

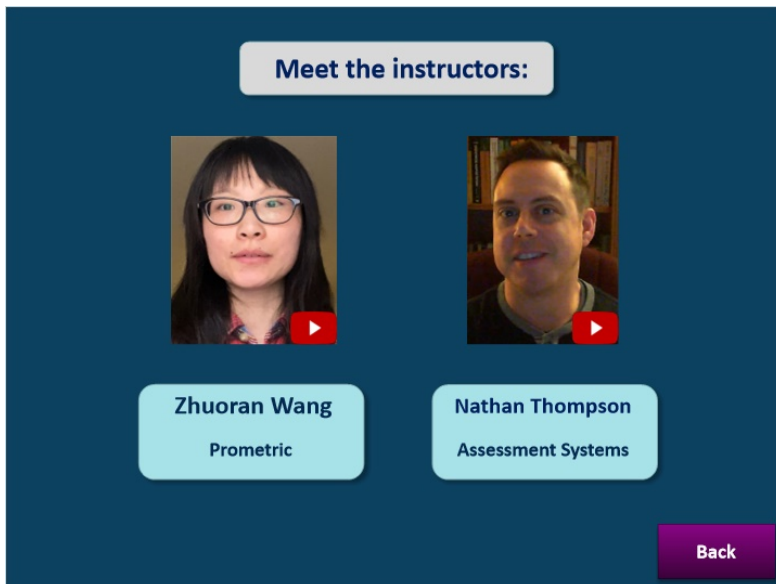
DM19 SLIDES (IRT Estimation, Version 1.0)

1. Module Overview

1.1 Module Cover (START)



1.2 Instructors



1.3 Designers

Meet the designers:


Jon Lehrfeld
ETS


André A. Rupp
Mindful Measurement

Back

1.4 Welcome



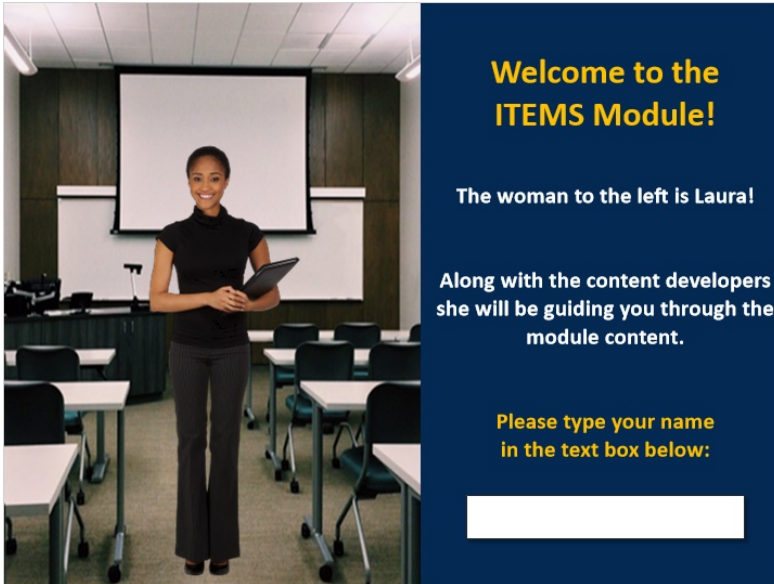
**Welcome to the
ITEMS Module!**

The woman to the left is Laura!

Along with the content developers
she will be guiding you through the
module content.

Please type your name
in the text box below:

Untitled Layer 1 (Slide Layer)



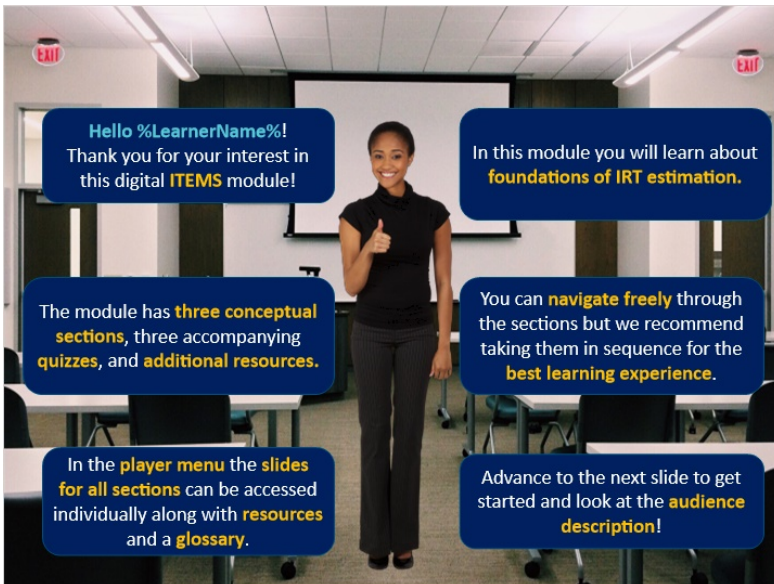
Welcome to the ITEMS Module!

The woman to the left is Laura!

Along with the content developers she will be guiding you through the module content.

Please type your name in the text box below:

1.5 Overview



Hello %LearnerName%!
Thank you for your interest in this digital **ITEMS** module!

The module has **three conceptual sections**, three accompanying **quizzes**, and **additional resources**.

In the **player menu** the **slides for all sections** can be accessed individually along with **resources** and a **glossary**.

In this module you will learn about **foundations of IRT estimation**.

You can **navigate freely** through the sections but we recommend taking them in sequence for the **best learning experience**.

Advance to the next slide to get started and look at the **audience description!**

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



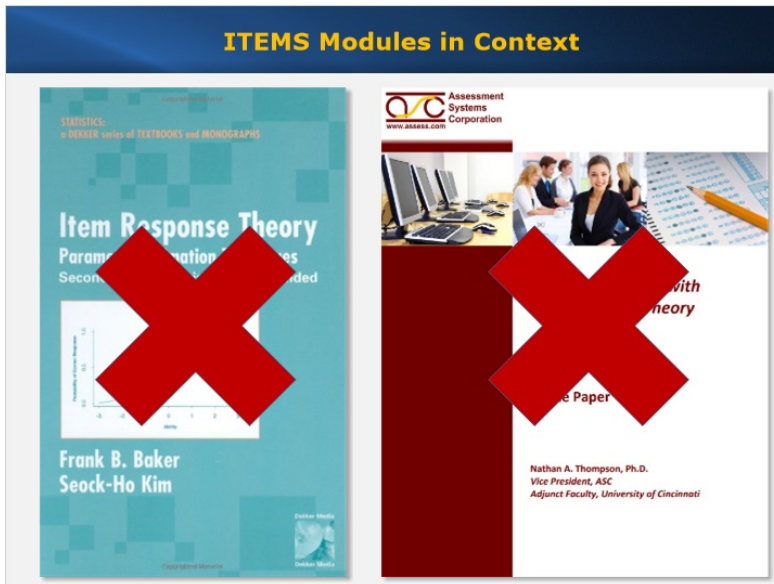
However, we hope that you find the information in this module useful no matter what your official title or role in an organization is!

1.7 Expectations (I)



Let's discuss expectations....

1.8 Expectations (II)



1.9 Learning Objectives

Learning Objectives



1. Demonstrate a preliminary understanding of item calibration and examinee scoring algorithms
2. Conduct item calibration and examinee scoring using R
3. Read and interpret item calibration and examinee scoring results
4. Evaluate estimation accuracy
5. Identify the potential need to increase estimation accuracy
6. Use multiple strategies to facilitate estimation

1.10 Prerequisites

Prerequisites

- **Foundational IRT concepts:**
 - ✓ Item response functions for 1/2/3PL models
 - ✓ Goal and logic of item calibration and examinee scoring
 - ✓ Parameter recovery using response matrix
- **Probability**
 - ✓ Prior and posterior distributions
- **R statistical software**
 - ✓ Useful for Section 3: Factors Affecting Estimation Accuracy

1.11 Resources

Resources

Module Citation

Wang, Z., & Thompson, N. (2020). Foundations of IRT estimation (Digital ITEMS Module 19). *Educational Measurement: Issues and Practice*, 39(4), 134-135.



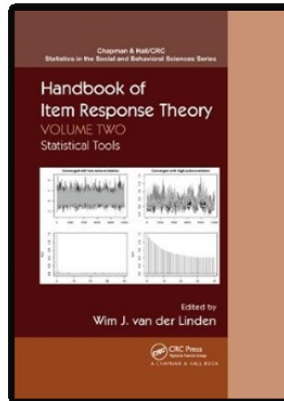
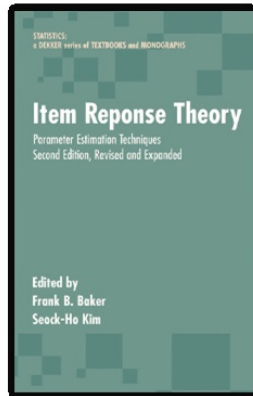
Sample R Code



Additional Resources

References (Slide Layer)

Resources



Click on the images to go to the publisher websites.



Back

1.12 Main Menu

Main Menu

Theory

- 01 IRT Basics [10 Minutes]
- 02 Calibration and Scoring [15 Minutes]
- 03 Factors Affecting Estimation Accuracy [20 Minutes]

Practice

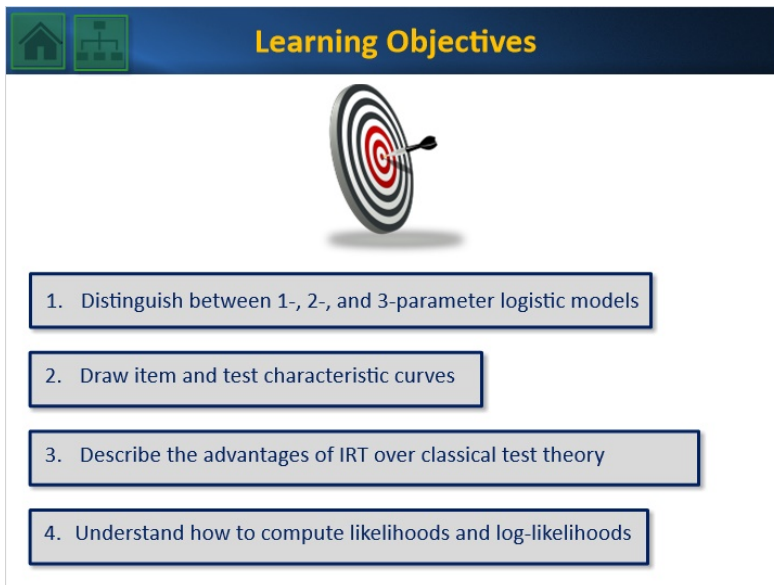
- 04 Example using [10 Minutes] [external website]
- 05 Quizzes [10 Minutes]

2. Section 1: IRT Basics

2.1 Cover: IRT Basics



2.2 IRT Basics: Learning Objectives

The slide has a dark blue header with a small icon of a house with a tree on the left and the text "Learning Objectives" in yellow. Below the header is a large target icon with an arrow in the bullseye. Below the target are four numbered learning objectives, each in a light blue box with a dark blue border.

1. Distinguish between 1-, 2-, and 3-parameter logistic models
2. Draw item and test characteristic curves
3. Describe the advantages of IRT over classical test theory
4. Understand how to compute likelihoods and log-likelihoods

2.3 Item Response Theory



Item Response Theory

- ☐ Theoretical approach of item response theory (IRT)
 - Relates test performance to examinee ability levels
 - Focuses on item-level performance
 - Models the performance of examinees at each ability level to each item on the test

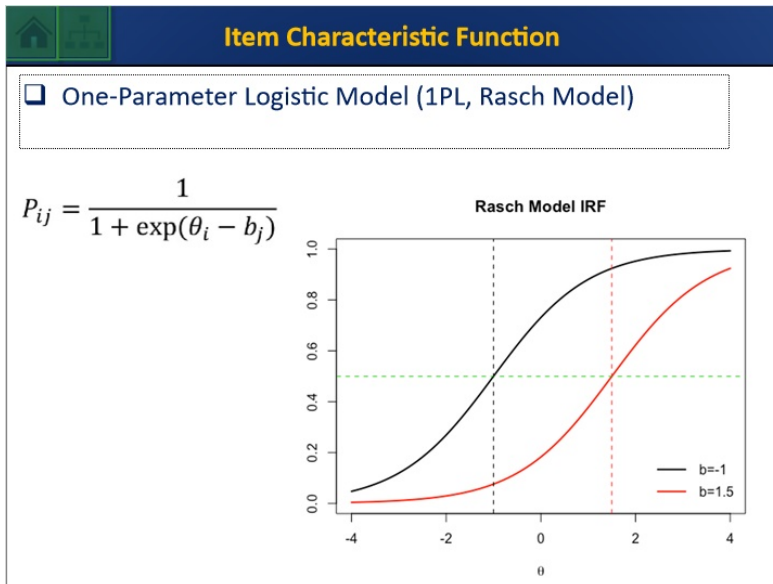
2.4 Advantages of IRT



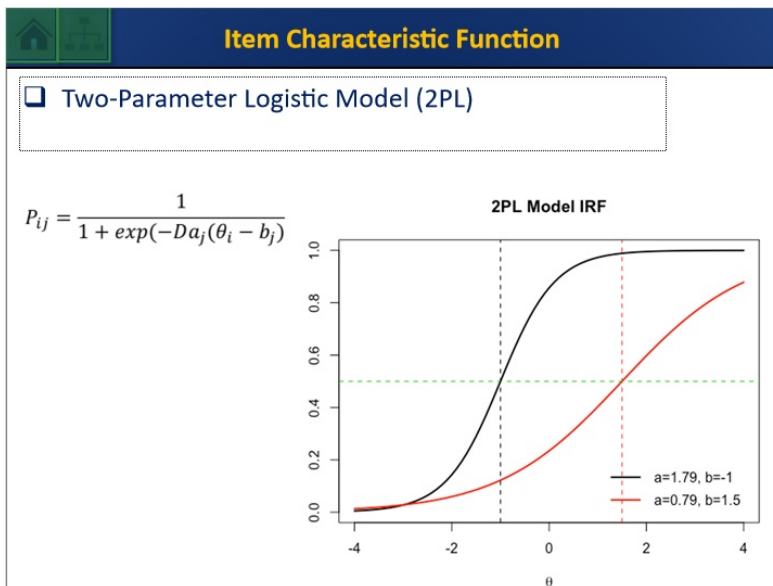
Advantages of IRT

- ☐ IRT overcomes certain issues in classical test theory (CTT)
 - Examinees and items are on the same scale (independent of sample)
 - Once linking of item parameters takes place, equating automatically occurs
 - Individualized standard error of measurement
 - Base of advanced testing schemes (e.g., computerized adaptive testing, multistage testing)

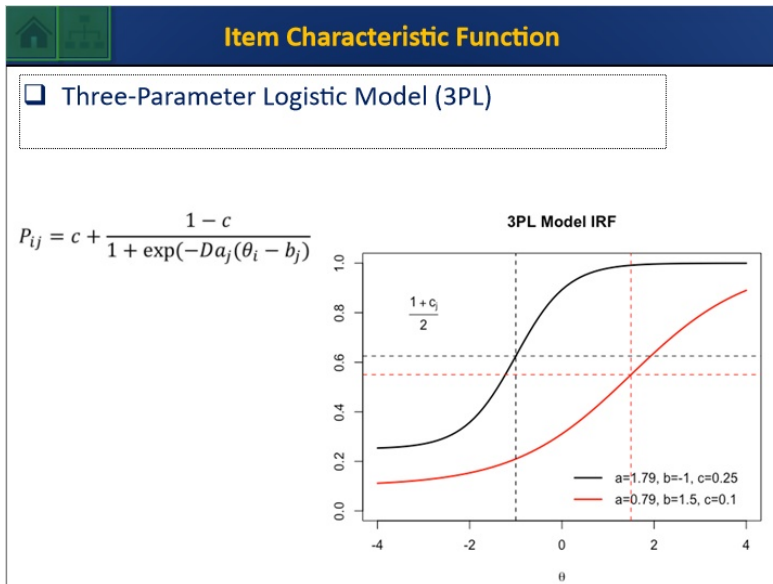
2.5 Item Characteristic Functions I



2.6 Item Characteristic Functions II



2.7 Item Characteristic Functions III



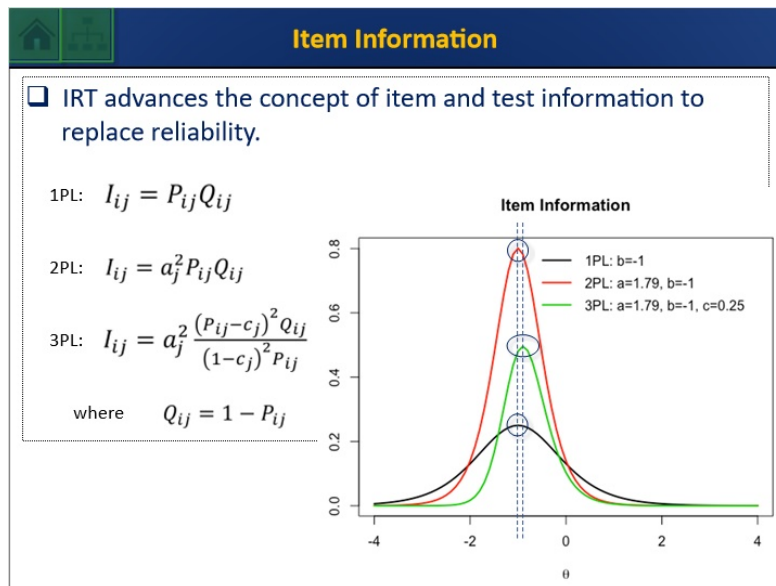
2.8 Simulation Example

Simulation Example

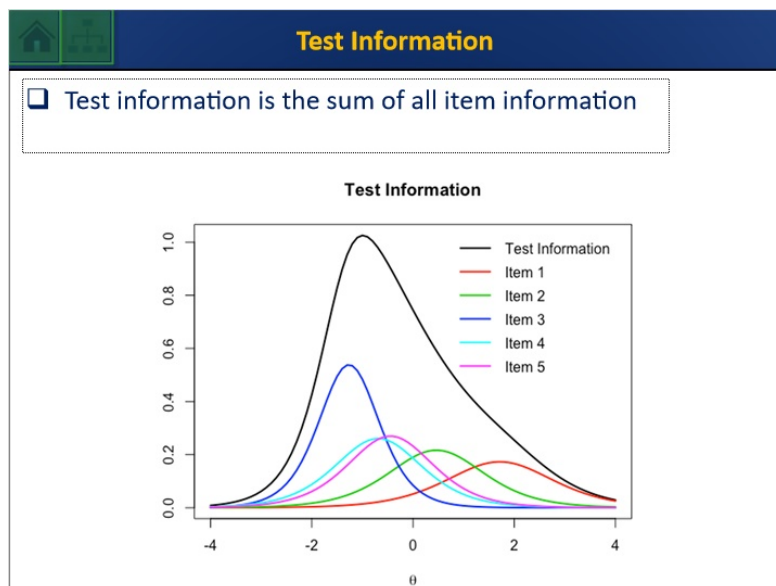
☐ 2PL Example

- 5 items
- $a \sim \text{Lognormal}(0,1)$
- $b \sim \text{Normal}(0,1)$
- $\theta \sim \text{Normal}(0,1)$
- Response for examinee #1 is (0, 1, 0, 0, 1) ["Mixed" response]

2.9 Item Information



2.10 Test Information



2.11 Standard Error of Measurement

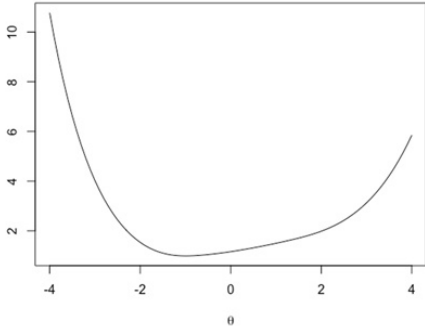


Standard Error of Measurement

☐ Standard error of measurement (SE) is the reciprocal of the test information at a given ability level.

$$SE_i = \frac{1}{\sqrt{I_i}}$$

SE



θ

2.12 Likelihood

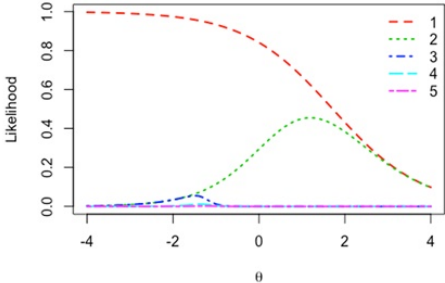


Likelihood

☐ Likelihood is the plausibility of a response pattern

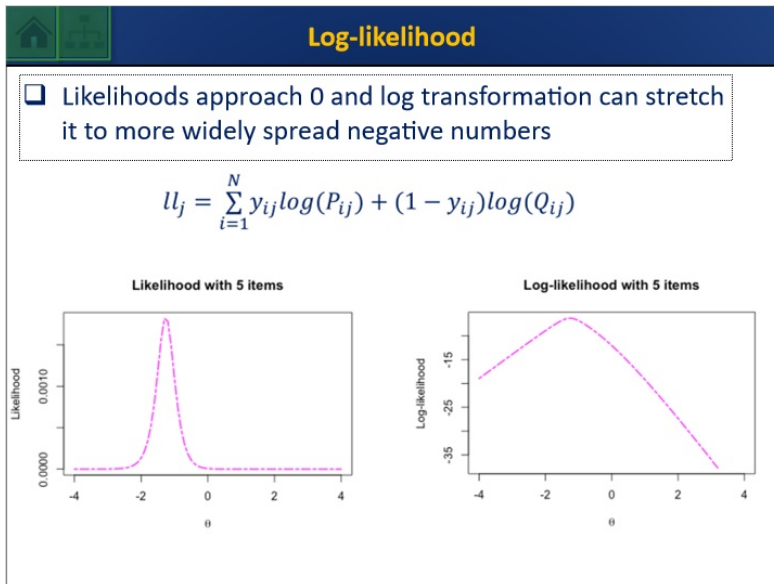
$$L_i = \prod_{j=1}^J P_{ij}^{y_{ij}} Q_{ij}^{1-y_{ij}} = Q_{i1} \times P_{i2} \times Q_{i3} \times P_{i4} \times P_{i5}$$

Likelihood of 1 to 5 items (0,1,0,1,1)



where y_{ij} is examinee i 's response on item j

2.13 Log-Likelihood

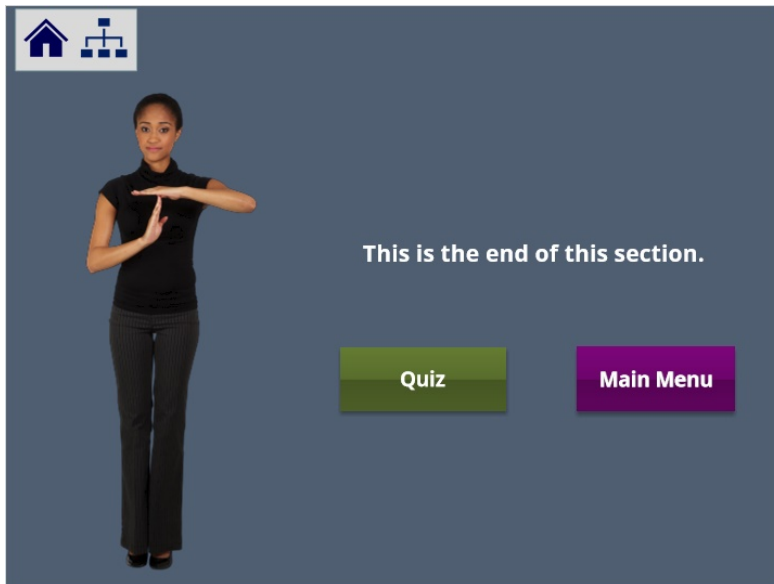


2.14 Summary

  **Summary**

- ☐ IRT puts item difficulty and examinee ability on the same scale, thus exerts advantages over CTT
- ☐ With item parameters, item response functions as well as likelihood functions can be derived
- ☐ Log-likelihood function is used more in computation than likelihood function, as it has wider spread

2.15 Bookend: IRT Basics



3. Section 2: Calibration and Scoring

3.1 Cover: Calibration and Scoring



3.2 Estimation Algorithms: Learning Objectives



Learning Objectives



1. Understand the difference between scoring and calibration
2. Learn about different scoring and calibration procedures
3. Learn how to find the maximum of a function
4. Understand the logic behind the EM algorithm

3.3 Scoring vs. Calibration I



Scoring vs. Calibration



- ☐ In IRT, calibration is usually referred to the estimation of item parameters
- ☐ On the other hand, scoring is the estimation of person parameters

3.4 Topic Selection



Scoring vs. Calibration



Click on each button to learn more 

End Section

3.5 Bookmark: Scoring





Scoring

3.6 Scoring: MLE

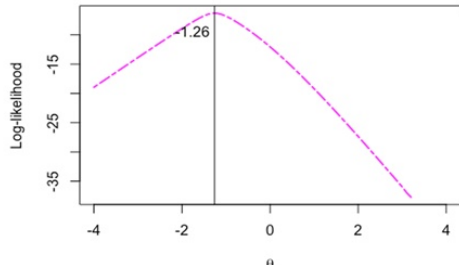


Scoring

☐ Scoring is always performed when the item parameters are known. When they are unknown, calibration is performed first.

☐ Using likelihood $\rightarrow$ Maximum likelihood estimation (MLE)

Maximum Likelihood Estimation (MLE)



3.7 Finding a Function Maximum

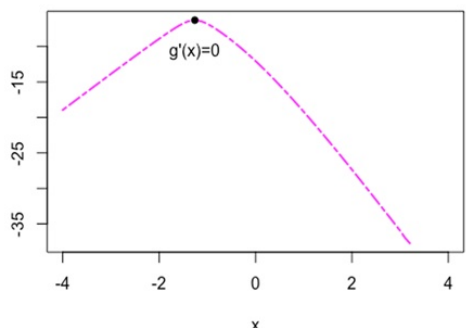


Finding a Function Maximum

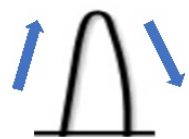
☐ How to find maximum of a function $g(x)$?

$g'(x) = 0$

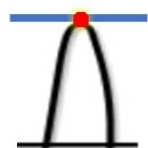
$g(x)$



Concave



First derivative (slope) = 0



3.8 Newton-Raphson Algorithm



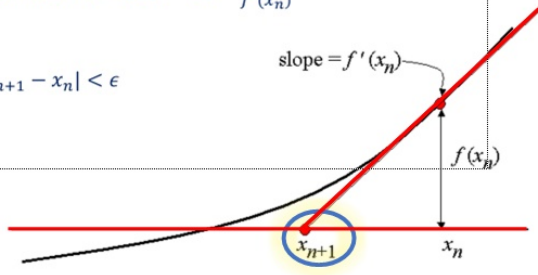
Newton-Raphson Algorithm

☐ How to find $g'(x) = 0$?

Newton Raphson iteration

$f(x) = g'(x)$

1. Start with initial value x_0 (becomes x_n on first iteration)
2. A better approximation is $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$
- ...
3. Stop, when $|x_{n+1} - x_n| < \epsilon$



3.9 Posterior Distributions

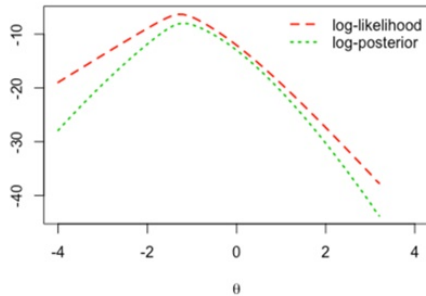


Posterior Distributions

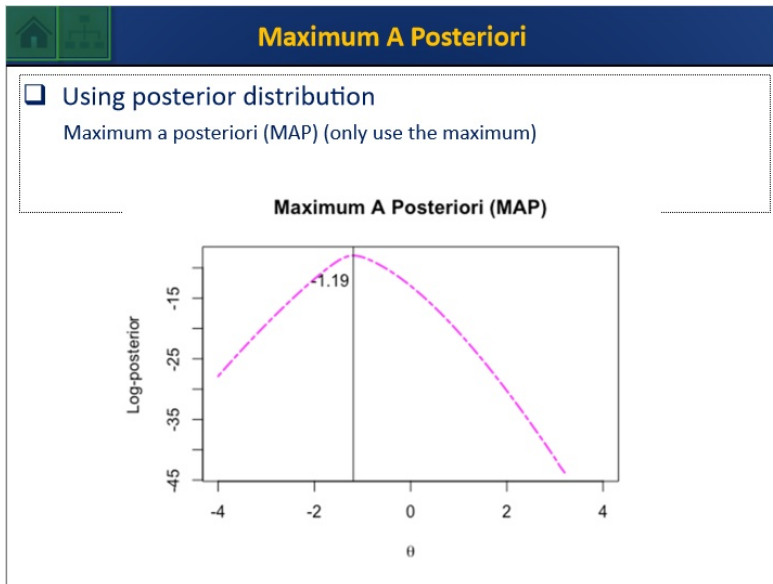
☐ When response pattern consists of all 0 or all 1, likelihood function doesn't have a maximum, thus MLE does not work.

☐ Using posterior distribution can solve this problem

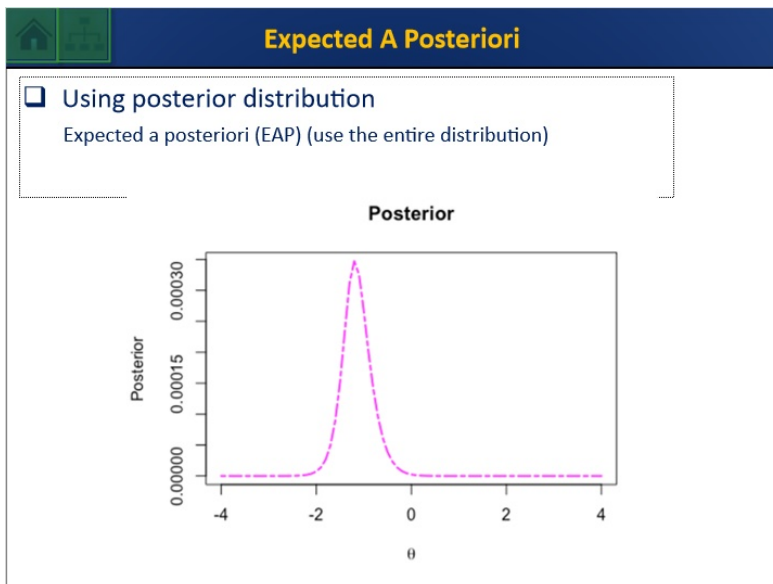
Log-likelihood and log-posterior comparison



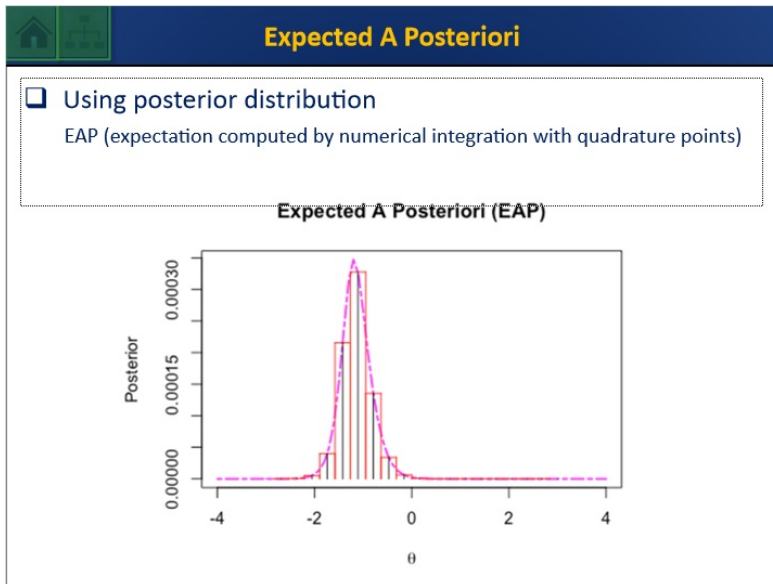
3.10 Maximum A Posteriori



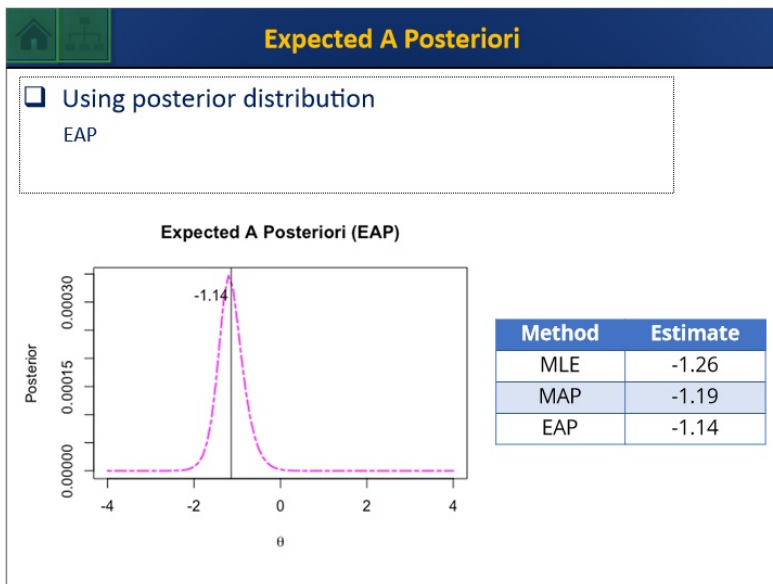
3.11 Expected A Posteriori I



3.12 Expected A Posteriori II



3.13 Expected A Posteriori III



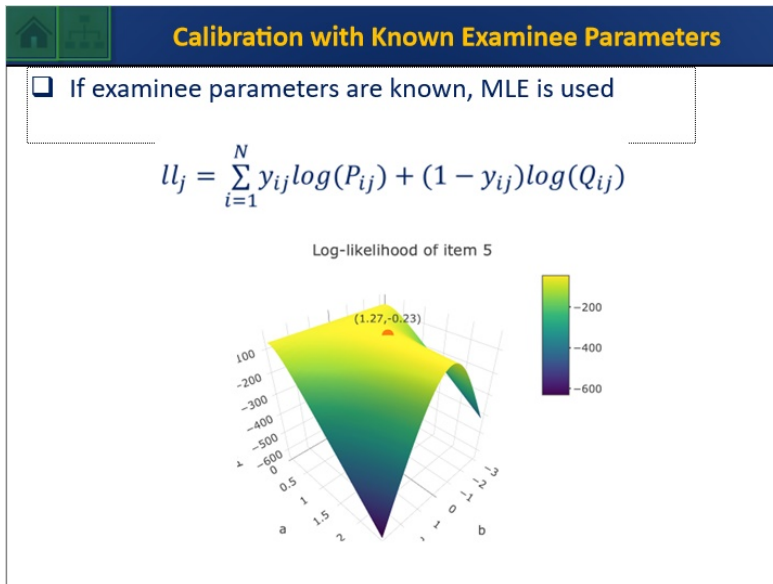
3.14 Bookend: Scoring



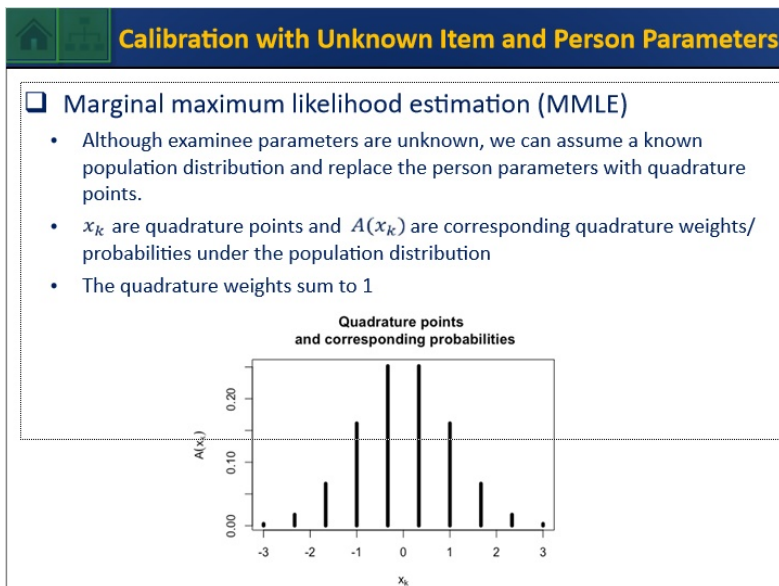
3.15 Bookmark: Calibration



3.16 Calibration with Unknown Item Parameters



3.17 Calibration with Unknown Item and Person Parameters



3.18 MMLE Log-Likelihood



MMLE Log-Likelihood

☐ MMLE

- Replace single response y_{ij} with $\bar{r}_{kj}$ which is number of correct answer in group k ; replace $1-y_{ij}$ with $\bar{n}_k - \bar{r}_{kj}$, where $\bar{n}_k$ is the size of group k .
- $\bar{n}_k$ and $\bar{r}_{kj}$ are called artificial data
- The log-likelihood function for item j becomes:

$$ll_j = \sum_{i=1}^N \bar{r}_{kj} \log(P_{kj}) + (\bar{n}_k - \bar{r}_{kj}) \log(Q_{kj})$$

3.19 MMLE Log-Likelihood II



MMLE Log-Likelihood

☐ MMLE

- $L_{ik} = \prod_{j=1}^J P_{kj}^{y_{ij}} Q_{kj}^{1-y_{ij}}$ is the likelihood of an examinee i with ability x_k
- Compute $\bar{n}_k$ and $\bar{r}_{jk}$ based on known item parameters and responses

$$\bar{n}_k = \sum_{i=1}^N \frac{L_{ik} A(x_k)}{\sum_{k=1}^K L_{ik} A(x_k)}$$
$$\bar{r}_{jk} = \sum_{i=1}^N \frac{y_{ij} L_{ik} A(x_k)}{\sum_{k=1}^K L_{ik} A(x_k)}$$

3.20 EM Algorithm



Expectation-Maximization (EM) Algorithm

- ❑ In reality, item parameters are usually unknown
- ❑ The EM algorithm iteratively estimates item parameters
- ❑ Begin by assigning starting values for item parameters, then cycle between E-steps and M-steps until the estimates converge
 - E-step: assume item parameters are known and compute $\bar{n}_k$ and $\bar{r}_{kj}$
 - M-step: using previously calculated values of $\bar{n}_k$ and $\bar{r}_{kj}$, and compute MLE of item parameters

3.21 Bookend: Calibration





This is the end of this part.

Topic Selection

3.22 Summary



Summary

- ☐ In addition to scoring, MLE can also be used to conduct calibration when examinee parameters are known
- ☐ Posterior distribution-based scoring methods (MAP and EAP) overcome the non-converge issue of MLE
- ☐ MMLE is used when neither item parameters nor examinee parameters are known; EM algorithm is used to iteratively update item parameters

3.23 Bookend: IRT Basics





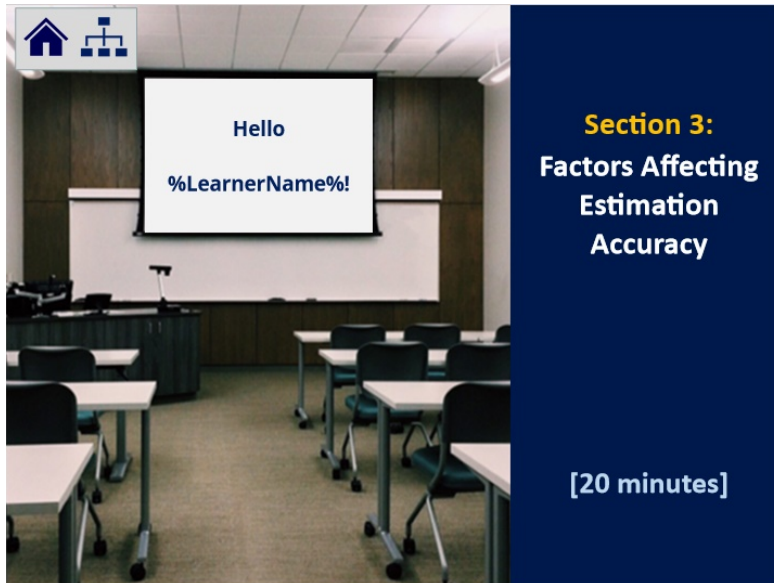
This is the end of this section.

[Quiz](#)

[Main Menu](#)

4. Section 3: Factors Affecting Estimation Accuracy

4.1 Cover: Factors Affecting Estimation Accuracy



4.2 Factors Affecting Estimation Accuracy: Learning Objectives

Learning Objectives

1. Understand how estimation accuracy is quantified
2. Recognize common factors affecting estimation accuracy
3. Figure out strategies to improve estimation accuracy
4. Understand how multiple factors can affect each other

4.3 Estimation Accuracy Evaluation Criteria



Estimation Accuracy Evaluation Criteria

- Bias and root mean square error (RMSE) are used to evaluate the estimation accuracy of calibration and scoring
- π is the true value of a parameter and $\hat{\pi}$ is the estimated value of this parameter

Bias:
$$\frac{\sum_{i=1}^n (\hat{\pi}_i - \pi_i)}{n}$$

RMSE:
$$\sqrt{\frac{\sum_{i=1}^n (\hat{\pi}_i - \pi_i)^2}{n}}$$

- The *proportional* bias and RMSE are computed as the proportion of bias and RMSE of a parameter over the SD of that parameter.

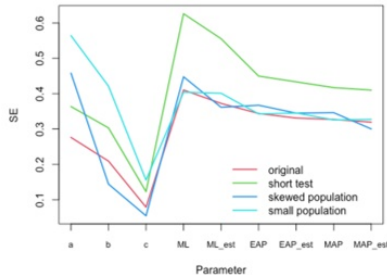
4.4 Identify the Need to Improve Estimation Accuracy



Identify the Need to Improve Estimation Accuracy

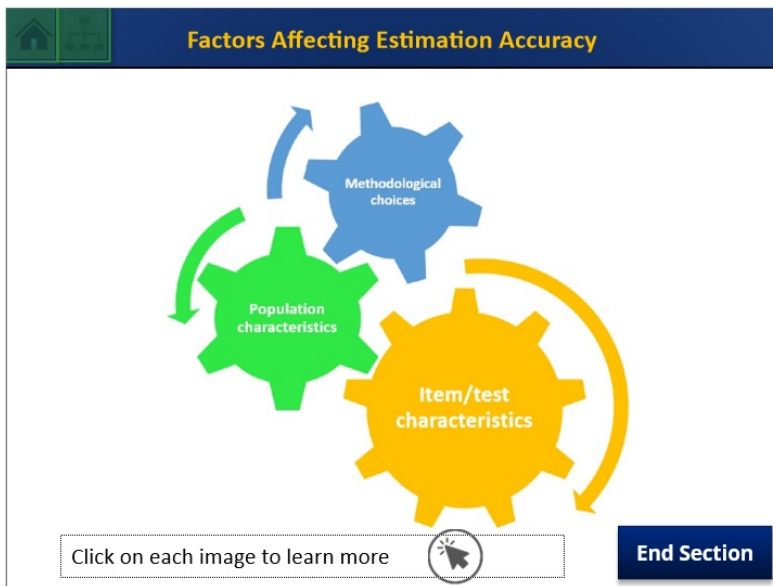
- How to identify the need of improving estimation accuracy?
 - Look at population size.
 - 500 for 2PL
 - 1000 for 3PL
 - Look at SE of parameter estimation (especially a parameter)

Estimation Accuracy



Parameter	original	short test	skewed population	small population
a	0.28	0.35	0.45	0.55
b	0.18	0.25	0.35	0.45
c	0.12	0.18	0.28	0.38
ML	0.40	0.60	0.45	0.45
ML_est	0.38	0.55	0.42	0.42
EAP	0.35	0.45	0.38	0.38
EAP_est	0.32	0.42	0.35	0.35
MAP	0.30	0.40	0.33	0.33
MAP_est	0.28	0.38	0.31	0.31

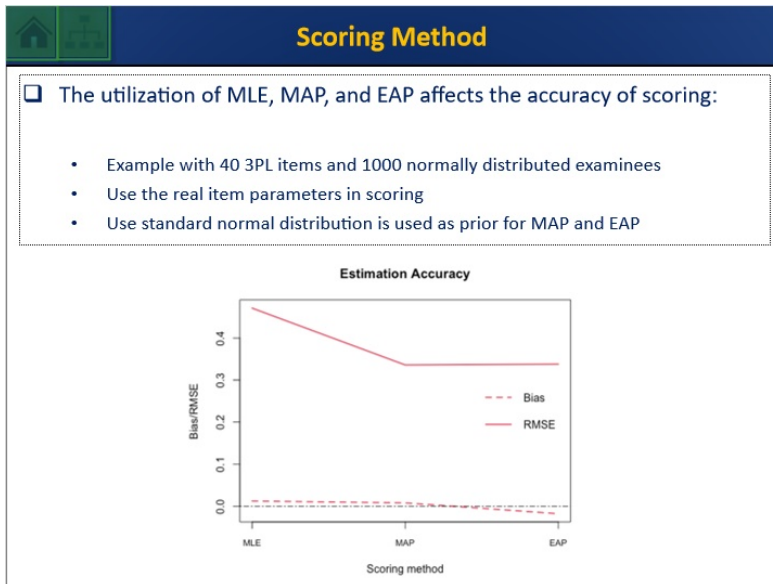
4.5 Topic Selection



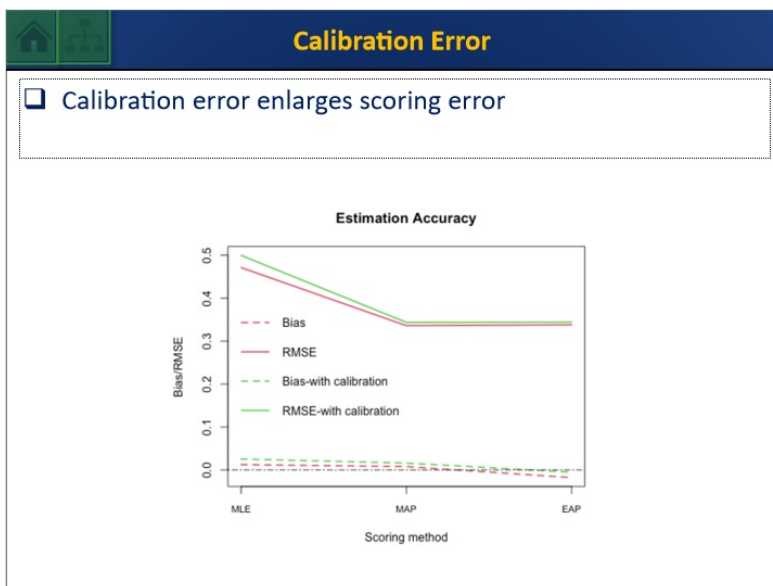
4.6 Bookmark: Methodological Choices



4.7 Scoring Method



4.8 Calibration Error



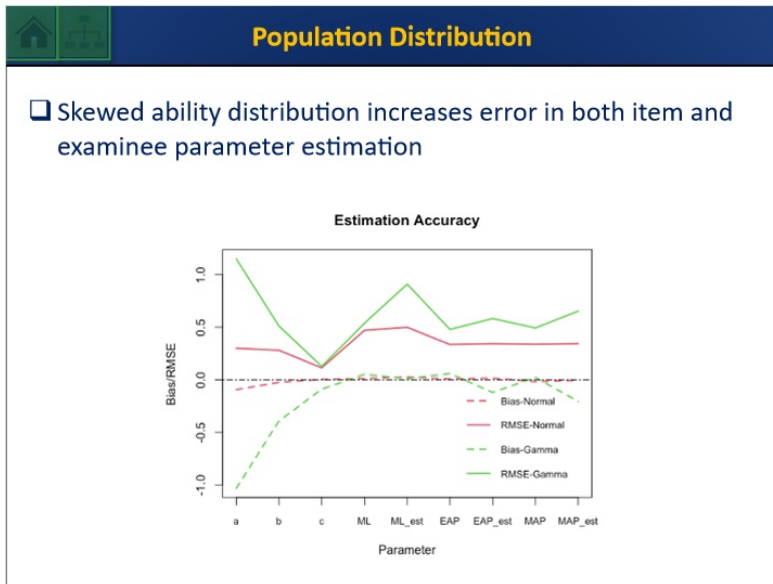
4.9 Bookend: Calibration



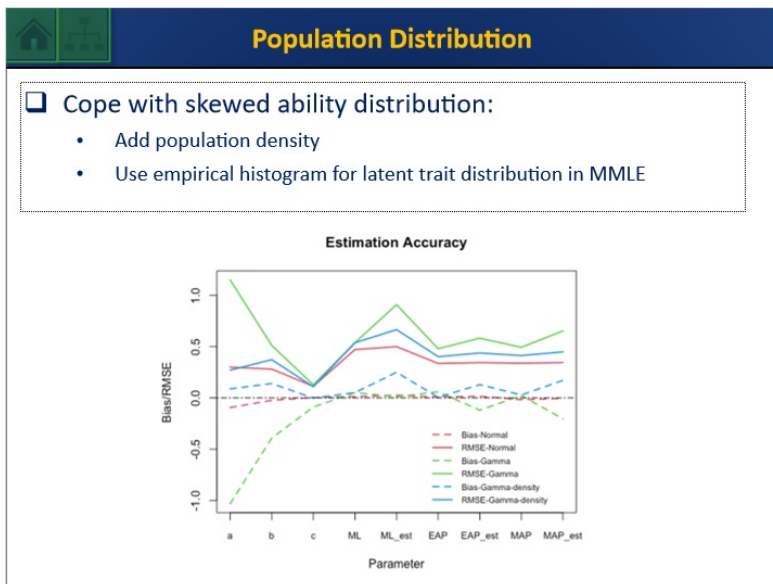
4.10 Bookmark: Population Characteristics



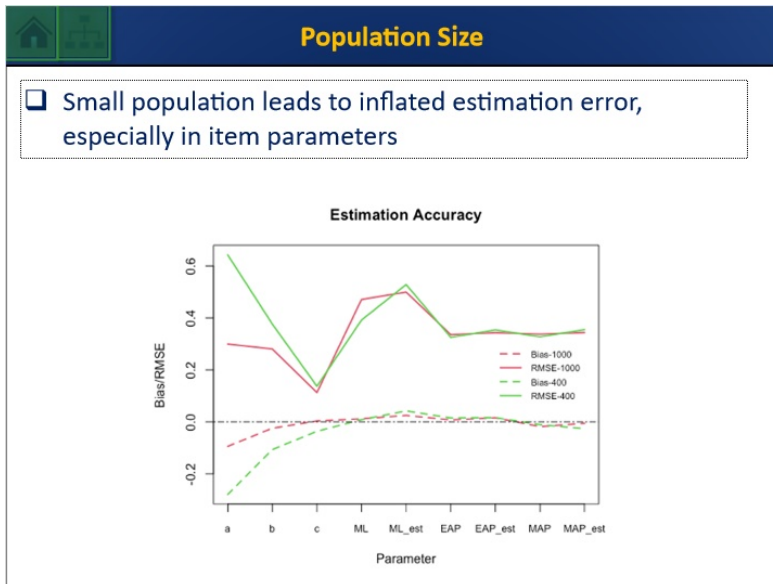
4.11 Population Distribution I



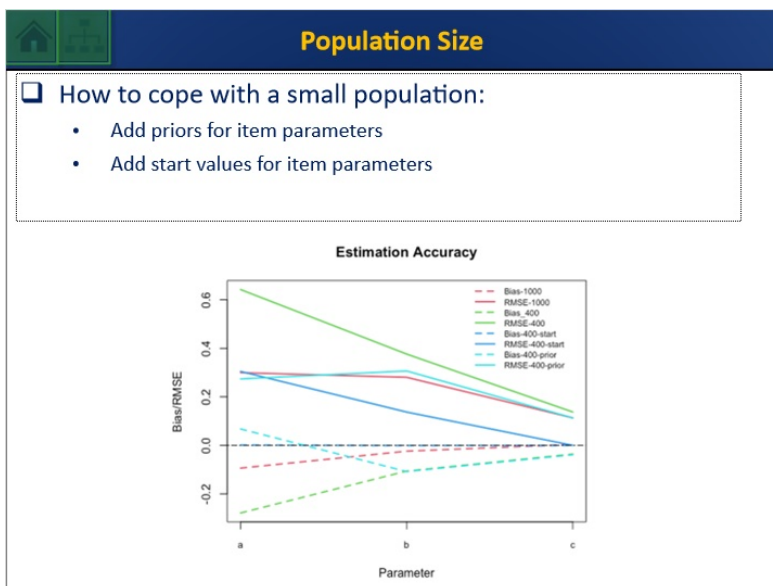
4.12 Population Distribution II



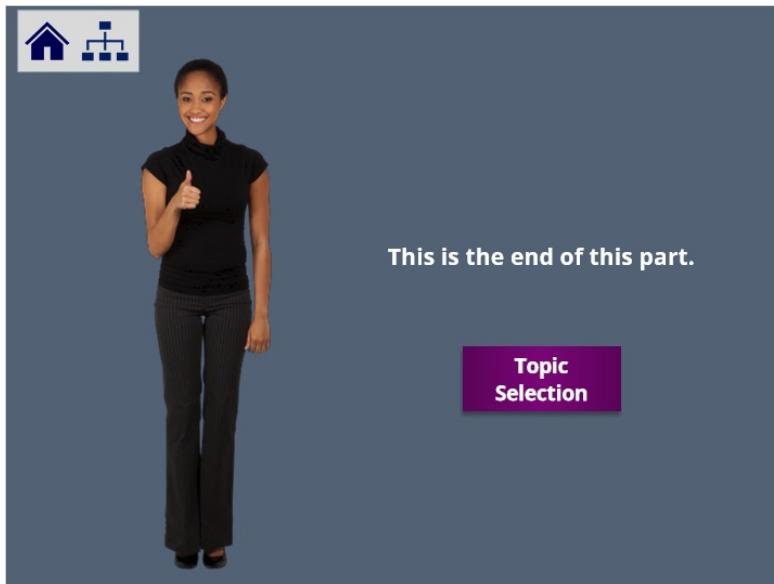
4.13 Population Size I



4.14 Population Size II



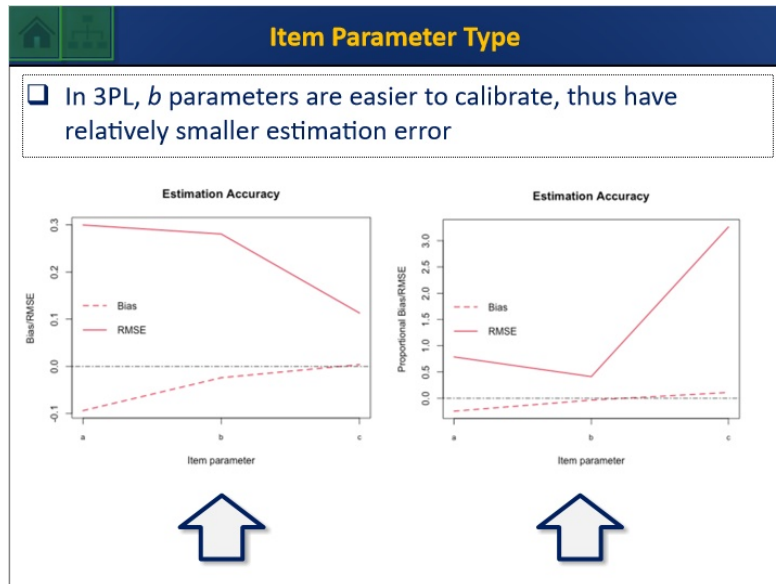
4.15 Bookend: Calibration



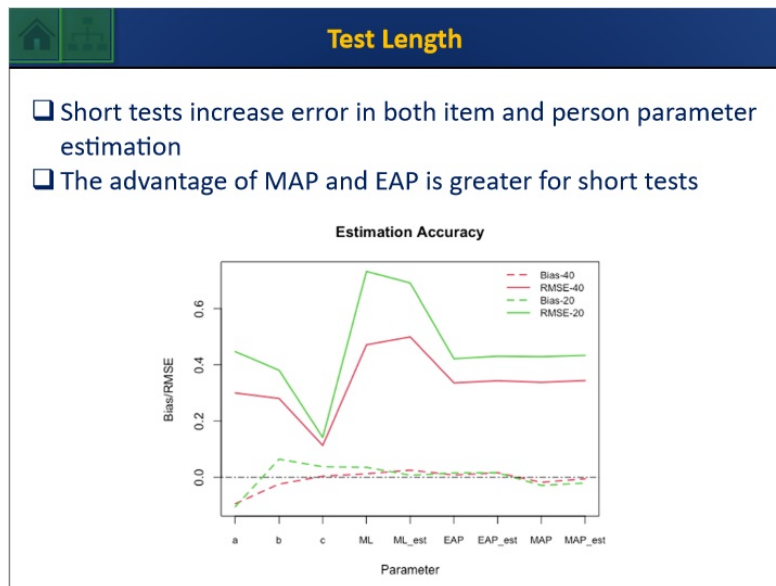
4.16 Bookmark: Item and Test Characteristics



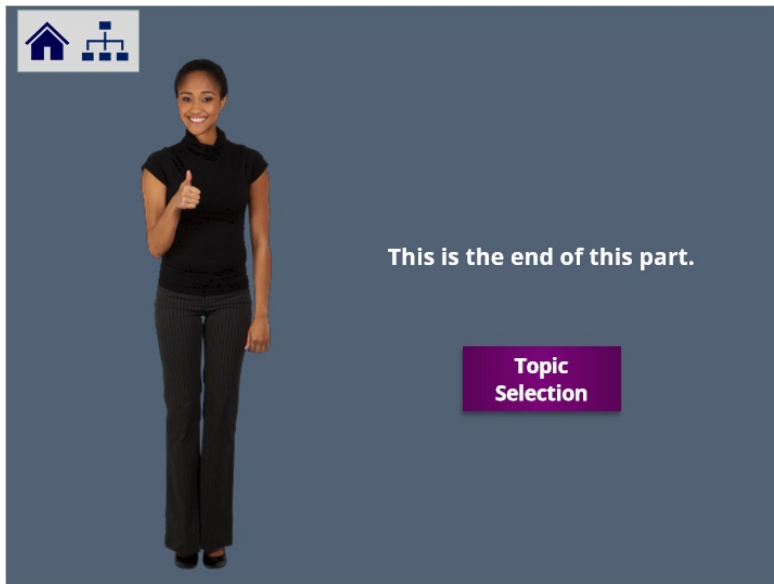
4.17 Item Parameter Types



4.18 Test Length



4.19 Bookend: Calibration



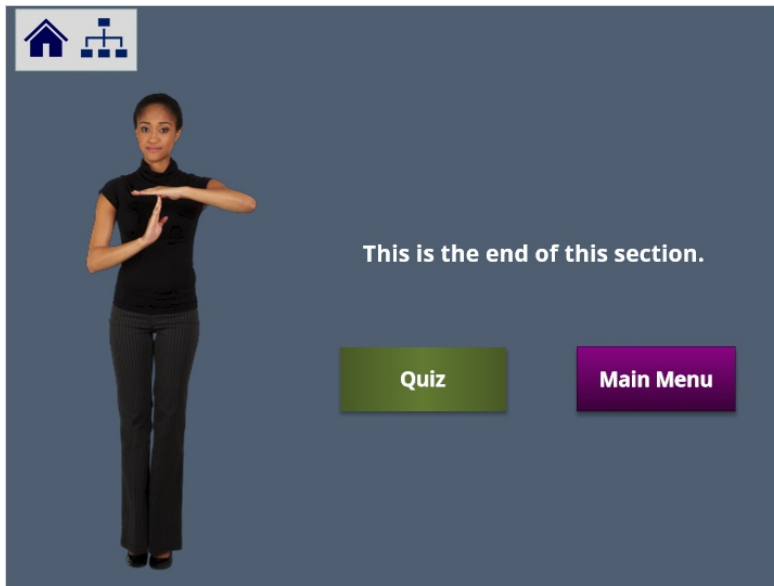
4.20 Summary



Summary

- ☐ Various factors can influence estimation accuracy
- ☐ SE of parameter estimation is a good resource to identify potential problems in estimation
- ☐ Potential remedies for estimation issues:
 - Adding population density prior
 - Adding item parameter prior
 - Adding item parameter start values
 - Using empirical histogram for latent trait in MMLE

4.21 Bookend: IRT Basics



4.22 Module Cover (START)

