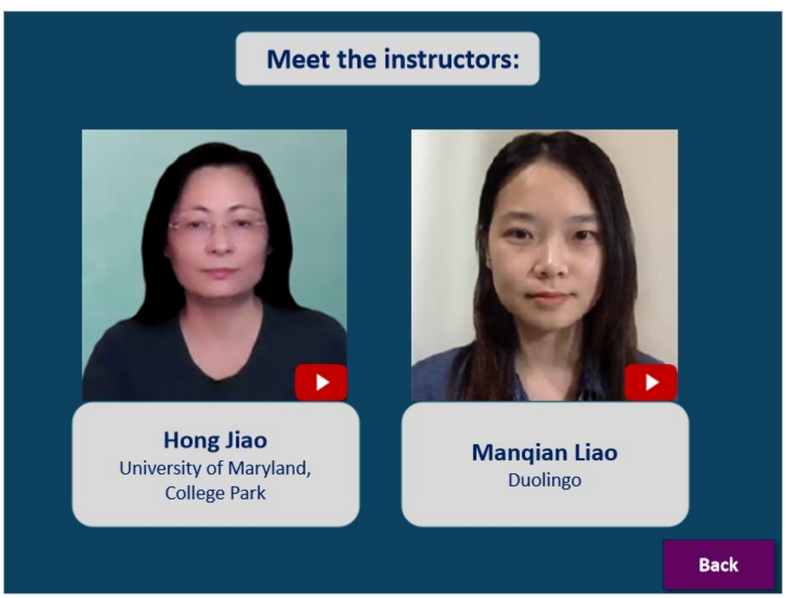
DM25 SLIDES (Testlet Models, Version 1.0)

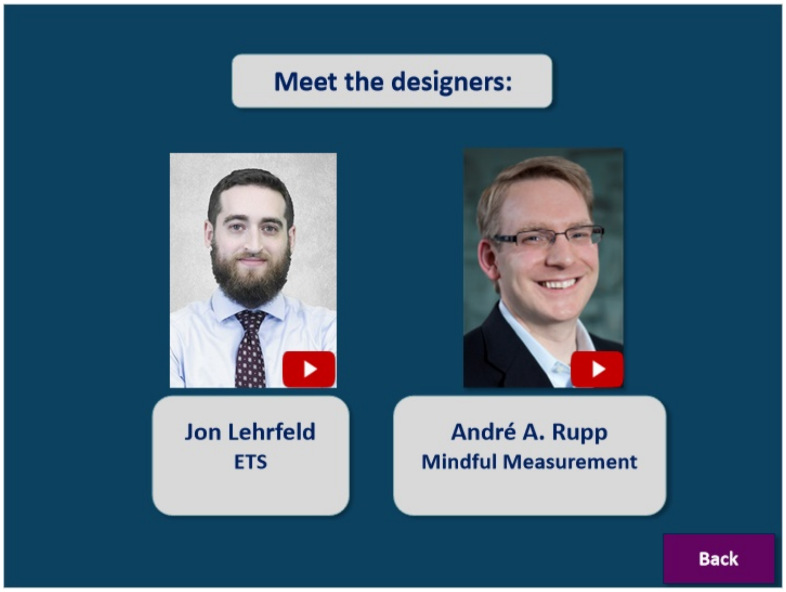
1. Module Overview

1.1 Module Cover (START)

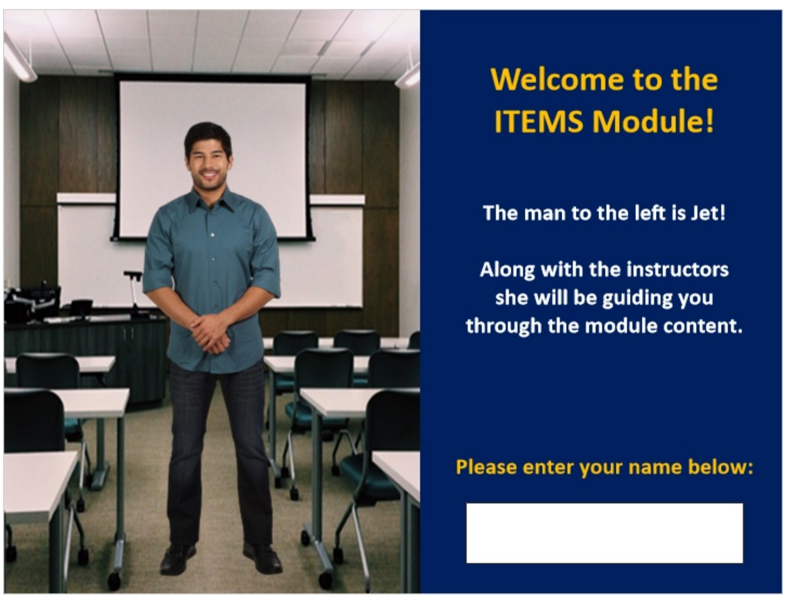
1.2 Instructors



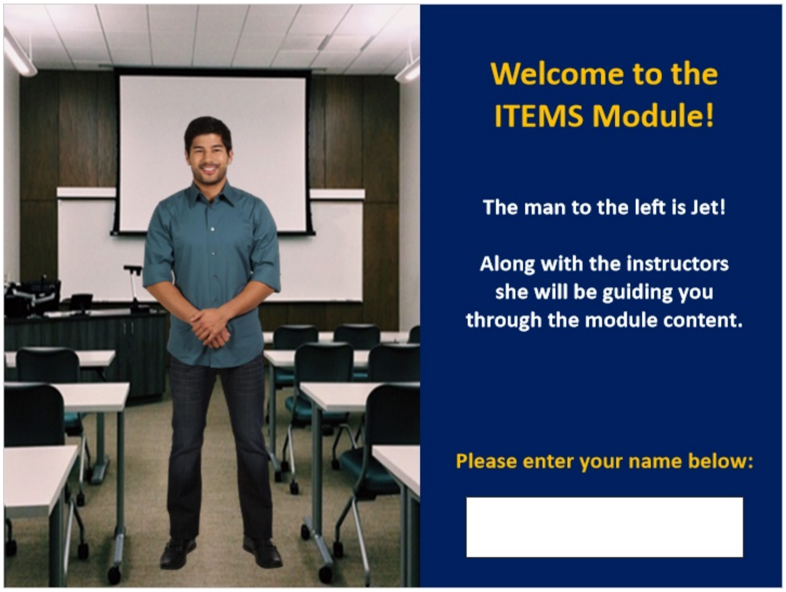
1.3 Designers



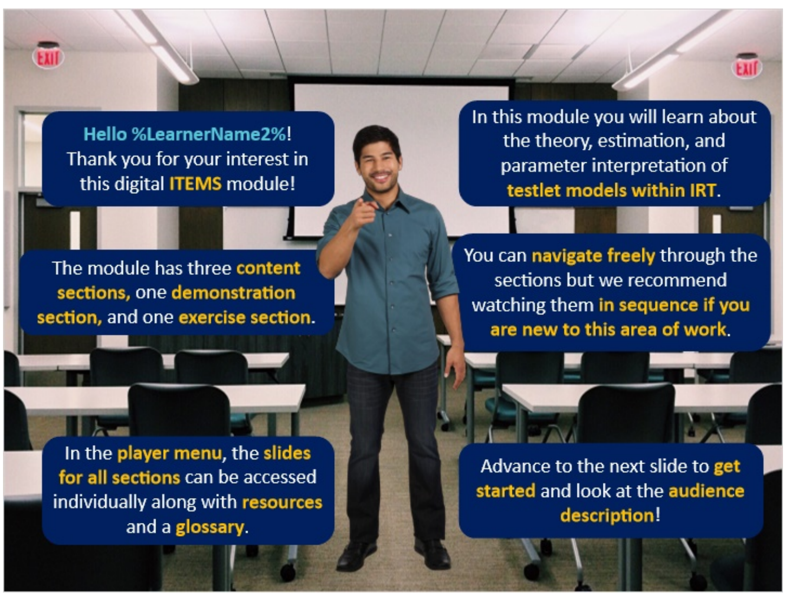
1.4 Welcome



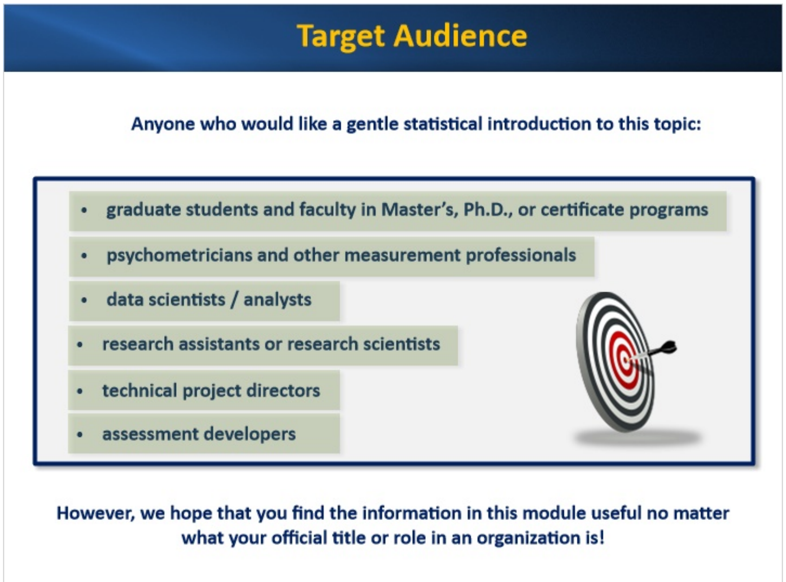
Untitled Layer 1 (Slide Layer)



1.5 Overview



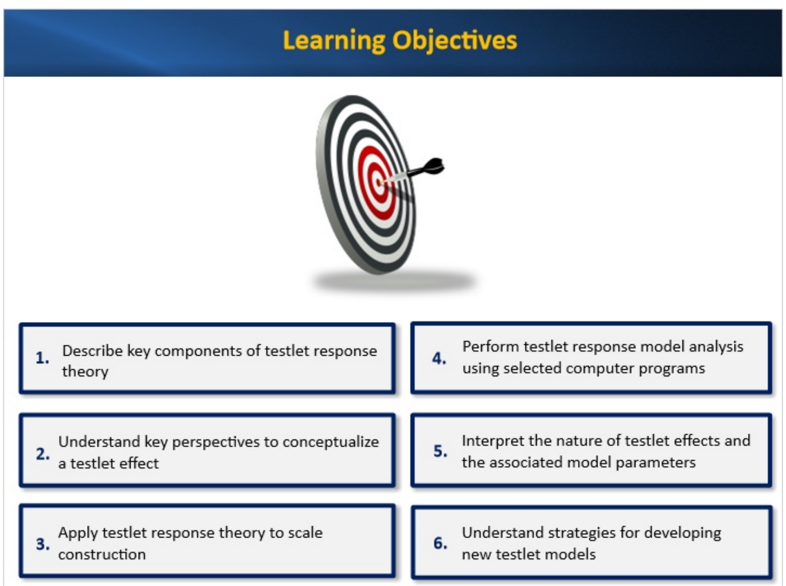
1.6 Target Audience



1.7 Expecations (I)

1.8 Expectations (II)

1.9 Learning Objectives



1.10 Prerequisites

Prerequisites

1. Foundations of unidimensional item response theory

2. Basic mathematical transformation functions (e.g., logarithms)

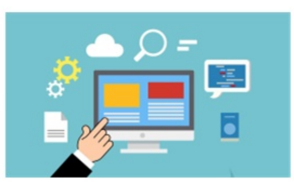
3. Foundations of Bayesian estimation

1.11 Module Citation

Module Citation

Module Citation

Jiao, H., & Liao, M. (2021). Testlet response models (Digital ITEMS Module 25). *Educational Measurement: Issues and Practice, 40*(3).

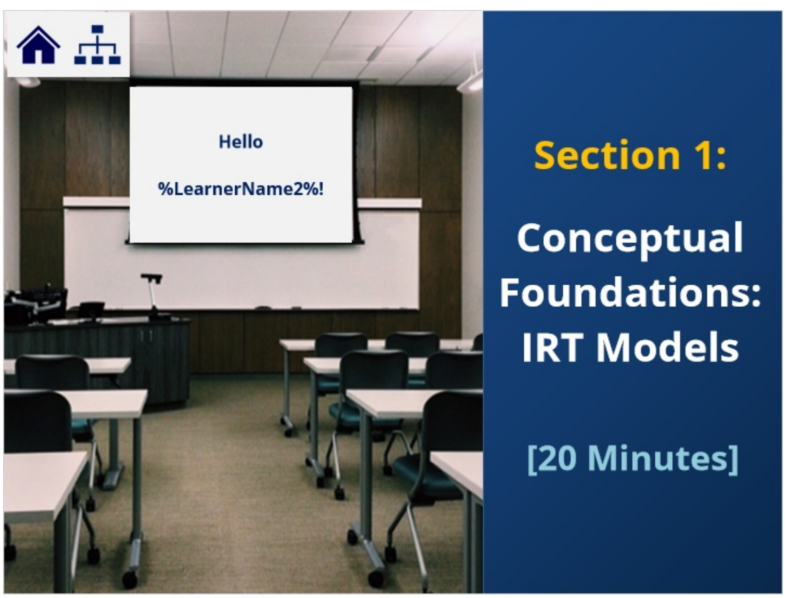


FREE WEB RESOURCES

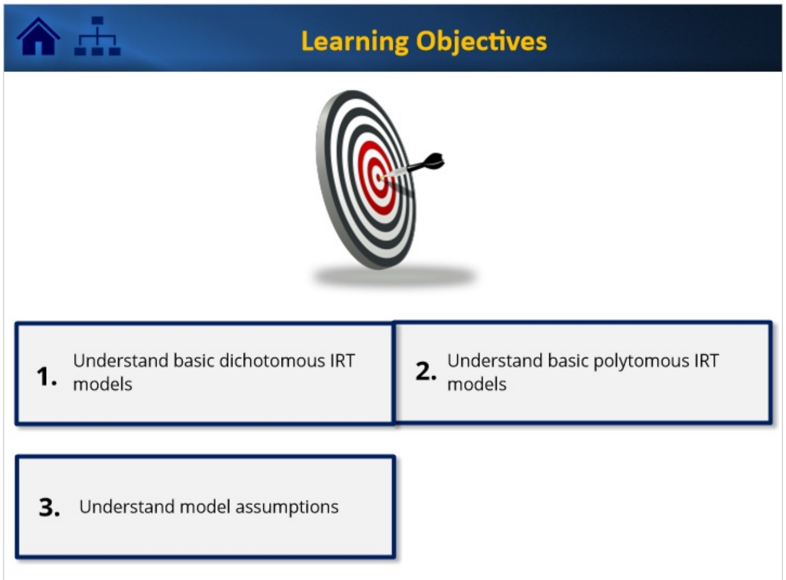
1.12 Main Menu

2. Section 1: Conceptual Foundations: IRT Models

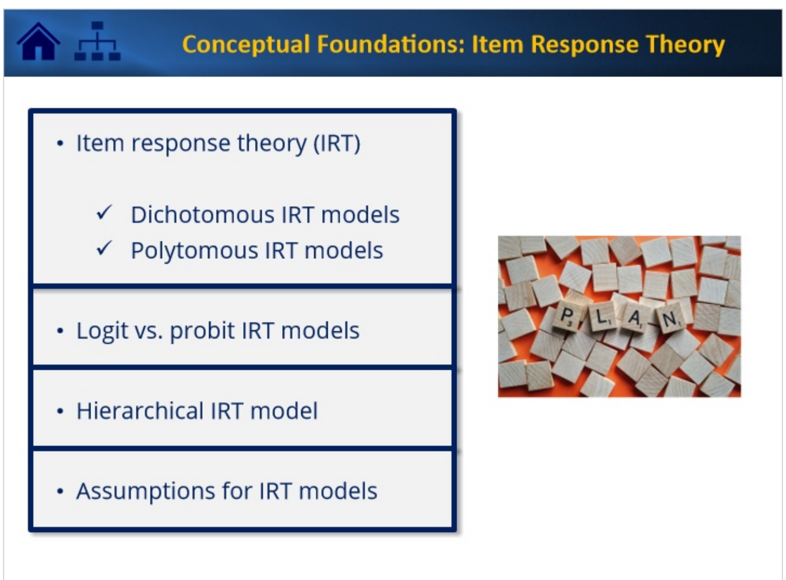
2.1 Cover: Section 1



2.2 Learning Objectives: Section 1



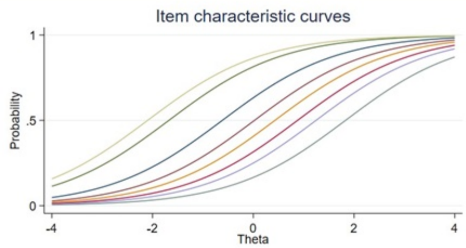
2.3 Conceptual Foundations-Item Response Theory



2.4 Item Response Theory

Item Response Theory (IRT)

- A modern measurement framework that quantifies the relationship among an item response and a person's latent ability given item characteristics.



adapted from https://www.stata.com/features/overview/irt/

- A mathematical representation modeling the nonlinear relationship between the probability of an item response, a latent ability level, and item parameters.

2.5 Item Response Theory Models

IRT Models

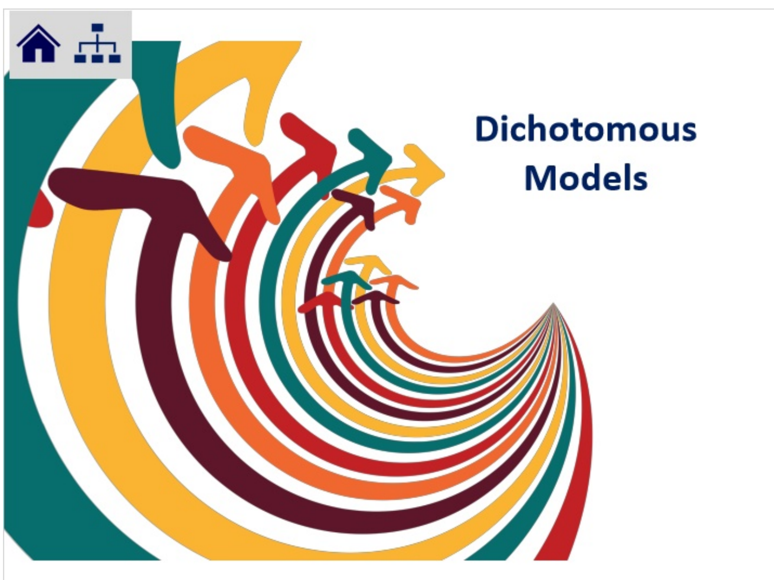
- The common IRT models include models for dichotomous responses and polytomous responses.

- Common IRT models for dichotomous item responses:
  - ✓ the Rasch measurement model
  - ✓ the two-parameter IRT model
  - ✓ the three-parameter IRT model

- Common IRT models for polytomous item responses:
  - ✓ the partial credit model
  - ✓ the generalized partial credit model
  - ✓ the graded response model

2.6 Topic Selection

2.7 Bookmark: Dichotomous Models



2.8 The Rasch Measurement Model I

The Rasch Measurement Model

- A **latent logistic model** modeling the **nonlinear relationship** between the probability of an item response and a latent ability level given an item difficulty parameter

- Such a nonlinear relationship can be graphically represented in an **item characteristic curve** (ICC) to show how an increase in latent ability leads to an increase in the probability of a correct item response given the specific item difficulty

- The **logit of a correct response** is linearly related to the interaction between a person's latent ability, ***θ***, and the item difficulty, *b*

2.9 The Rasch Measurement Model II

The Rasch Measurement Model

Logistic model

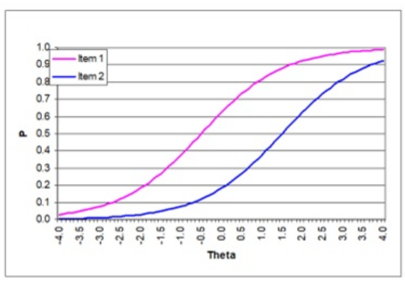
$$p_{ij}(x_{ij}=1|\theta_j)=\frac{\exp(\theta_j-b_i)}{1+\exp(\theta_j-b_i)}$$

$$p_{ij}(x_{ij}=1|\theta_j)=\frac{1}{1+\exp[-(\theta_j-b_i)]}$$

Logit format

$$\ln\left(\frac{P_{ij}}{1-P_{ij}}\right)=(\theta_j-b_i)$$

Item Characteristic Curve



2.10 The Two-Parameter Logistic Model

The Two-Parameter Logistic Model

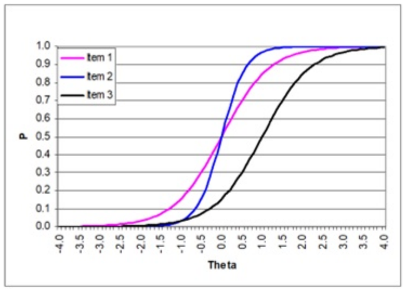
Logistic model

$$p_{ij}(x_{ij}=1|\theta_j)=\frac{1}{1+\exp[-a_i(\theta_j-b_i)]}$$

Logit format

$$\ln\left(\frac{P_{ij}}{1-P_{ij}}\right)=a_i(\theta_j-b_i)$$

Item Characteristic Curve



2.11 The Three-Parameter Logistic Model

The Three-Parameter Logistic Model

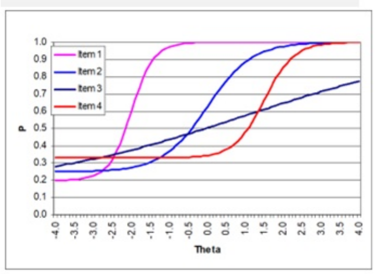
Logistic model

$$p_{ij}(x_{ij}=1|\theta_j)=c_i+\frac{1-c_i}{1+\exp[-a_i(\theta_j-b_i)]}$$

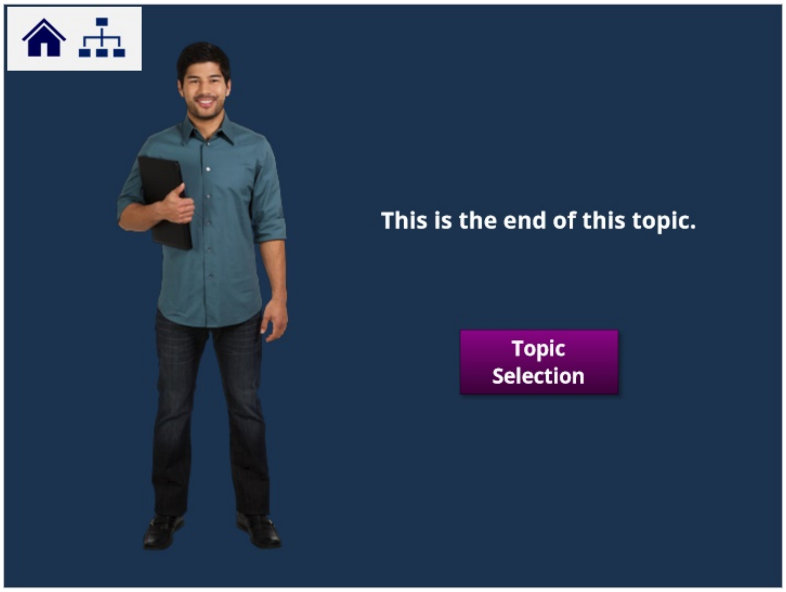
Logit format

$$\ln\left(\frac{P_{ij}-c_i}{1-P_{ij}}\right)=a_i(\theta_j-b_i)$$

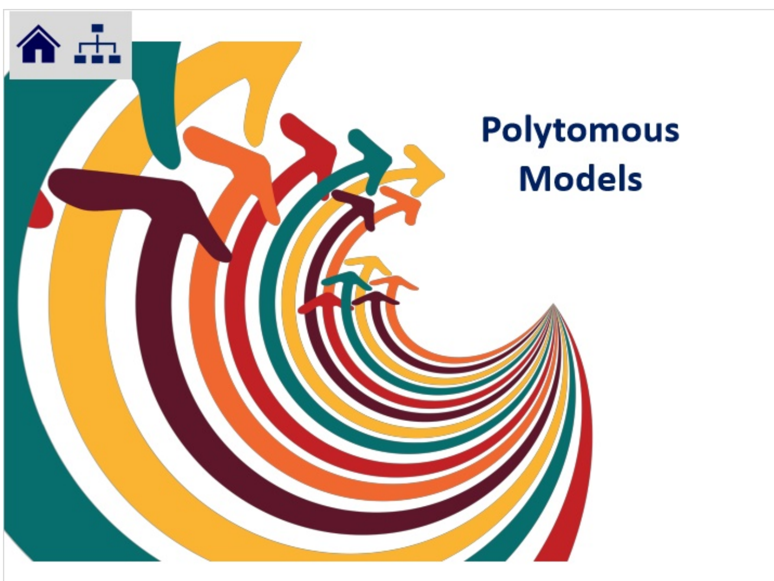
Item Characteristic Curve



2.12 Bookend: Dichotomous Models



2.13 Bookmark: Polytomous Models



2.14 Overview of Different Polytomous Models

Overview of Polytomous Models

Adjacent-category models

Graded response model	(e.g., Samejima, 1969)
Modified graded response model	(e.g., Muraki, 1990)

Divide-by-total models

Partial credit model	(e.g., Masters, 1982)
Generalized partial credit model	(e.g., Muraki, 1992)
Rating scale model	(e.g., Andrich, 1978)
Modified rating scale model	(e.g., du Toit et al., 1999)
Nominal response model	(e.g., Bock, 1972)

2.15 The Graded Response Model I

The Graded Response Model

2PL IRT model

$$p_{ij}(x_{ij} = 1|\theta_j) = \frac{\exp[a_i(\theta_j - b_i)]}{1 + \exp[a_i(\theta_j - b_i)]}$$

Graded Response Model

- Probability of a person's raw item score x falling in or above a given category k = 1, 2, ..., m conditional on theta
- Operating characteristic curve

$$p^*_{ix}(\theta) = \frac{\exp[a_i(\theta_j - b_{ik})]}{1 + \exp[a_i(\theta_j - b_{ik})]}$$

- Category threshold represents the trait level to respond at or above threshold k with 0.5 probability.

2.16 The Graded Response Model II

The Graded Response Model

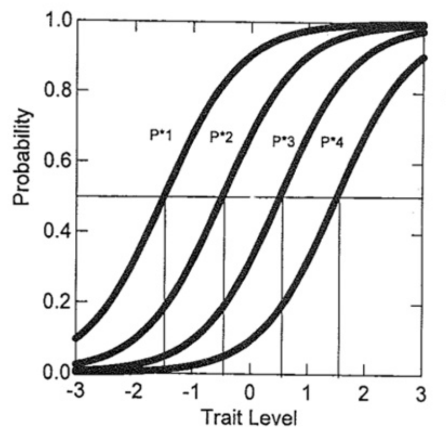


FIG. 5.1. Operating characteristic curves for five-category example item under the graded response model.

2.17 The Graded Response Model III

The Graded Response Model

Probability of a person's raw item score-x fall in or above a given category k=1, 2,..m conditional on theta

$$P_{ix}^{*}(\theta) = \frac{\exp[a_i(\theta_j - b_{ik})]}{1 + \exp[a_i(\theta_j - b_{ik})]}$$

Operating characteristic curve

Given three scores x=0, 1, 2 with 3 response options, the item is treated as a series of 3-1=2 dichotomies: 0 vs 1, 2; 0, 1 vs 2

$$P_{ix}(\theta) = P_{ix}^{*}(\theta) - P_{i(x+1)}^{*}(\theta)$$
$$P_{i0}(\theta) = 1 - P_{i1}^{*}(\theta)$$
$$P_{i1}(\theta) = P_{i1}^{*}(\theta) - P_{i2}^{*}(\theta)$$
$$P_{i2}(\theta) = P_{i2}^{*}(\theta) - 0$$

Category response curves: the probability of an examinee responding in a particular category conditional on trait level.

2.18 The Graded Response Model IV

The Graded Response Model

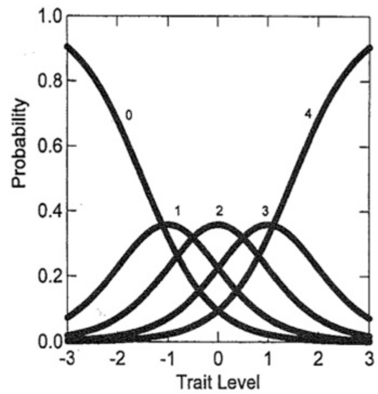


FIG. 5.2. Category response curves for five-category example item under the graded response model.

For any given theta value, the sum of the response probabilities is equal to 1.

2.19 The Partial Credit Model

The Partial Credit Model

$$p_{ix}(\theta) = \frac{\exp\left[\sum_{s=0}^{x}(\theta_j - b_{is})\right]}{\sum_{k=0}^{K}\exp\left[\sum_{s=0}^{k}(\theta_j - b_{is})\right]}$$

where $\sum_{s=0}^{0}(\theta - b_{is}) = 0$

- The point on the latent trait scale at which two consecutive category response curves intersect.
- Relative difficulty of each step.
- Indicates where on the latent trait scale the response of one category becomes relatively more likely than the previous category.

Masters (1982)

2.20 The Reparameterized Partial Credit Model I

The Reparameterized Partial Credit Model

Assume three score categories: 0, 1, 2

The probability of getting a certain score is

$$t_1 = \exp(\theta_j - b_i - d_{i1}) \qquad t_2 = \exp(2\theta_j - 2b_i - d_{i1} - d_{i2})$$

$$p(x_{ji} = 0) = \frac{1}{1 + t_1 + t_2}$$

$$p(x_{ji} = 1) = \frac{t_1}{1 + t_1 + t_2}$$

$$p(x_{ji} = 2) = \frac{t_2}{1 + t_1 + t_2}$$

2.21 The Reparameterized Partial Credit Model II

Category Response Curves

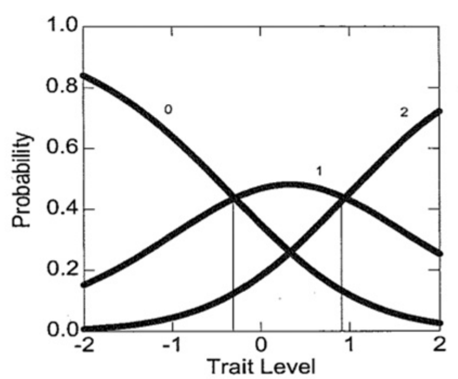
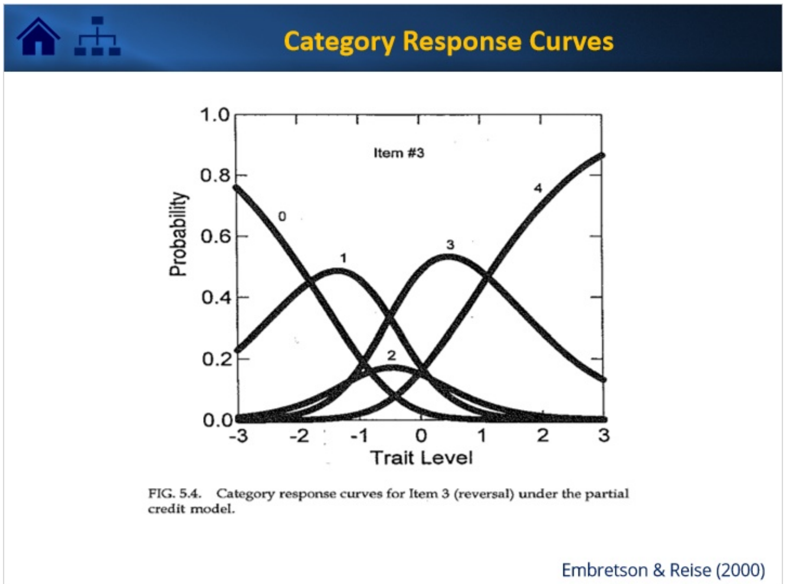


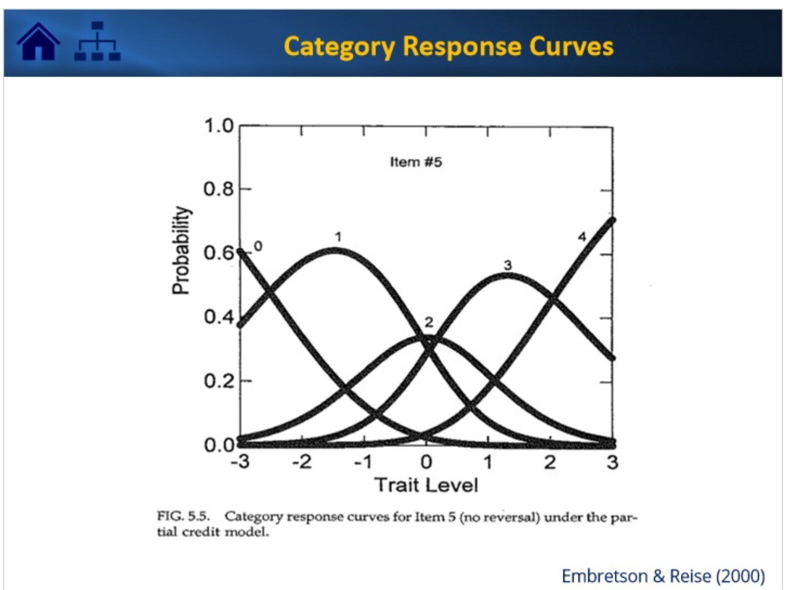
FIG. 5.3. Category response curves for the example items under the partial credit model.

Embretson & Reise (2000)

2.22 The Reparameterized Partial Credit Model III



2.23 The Reparameterized Partial Credit Model IV



2.24 The Rating Scale Model

The Rating Scale Model

$$p_{ix}(\theta) = \frac{\exp\left[\sum_{s=0}^{x}\left(\theta_j - (d_i + c_s)\right)\right]}{\sum_{k=0}^{K}\exp\left[\sum_{s=0}^{k}\left(\theta_j - (d_i + c_s)\right)\right]}$$

Category intersection parameter (c_s)

Item location parameter (d_i)

where $\sum_{s=0}^{0}(\theta - (d_i + c_s)) = 0$

Andrich (1978a, 1978b)

2.25 The Generalized Partial Credit Model I

The Generalized Partial Credit Model

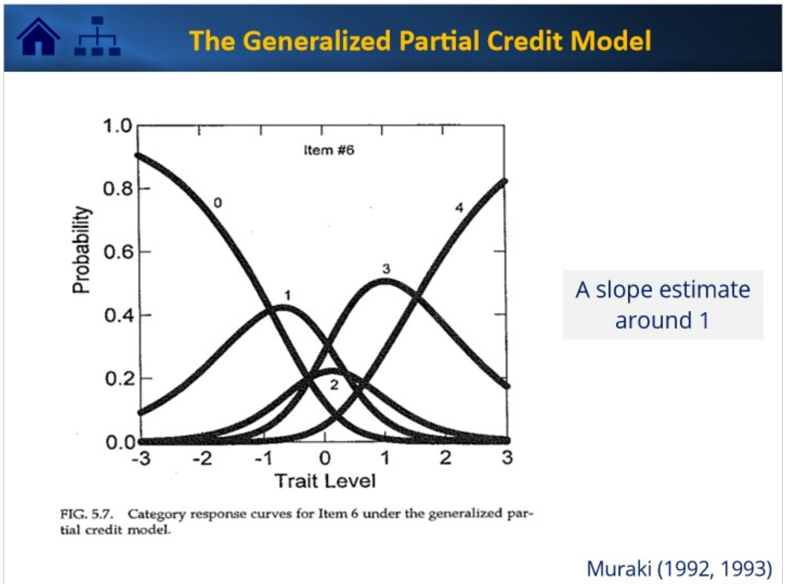
$$p_{ix}(\theta) = \frac{\exp\left[\sum_{s=0}^{x} a_i\left(\theta_j - b_{is}\right)\right]}{\sum_{k=0}^{K}\exp\left[\sum_{s=0}^{k} a_i\left(\theta_j - b_{is}\right)\right]}$$

where $\sum_{s=0}^{0} a_i(\theta - b_{is}) = 0$

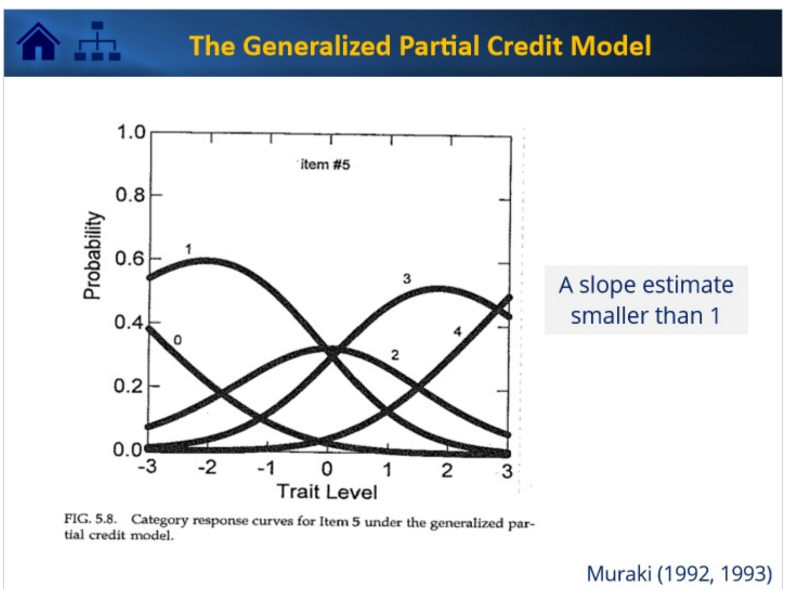
- The point on the latent trait scale at which two consecutive category response curves intersect.
- Indicates where on the latent trait scale the response of one category becomes relatively more likely than the previous category.
- The slope parameter indicate the degree to which categorical responses vary among items as theta level changes.

Muraki (1992, 1993)

2.26 The Generalized Partial Credit Model II



2.27 The Generalized Partial Credit Model III



2.28 The Generalized Partial Credit Model IV

The Generalized Partial Credit Model

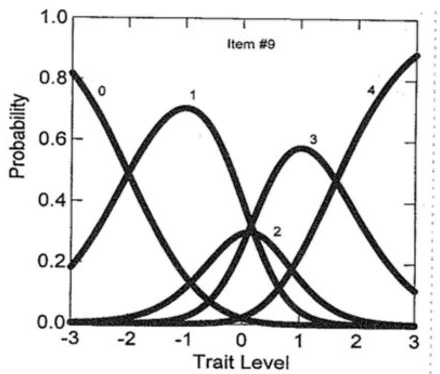


FIG. 5.9. Category response curves for Item 9 under the generalized partial credit model.

A slope estimate larger than 1

Muraki (1992, 1993)

2.29 Logit vs. Probit Models I

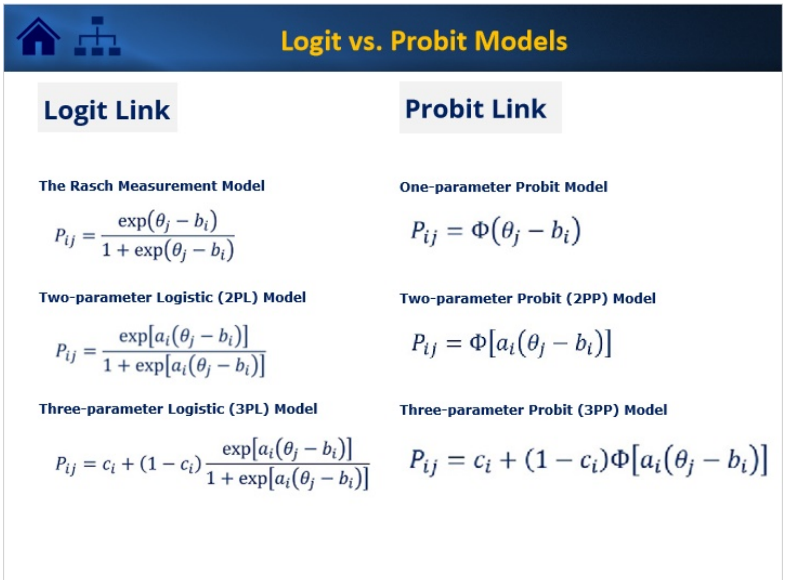
Logit vs. Probit Models

- Both logit and probit link functions can be used in IRT modeling.
- The difference between the probit and logit probabilities are minimal.

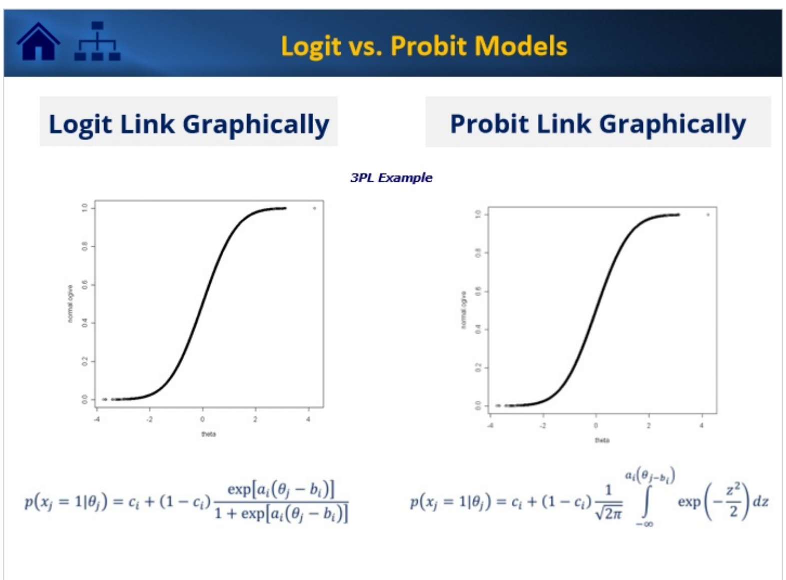
$$|\Phi(x) - \Psi(1.7x)| < .01 \qquad \text{for all } x$$

- By adding a scaling constant of 1.7, the probit link function can approximate the probit link function.

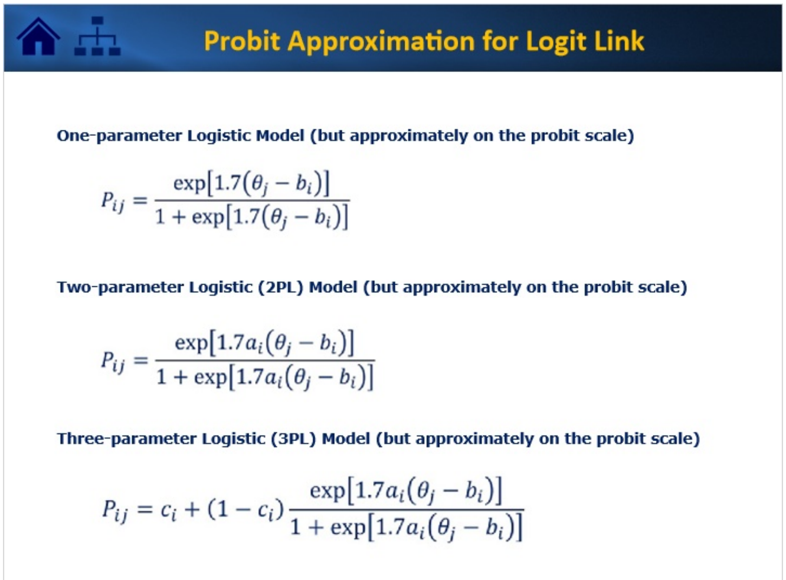
2.30 Logit vs. Probit Models II



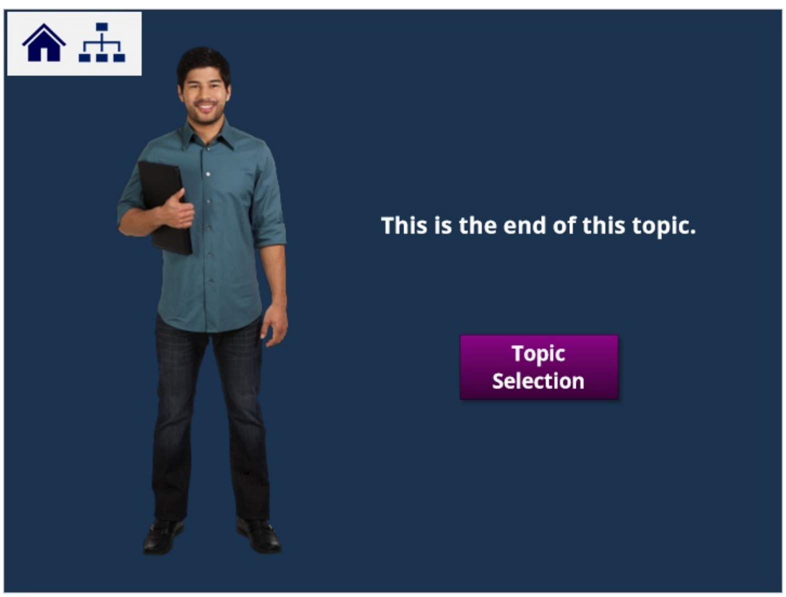
2.31 Logit vs. Probit Models III



2.32 Probit Approximation for Logit Link

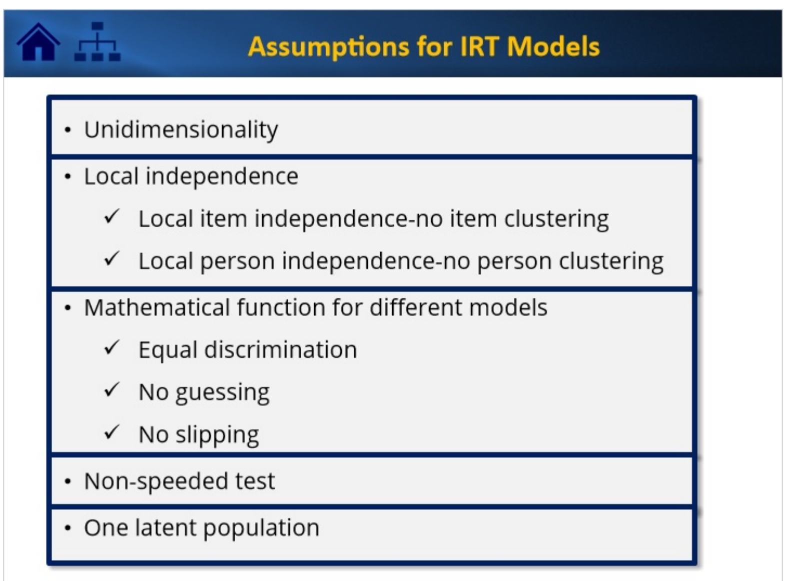


2.33 Bookend: Polytomous Models

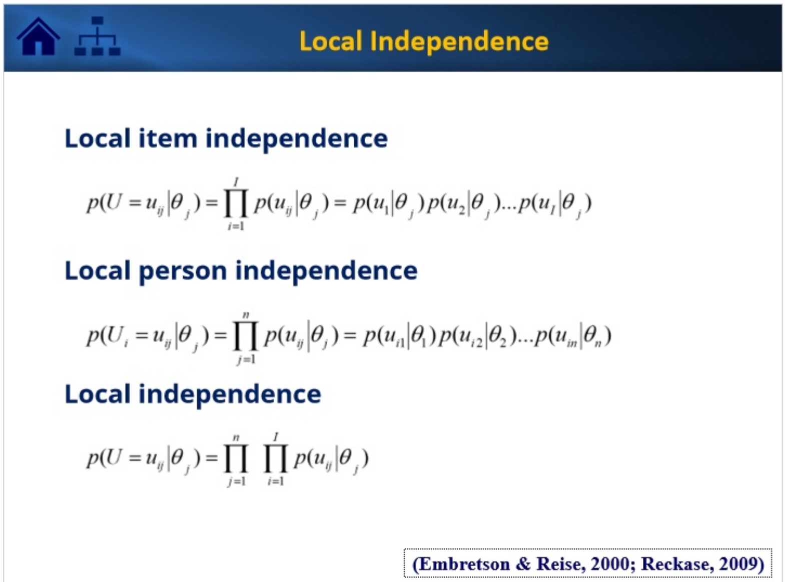


2.34 Bookmark: Model Assumptions

2.35 Assumptions for IRT Models

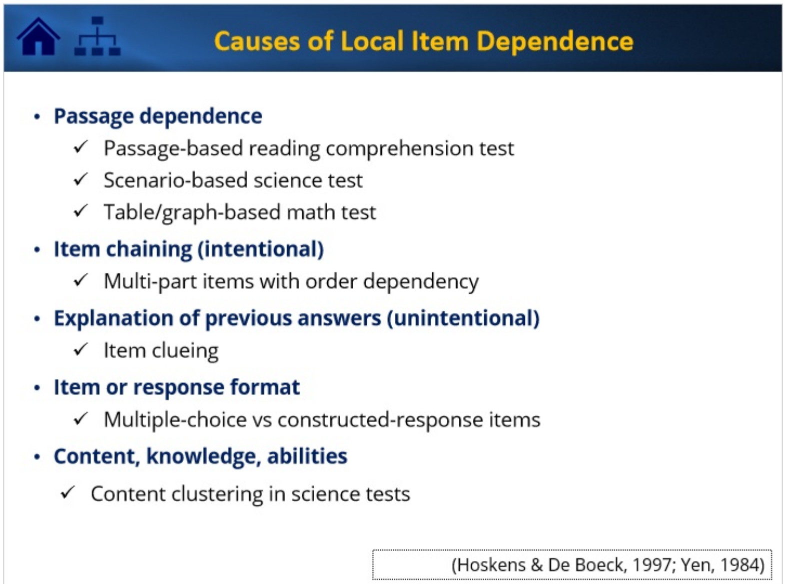


2.36 Local Independence

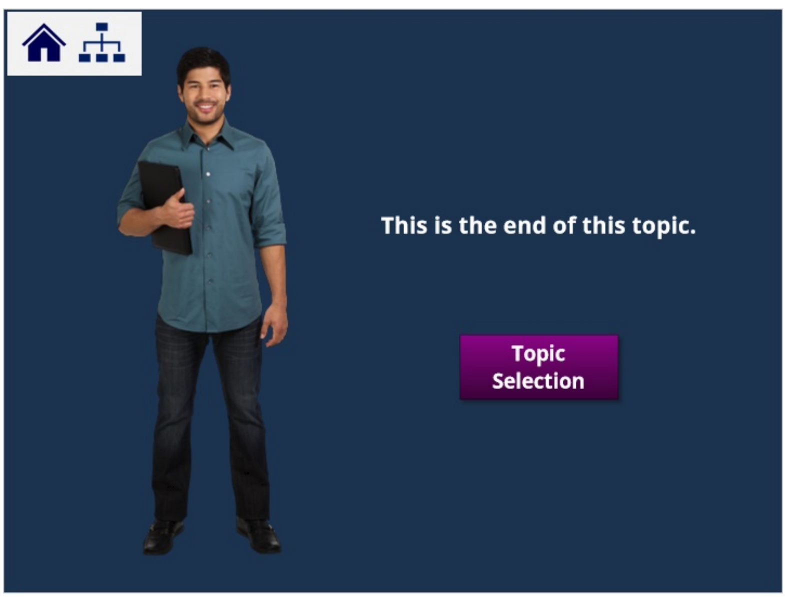


2.37 Violations of Local Independence

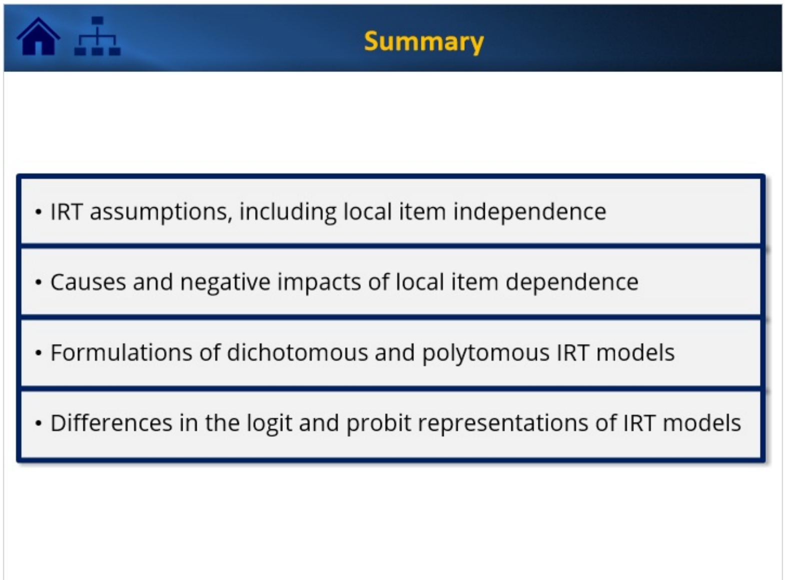
2.38 Causes of Local Item Dependence



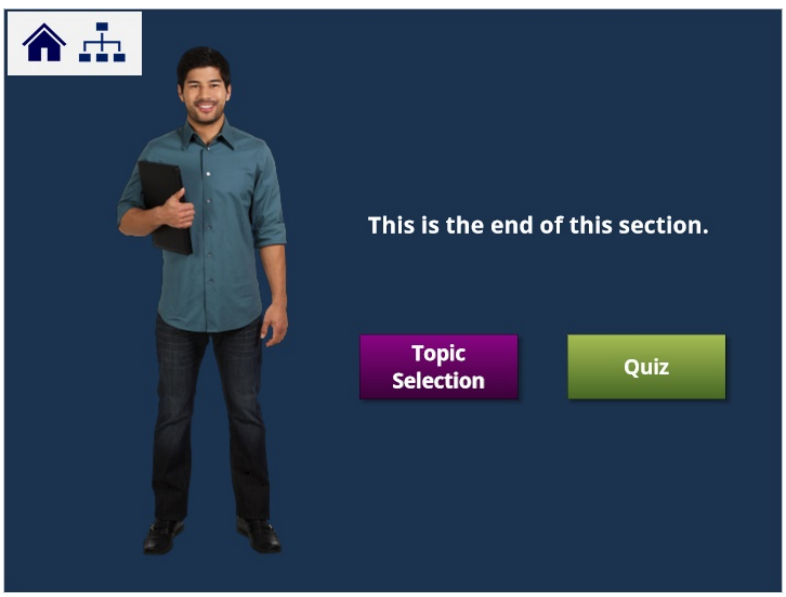
2.39 Bookend: Model Assumptions



2.40 Summary



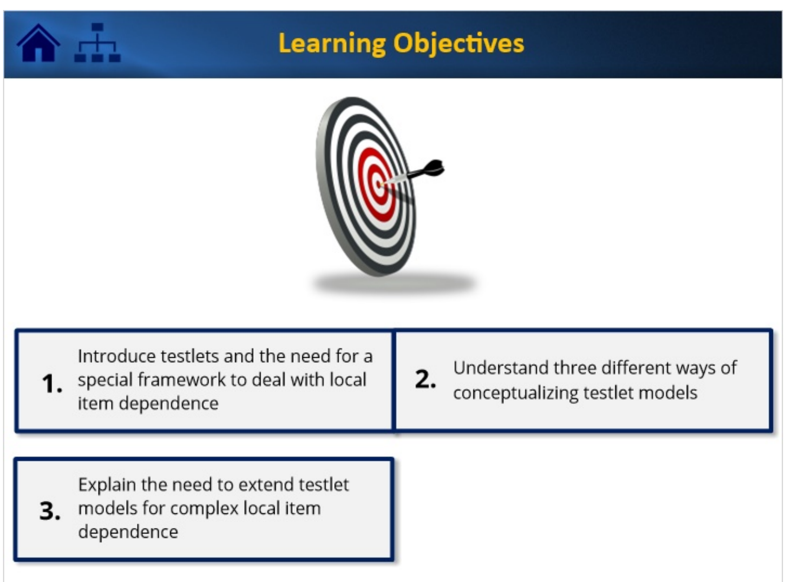
2.41 Bookend: Section 1



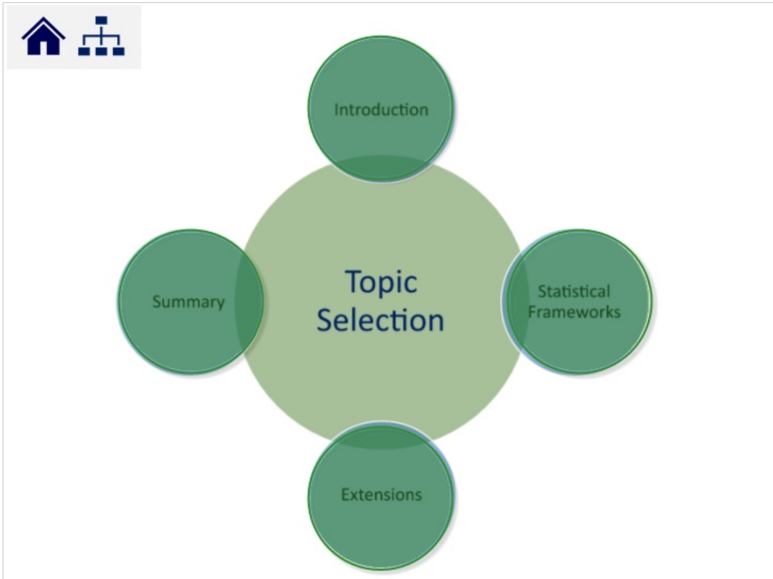
3. Section 2: Testlet Models

3.1 Cover: Section 3

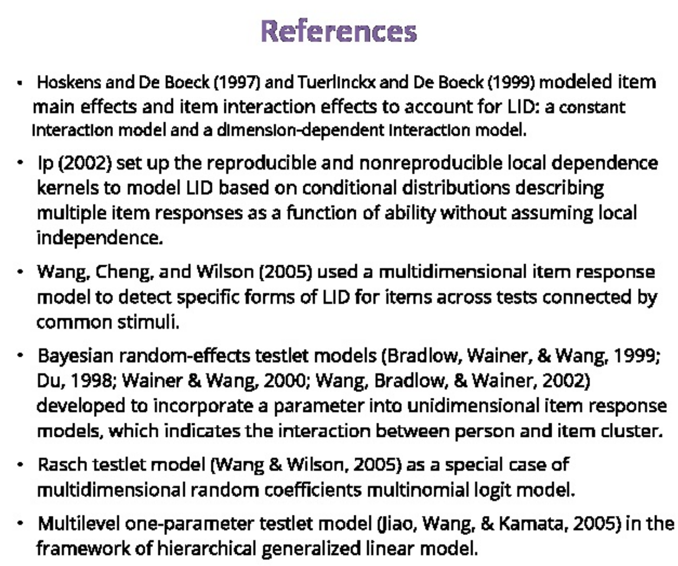
3.2 Learning Objectives: Section 3



3.3 Topic Selection

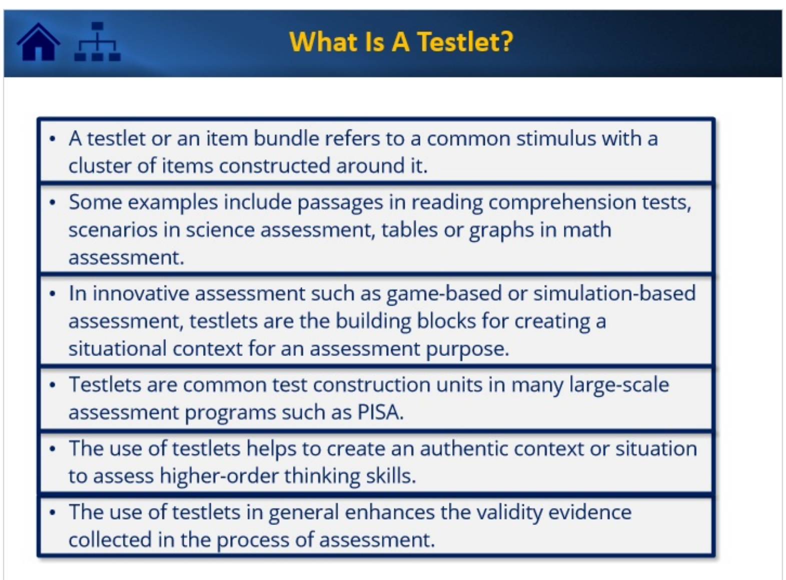


References (Slide Layer)

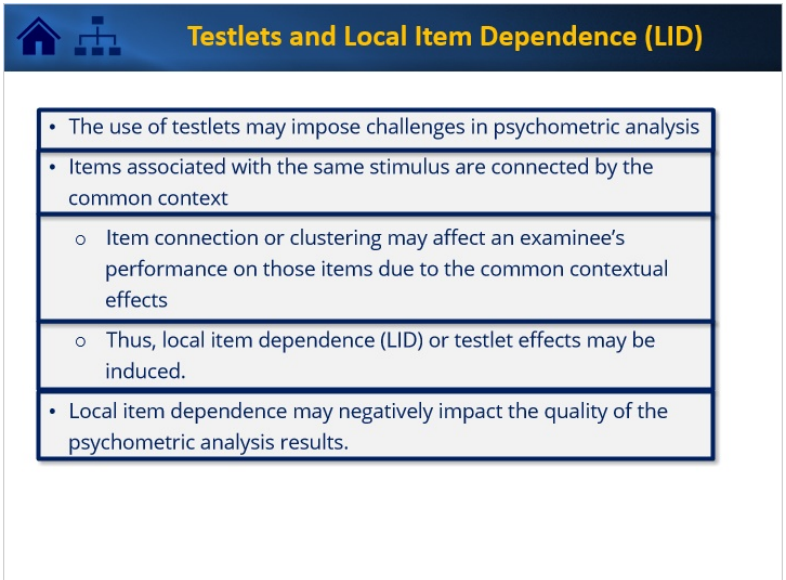


3.4 Bookmark: Introduction

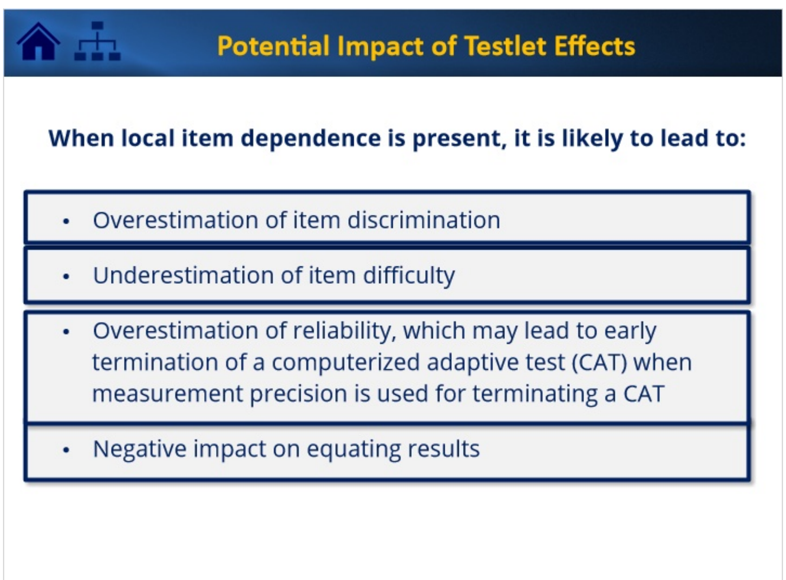
3.5 What Is A Testlet



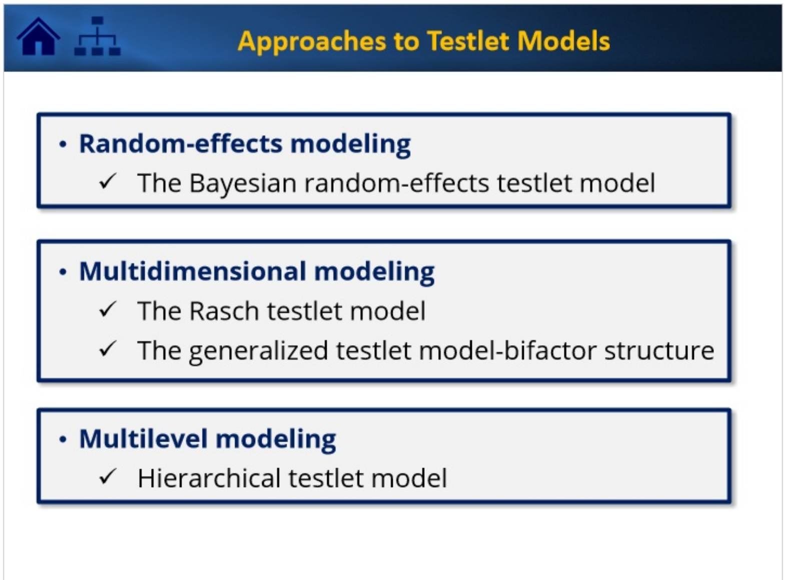
3.6 Testlets and LID



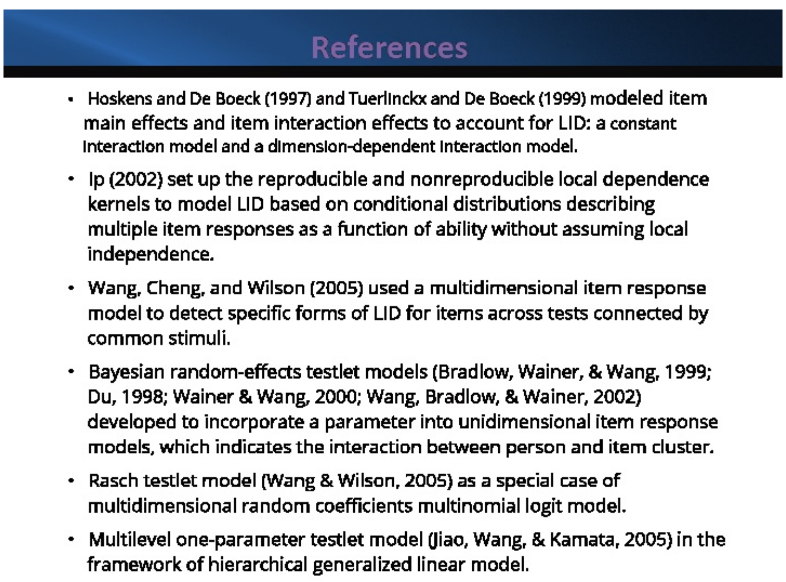
3.7 Potential Impact of Testlet Effects



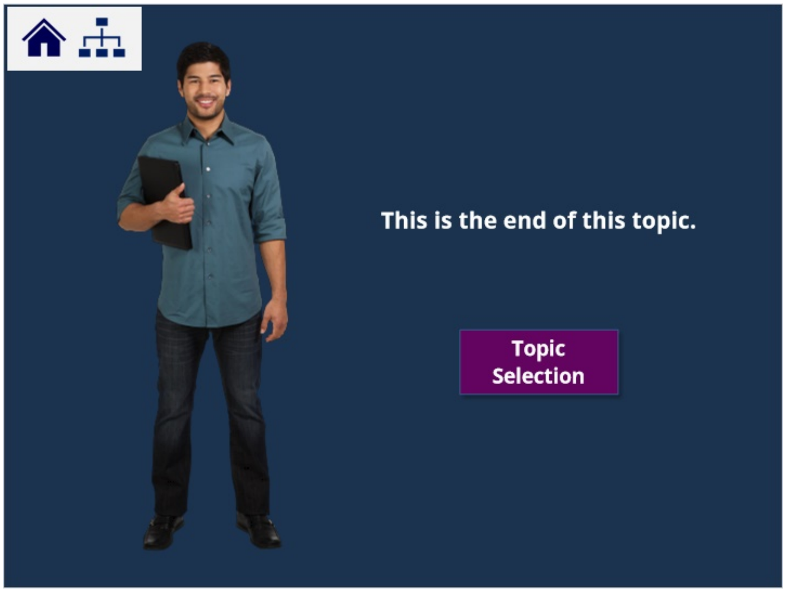
3.8 Approaches to Testlet Models



References (Slide Layer)

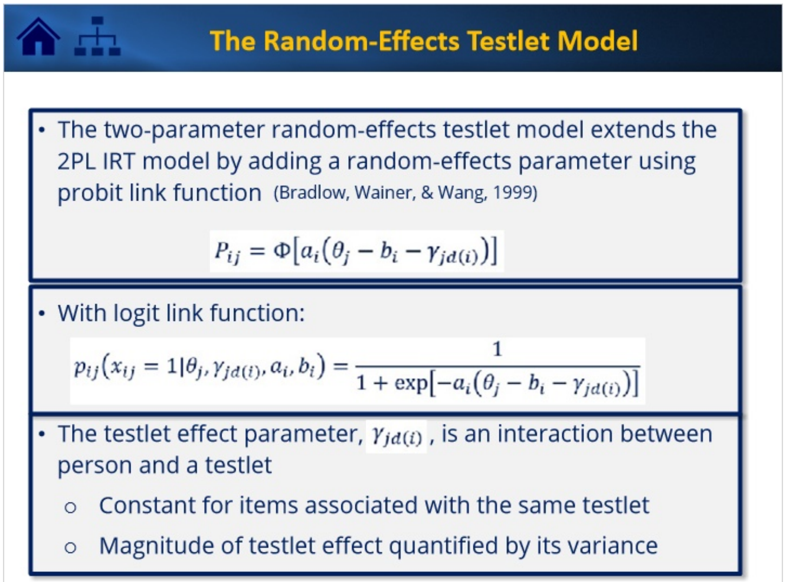


3.9 Bookend: Introduction

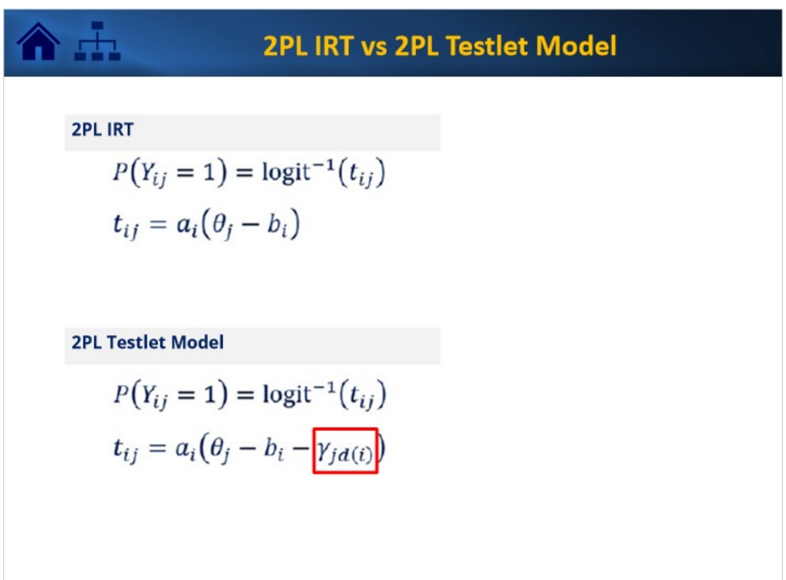


3.10 Bookmark: Statistical Frameworks

3.11 The Random-Effects Testlet Model



3.12 The Two-Parameter Logistic Testlet Model



3.13 The Rasch Testlet Model I

The Rasch Testlet Model

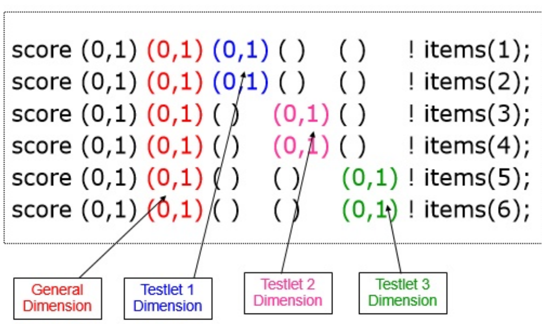
- Formulated from a multidimensional perspective
- Proposed as a special case of multidimensional random coefficients multinomial logit model by adding an additional dimension related to each testlet (Wang & Wilson, 2005)

$$p_{ij}(x_{ij} = 1|\theta_j, \gamma_j, b_i) = \frac{1}{1 + \exp[-(\theta_j + \gamma_{jd(i)} - b_i)]}$$

- The ConQuest syntax on the next slide graphically shows the formulation of the multidimensional structure

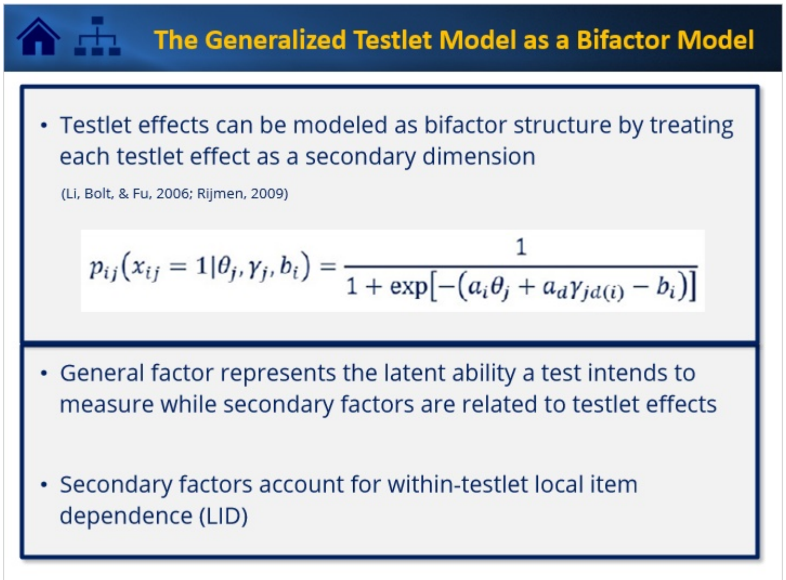
3.14 The Rasch Testlet Model II

The Rasch Testlet Model

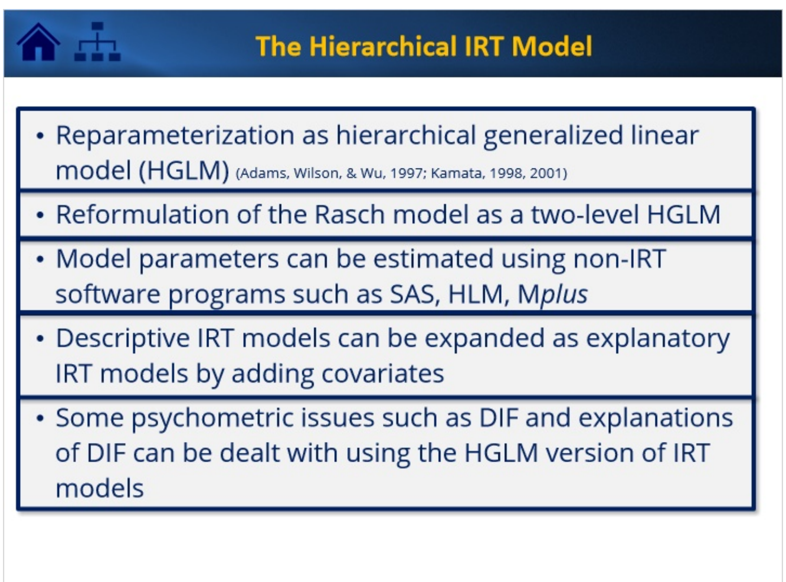


ConQuest syntax

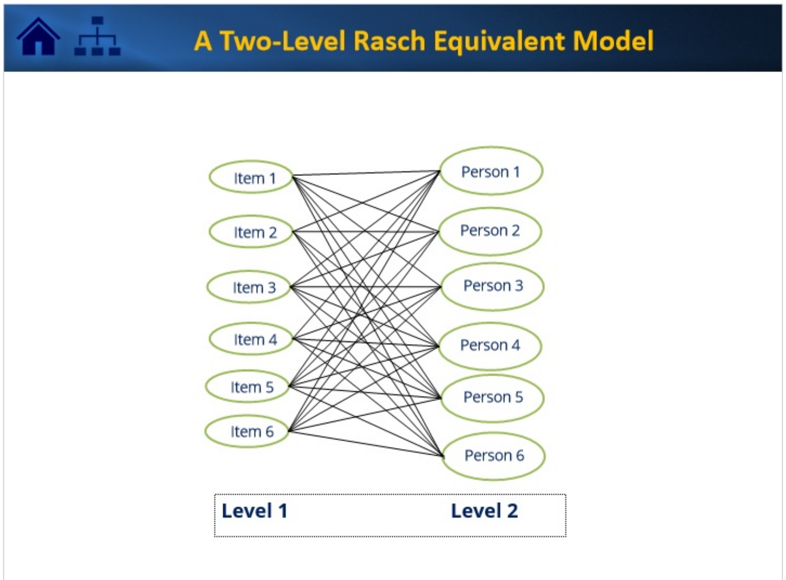
3.15 The Generalized Testlet Model as a Bifactor Model



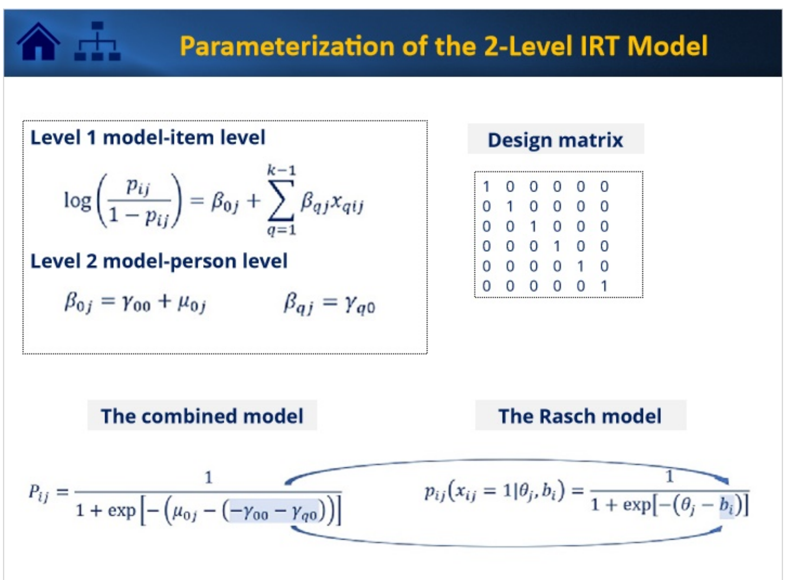
3.16 The Hierarchical IRT Model



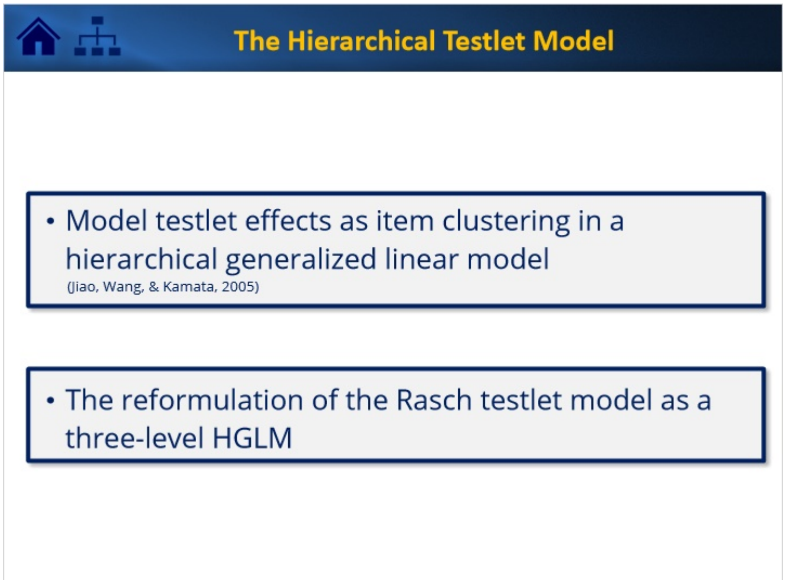
3.17 A Two-Level Rasch Equivalent Model



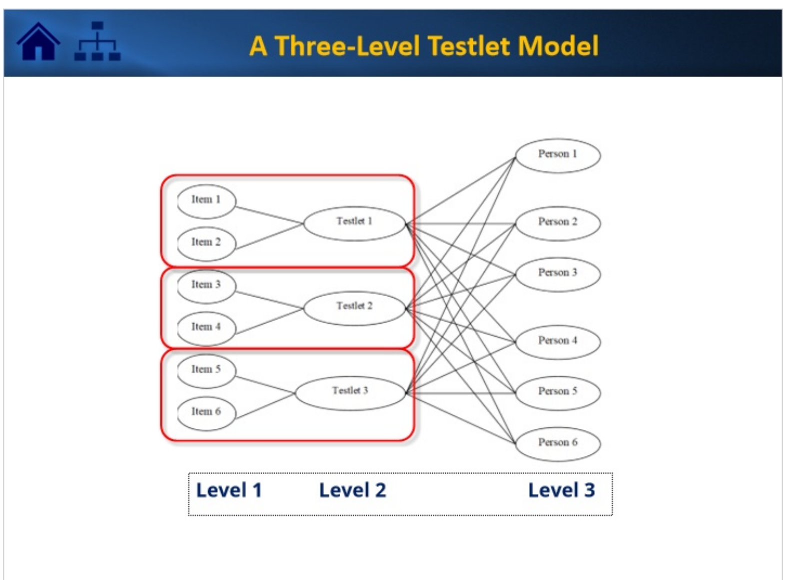
3.18 Parameterization of the Two-Level IRT Model



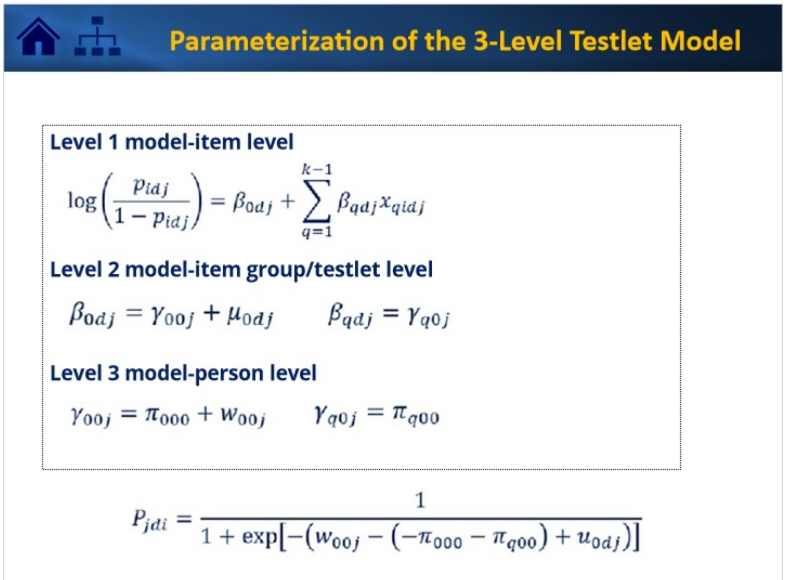
3.19 The Hierarchical Testlet Model



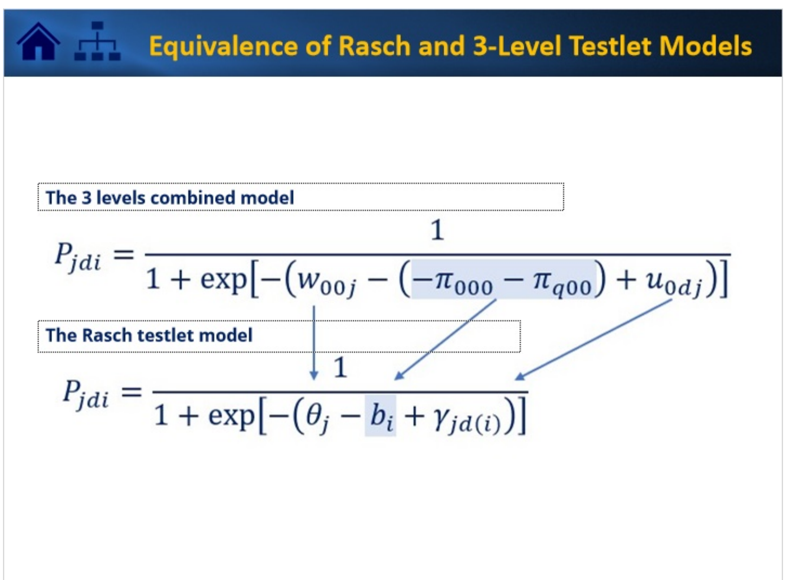
3.20 A Three-Level Testlet Model



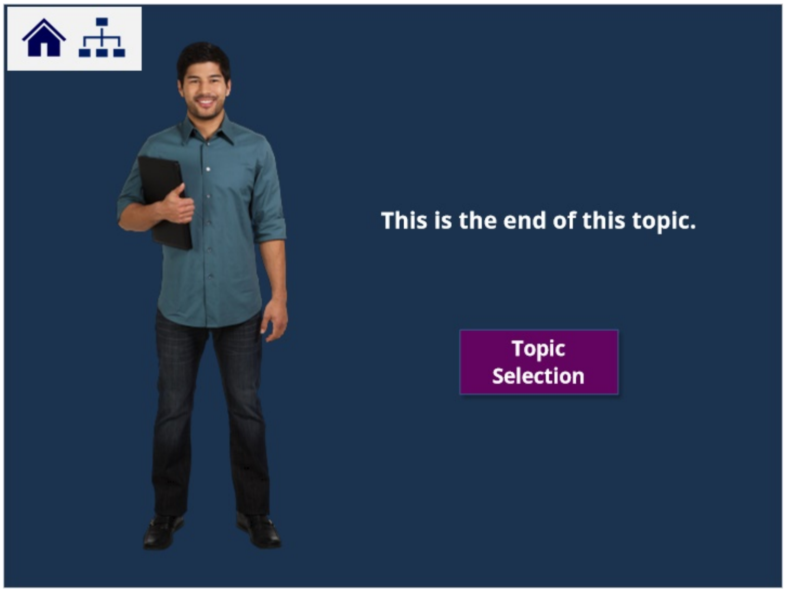
3.21 Parameterization of the Three-Level Testlet Model



3.22 Equivalence of Rasch and Three-Level Testlet Models

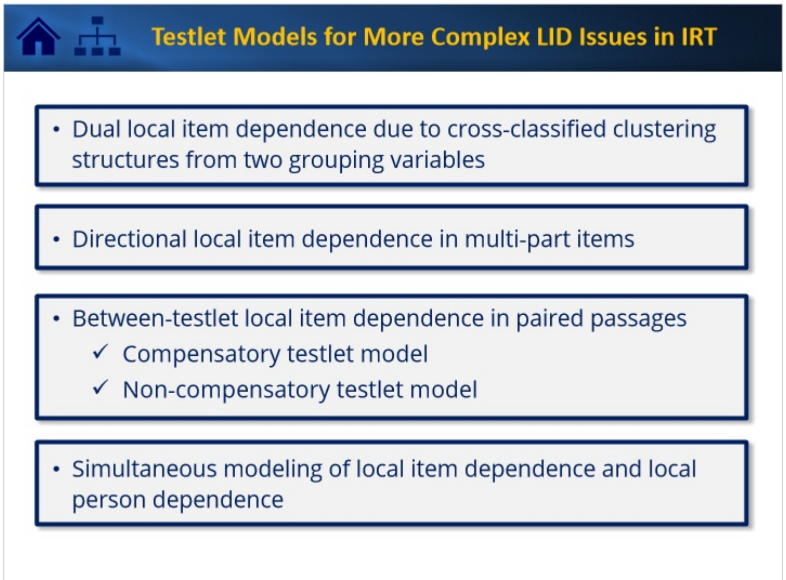


3.23 Bookend: Statistical Frameworks

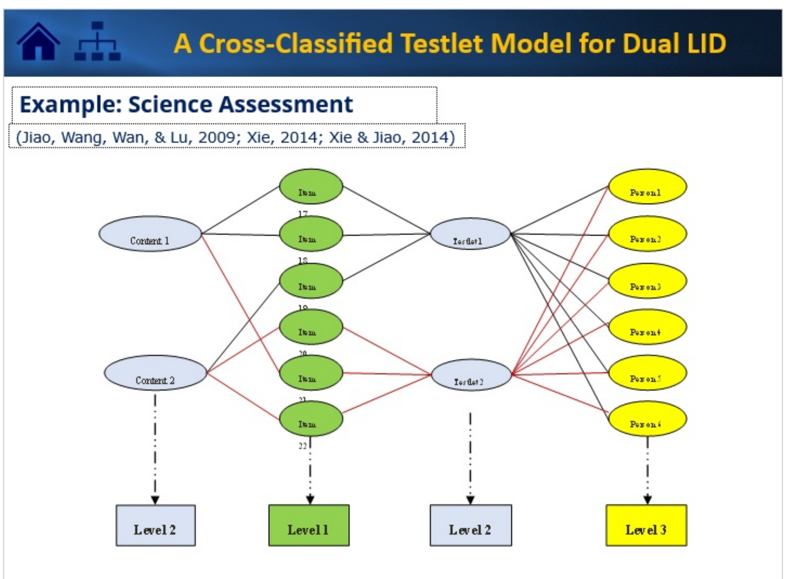


3.24 Bookmark: Extensions

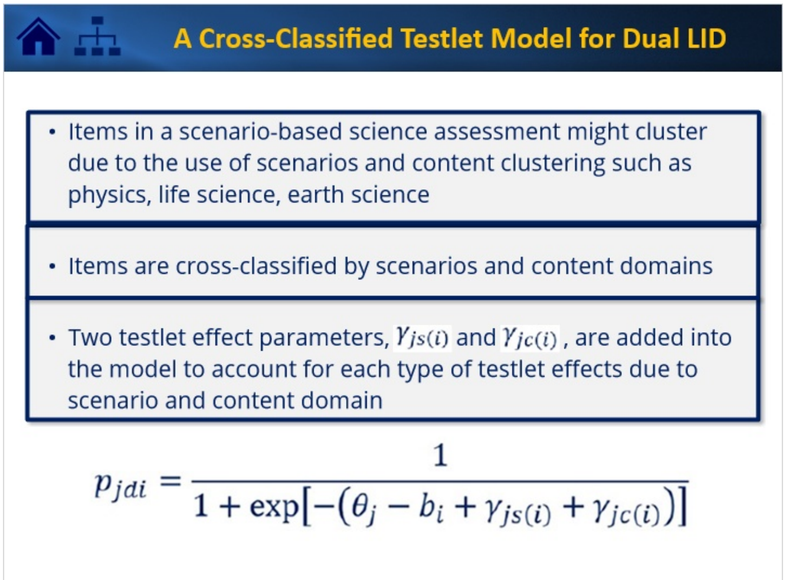
3.25 Testlet Models for More Complex Local Item Dependence Issues in Item Response Modeling



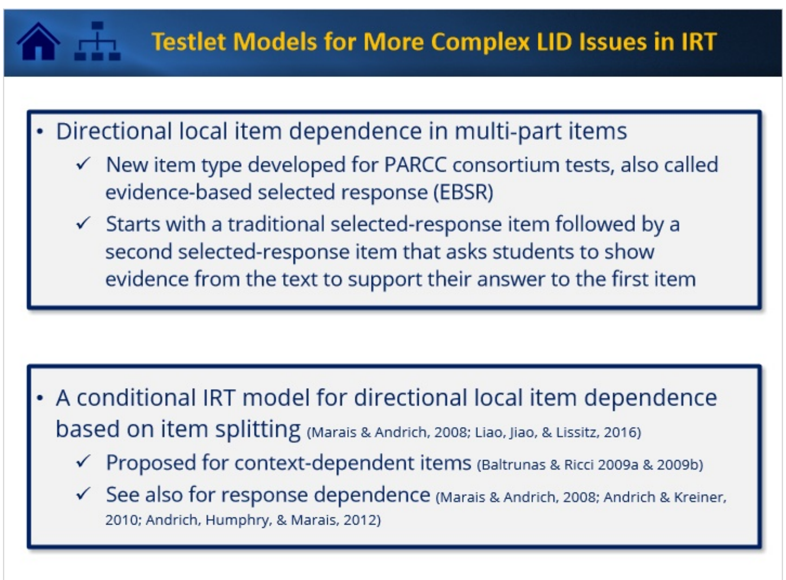
3.26 A Cross-Classified Testlet Model for Dual LID I



3.27 A Cross-Classified Testlet Model for Dual LID II



3.28 Testlet Models for More Complex LID Issues in IRT



3.29 Item Splitting I

Item Splitting

Student	Item 1	Item 2 – Original	Item 2 – Recoded	
			Item 2a	Item 2b
1	1	1	1	x
2	0	1	x	1
3	0	0	x	0
4	0	1	x	1
5	1	0	0	x
6	1	1	1	x
7	1	1	1	x
8	0	0	x	0
9	1	0	0	x
10	0	1	x	1

3.30 Item Splitting II

Item Splitting

- IRT model for directional LID based on item splitting
  (Marais & Andrich, 2008; Liao, Jiao, & Lissitz, 2016)

$$P(x_{2j} = 1 | x_{1j} = x) = \frac{\exp[a_2(\theta_j - (b_2 - xd))]}{1 + \exp[a_2(\theta_j - (b_2 - xd))]}$$

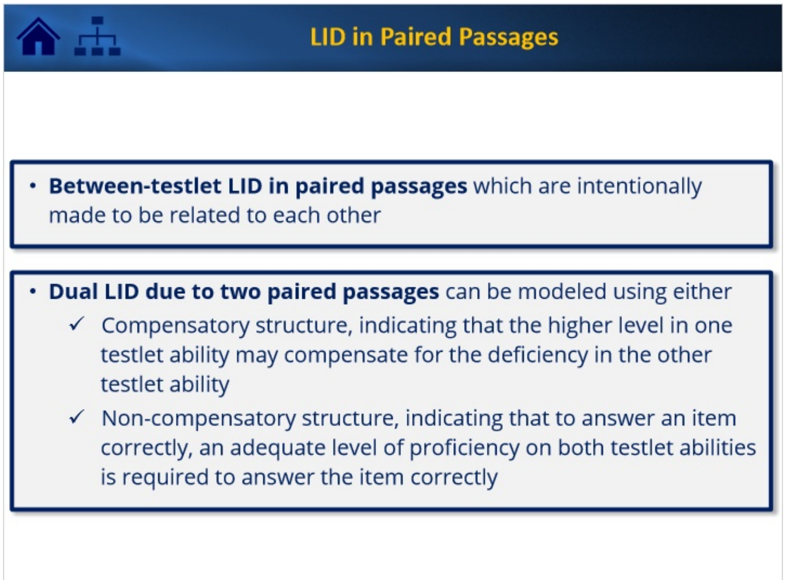
- Probability of second item conditional on getting first item wrong

$$P(x_{2j} = 1 | x_{1j} = 0) = \frac{\exp[a_2(\theta_j - b_2)]}{1 + \exp[a_2(\theta_j - b_2)]}$$

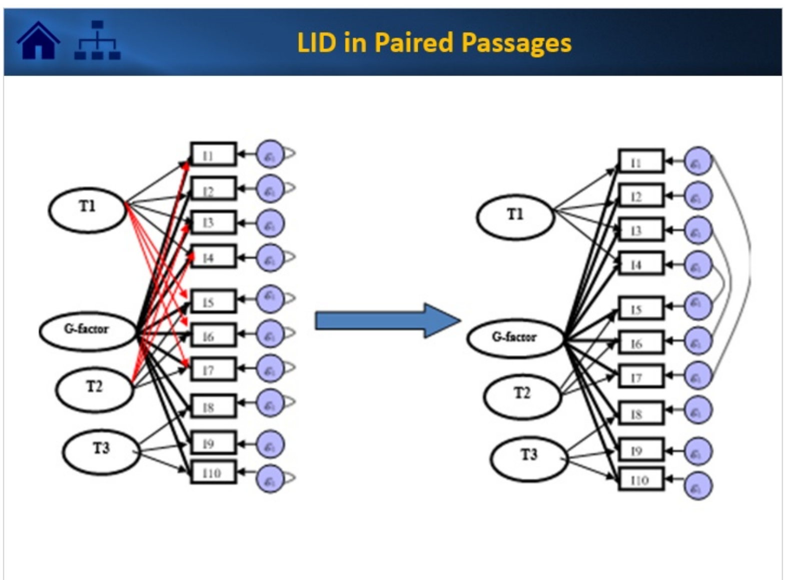
- Probability of second item conditional on getting first item right
  (Includes non-negative parameter *d*)

$$P(x_{2j} = 1 | x_{1j} = 1) = \frac{\exp[a_2(\theta_j - (b_2 - d))]}{1 + \exp[a_2(\theta_j - (b_2 - d))]}$$

3.31 LID in Paired Passages I



3.32 LID in Paired Passages II



3.33 Compensatory and Non-Compensatory Models

Compensatory and Non-Compensatory Models

A Non-Compensatory Two-Parameter Testlet Model

$$P(x_{ij}|a_{1i}, a_{2i}, b_i, \theta_j, \gamma_{jt}) = \prod \frac{1}{1 + \exp[-(a_{1i} * \theta_j + a_{2i} * \gamma_{jt} - b_i)]}$$

A Compensatory Two-Parameter Testlet Model (Jiao & Shu, 2014)

$$P(x_{ij}|a_{1i}, a_{2i}, a_{3i}, b_i, \theta_j, \gamma_{jt1}, \gamma_{jt2}) = \frac{1}{1 + \exp[-(a_{1i} * \theta_j + a_{2i} * \gamma_{jt1} + a_{3i} * \gamma_{jt2} - b_i)]}$$

3.34 Dual LID and LPD

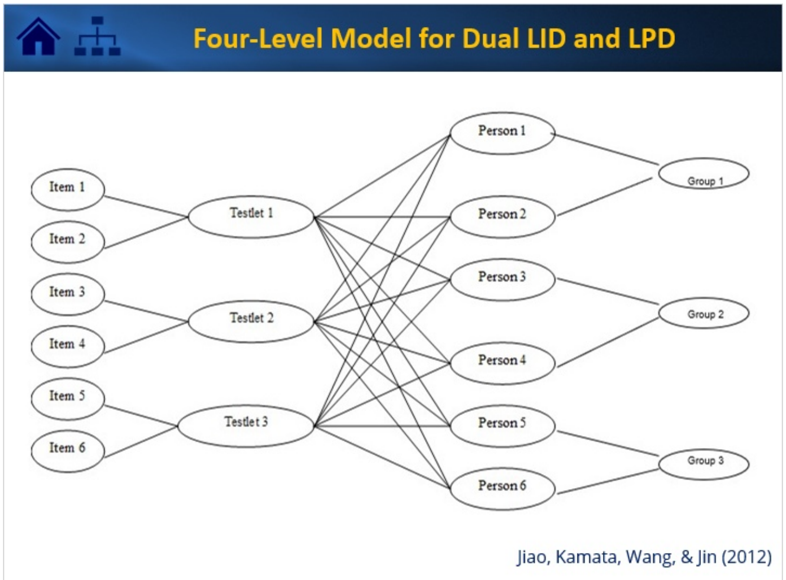
Dual LID and LPD

Simultaneous modeling of local item dependence (LID) and local person dependence (LPD):

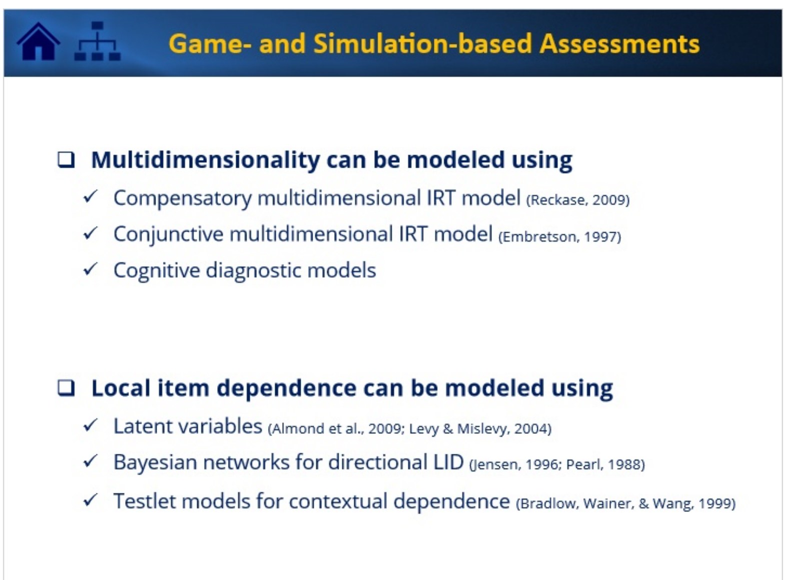
- Modeling of local item dependence and local person dependence simultaneously
- Person clustering effect is modeled by adding a group-specific ability parameter, θ_g
- Testlet effects are modeled as in the standard testlet model

$$P(x_{ij}|b_i, \theta_j, \gamma_{jt}, \theta_g) = \frac{1}{1 + \exp[-(\theta_j + \gamma_{jt(i)} + \theta_g - b_i)]}$$

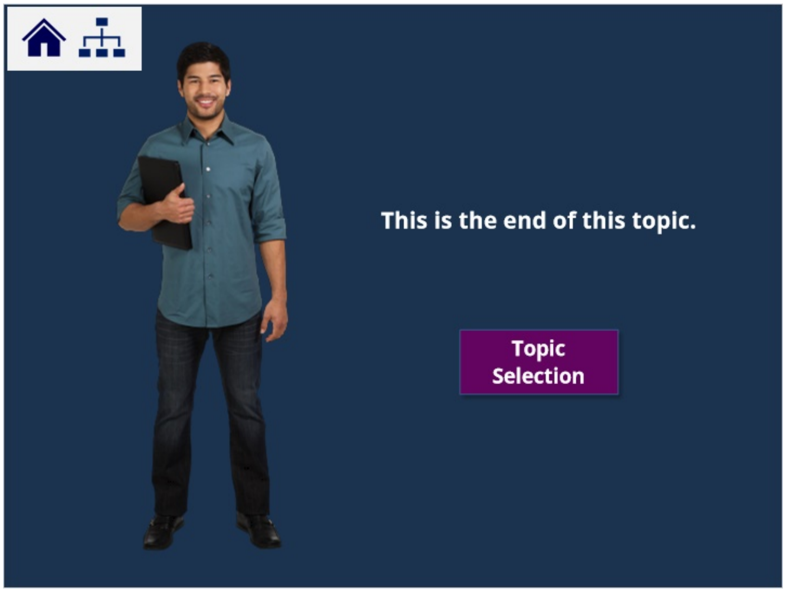
3.35 Four-Level Model for Dual LID and LPD



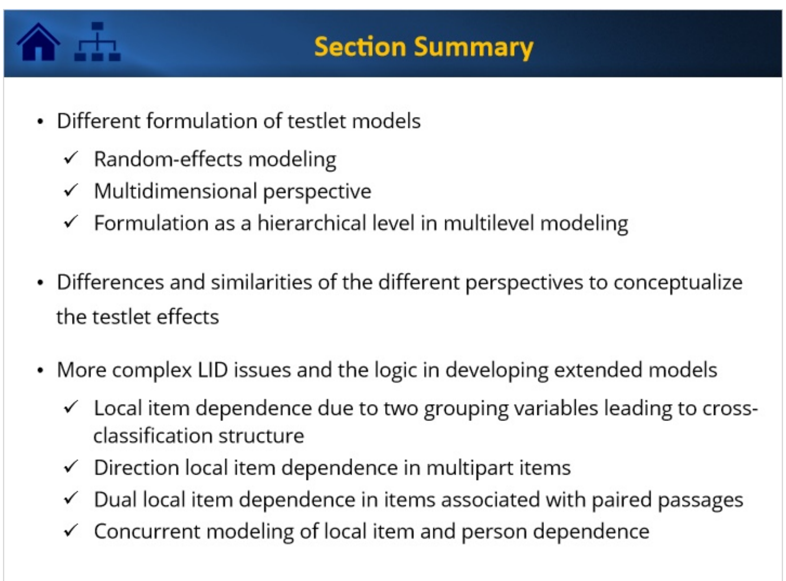
3.36 Cutting-Edge Examples



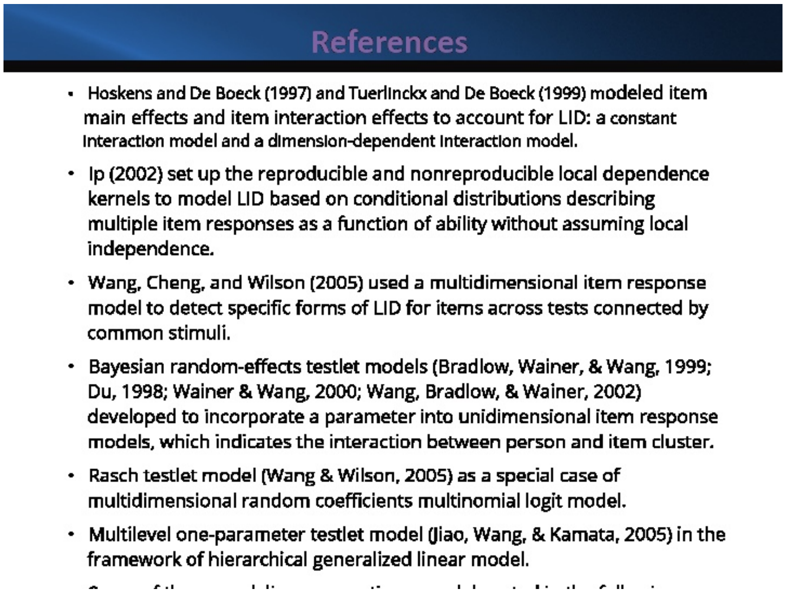
3.37 Bookend: Extensions



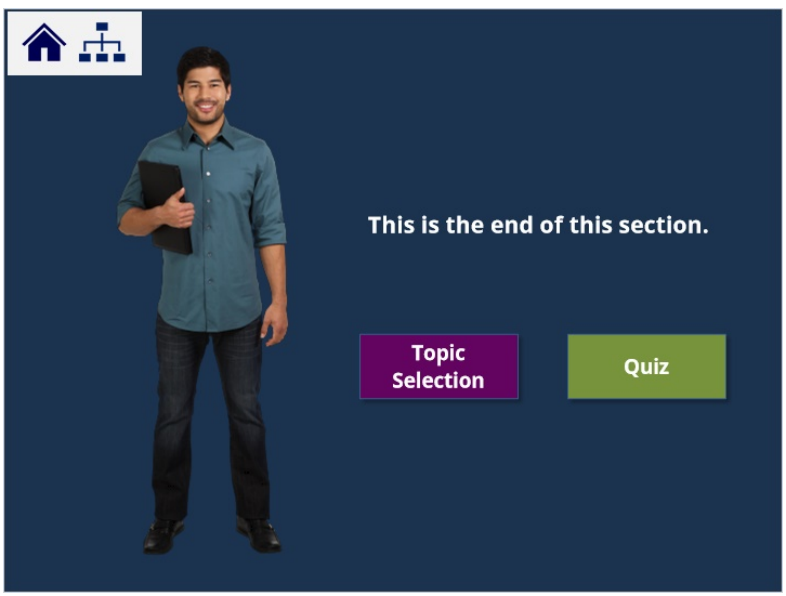
3.38 Section Summary



References (Slide Layer)

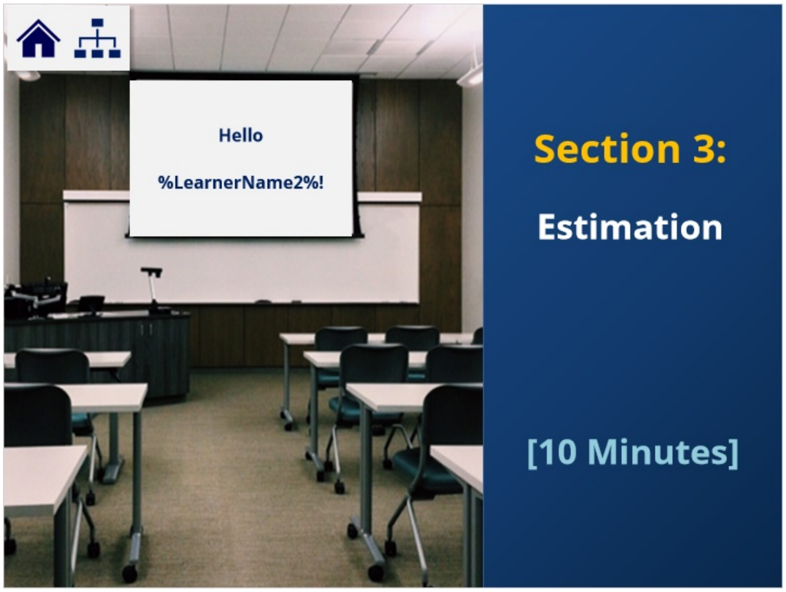


3.39 Bookend: Section 3

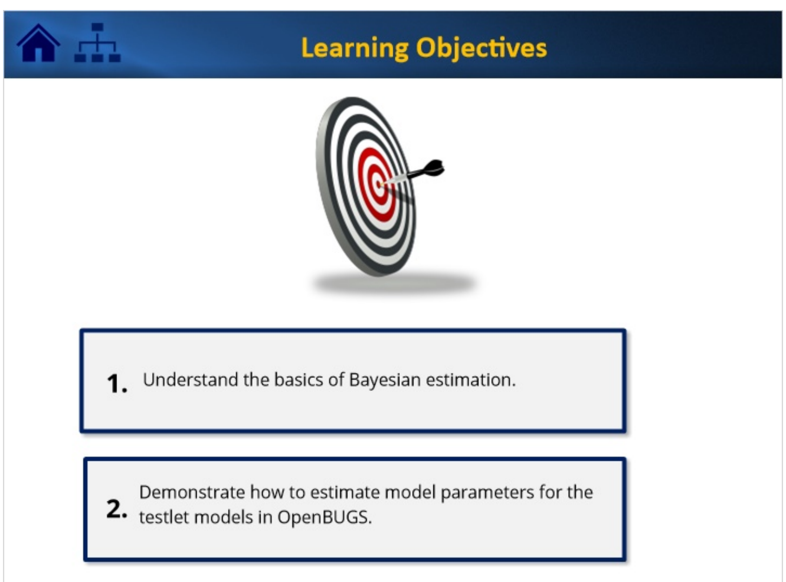


4. Section 3: Estimation

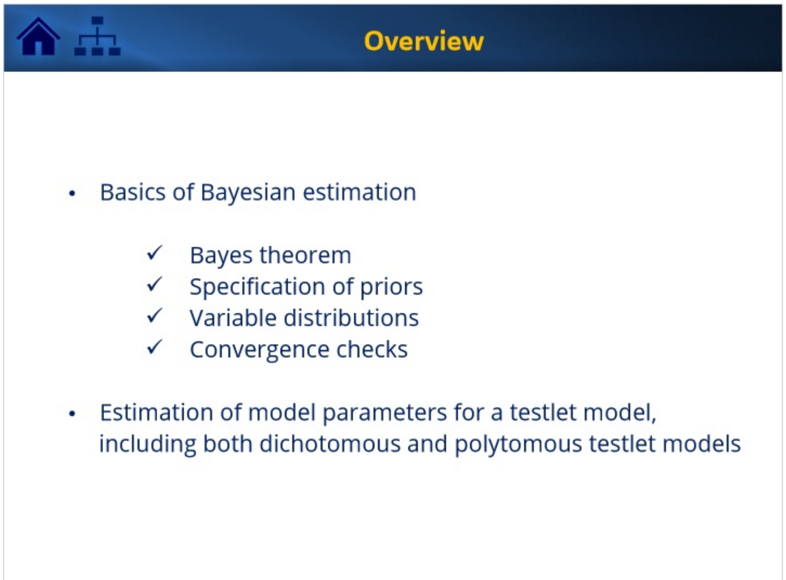
4.1 Cover: Section 3



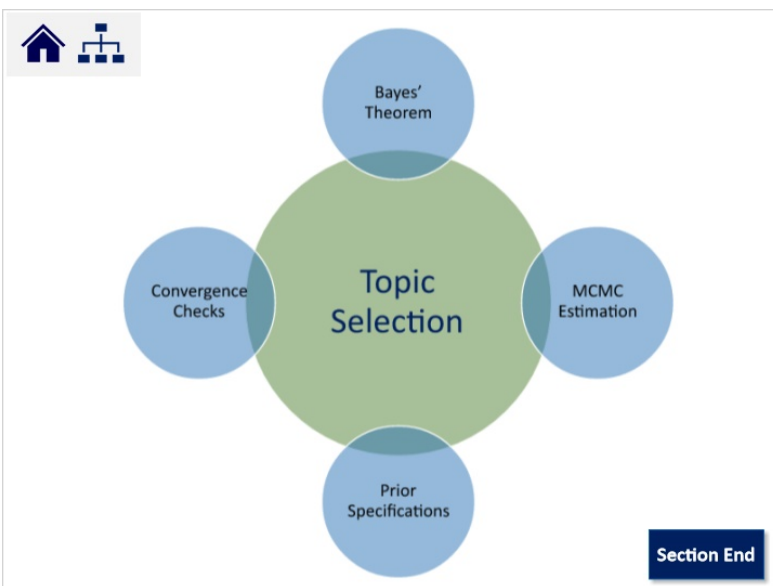
4.2 Learning Objectives: Section 4



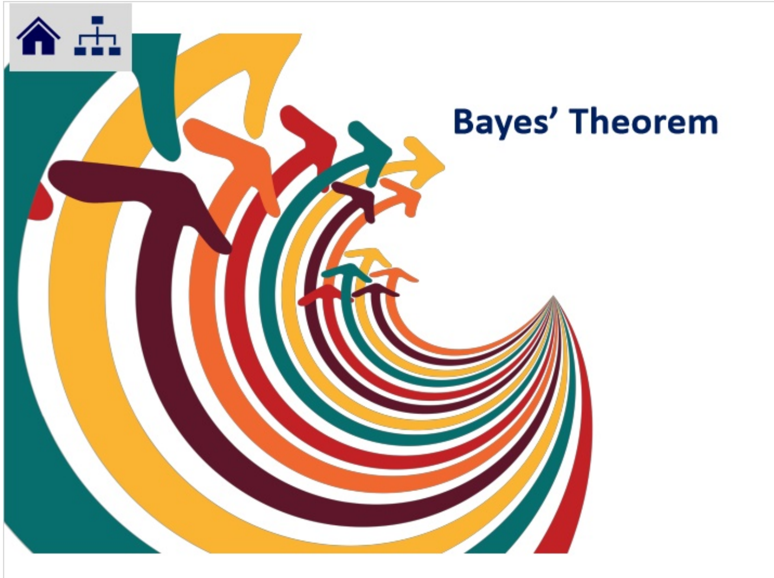
4.3 Summary



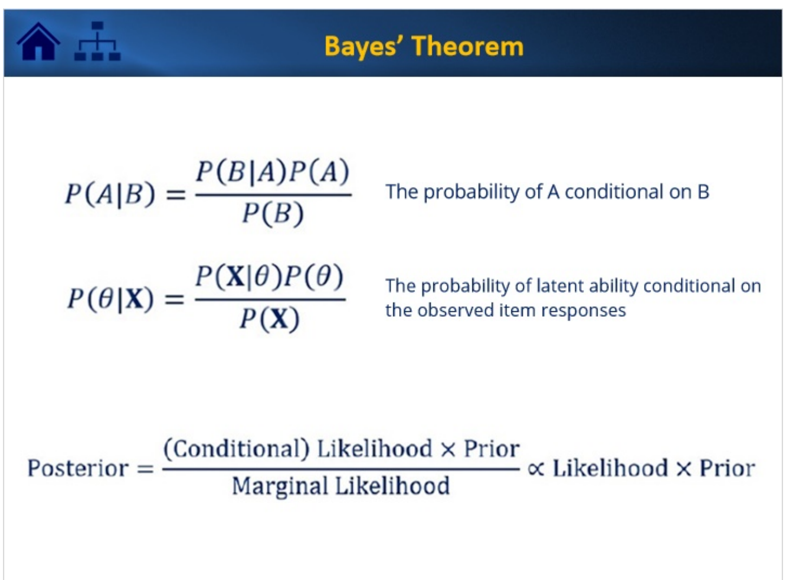
4.4 Topic Selection



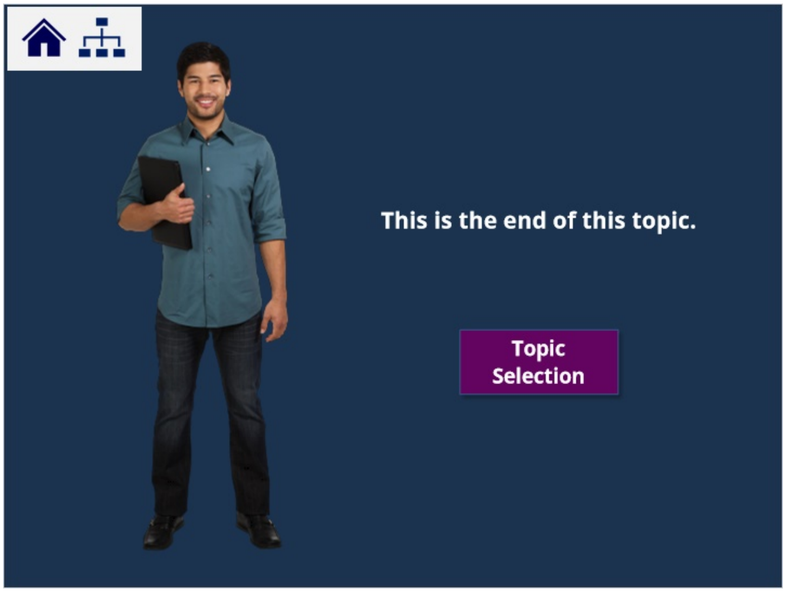
4.5 Bookmark: Bayes' Theorem



4.6 Bayes' Theorem

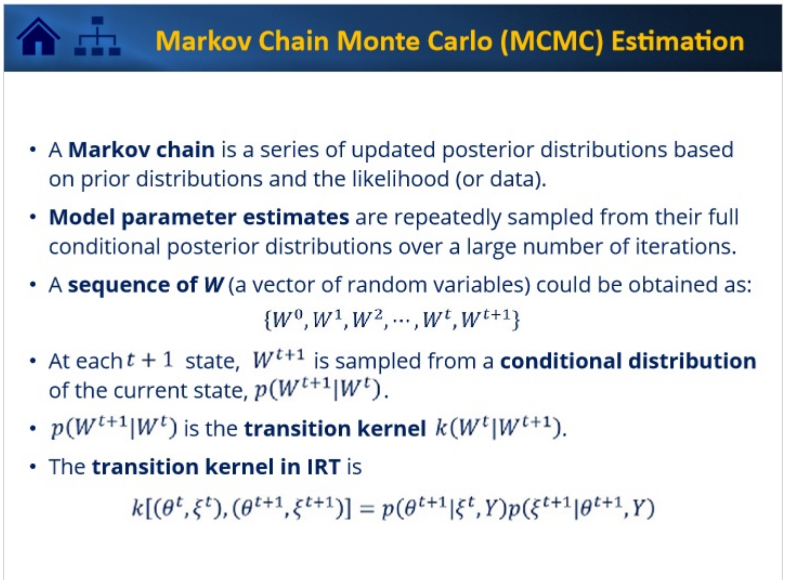


4.7 Bookend: Bayes' Theorem

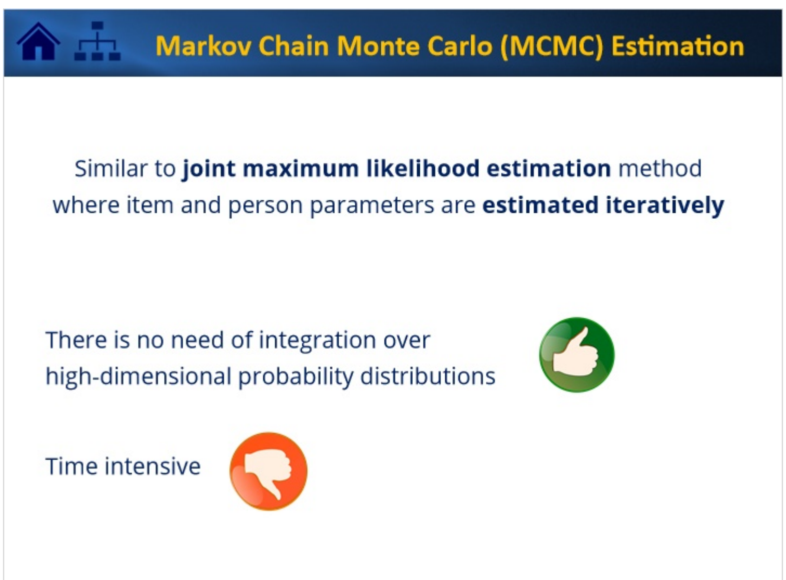


4.8 Bookmark: MCMC Estimation

4.9 Markov Chain Monte Carlo (MCMC) Estimation I



4.10 Markov Chain Monte Carlo (MCMC) Estimation II



4.11 Gibbs Sampler

Gibbs Sampler

Step 1: Initiate $\theta_{1,0}, \cdots, \theta_{p,0}$
Step 2: Generate new values at iteration i as follows:

$$\theta_{1,i} \sim \pi(\theta_1 | \theta_{2,i-1}, \theta_{3,i-1}, \cdots, \theta_{p,i-1})$$
$$\theta_{2,i} \sim \pi(\theta_2 | \theta_{1,i}, \theta_{3,i-1}, \cdots, \theta_{p,i-1})$$
$$\vdots$$
$$\theta_{p,i} \sim \pi(\theta_p | \theta_{1,i}, \theta_{2,i}, \cdots, \theta_{p-1,i})$$

Step 3: Repeat until convergence.

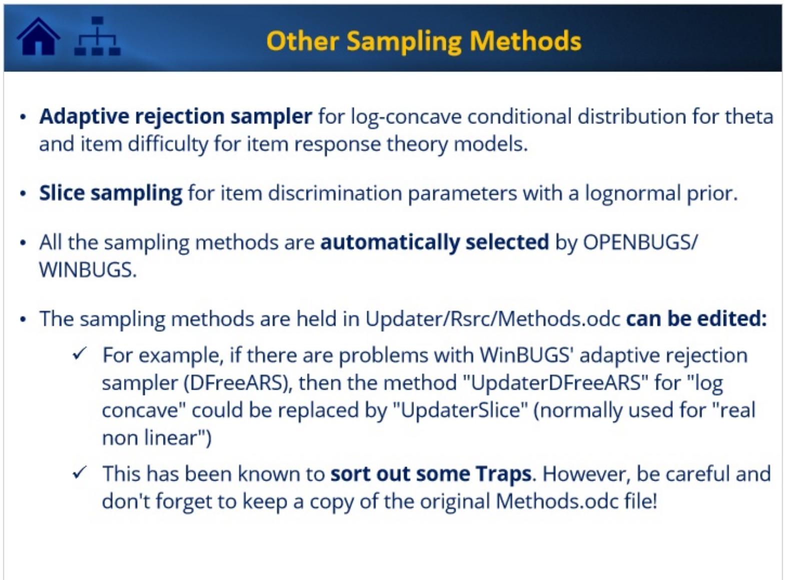
θ may denote an unknown parameter of interest, not necessarily the ability parameter in IRT

4.12 Metropolis Hastings Algorithm

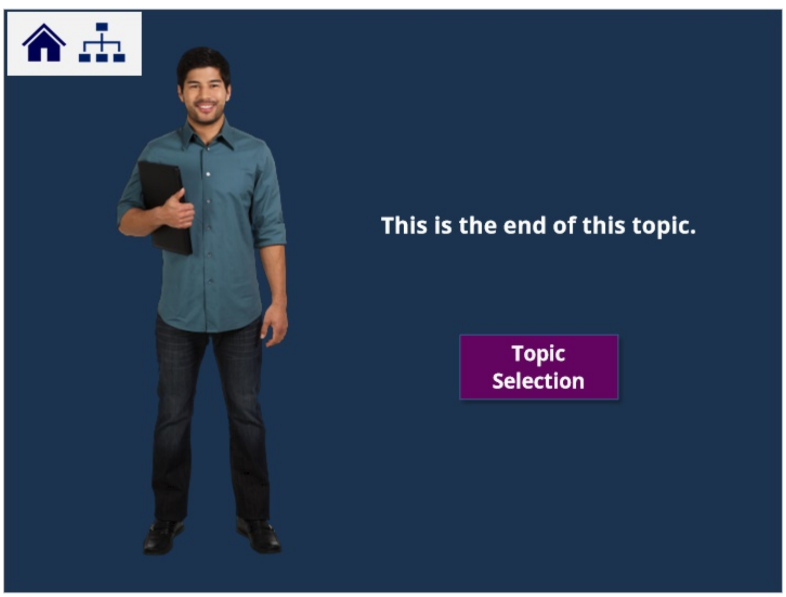
Metropolis Hastings Algorithm

Step 1: Establish the set of starting values for the parameters: $\boldsymbol{\theta}^{j=0}$
Step 2: Set $j = j + 1$, and draw a set of candidate values, $\boldsymbol{\theta}^*$, from a proposal distribution
Step 3: Compute the ratio $r = \frac{\pi(\theta^* | y)}{\pi(\theta^{t-1} | y)} \frac{J_t(\theta^{t-1} | \theta^*)}{J_t(\theta^* | \theta^{t-1})}$ and determine whether it is smaller or larger than 1. Choose the smaller of the two and denote it by α
Step 4: Draw a random value u from a uniform distribution defined over (0,1)
Step 5: Compare α with u. If $\alpha > u$, then set $\boldsymbol{\theta}^j = \theta^*$; otherwise, set $\boldsymbol{\theta}^j = \boldsymbol{\theta}^{j-1}$
Step 6: Return to Step 2 until enough candidate values are sampled

4.13 Other Sampling Methods

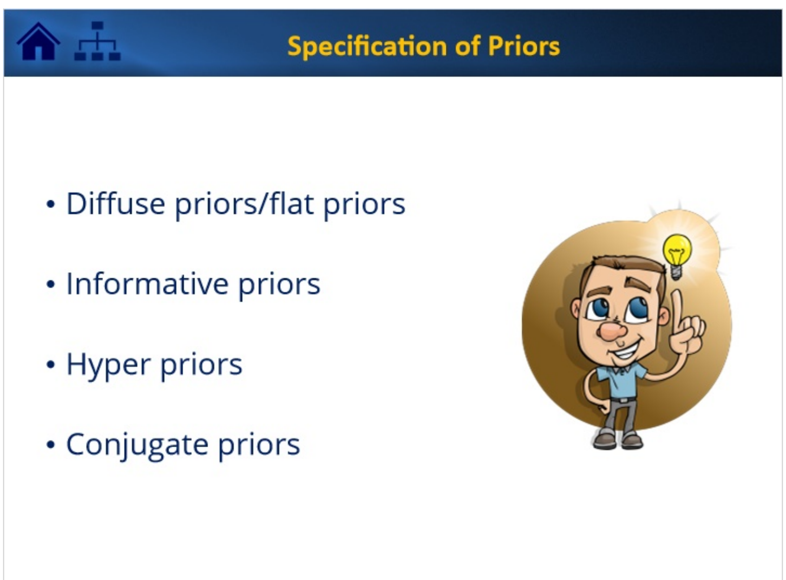


4.14 Bookend: MCMC Estimation

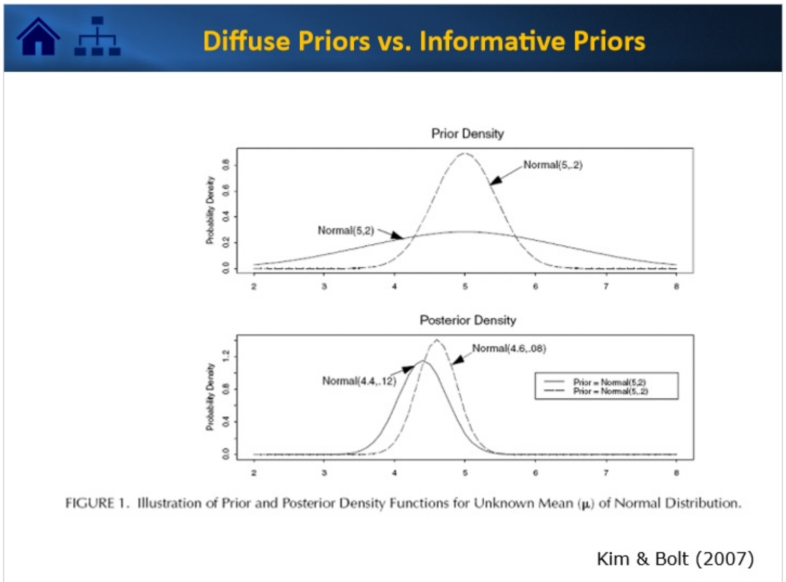


4.15 Bookmark: Prior Specifications

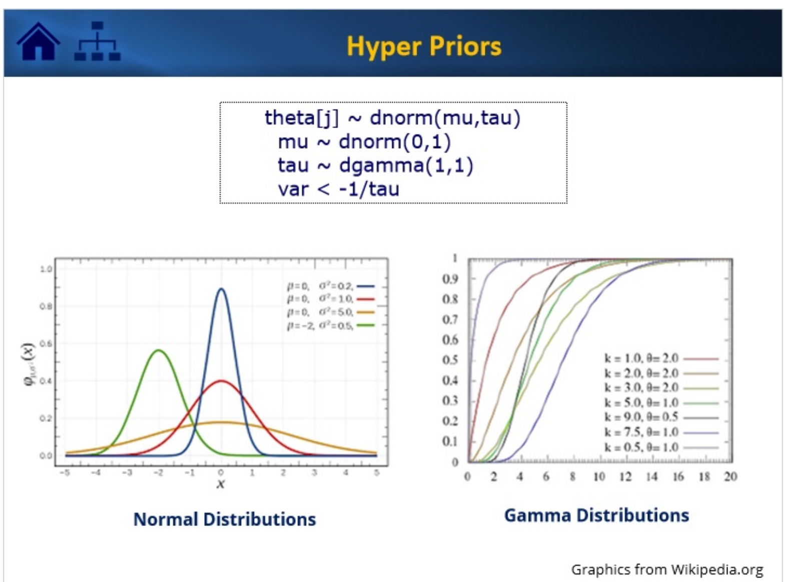
4.16 Specification of Priors I



4.17 Specification of Priors II



4.18 Specification of Priors III



4.19 Specification of Priors IV

Conjugate Priors

TABLE 3
Some Familiar Univariate Conjugate Priors in Exponential Families[a]

Prior $\pi(\theta)$	*Data*	*Posterior* $\pi(\theta\|y)$
$\pi(\theta) \equiv N(\mu, \tau^2)$	$Y_i \sim N(\theta, \sigma^2)$	$\pi(\theta \mid y) \equiv N(\frac{\tau^2 n\bar{y} + \sigma^2 \mu}{n\tau^2 + \sigma^2}, \frac{\tau^2\sigma^2}{n\tau^2 + \sigma^2})$
$\pi(\theta) \equiv Beta(\alpha, \beta)$	$Y \sim Bin(n, \theta)$	$\pi(\theta \mid y) \equiv Beta(\alpha + y, n + \beta - y)$
$\pi(\theta) \equiv Gamma(\alpha, \beta)$	$Y_i \sim Poisson(\theta)$	$\pi(\theta \mid y) \equiv Gamma(\alpha + n\bar{y}, (n + \beta^{-1})^{-1}$
$\pi(\theta) \equiv InvGamma(\alpha, \beta)$	$Y_i \sim N(\mu, \theta)$	$\pi(\theta \mid y) \equiv InvGamma(\alpha + \frac{n-1}{2}, (\frac{(n-1)s^2}{2} + \beta^{-1})^{-1})$

Note. In general, probability distributions that belong to an exponential family are the only ones with natural conjugate prior distributions and the posterior will involve the sufficient statistic for θ (see Gelman, Carlin, Stern, & Rubin, 1995, p. 38, for details).

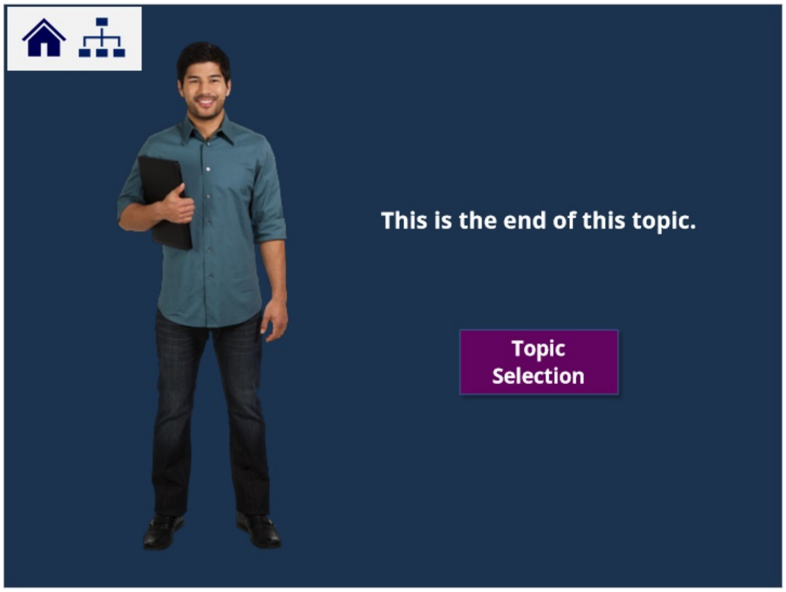
Rupp, Dey, & Zumbo (2004)

4.20 Specification of Priors V

Common Priors for IRT Models

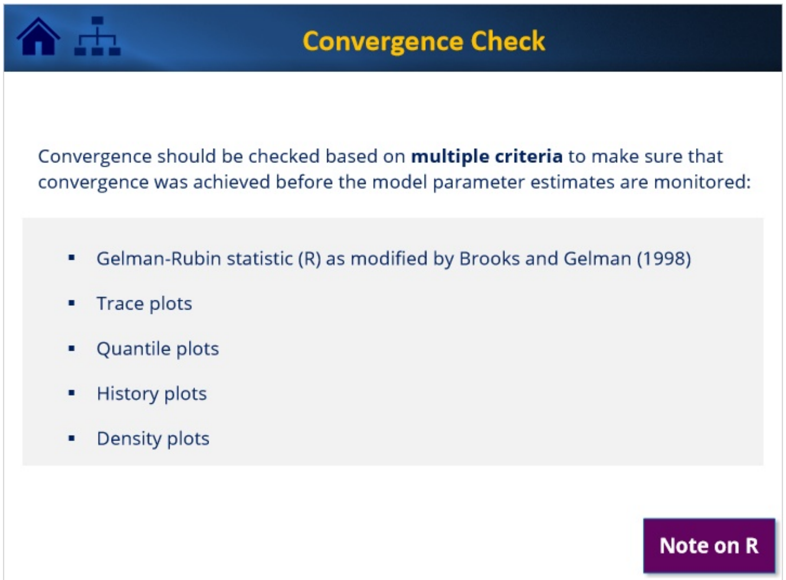
- **dnorm(** μ **,** τ **)** is the normal distribution or truncated normal distribution with parameters μ and $\tau = 1/\sigma^2$, (0, 0.001), as flat prior: θ , b
- **dnorm(** μ **,** τ **) I(L, U)** is the truncated normal distribution with parameters μ and $\tau = 1/\sigma^2$: a
- **Lognorm (** μ **,** τ **)** is the log-normal distribution with parameters μ and $\tau = 1/\sigma^2$, (0, 0.001), as flat prior: a
- **dbeta(**a**,**b**)** is the beta distribution with parameters a and b , (0.001, 0.001), as flat prior: c , d
- **dgamma(**a**,**s**)** is the gamma distribution, (0.001, 0.001), as flat prior: σ^2

4.21 Bookend: Prior Specification

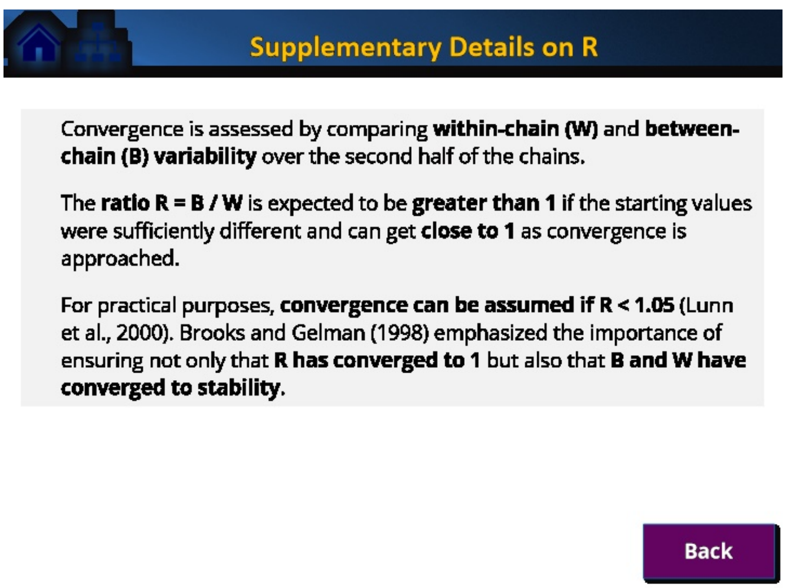


4.22 Bookmark: Convergence Checks

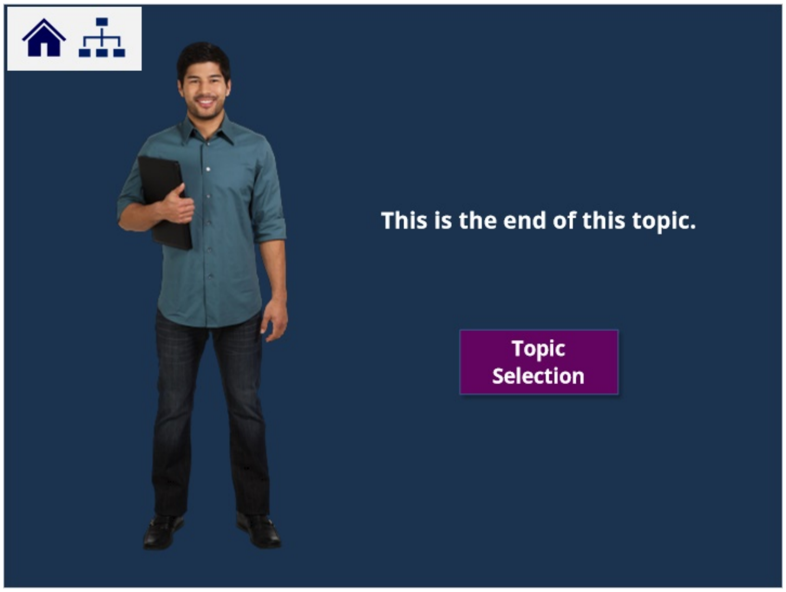
4.23 Convergence Checks



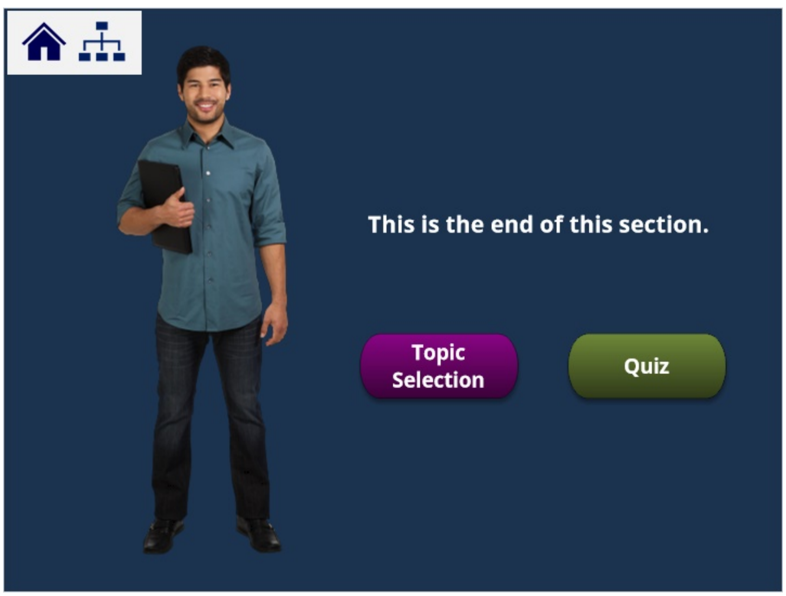
Details on R (Slide Layer)



4.24 Bookend: Convergence Checks

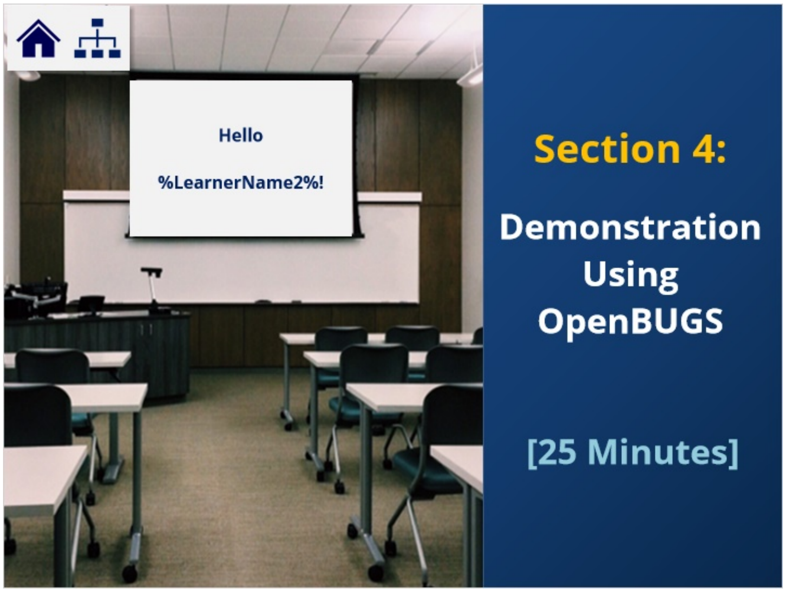


4.25 Bookend: Section 4

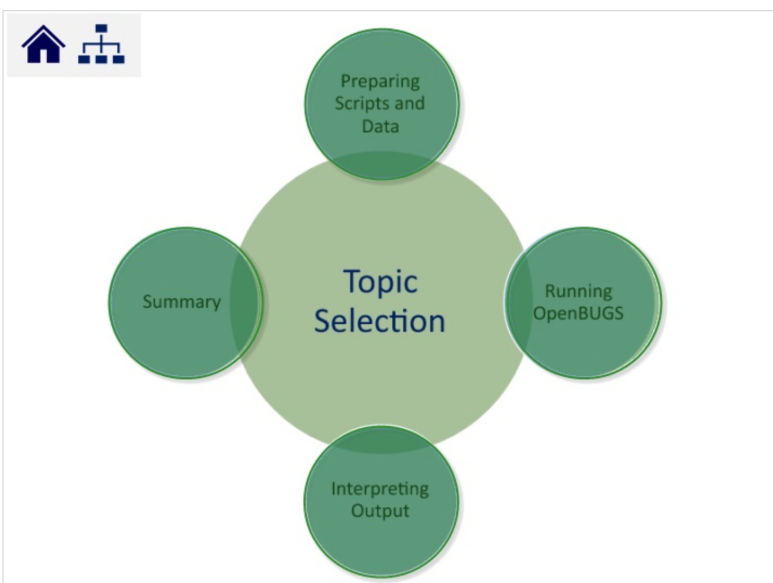


5. Section 4: Demonstration Using OpenBUGS

5.1 Cover: Section 5

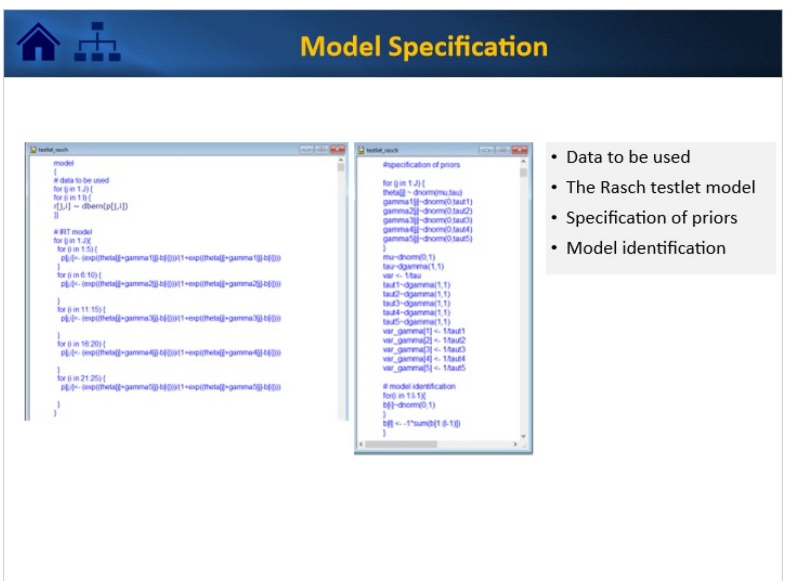


5.2 Topic Selection

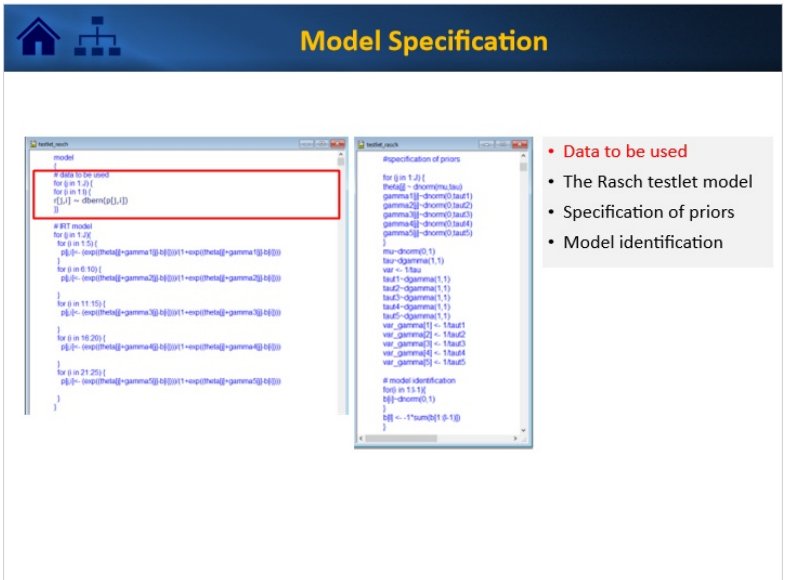


5.3 Bookmark: Preparing Script and Data

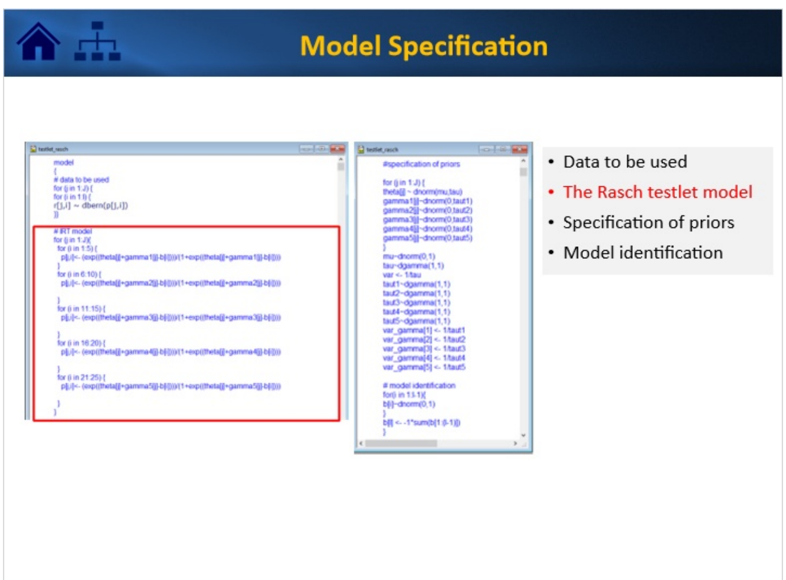
5.4 Model Specification I



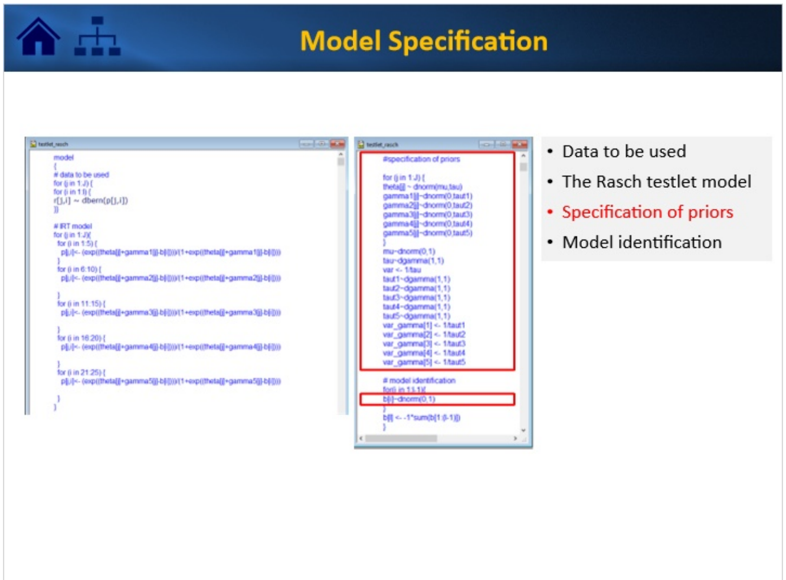
5.5 Model Specification III



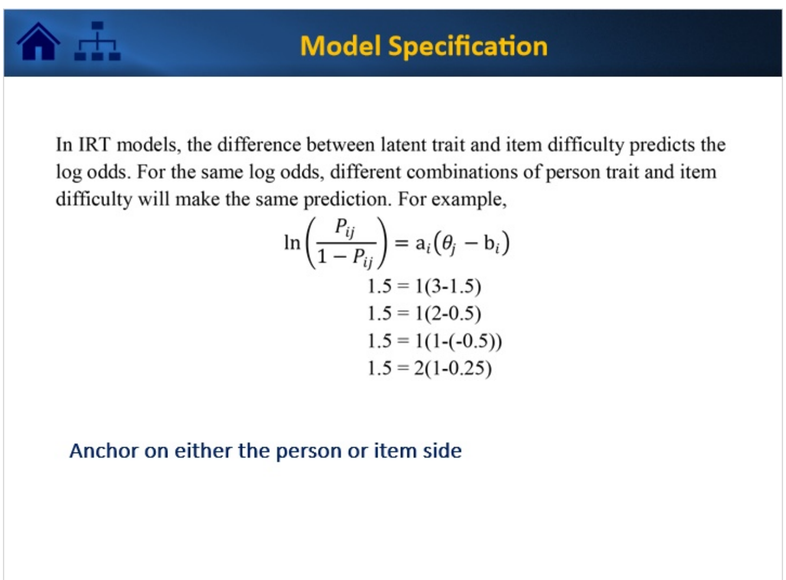
5.6 Model Specification III



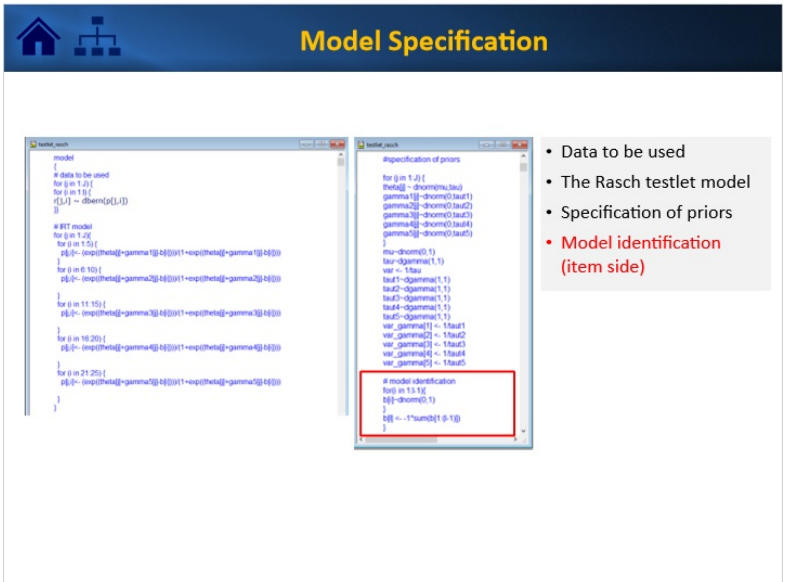
5.7 Model Specification IV



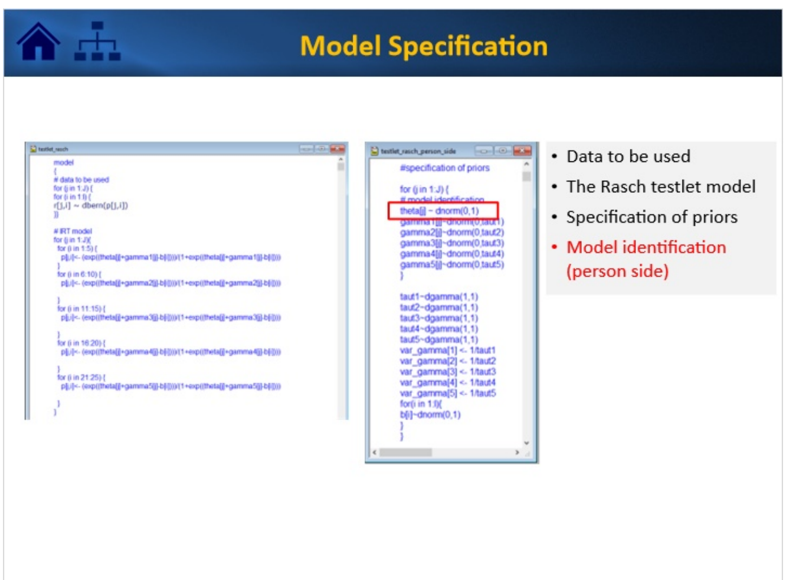
5.8 Model Specification V



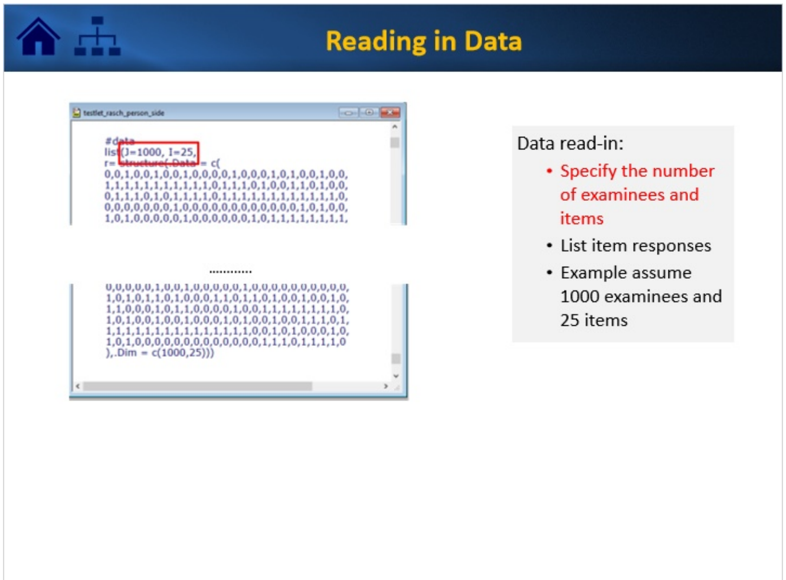
5.9 Model Specification VI



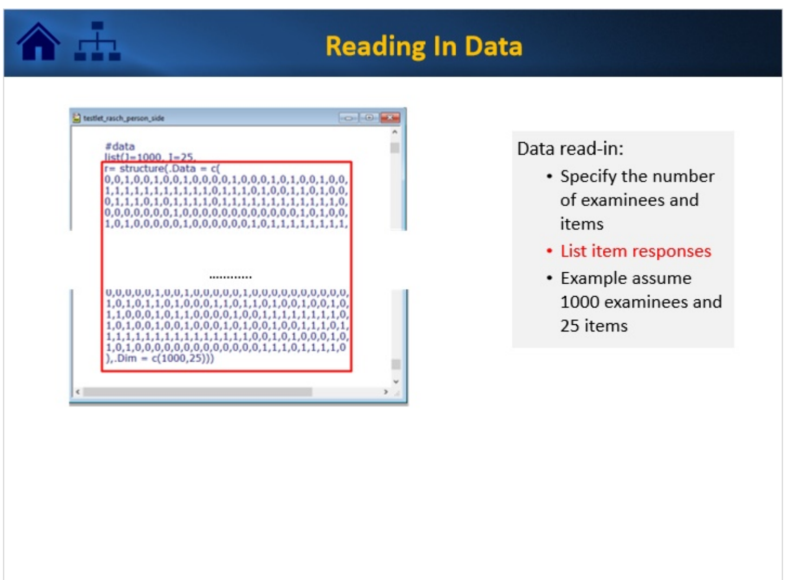
5.10 Model Specification VII



5.11 Reading in Data I



5.12 The Partial Credit Testlet Model I



5.13 The Partial Credit Testlet Model I

The Partial Credit Testlet Model

Assume four score categories: 0, 1, 2, 3
Recode scores into 1, 2, 3, 4

The probability of getting a certain score is:

$$t_1 = \exp(\theta_j + \gamma_{jd(i)} - b_i - d_{i1}), \quad t_2 = \exp(2\theta_j + 2\gamma_{jd(i)} - 2b_i - d_{i1} - d_{i2}),$$

and

$$t_3 = \exp(3\theta_j + 3\gamma_{jd(i)} - 3b_i - d_{i1} - d_{i2} - d_{i3})$$

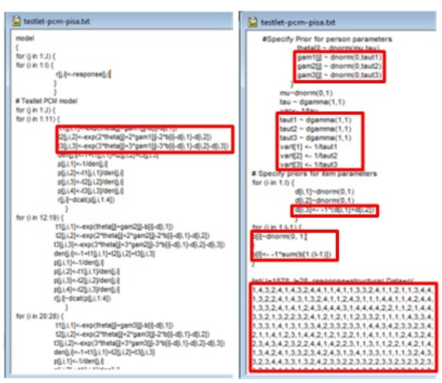
then:

$$pr(X_{ji} = 0) = \frac{1}{1 + t_1 + t_2 + t_3} \qquad pr(X_{ji} = 1) = \frac{t_1}{1 + t_1 + t_2 + t_3}$$

$$pr(X_{ji} = 2) = \frac{t_2}{1 + t_1 + t_2 + t_3} \qquad pr(X_{ji} = 3) = \frac{t_3}{1 + t_1 + t_2 + t_3}$$

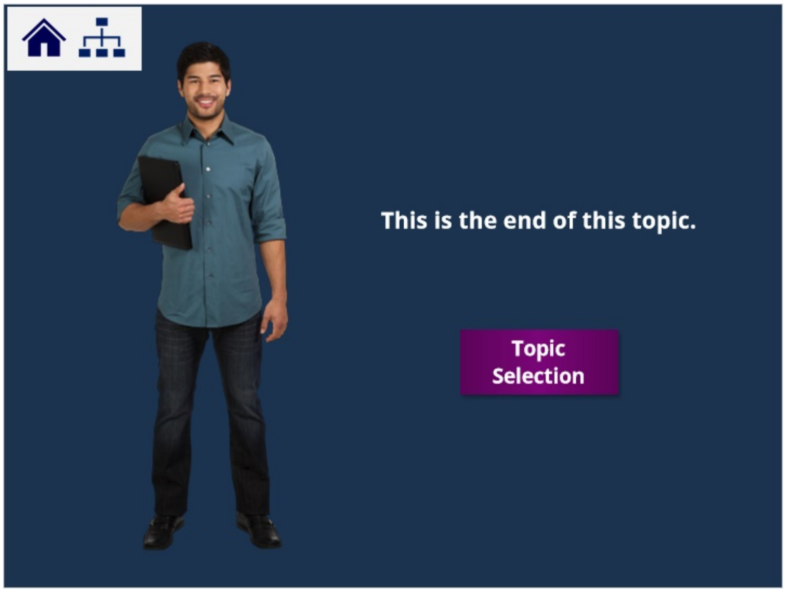
5.14 The Partial Credit Testlet Model II

The Partial Credit Testlet Model



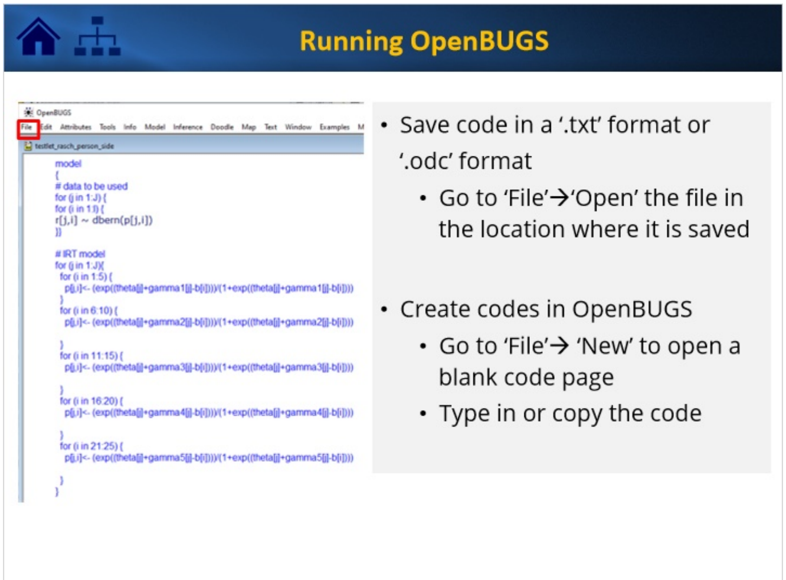
- Testlet variances
  - Items nested with testlets as
    - Items 1 to 11----testlet 1
    - Items 12 to 19----testlet 2
    - Items 20 to 28----testlet 3
  - Testlet effect parameters follow a normal distribution with a mean of 0 an a standard deviation to be estimated
  - Testlet variances follow an inverse gamma distribution.
- Model identification
  - Item side
  - Step parameters to be constrained
- Item responses
  - Starts at 1

5.15 Bookend: Preparing the Script and Data

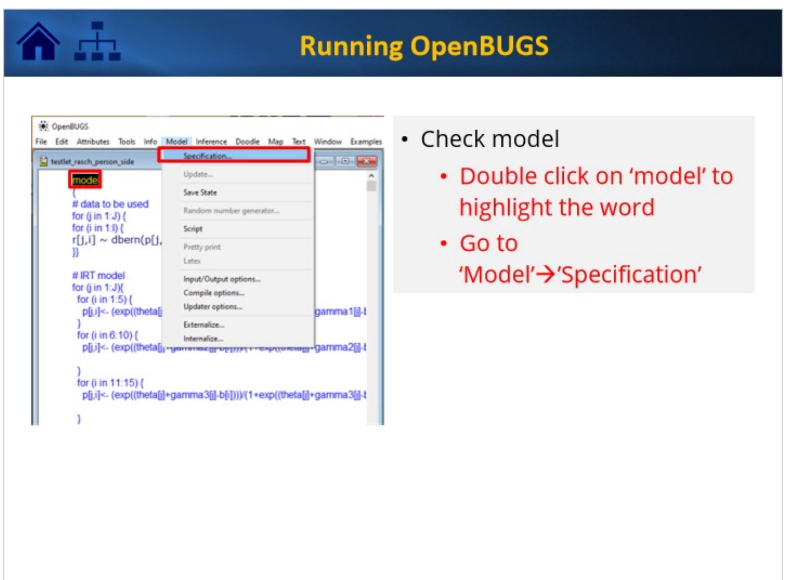


5.16 Bookmark: Running OpenBUGS

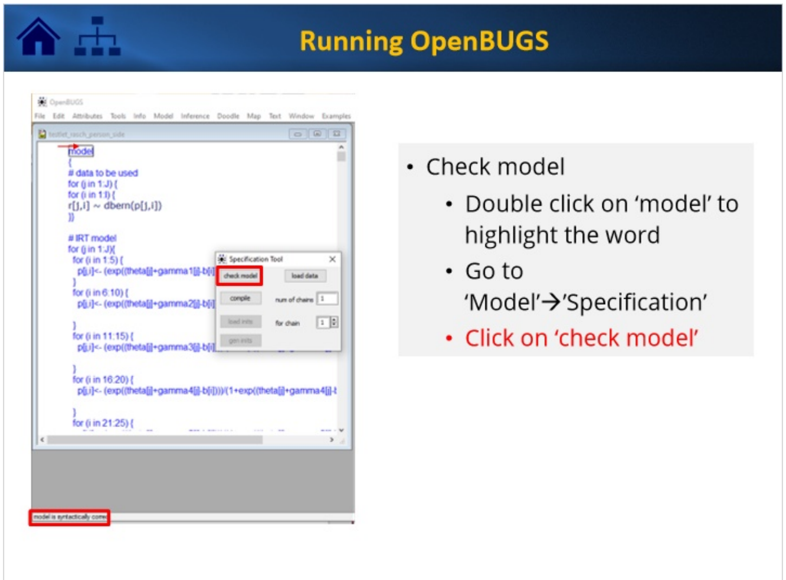
5.17 Running OpenBUGS I



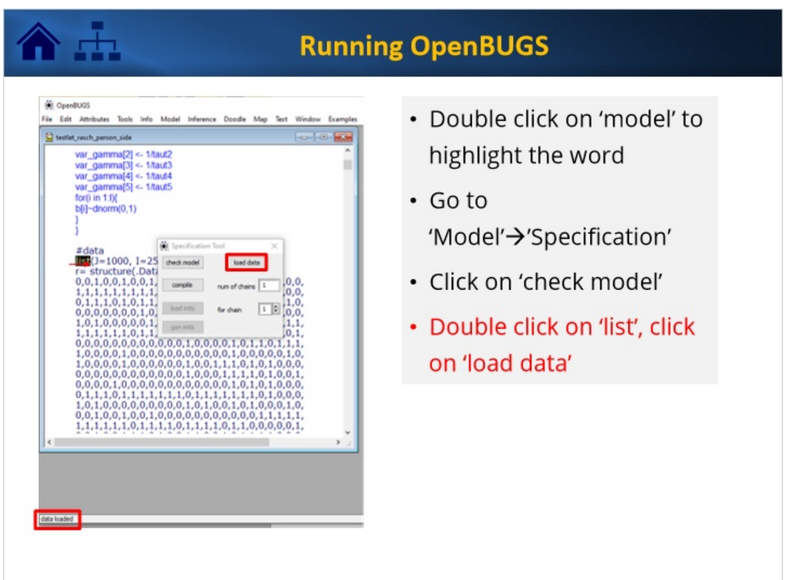
5.18 Running OpenBUGS II



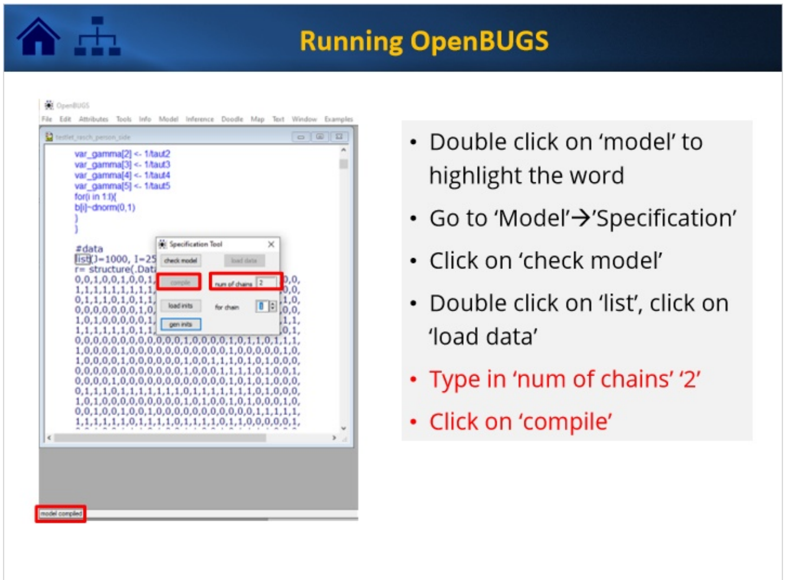
5.19 Running OpenBUGS III



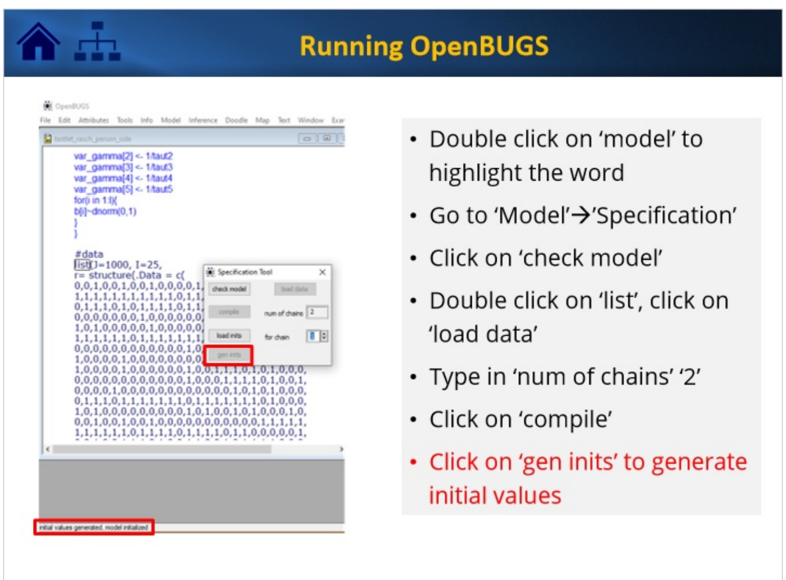
5.20 Running OpenBUGS V



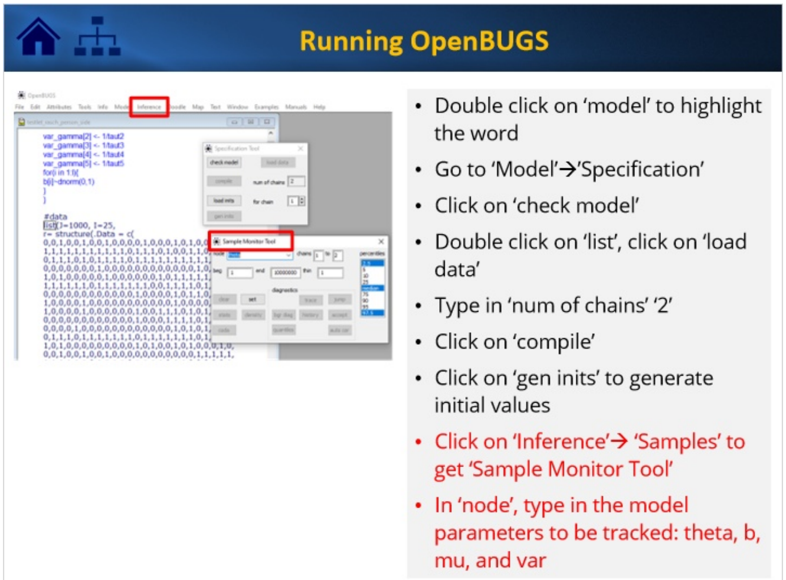
5.21 Running OpenBUGS VI



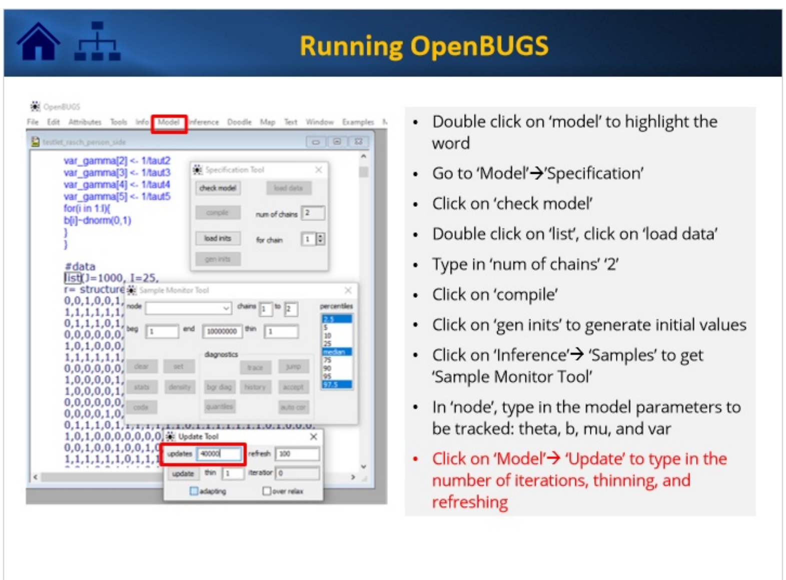
5.22 Running OpenBUGS VII



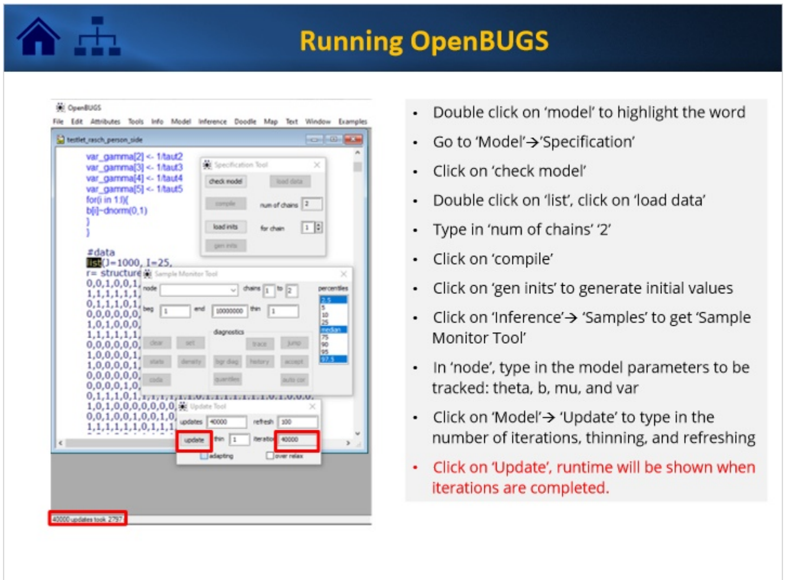
5.23 Running OpenBUGS VIII



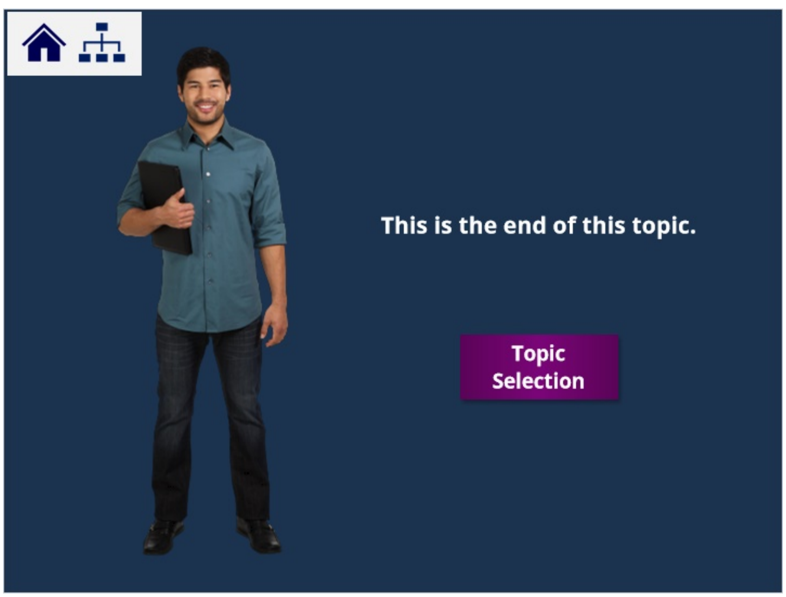
5.24 Running OpenBUGS IX



5.25 Running OpenBUGS X

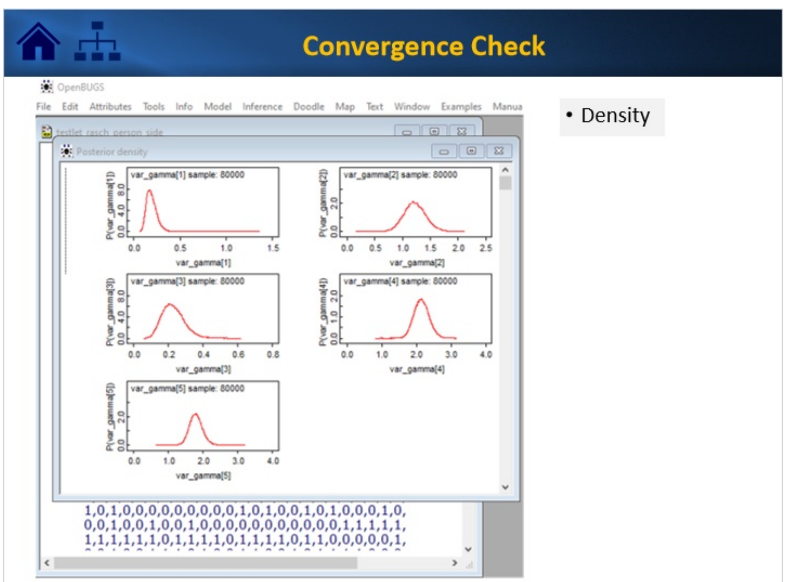


5.26 Bookend: Running OpenBUGS

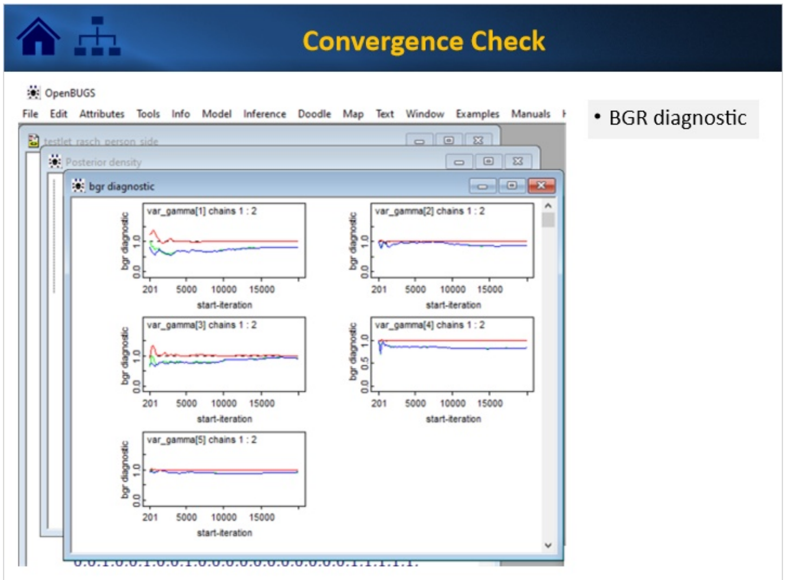


5.27 Bookmark: Interpreting Output

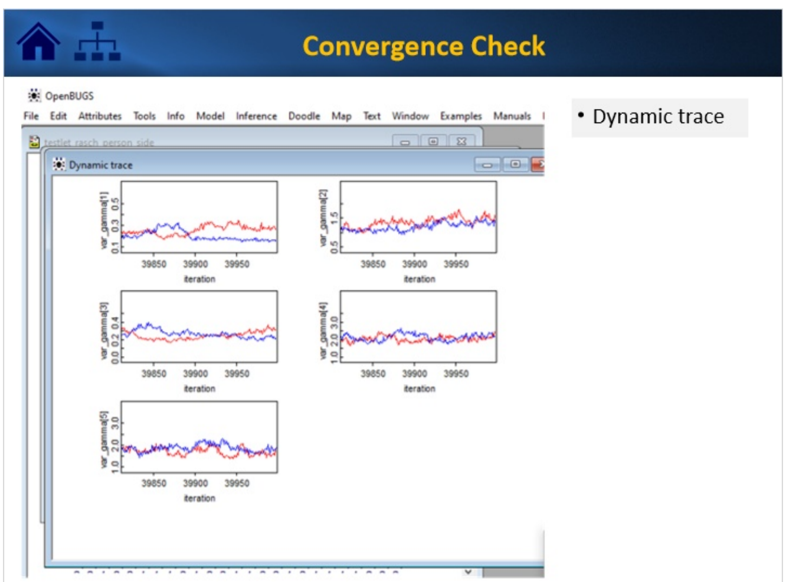
5.28 Convergence Check I



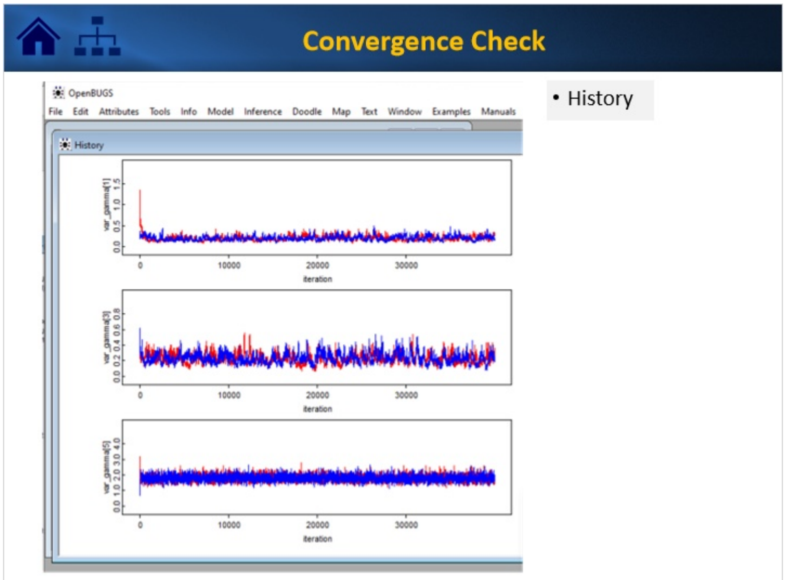
5.29 Convergence Check II



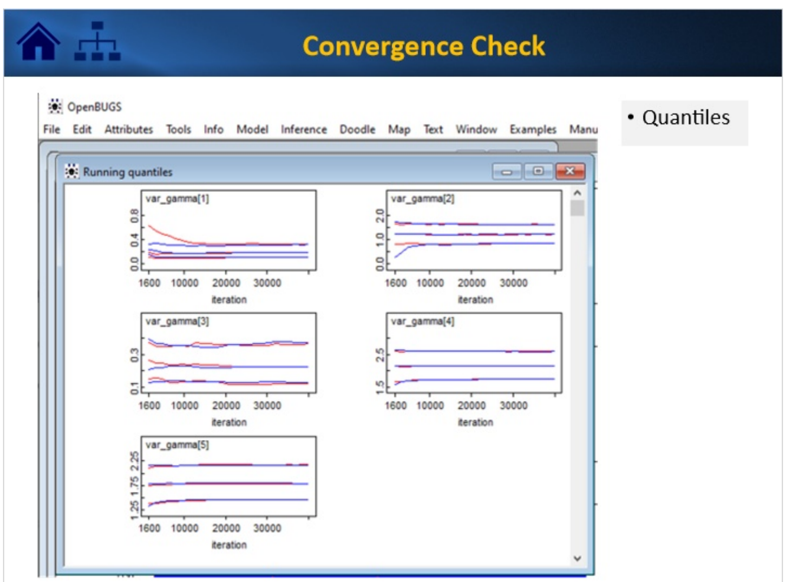
5.30 Convergence Check III



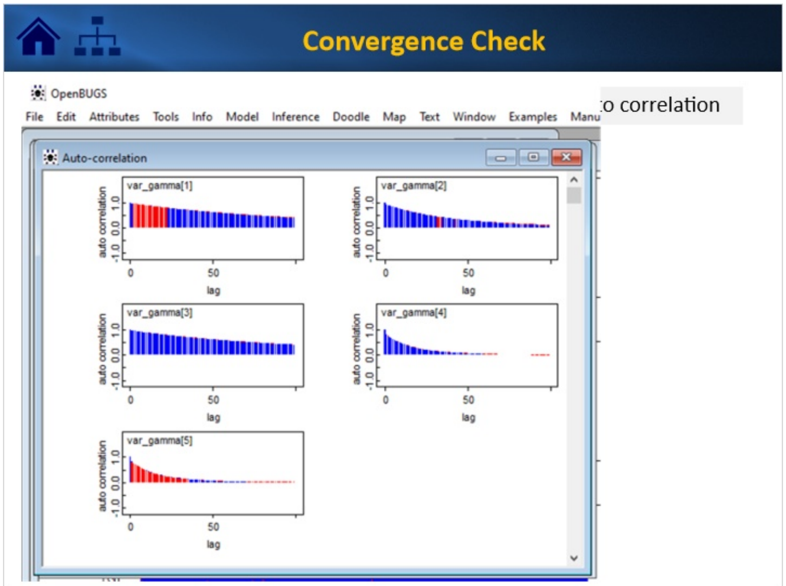
5.31 Convergence Check IV



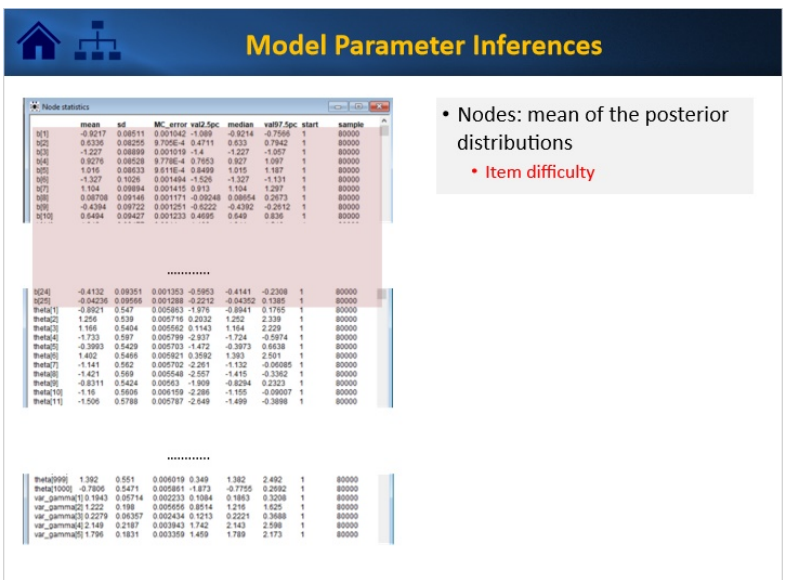
5.32 Convergence Check V



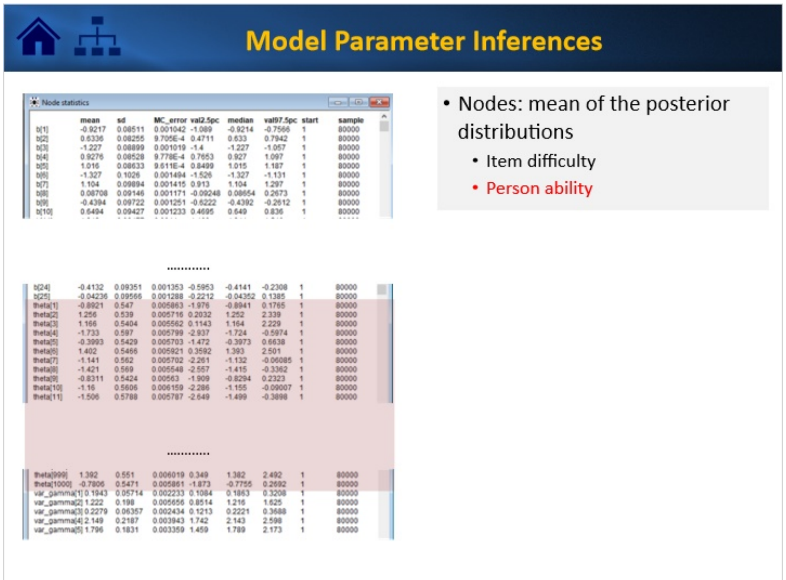
5.33 Convergence Check VI



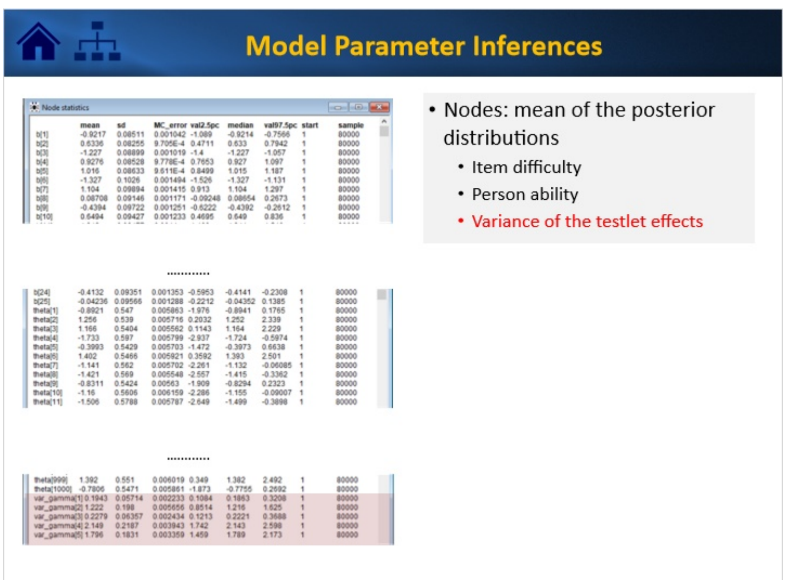
5.34 Model Parameter Inferences I



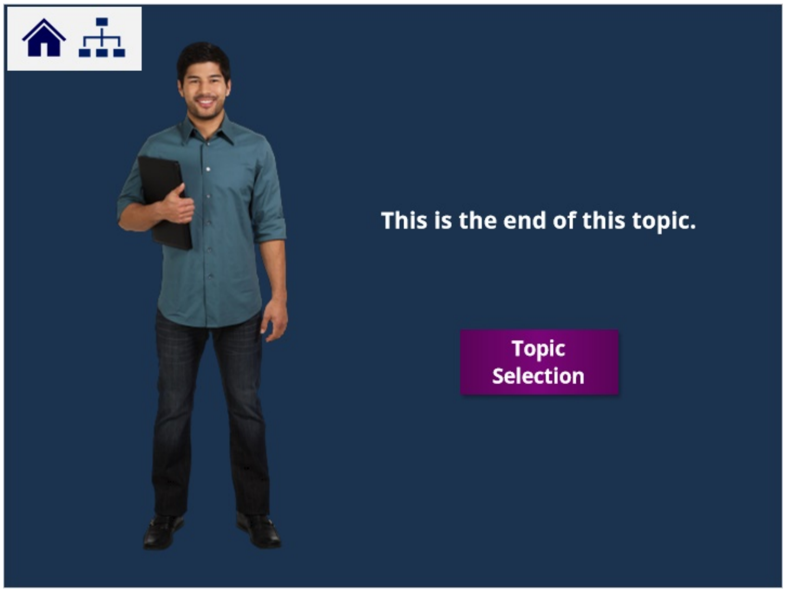
5.35 Model Parameter Inferences II



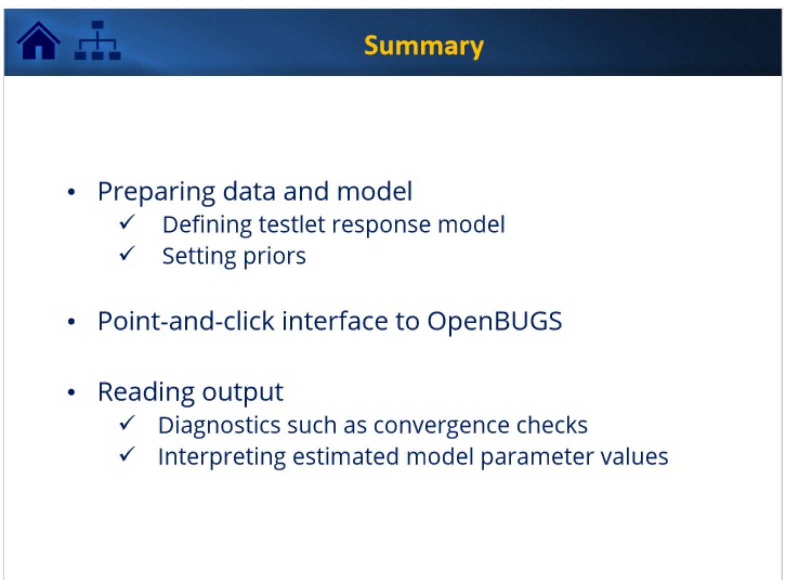
5.36 Model Parameter Inferences III



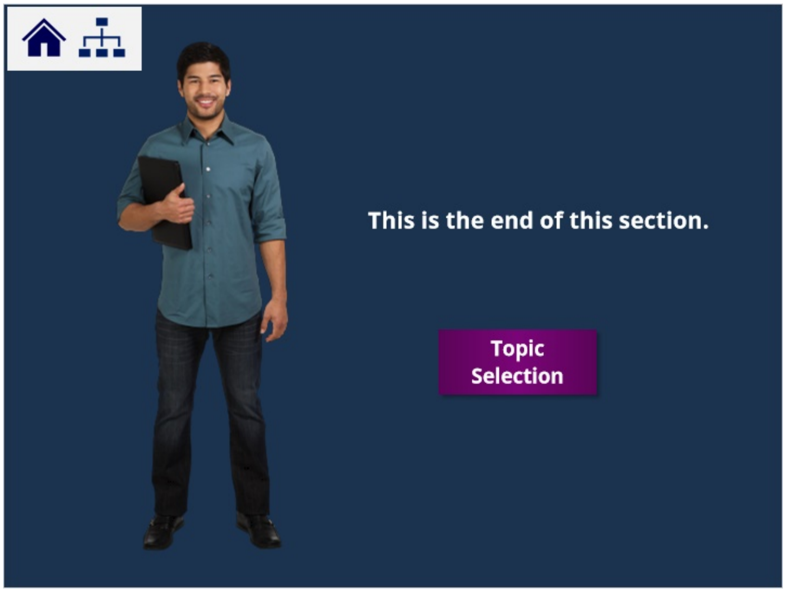
5.37 Bookend: Interpreting Output



5.38 Summary



5.39 Bookend: Section 5



5.40 Module Cover (END)