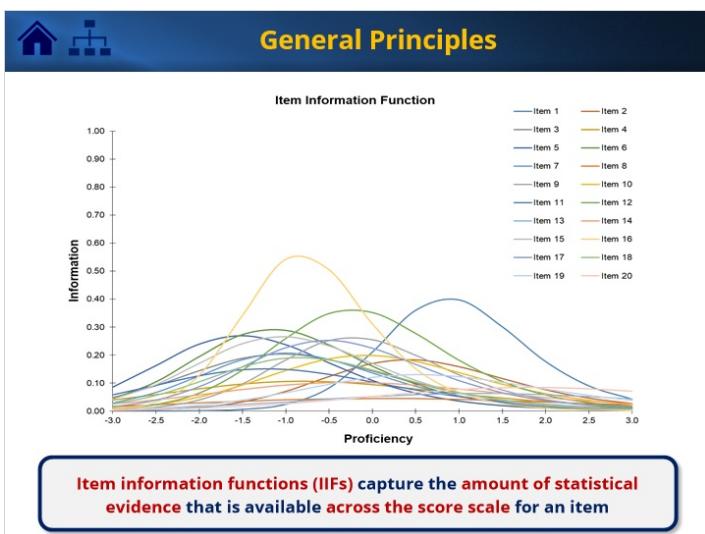
4.23 Bookend: ICCs



4.24 Bookmark: IIFs



4.25 IIFs (I)



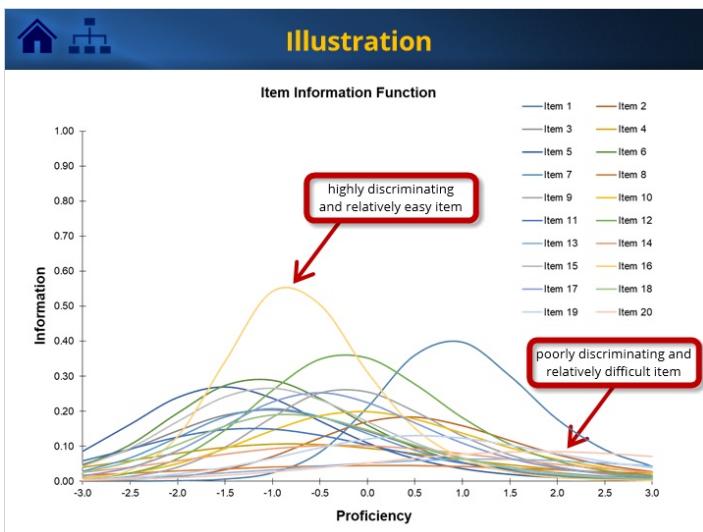
4.26 IIFs (II)

Key Properties

- IIFs are **more peaked** when items **discriminate highly**.
- IIFs are **generally lower** when **pseudo-guessing** is present.
- IIFs peak at the **lower end** of the score scale when items are **easy**
- IIFs peak at the **higher end** of the score scale when items are **difficult**.



4.27 IIFs (III)



4.28 IIfs (V)

Formula and Example

$$I_i(\theta) = \frac{2.89a_i^2(1 - c_i)}{[c_i + e^{Da_i(\theta - b_i)}][1 + e^{-Da_i(\theta - b_i)}]}$$

- a_i , b_i , and c_i imply discrimination, difficulty, and pseudo-guessing parameter of item i
- D is a scaling factor as 1.702

A	B	C	D	E	F	G	H	I	J	K	L
1											
2											
3											
4											
5											
6											
7											

Parameters				I(θ)
a	b	c	θ	
1.00	0.00	0.20	0.00	0.48
$=((1.7^2)*(B4^2)*(1-D4))/((D4*(2.71828^(1.7*(B4^(S54-C4)))))*(1+(2.71828^(-1*1.7*(B4^(S54-C4)))))^2)$				

Maximum Information

Maximum Info (Slide Layer)

Maximum Item Information

$$\theta_{max} = b_i + \frac{1}{Da_i} \ln[0.5(1 + \sqrt{1 + 8c_i})]$$

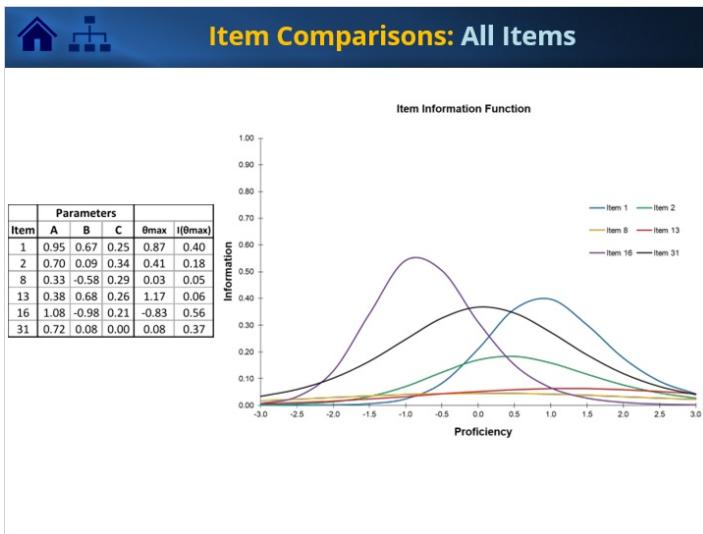
- a_i , b_i , and c_i imply discrimination, difficulty, and pseudo-guessing parameter of item i
- D is a scaling factor as 1.702

A	B	C	D	E	F	G	H	I	J	K	L	M	N
1													
2													
3													
4													
5													
6													
7													
8													

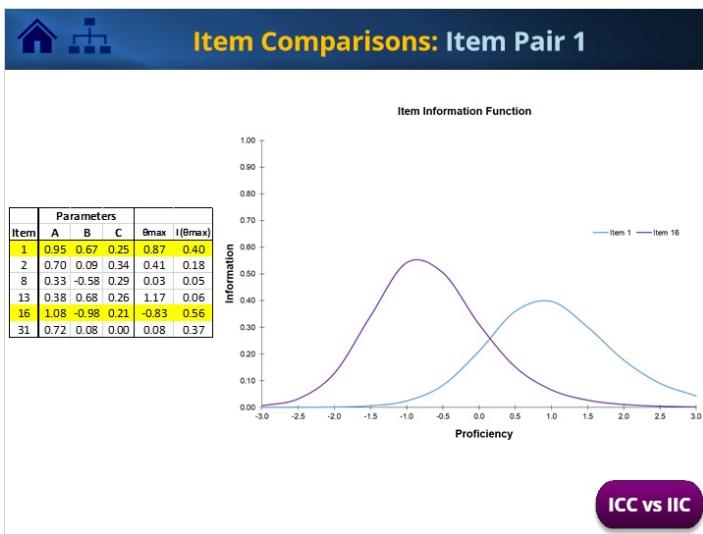
Parameters		
a	b	c
1.000	0.000	0.200
θ_{max} 0.157 $I(\theta_{max})$ 0.492		
$C4+((1/(1.7*B4))*LN(0.5*(1+SQRT(1+B4*D4))))$ $=((1.7^2)*(B4^2)*(1-D4))/((D4*(2.71828^(1.7*(B4^(S56-C4)))))*(1+(2.71828^(-1*1.7*(B4^(S56-C4)))))^2)$		

Back

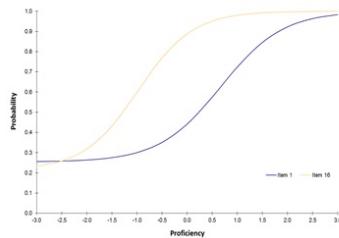
4.29 IIFs (VI)



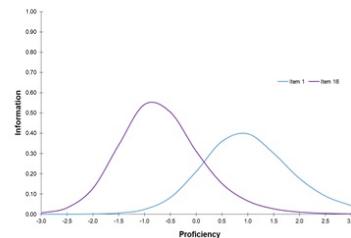
4.30 IIFs (VII)



ICC vs IIF (Slide Layer)



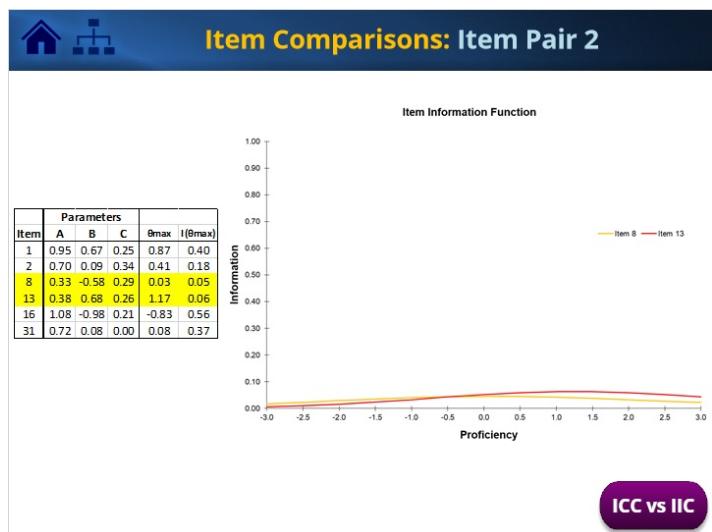
Item Characteristic Curve



Item Information Curve

Back

4.31 IIFs (VIII)



ICCs vs IIFs (Slide Layer)

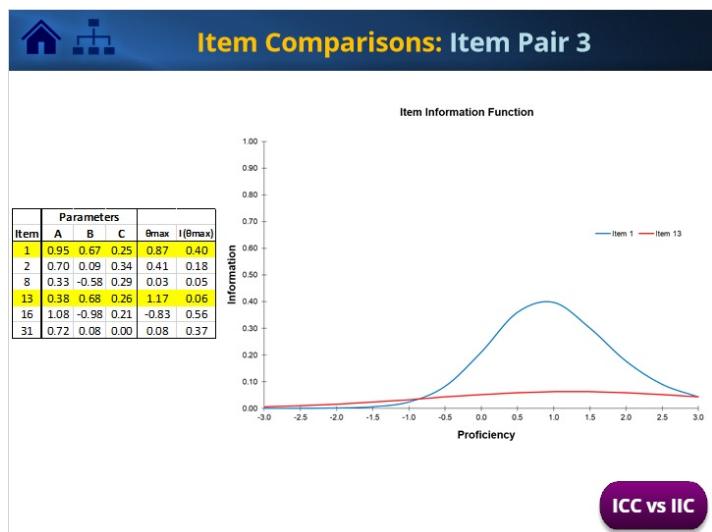


Item Characteristic Curve

Item Information Curve

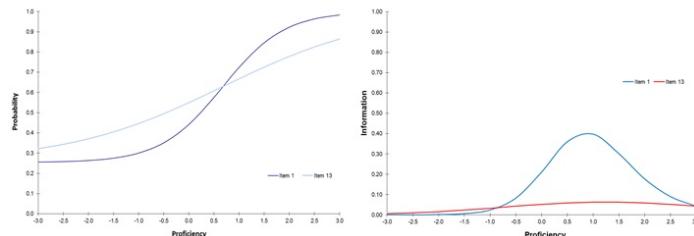
Back

4.32 IIFs (VII)



ICC vs IIC

ICC vs IIF (Slide Layer)

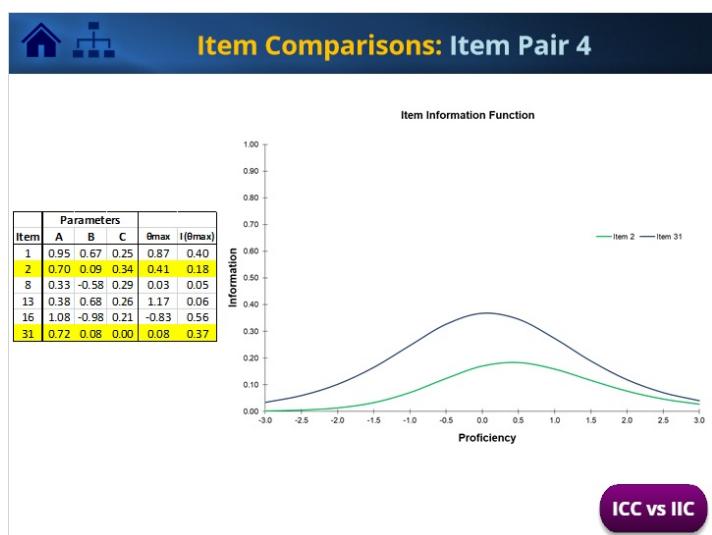


Item Characteristic Curve

Item Information Curve

Back

4.33 IIFs (IX)



ICCs vs IIFs (Slide Layer)

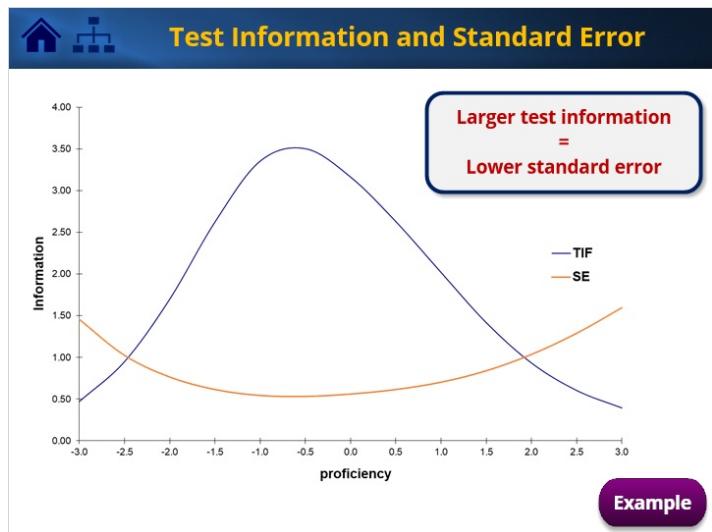


Item Characteristic Curve

Item Information Curve

Back

4.34 IIFs (XI)



Example (Slide Layer)

Example

Item	Parameters			Information Function												θ_{\max}	θ_{\min}	
	A	B	C	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0	2.5			3.0
1	0.95	0.67	0.25	0.00	0.00	0.00	0.01	0.02	0.08	0.23	0.36	0.40	0.40	0.18	0.09	0.04	0.84	0.40
2	0.70	0.09	0.34	0.00	0.00	0.01	0.03	0.07	0.12	0.17	0.18	0.16	0.12	0.08	0.05	0.03	0.45	0.18
3	0.65	-2.30	0.20	0.05	0.09	0.14	0.19	0.20	0.18	0.14	0.10	0.07	0.04	0.02	0.01	0.01	-1.05	0.20
4	0.47	-1.16	0.21	0.04	0.06	0.08	0.10	0.11	0.11	0.10	0.08	0.06	0.05	0.03	0.02	0.03	0.83	0.11
5	0.75	-2.74	0.22	0.06	0.16	0.24	0.27	0.24	0.17	0.11	0.06	0.04	0.02	0.01	0.00	0.00	-1.52	0.11
6	0.78	-1.37	0.22	0.02	0.10	0.19	0.27	0.29	0.24	0.16	0.10	0.05	0.03	0.02	0.01	0.00	1.16	0.29
7	0.72	-2.30	0.31	0.03	0.07	0.13	0.18	0.21	0.18	0.14	0.09	0.05	0.03	0.02	0.01	0.01	-1.03	0.21
8	0.33	0.58	0.29	0.07	0.01	0.03	0.04	0.04	0.04	0.05	0.04	0.04	0.03	0.03	0.07	0.03	0.05	
9	0.80	-0.49	0.30	0.00	0.01	0.04	0.10	0.18	0.25	0.26	0.20	0.12	0.08	0.04	0.02	0.01	0.22	0.26
10	0.64	-0.34	0.21	0.01	0.00	0.05	0.09	0.14	0.19	0.20	0.18	0.14	0.10	0.06	0.04	0.07	0.09	0.20
11	0.57	-1.53	0.23	0.06	0.09	0.13	0.15	0.15	0.13	0.10	0.08	0.05	0.03	0.02	0.01	0.01	-1.23	0.15
12	0.81	-0.38	0.13	0.00	0.03	0.06	0.15	0.36	0.35	0.38	0.18	0.11	0.06	0.08	0.03	0.23	0.36	
13	0.38	0.68	0.26	0.01	0.01	0.02	0.02	0.03	0.04	0.05	0.06	0.06	0.06	0.05	0.04	0.17	0.06	
14	0.46	-0.67	0.20	0.03	0.04	0.06	0.08	0.09	0.10	0.10	0.09	0.08	0.06	0.05	0.04	0.03	-0.77	0.10
15	0.73	-1.26	0.20	0.04	0.09	0.17	0.24	0.27	0.23	0.17	0.11	0.07	0.04	0.02	0.01	0.01	1.05	0.27
16	1.08	-0.98	0.71	0.01	0.00	0.13	0.34	0.54	0.51	0.51	0.51	0.15	0.07	0.03	0.01	0.00	0.00	0.56
17	0.72	0.75	0.21	0.02	0.04	0.09	0.16	0.23	0.25	0.22	0.17	0.11	0.07	0.04	0.02	0.01	0.52	0.25
18	0.68	-1.14	0.29	0.03	0.00	0.10	0.16	0.19	0.18	0.15	0.10	0.07	0.04	0.02	0.01	0.01	-0.84	0.19
19	0.55	0.17	0.28	0.00	0.01	0.02	0.04	0.07	0.14	0.12	0.13	0.12	0.10	0.08	0.06	0.04	0.53	0.13
20	0.40	1.48	0.17	0.00	0.01	0.02	0.03	0.04	0.05	0.07	0.08	0.08	0.08	0.08	0.07	0.07	1.82	0.08
25	TIF	0.47	0.95	1.70	2.02	3.85	3.50	3.46	2.83	2.02	1.42	0.93	0.60	0.39				
26	SE	1.46	1.03	0.77	0.82	0.55	0.53	0.56	0.62	0.70	0.84	1.04	1.29	1.59				

Back

4.35 IIFs (XII)

Uses of Information Functions

Building a test to meet the desired statistical specifications

- Chapter 7: Test Construction
(Hambleton, Swaminathan, & Rogers, 1990)



Revising an existing test

- Chapter 7: Test Construction
(Hambleton, Swaminathan, & Rogers, 1990)



4.36 IIFs (XII)

 **Uses of Information Functions**

Comparing two tests (Relative Efficiency)

- Chapter 6: The Relative Efficiency of Two Tests
(Lord, 1980)



Providing standard errors of ability estimation

- Impact of item parameter estimation errors on test development
(Hambleton & Jones, 1994; Hambleton, Jones, & Rogers, 1993)



Comparing the scoring methods

- Chapter 3: IRT for Items Scored in Two Categories
(Thissen & Wainer, 2001)



4.37 Bookend: IIFs





This is the end of this part.

4.38 Bookmark: Considerations



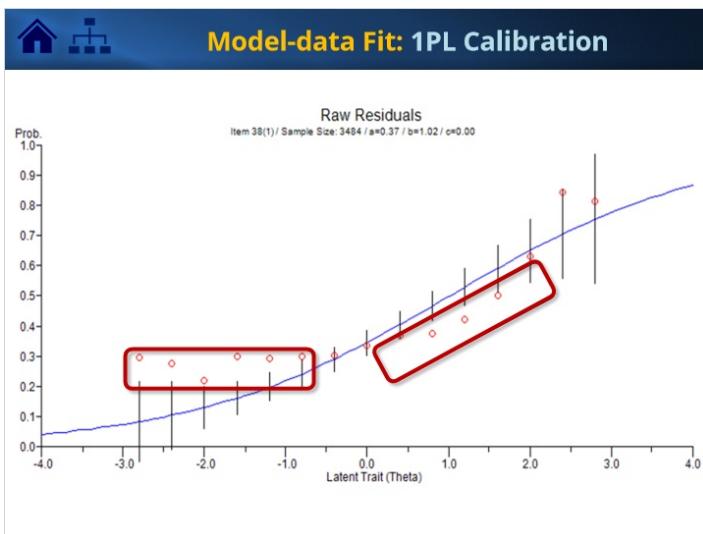
4.39 Considerations (I)

Model-data Fit

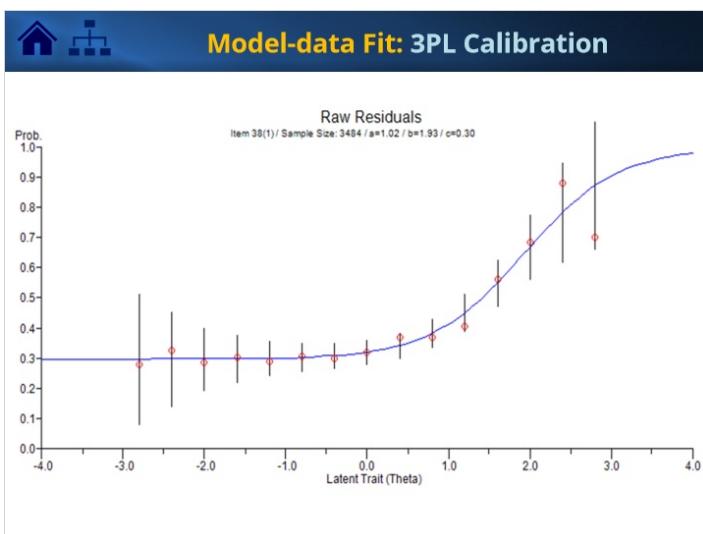
Three types of statistical evidence to support the use of an IRT model

- Determine if the model assumptions are met to a reasonable degree by the test data (e.g., unidimensionality, local independence)
- Determine the closeness between predictions from the IRT model and actual test results (e.g., Q_1 , residuals, test score distributions)
- Address practical consequences of the model misfit in relation to intended purpose such as equating and proficiency estimation.

4.40 Considerations (II)



4.41 Considerations (III)



4.42 Considerations (IV)

Model-data Fit: Convergence Issues

- 1** Unstable guessing parameter estimates for 3PL items
- 2** Items with large proportions of omit and/or not reached
- 3** Many examinees with perfect (right or wrong) scores (MLE)
- 4** Violating local independence assumption

4.43 Considerations (V)

Model-data Fit: Post-hoc Adjustments

- ✓ **Removing** problematic items via CTT item analysis
- ✓ **Providing** reasonable item prior or starting values
- ✓ **Fixing** guessing parameters for multiple-choice items



4.44 Bookend: Considerations



This is the end of this part.

4.45 Module Cover (END)

