

# Latent Trait Measurement Models for Other (not Binary) Responses

## Topics:

- Review of generalized models for categorical outcomes
- Ordered categories → Graded Response
- Maybe ordered categories → Partial Credit
- Unordered categories → Nominal Response

# Too Logit to Quit\*...

<https://www.youtube.com/watch?v=HFCv86Olk8E>

- The **logit** (and the probit) are also the basis for many other generalized models for predicting categorical outcomes
- Next we'll see how  $C$  possible response categories can be predicted using  $C - 1$  binary "submodels" that carve up the categories in different ways based on ordering, in which each binary submodel uses a logit (or probit, if limited info) link to predict its outcome
- Types of categorical outcomes:

Let's start with a single outcome,  $y_i$ , that varies over persons...

  - Definitely ordered categories: "**cumulative logit/probit**"
  - Maybe ordered categories: "**adjacent category logit/probit**"
  - Not-ordered categories: "**baseline category (multinomial) logit/probit**"

\* Starts about 8 minutes into 15-minute video—and MY joke for over 10+ years!

# Generalized Models for $C$ Ordinal Categories

- Known as “**cumulative logit**” or “**proportional odds**” model in generalized models; known as “**graded response model**” in IRT
  - Also still known as “CFA for categorical outcomes” → using IFA parameterization
- Predicts the probability of **lower vs. higher** cumulative categories via  $C - 1$  submodels (e.g., if  $C = 4$  possible responses of  $c = 0,1,2,3$ ):

**0** vs. **1,2,3**  
Submodel<sub>1</sub>

**0,1** vs. **2,3**  
Submodel<sub>2</sub>

**0,1,2** vs. **3**  
Submodel<sub>3</sub>

I've named these submodels based on what they predict, but programs will use their own names in the output

- What the binary submodels predict depends on whether the model is predicting **DOWN** ( $y_i = 0$ , **the default**) or **UP** ( $y_i = 1$ ) **cumulatively**
- **Example predicting UP in an empty model (subscripts=parm,submodel)**
- Submodel 1:  $\text{Logit}[p(y_i > 0)] = \beta_{01} \rightarrow p(y_i > 0) = \exp(\beta_{01})/[1 + \exp(\beta_{01})]$
- Submodel 2:  $\text{Logit}[p(y_i > 1)] = \beta_{02} \rightarrow p(y_i > 1) = \exp(\beta_{02})/[1 + \exp(\beta_{02})]$
- Submodel 3:  $\text{Logit}[p(y_i > 2)] = \beta_{03} \rightarrow p(y_i > 2) = \exp(\beta_{03})/[1 + \exp(\beta_{03})]$
- Intercepts vary by submodel, but predictor slopes are usually held constant!

# Generalized Models for $C$ Ordinal Categories

- Models the probability of **lower vs. higher** cumulative categories via  $C - 1$  submodels (e.g., if  $C = 4$  possible responses of  $c = 0,1,2,3$ ):

**0 vs. 1,2,3**  
Submodel<sub>1</sub>  
→ Prob<sub>1</sub>

**0,1 vs. 2,3**  
Submodel<sub>2</sub>  
→ Prob<sub>2</sub>

**0,1,2 vs. 3**  
Submodel<sub>3</sub>  
→ Prob<sub>3</sub>

$$\text{Logit}[p(y_i > 2)] = \beta_{03}$$
$$\rightarrow p(y_i > 2) = \frac{\exp(\beta_{03})}{1 + \exp(\beta_{03})}$$

- What the binary submodels predict depends on whether the model is predicting **DOWN** ( $y_i = 0$ , the default) or **UP** ( $y_i = 1$ ) **cumulatively**
  - Either way, the model predicts the middle category responses *indirectly*

- Example if predicting UP with an empty model (intercept only):**

- Probability of 0 =  $1 - \text{Prob}_1$
- Probability of 1 =  $\text{Prob}_1 - \text{Prob}_2$
- Probability of 2 =  $\text{Prob}_2 - \text{Prob}_3$
- Probability of 3 =  $\text{Prob}_3 - 0$

The cumulative submodels that create these probabilities are each estimated using **all the data** (good, especially for categories not chosen often), but **assume order in doing so** (may be bad or ok, depending on your response format)

# Generalized Models for $C$ Categories

- Uses **multinomial distribution**: e.g., PDF for  $C = 4$  categories of  $c = 0,1,2,3$ ; an observed  $y_i = c$ ; and indicators  $I$  if  $c = y_i$

$$f(y_i = c) = p_{i0}^{I[y_i=0]} p_{i1}^{I[y_i=1]} p_{i2}^{I[y_i=2]} p_{i3}^{I[y_i=3]}$$

Only  $p_{ic}$  for the response  $y_i = c$  gets used

- ML or WLSMV estimation is then used to find the most likely model parameters to predict the probability of each response through the logit (or probit) link function; probabilities sum to 1:  $\sum_{c=1}^C p_{ic} = 1$
- Other models for categorical data that use the multinomial:
  - Adjacent category logit (partial credit): Models the probability of **each next highest** category via  $C - 1$  submodels (e.g., if  $C = 4$ ):

**0** vs. **1**

**1** vs. **2**

**2** vs. **3**

- Baseline category logit (nominal): Models the probability of **reference vs. other** category via  $C - 1$  submodels (e.g., if  $C = 4$  and  $3 = \text{ref}$ ):

**3** vs. **0**

**3** vs. **1**

**3** vs. **2**

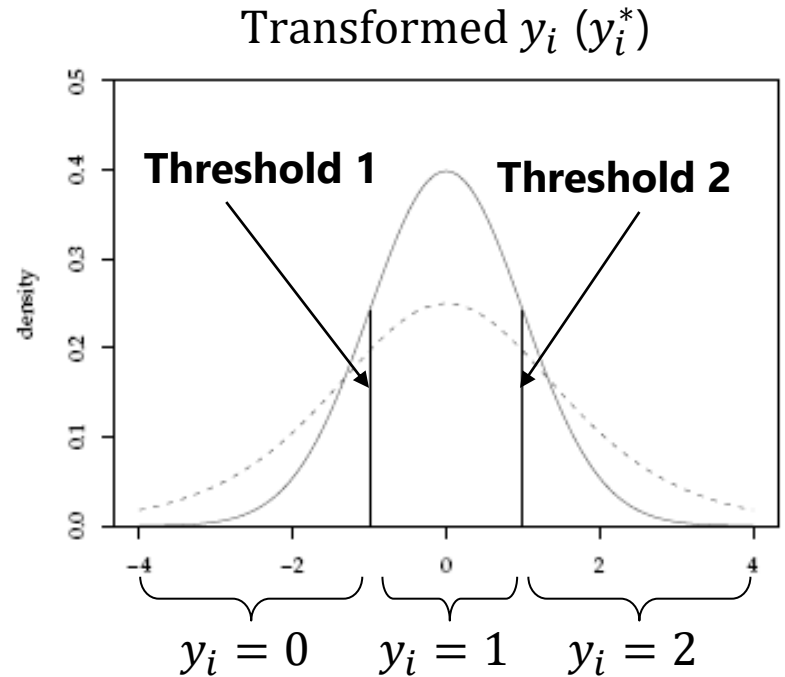
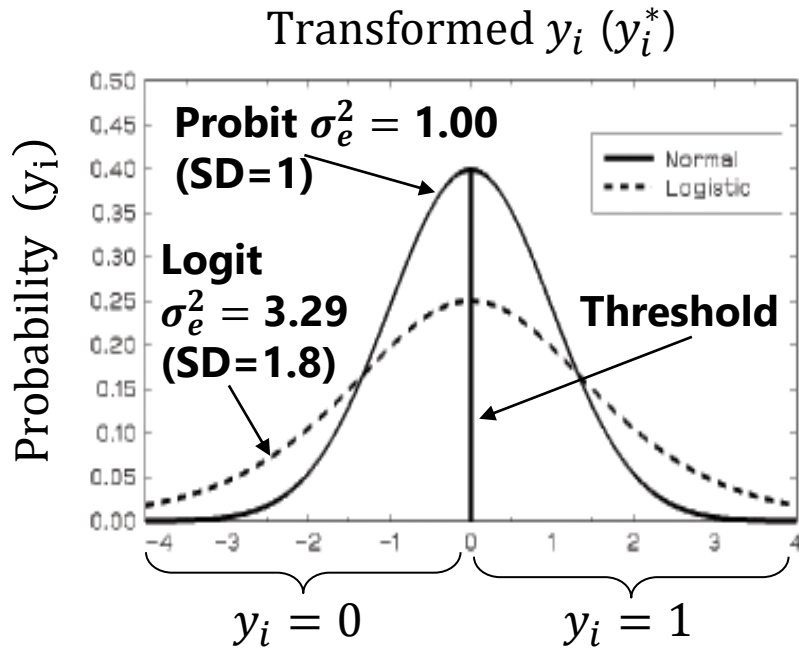
All parameters are estimated **separately** per nominal submodel

# Polytomous Items = Categorical Outcomes

- **Polytomous = more than 2 response options**
- Polytomous models are not numbered like binary models, but instead get called different names (or acronyms)
  - Most have a 1-PL vs. 2-PL analog that go by different names (i.e., where 1PL → constrain loadings equal; 2PL → allow item-specific loadings)
- **3 main kinds of polytomous models (e.g., in logits here):**
  - Response options are ordered for sure → Cumulative Logit
    - Graded Response or Modified Graded Response Model
  - Response options partially ordered → Adjacent Category Logit
    - (Generalized) Partial Credit Model or Rating Scale Model
  - Response options are unordered (to be found) → Baseline Category Logit
    - Nominal Response Model: Notably, this can be constrained to fit combo of partial credit and nominal responses (e.g., for “not applicable” responses)

# The Threshold Concept for Ordinal Items

$$y_i^* = \text{threshold} + \text{rest of your model} + e_i$$



- Each categorical variable is **really** the chopped-up version of a pretend underlying continuous variable ( $y_i^*$ ) with mean = 0 (variance = 1.00 or 3.29)
- Polytomous models will differ in how they make use of multiple thresholds per item in which  $C = \#$  categories, so **# thresholds per item =  $k = C - 1$**

# Graded Response Model for Ordinal Categories

- Ideal for items with clear underlying response continuum (e.g., Likert)
- **# response options DO NOT have to be the same across items!**
- GRM is an “indirect” or “difference” model
  - Compute difference between models to get probability of each response
- Estimate 1  $a_i$  per item and  $k = C - 1$  difficulties (4 options  $\rightarrow$  3 difficulties)
- Models the probability of **lower vs. higher** cumulative categories via  $k$  submodels (e.g., if  $C = 4$  possible responses of  $c = 0,1,2,3$ ):

**0 vs. 1,2,3**  
Submodel<sub>1</sub>

**0,1 vs. 2,3**  
Submodel<sub>2</sub>

**0,1,2 vs. 3**  
Submodel<sub>3</sub>

Mplus will refer to these thresholds using \$1, \$2, \$3 after the item name

- Each submodel is estimated using all the data cumulatively (assumes order)
- As with binary items, Mplus estimates the IFA model directly (loadings and thresholds), but unlike binary items, it won't do the IRT conversion into discriminations and difficulties (but can use NEW parameters to get them)



# Example Graded Response Model (GRM)

**IRT version of the GRM: Note that  $a_i$  is the same across submodels...**

- **0** vs. **1,2,3**:  $\text{Logit}[p(y_{is} > 0)] = a_i(\theta_s - b_{1i}) \rightarrow p1(y_{is} > 0) = \frac{\exp[a_i(\theta_s - b_{1i})]}{1 + \exp[a_i(\theta_s - b_{1i})]}$
- **0,1** vs. **2,3**:  $\text{Logit}[p(y_{is} > 1)] = a_i(\theta_s - b_{2i}) \rightarrow p2(y_{is} > 1) = \frac{\exp[a_i(\theta_s - b_{2i})]}{1 + \exp[a_i(\theta_s - b_{2i})]}$
- **0,1,2** vs. **3**:  $\text{Logit}[p(y_{is} > 2)] = a_i(\theta_s - b_{3i}) \rightarrow p3(y_{is} > 2) = \frac{\exp[a_i(\theta_s - b_{3i})]}{1 + \exp[a_i(\theta_s - b_{3i})]}$

**IFA version of the GRM—what is estimated directly in Mplus:**

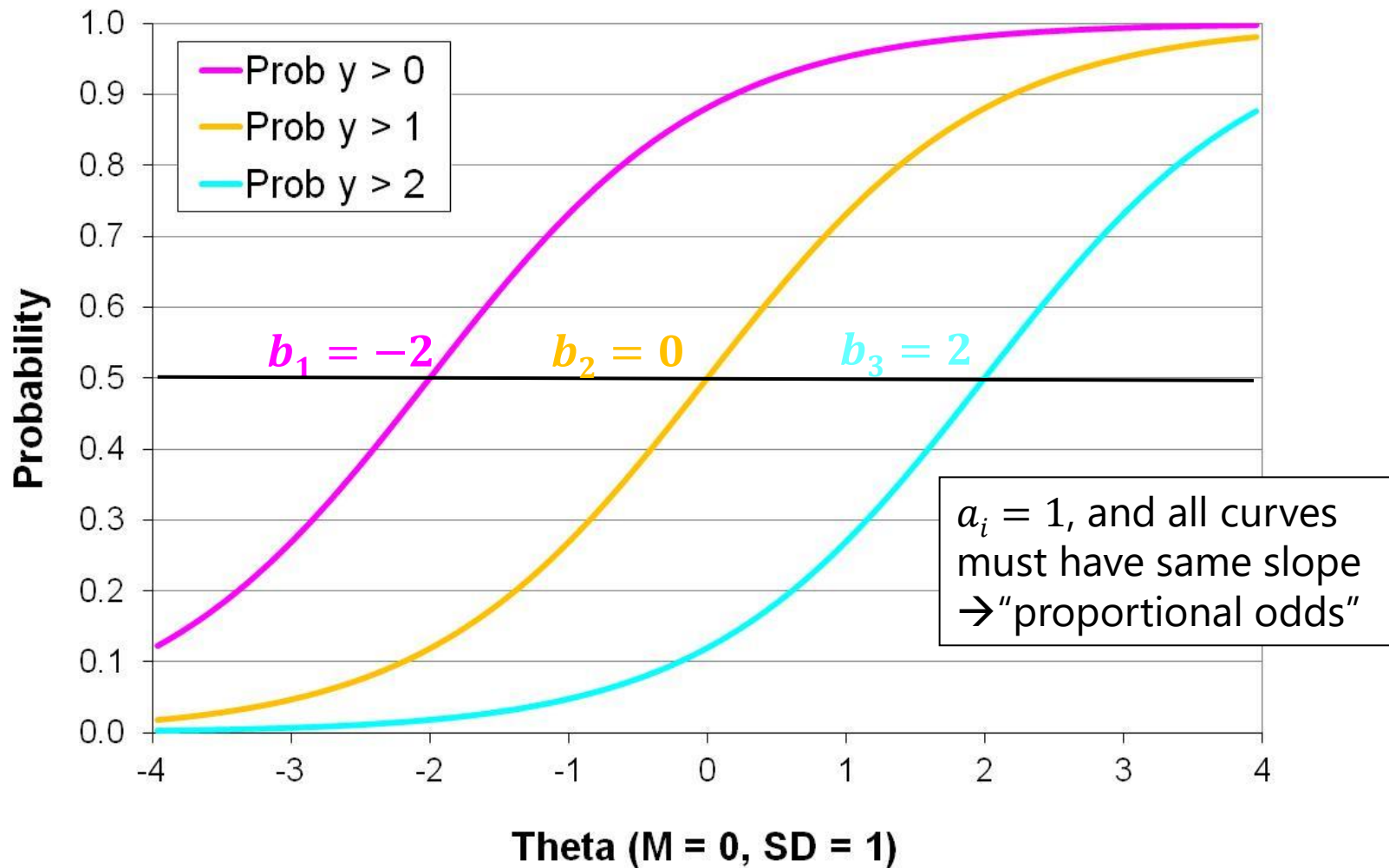
- **0** vs. **1,2,3**:  $\text{Logit}[p(y_{is} > 0)] = -\tau_{1i} + \lambda_i F_s \rightarrow p1(y_{is} > 0) = \frac{\exp(-\tau_{1i} + \lambda_i F_s)}{1 + \exp(-\tau_{1i} + \lambda_i F_s)}$
- **0,1** vs. **2,3**:  $\text{Logit}[p(y_{is} > 1)] = -\tau_{2i} + \lambda_i F_s \rightarrow p2(y_{is} > 1) = \frac{\exp(-\tau_{2i} + \lambda_i F_s)}{1 + \exp(-\tau_{2i} + \lambda_i F_s)}$
- **0,1,2** vs. **3**:  $\text{Logit}[p(y_{is} > 2)] = -\tau_{3i} + \lambda_i F_s \rightarrow p3(y_{is} > 2) = \frac{\exp(-\tau_{3i} + \lambda_i F_s)}{1 + \exp(-\tau_{3i} + \lambda_i F_s)}$

**GRM indirectly predicts probability of each category**

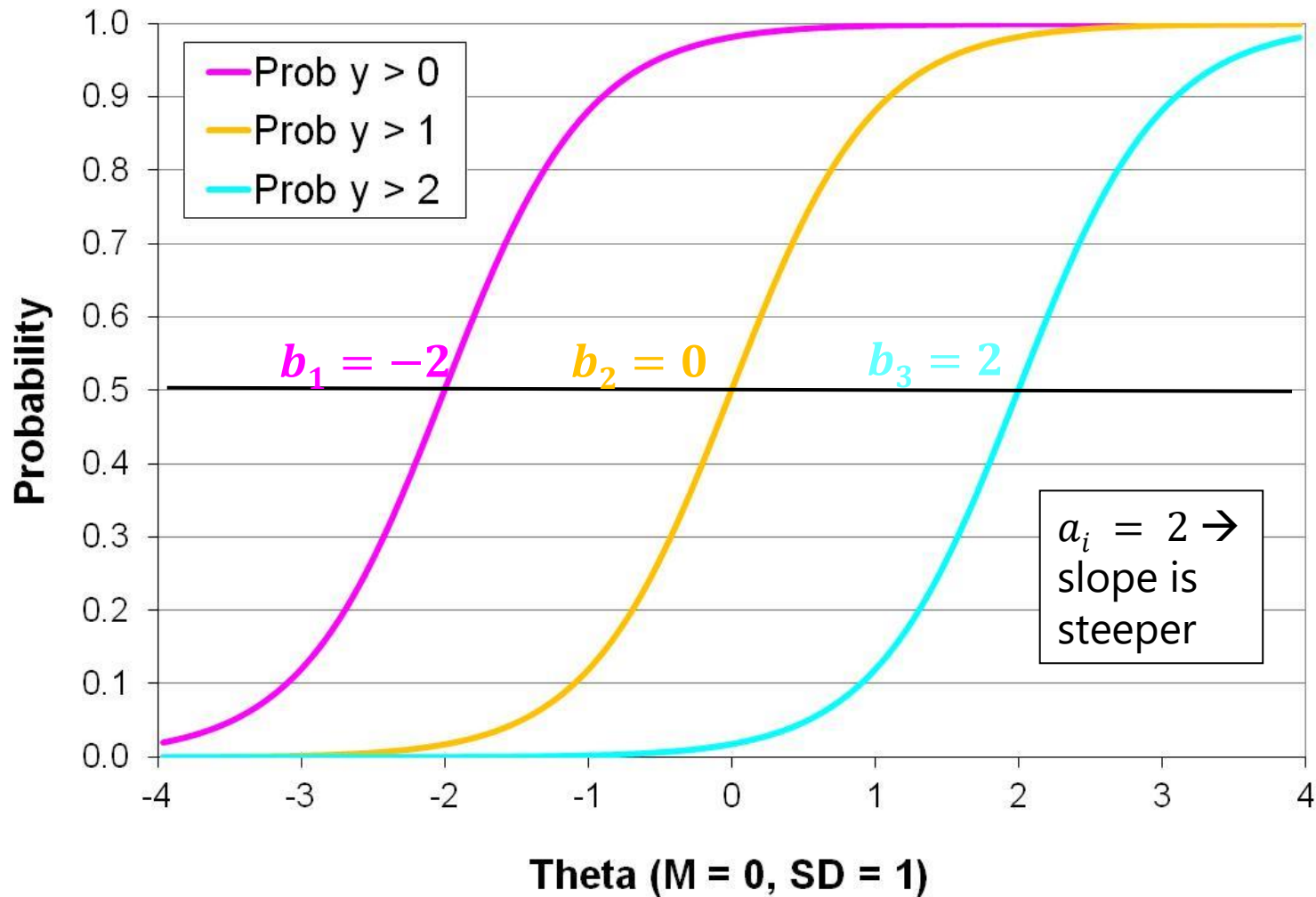
$$\begin{aligned} \triangleright p(y_{is} = 0) &= 1 - p1 \\ p(y_{is} = 1) &= p1 - p2 \\ p(y_{is} = 2) &= p2 - p3 \\ p(y_{is} = 3) &= p3 - 0 \end{aligned}$$

**IRT  $b_{ki}$**  = "difficulty" = trait needed for a 50% probability (logit=0) of **higher binary category**  
**IFA  $\tau_{ki}$**  = "threshold" = logit of the probability of **lower binary category** when Factor=0

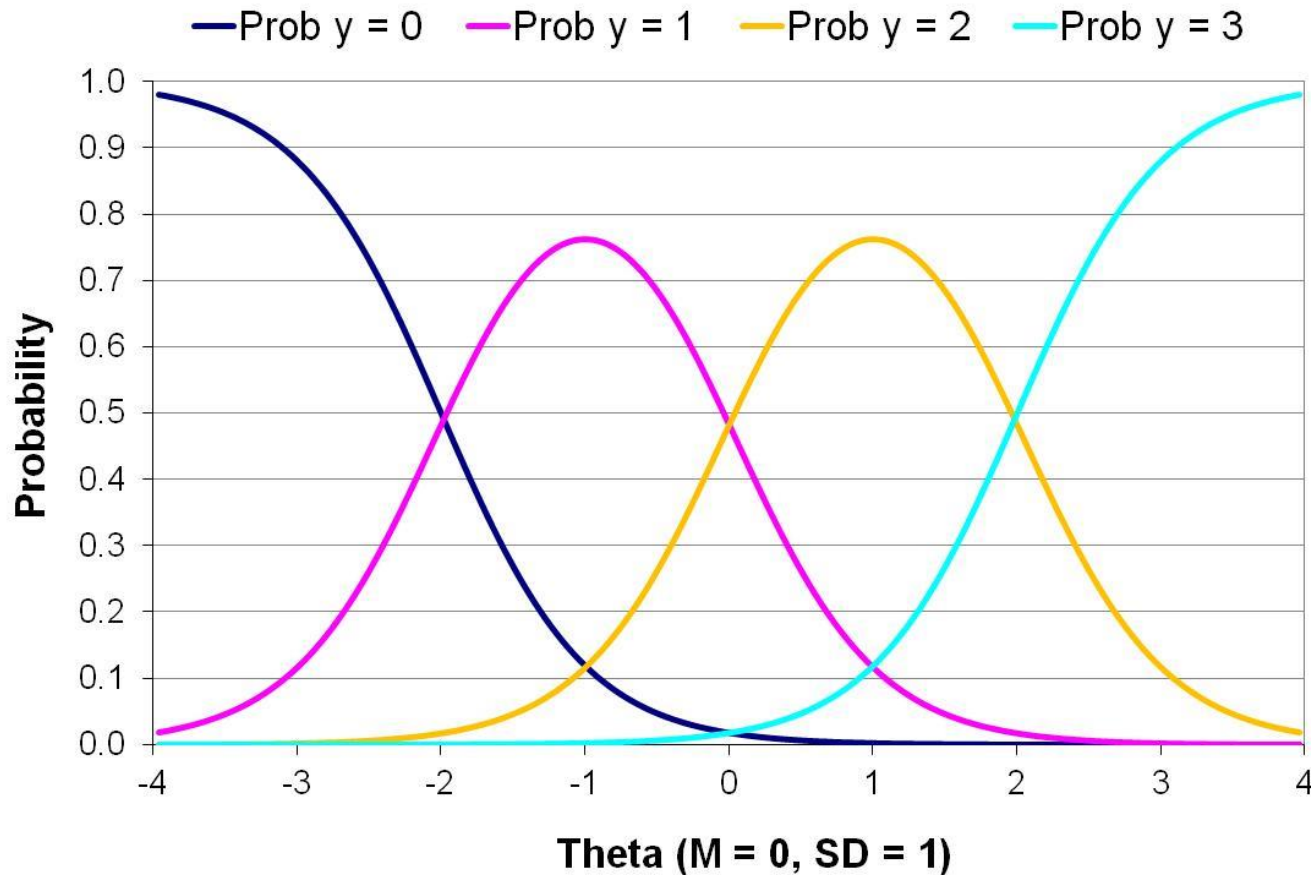
# Cumulative Item Response Curves: GRM for 4-Category (0 | 23) Item, $a_i = 1$



# Cumulative Item Response Curves: GRM for 4-Category (0|23) Item, $a_i = 2$



# Category Response Curves: GRM for 4-Category (0 | 23) Item, $a_i = 2$

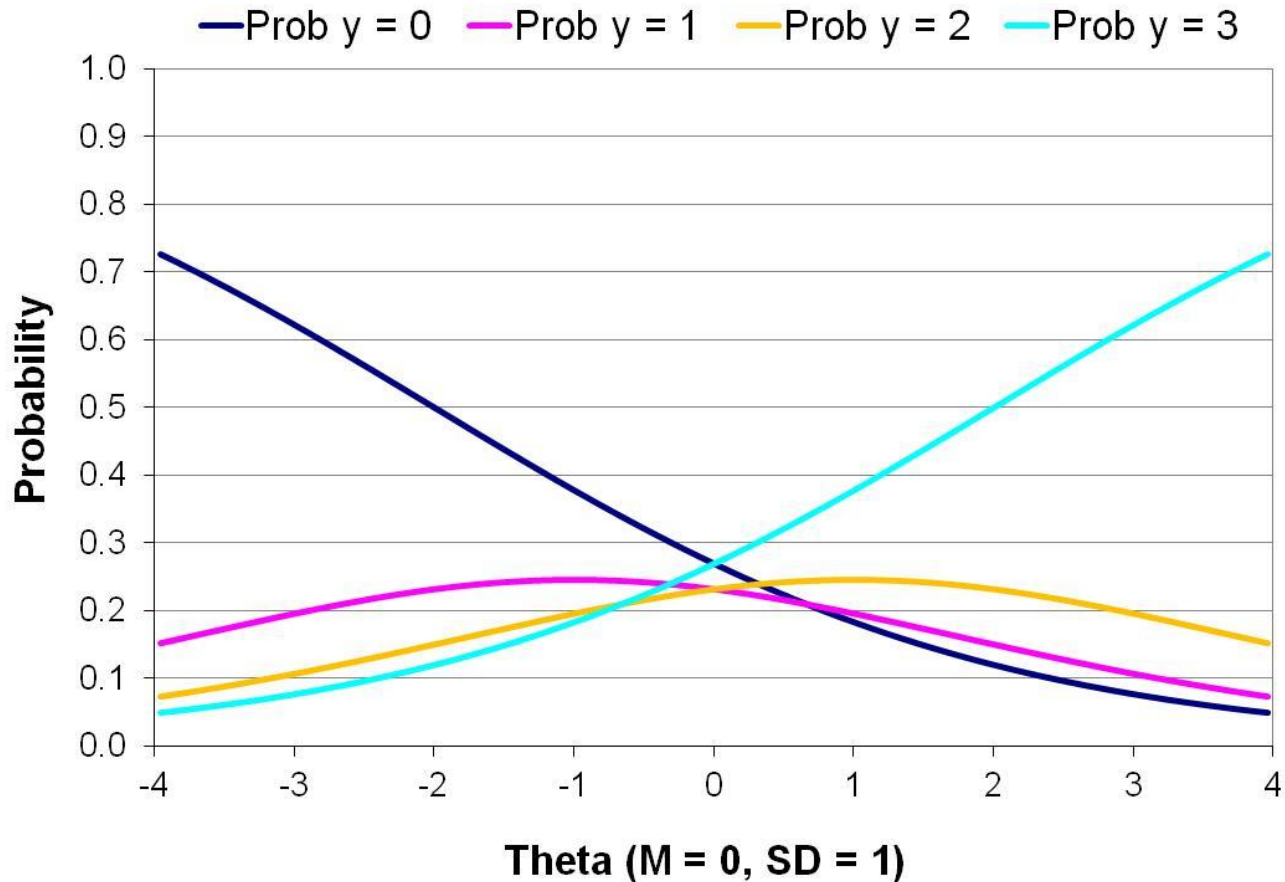


GRM indirectly predicts the probability of each category response across Theta:

$$\begin{aligned}p(y_{is} = 0) &= 1 - p_1 \\p(y_{is} = 1) &= p_1 - p_2 \\p(y_{is} = 2) &= p_2 - p_3 \\p(y_{is} = 3) &= p_3 - 0\end{aligned}$$

Mplus will make this type of Plot for you under the title "Item Characteristic Curve"

# Category Response Curves: GRM for 4-Category (0 | 23) Item, $a_i = .5$

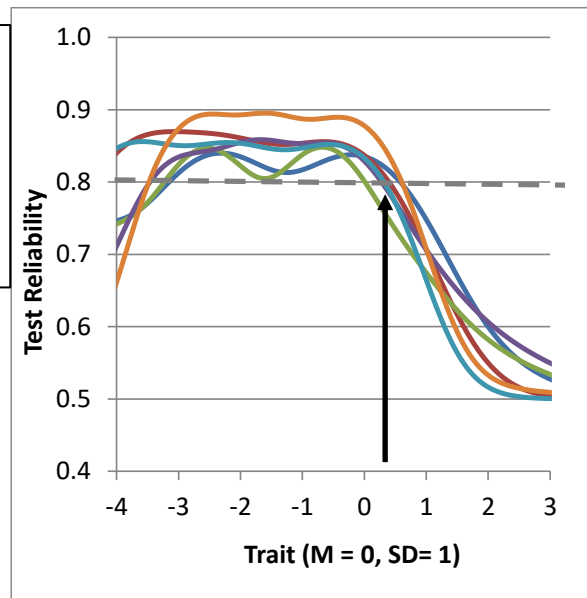
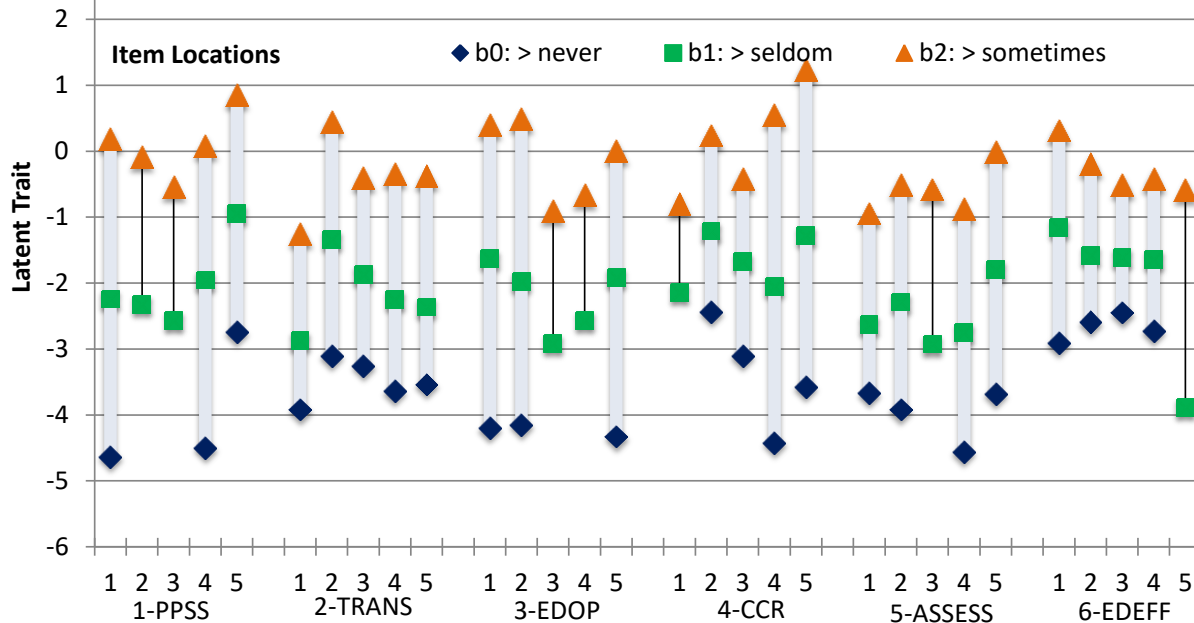
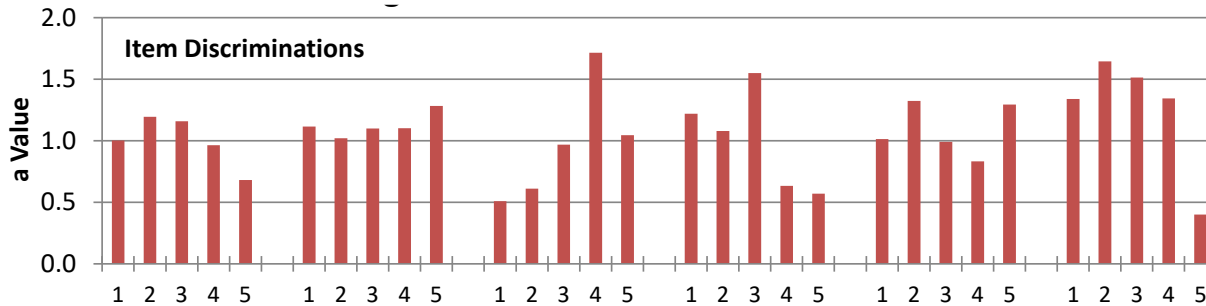


This is exactly what you do NOT want to see

Although they are ordered (as forced by the model), the middle categories are worthless (not differentiated)

Top Panel: Red bars display item discriminations.

Bottom Panel: Item locations for 4 response options:  
*Never, Seldom, Sometimes, and Often*



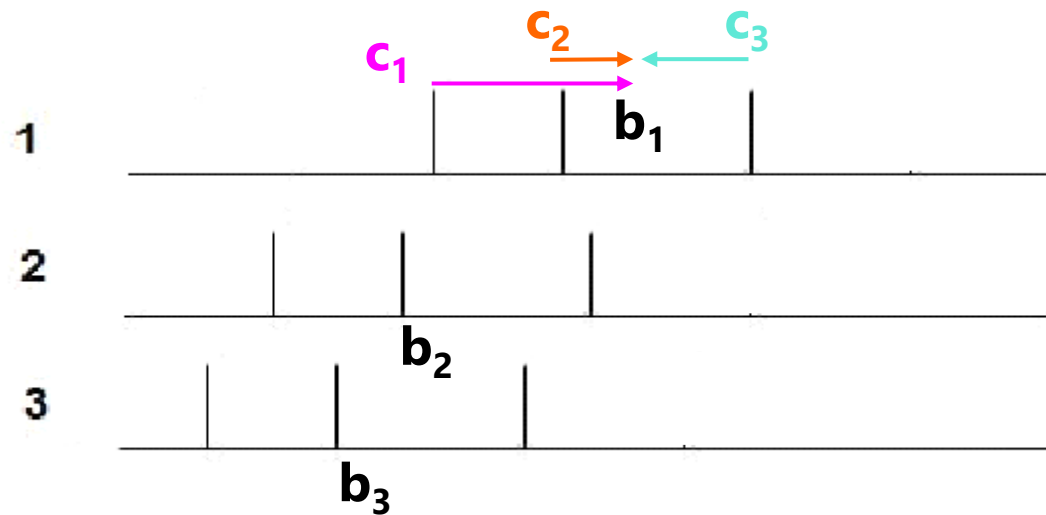
The location of  $y_{is} > \textit{sometimes}$  ranges from  $\sim 0$  to 0.5 across items, creating lower reliability for traits  $> 0.5$

Being able to select a higher option (e.g., *almost always*) would improve reliability for high-trait respondents

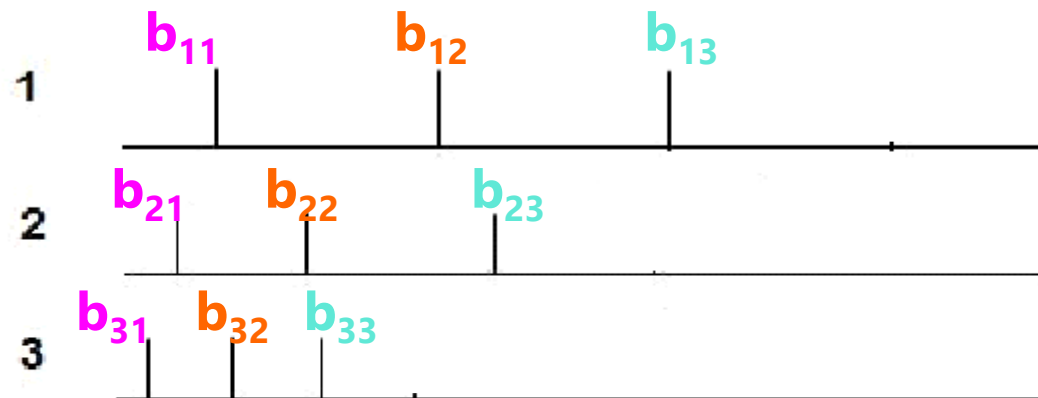
# “Modified” (or “Rating Scale”) Graded Response Model (MGRM)

- More parsimonious version of GRM for items with **same response options**
- In GRM,  $k \times \# \text{items}$  difficulties + 1 discrimination are estimated per item
- In MGRM, each item gets own slope and own overall “location” parameter of  $b_i$ , but the differences between categories around that location are constrained to be equal across items (get a  $c_k$  “shift” for each submodel)
  - So, different  $a_i$  and  $b_i$  per item, but same  $c_1$ ,  $c_2$ , and  $c_3$  across items
  - Not directly in Mplus, but I think it could be done using threshold constraints
- **0** vs. **1,2,3**:  $\text{Logit}[p(y_{is} > 0)] = a_i(\theta_s - b_i + c_1) \rightarrow p1(y_{is} > 0) = \frac{\exp[a_i(\theta_s - b_i + c_1)]}{1 + \exp[a_i(\theta_s - b_i + c_1)]}$
- **0,1** vs. **2,3**:  $\text{Logit}[p(y_{is} > 1)] = a_i(\theta_s - b_i + c_2) \rightarrow p2(y_{is} > 1) = \frac{\exp[a_i(\theta_s - b_i + c_2)]}{1 + \exp[a_i(\theta_s - b_i + c_2)]}$
- **0,1,2** vs. **3**:  $\text{Logit}[p(y_{is} > 2)] = a_i(\theta_s - b_i + c_3) \rightarrow p3(y_{is} > 2) = \frac{\exp[a_i(\theta_s - b_i + c_3)]}{1 + \exp[a_i(\theta_s - b_i + c_3)]}$

# Modified vs. Regular GRM



**Modified GRM** →  
 $k$  "shift"  $c_k$  terms:  
All category distances are same across items, and each item gets its own " $b_i$ " overall location



**Original GRM** →  
 $k$  locations per item: All category distances differ across items

Item Difficulty / Latent Ability



# Summary of Models for Ordered Categorical Responses

Available in Mplus via CATEGORICAL in VARIABLE: section	Difficulty Per Item Only (category distances equal)	Difficulty Per Category Per Item
<b>Equal discrimination across items (1-PLish)?</b>	(possible, but no special name)	(possible, but no special name)
<b>Unequal discriminations (2-PLish)?</b>	"Modified GRM" or "Rating Scale GRM" (same response options)	"Graded Response Model" "Cumulative Logit"

- GRM and Modified GRM are reliable models for ordered categorical data
  - Commonly used in real-world testing; most stable to use in practice
  - Least data demand because all responses get used in estimating each  $b_{ki}$
  - Only *major* deviations from the model will end up causing problems such that categories should be allowed to be unordered (perhaps bimodal distributions)

# (Generalized) Partial Credit Model (PCM)

- When you want to **test** an assumption of an ordered underlying continuum
- **# response options doesn't have to be same across items**; there is no guarantee that every category will be most likely at some point
- Is a "direct, divide-by-total" model (prob of each category is given directly)
- Estimate  $k$  difficulty-like  $\delta_{ki}$  "**step**" parameters, which are the Theta values at which the next category becomes **more likely** (not necessarily 50%)
- Models the probability of **adjacent categories** ("adjacent category logit")
  - Divide item into a series of binary items, but without order constraints beyond adjacent categories because it only uses those 2 categories ( $c = 0,1,2,3$ ):
- Models the probability of **lower vs. higher** cumulative categories via  $k$  submodels (e.g., if  $C = 4$  possible responses of  $c = 0,1,2,3$ ):

**0 vs. 1**  
Submodel<sub>1</sub>

**1 vs. 2**  
Submodel<sub>2</sub>

**2 vs. 3**  
Submodel<sub>3</sub>

Mplus will refer to these "steps" using \$1, \$2, \$3 after the item name

- Btw, if  $a_i$  is estimated separately across items, it's the "generalized" PCM

# Example Generalized Partial Credit Model (PCM)

**IRT version of the PCM: Note that  $a_i$  is the same across submodels...**

- **If 0** or **1**:  $\text{Logit}[p(y_{is} = 1)] = a_i(\theta_s - \delta_{1i}) \rightarrow p1(y_{is} = 1) = \frac{\exp[a_i(\theta_s - \delta_{1i})]}{1 + \exp[a_i(\theta_s - \delta_{1i})]}$
- **If 1** or **2**:  $\text{Logit}[p(y_{is} = 2)] = a_i(\theta_s - \delta_{2i}) \rightarrow p2(y_{is} = 2) = \frac{\exp[a_i(\theta_s - \delta_{2i})]}{1 + \exp[a_i(\theta_s - \delta_{2i})]}$
- **If 2** or **3**:  $\text{Logit}[p(y_{is} = 3)] = a_i(\theta_s - \delta_{3i}) \rightarrow p3(y_{is} = 3) = \frac{\exp[a_i(\theta_s - \delta_{3i})]}{1 + \exp[a_i(\theta_s - \delta_{3i})]}$

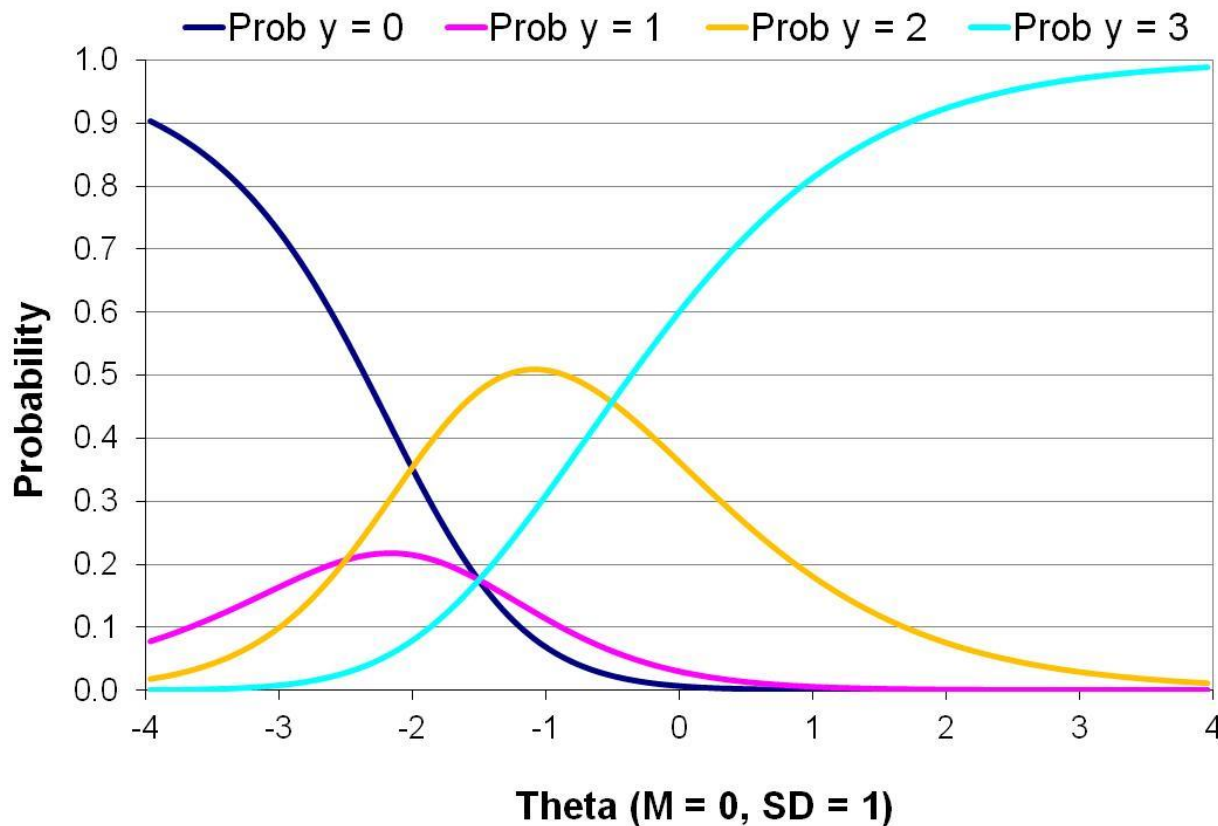
**IFA version of the PCM—what is estimated directly in Mplus:**

- **If 0** or **1**:  $\text{Logit}[p(y_{is} = 1)] = -\tau_{1i} + \lambda_i F_s \rightarrow p1(y_{is} = 1) = \frac{\exp(-\tau_{1i} + \lambda_i F_s)}{1 + \exp(-\tau_{1i} + \lambda_i F_s)}$
- **If 1** or **2**:  $\text{Logit}[p(y_{is} = 2)] = -\tau_{2i} + \lambda_i F_s \rightarrow p2(y_{is} = 2) = \frac{\exp(-\tau_{2i} + \lambda_i F_s)}{1 + \exp(-\tau_{2i} + \lambda_i F_s)}$
- **If 2** or **3**:  $\text{Logit}[p(y_{is} = 3)] = -\tau_{3i} + \lambda_i F_s \rightarrow p3(y_{is} = 3) = \frac{\exp(-\tau_{3i} + \lambda_i F_s)}{1 + \exp(-\tau_{3i} + \lambda_i F_s)}$

**IRT**  $\delta_{ki}$  = "location" = trait at which the **higher binary category** becomes more likely (not necessarily probability=.50, though)

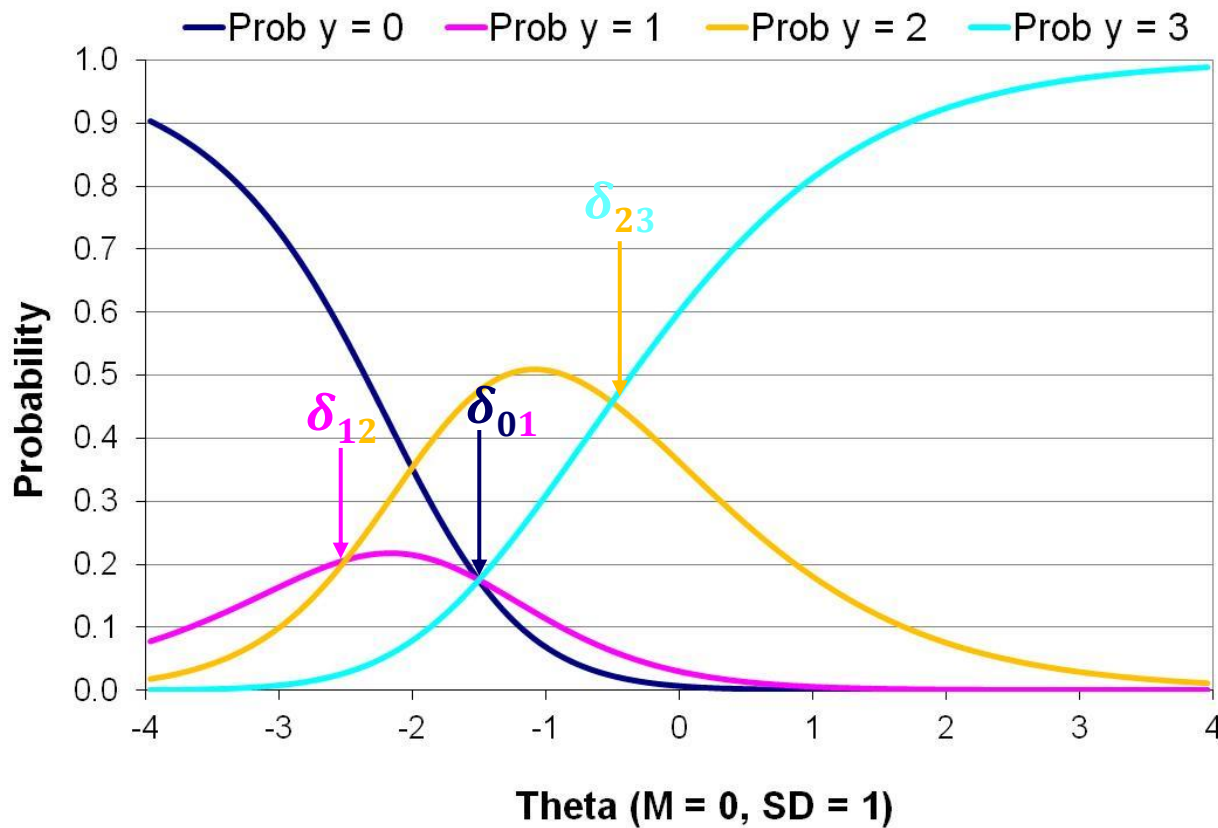
**IFA**  $\tau_{ki}$  = "step" = logit of the probability of the **lower binary category** when Factor=0 (yes, it's still a threshold)

# Category Response Curves: PCM for 4-Category (0|23) Item, $a_i = 1$



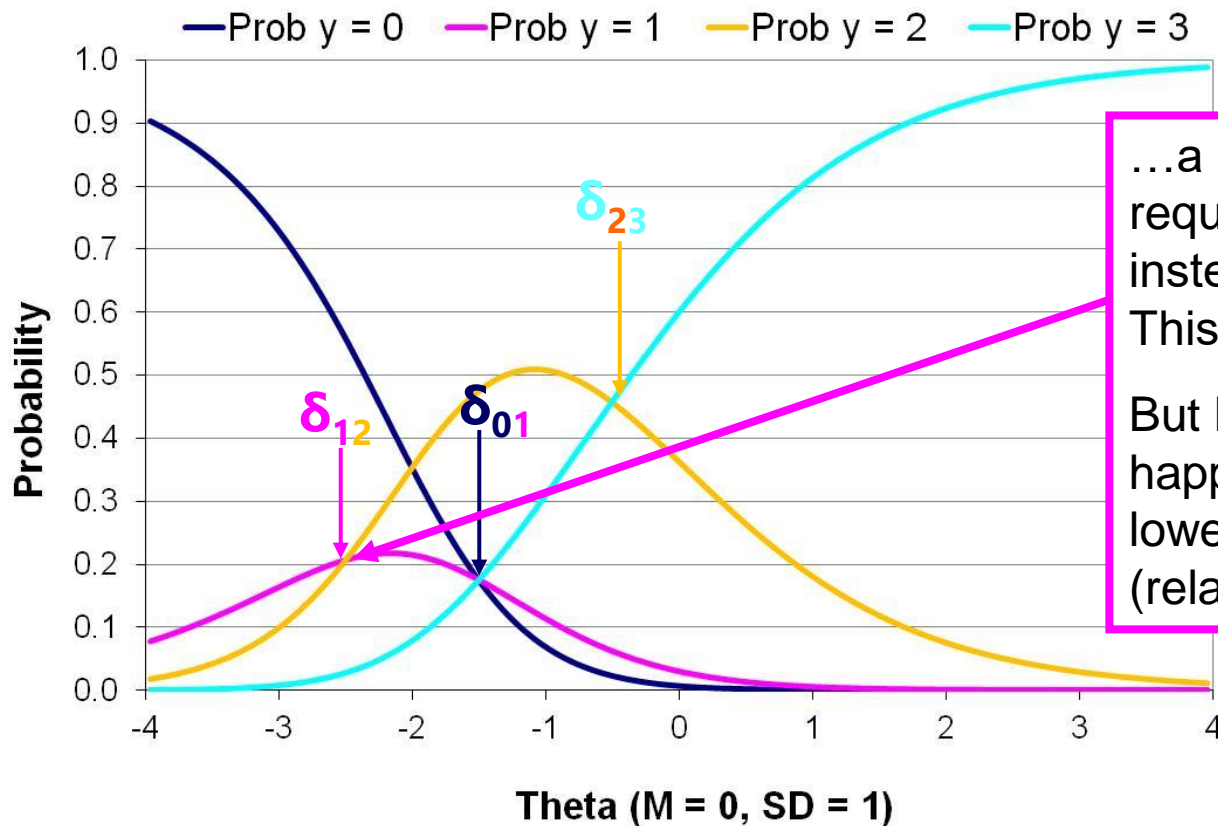
These curves look similar to the GRM, but the location parameters are interpreted differently because they are NOT cumulative, they are only adjacent...

# Category Response Curves: PCM for 4-Category (0|23) Item, $a_i = 1$



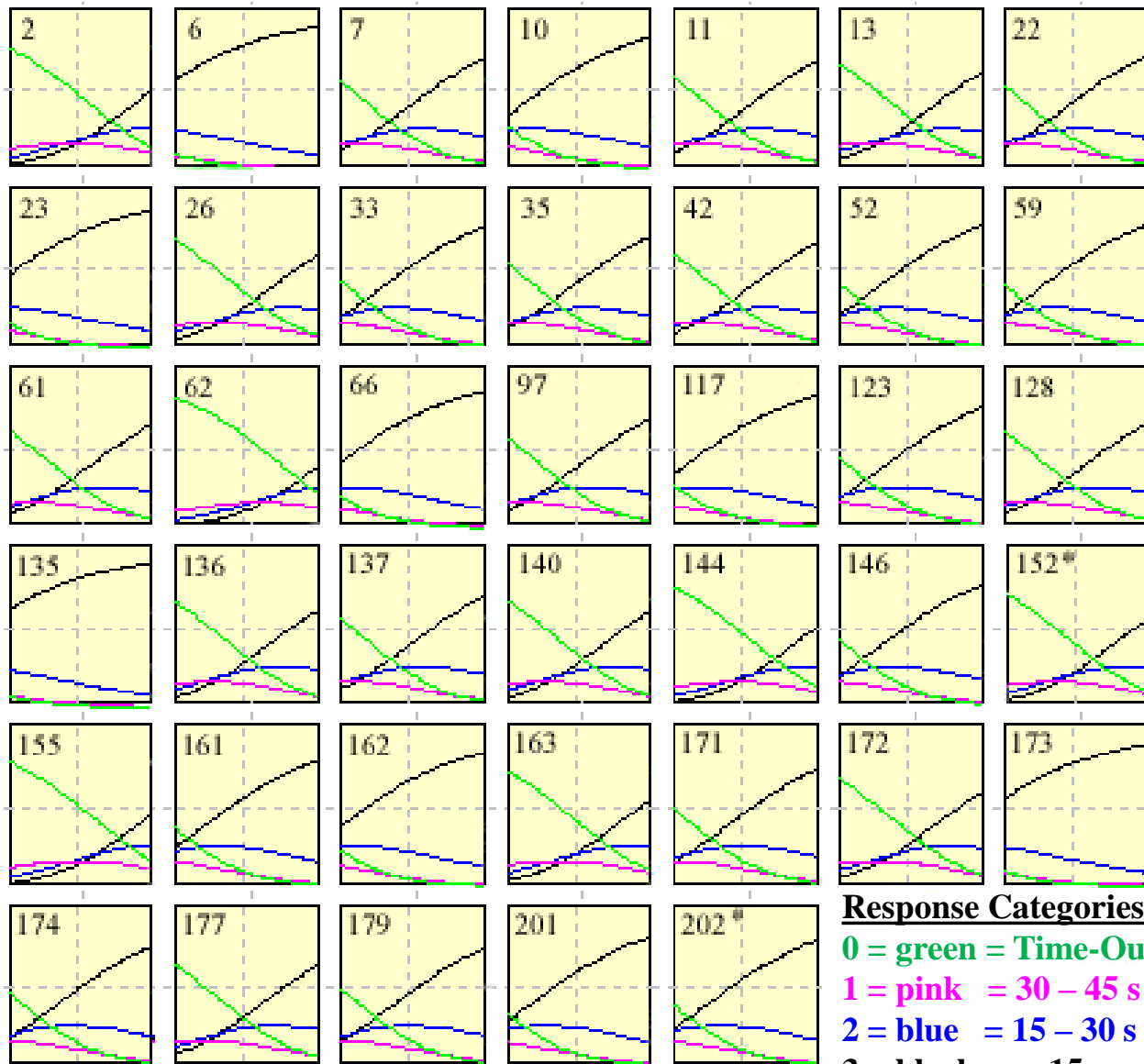
The  $\delta_{ki}$  terms are the location where the next category response becomes more likely (not necessarily 50%).

# Category Response Curves: PCM for 4-Category (0|23) Item, $a_i = 1$



...a score of 2 instead of 1 requires less Theta than 1 instead of 0 ... This is called a 'reversal'

But here, this likely happens because of a lower frequency of  $y_{is} = 1$  (relative to  $y_{is} = 0$ )



**Response Categories**

0 = green = Time-Out

1 = pink = 30 – 45 s

2 = blue = 15 – 30 s

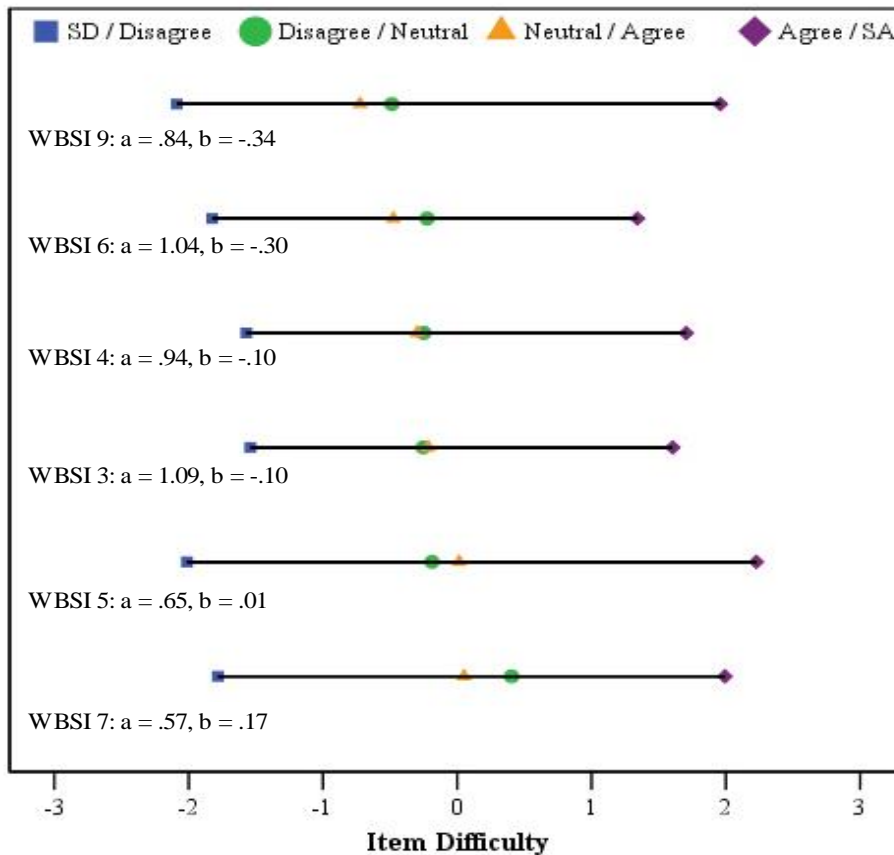
3 = black = < 15 s

\*Misfit ( $p < .05$ )

More of what you don't want to see... category response curves from a PCM where reversals are a plenty...

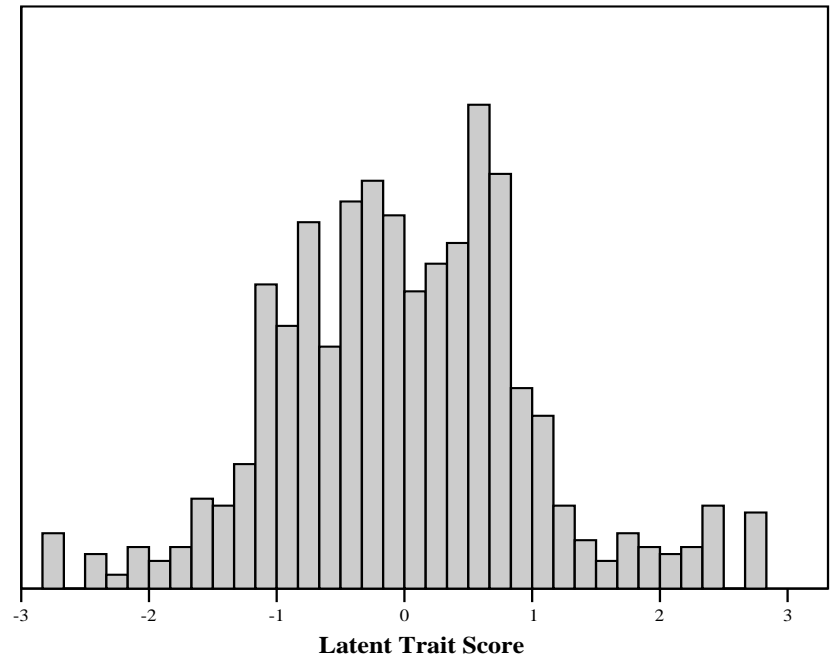
...and the middle categories are fairly useless.

# Generalized Partial Credit Model Example: General Intrusive Thoughts (5 options)

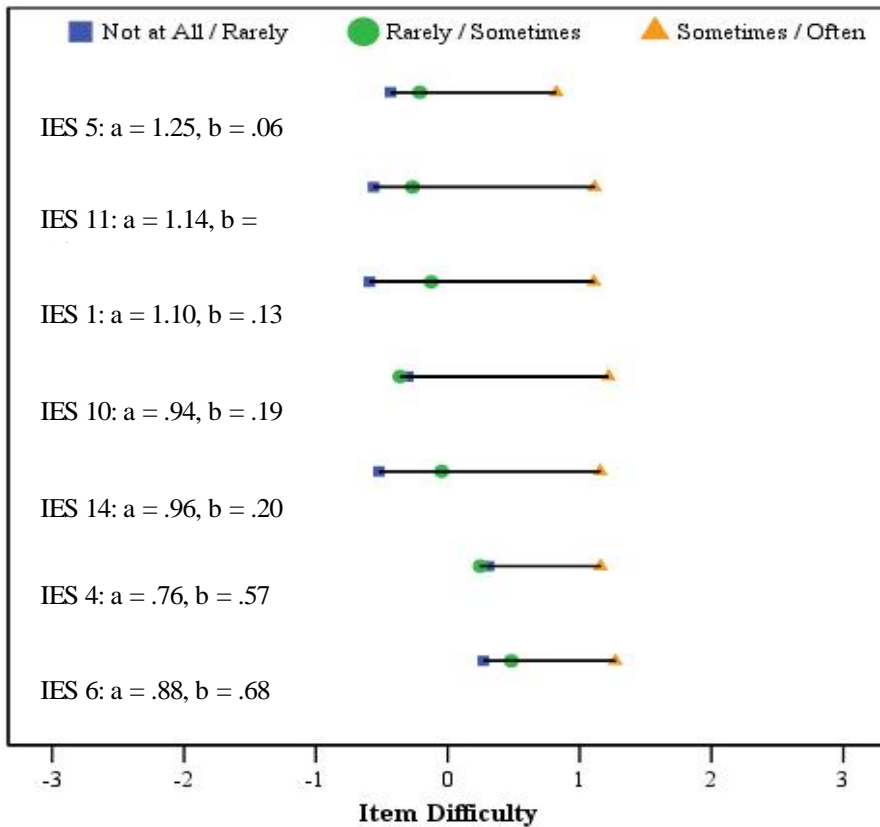


Note that the 4 thresholds **cover a wide range** of the latent trait, and what the distribution of Theta looks like as a result...

But the middle 3 categories are used infrequently and/or are not differentiable

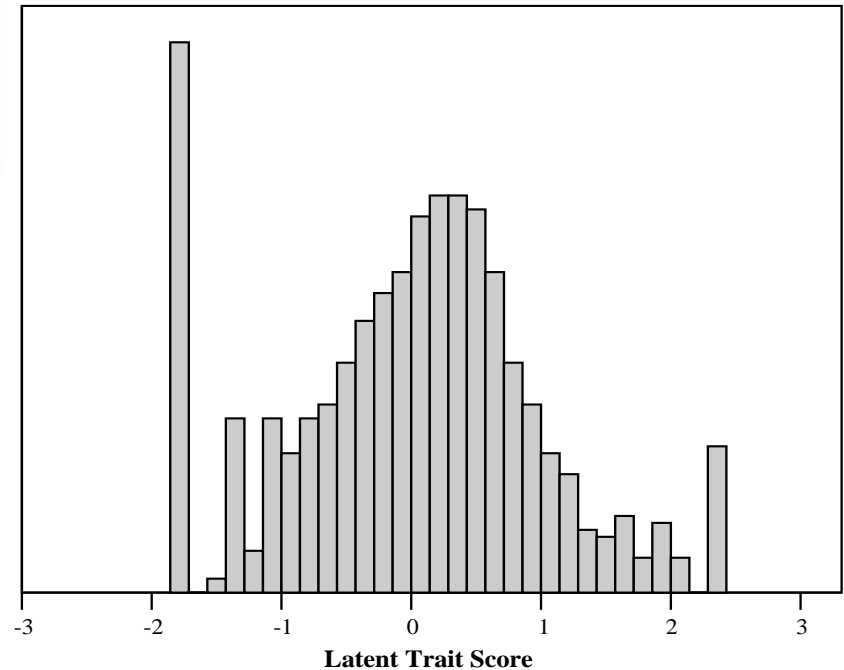






Note that the 3 thresholds **do not cover a wide range** of the latent trait, and what the distribution of theta looks like as a result...

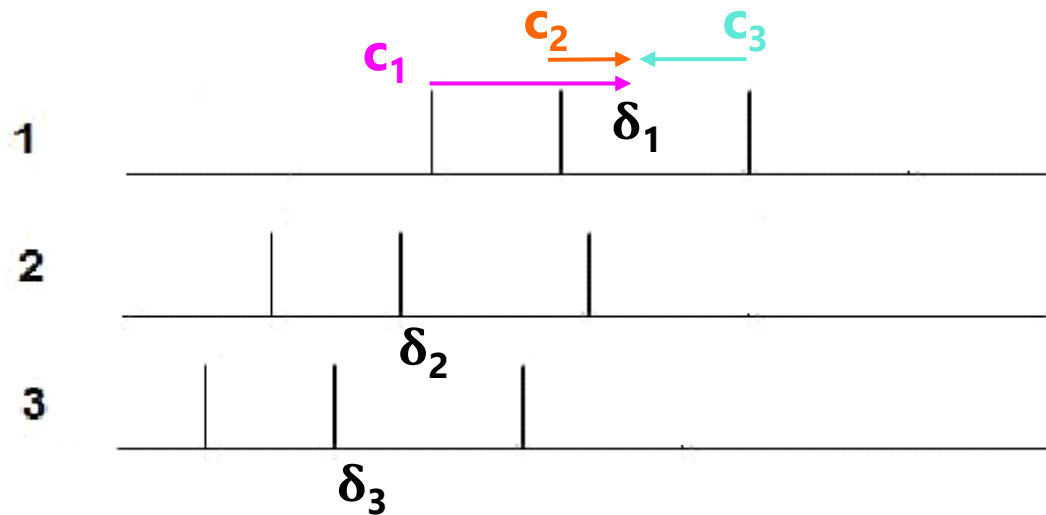
## Generalized Partial Credit Model Example: Event-Specific Intrusive Thoughts (4 options)



# Rating Scale Model (RSM)

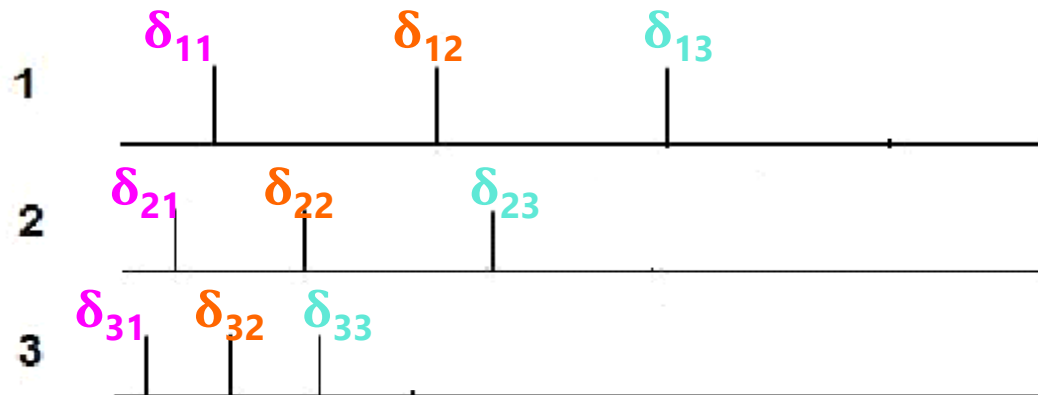
- More parsimonious version of PCM for items with **same response format**
- In PCM,  $k \times \# \text{items}$  step parameters + 1 discrimination are estimated per item
- In RSM, each item gets own slope and own overall 'location' parameter, but the differences between categories around that location are constrained to be equal across items (get a "c" shift for each threshold; **one c = 0**)
  - So, different  $b_i$  (and possibly  $a_i$ ) per item, but same  $c_1, c_2,$  and  $c_3$  across items
  - Also not available within Mplus (but perhaps could trick it using constraints)
- **If 0 or 1:**  $\text{Logit}[p(y_{is} = 1)] = a(\theta_s - \delta_i + c_1) \rightarrow p1(y_{is} = 1) = \frac{\exp[a(\theta_s - \delta_i + c_1)]}{1 + \exp[a(\theta_s - \delta_i + c_1)]}$
- **If 1 or 2:**  $\text{Logit}[p(y_{is} = 2)] = a(\theta_s - \delta_i + c_2) \rightarrow p2(y_{is} = 2) = \frac{\exp[a(\theta_s - \delta_i + c_2)]}{1 + \exp[a(\theta_s - \delta_i + c_2)]}$
- **If 2 or 3:**  $\text{Logit}[p(y_{is} = 3)] = a(\theta_s - \delta_i + c_3) \rightarrow p3(y_{is} = 3) = \frac{\exp[a(\theta_s - \delta_i + c_3)]}{1 + \exp[a(\theta_s - \delta_i + c_3)]}$

# Rating Scale vs. Regular PCM



## Rating Scale →

**$k$  "shift"  $c$  terms:** All category distances are same across items, and each item gets its own " $\delta_i$ " overall location



## Original PCM →

**$k$  locations per item:** All category distances differ across items

Item Difficulty / Latent Ability

# Summary of Models for Maybe-Ordered Categorical Responses

Available in Mplus via "CATEGORICAL" in "VARIABLE:" with (GPCM) option	<b>Difficulty Per Item Only (category distances equal)</b>	<b>Difficulty Per Category Per Item</b>
<b>Equal discrimination across items (1-PLish)?</b>	"Rating Scale PCM"	"Partial Credit Model"
<b>Unequal discriminations (2-PLish)?</b>	"Generalized Rating Scale PCM"?? (same response options)	"Generalized PCM" "Adjacent Category Logit"

- Partial Credit Models test the assumption of ordered categories
  - This can be useful for item screening, but perhaps not for actual analysis, because...
- These models have additional data demands relative to GRM
  - Only responses from that submodel get used (i.e., for 1 vs. 2, 0 and 3 don't contribute)
  - So larger sample sizes are needed to identify all model parameters
  - Sometimes categories may have to be consolidated to get the model to not blow up

# Nominal Response Model (NRM)

- Ideal for items with unknown ordering (e.g., what would you do? items)
- **# response options don't have to be the same across items**
- NRM is a "direct" model → prob of each category given directly
- Estimate  $k$   $a_i$  slopes per item and  $k$  "intercepts" per item
- Mplus parameterization (estimated as IFA model using NOMINAL ARE) models the probability of **each** category relative to a baseline category (highest by default) (sort of like dummy-coding, here baseline = 3):

**3 vs. 0**  
Submodel<sub>1</sub>

**3 vs. 1**  
Submodel<sub>2</sub>

**3 vs. 2**  
Submodel<sub>3</sub>

Mplus will refer to these "intercepts" using #1, #2, #3 after the item name

- Alternative parameterization sets sum of intercepts to 0 and sum of loadings to 0 (more useful if there is not an obvious reference category)
- Can be useful to examine utility of distractors in multiple choice tests
- Can also be constrained to become the GPCM for Likert-scale responses and stay nominal for unordered responses (like "not applicable")

# Example Nominal Response Model (NRM)

**IRT version of the NRM: Note that  $a_i$  is NOT the same across submodels...**

- **If 3** or **0**:  $\text{Logit}[p(y_{is} = 0)] = a_{1i}(\theta_s - b_{1i}) \rightarrow p1(y_{is} = 0) = \frac{\exp[a_{1i}(\theta_s + b_{1i})]}{1 + \exp[a_{1i}(\theta_s + b_{1i})]}$
- **If 3** or **1**:  $\text{Logit}[p(y_{is} = 1)] = a_{2i}(\theta_s - b_{2i}) \rightarrow p2(y_{is} = 1) = \frac{\exp[a_{2i}(\theta_s + b_{2i})]}{1 + \exp[a_{2i}(\theta_s + b_{2i})]}$
- **If 3** or **2**:  $\text{Logit}[p(y_{is} = 2)] = a_{3i}(\theta_s - b_{3i}) \rightarrow p3(y_{is} = 2) = \frac{\exp[a_{3i}(\theta_s + b_{3i})]}{1 + \exp[a_{3i}(\theta_s + b_{3i})]}$

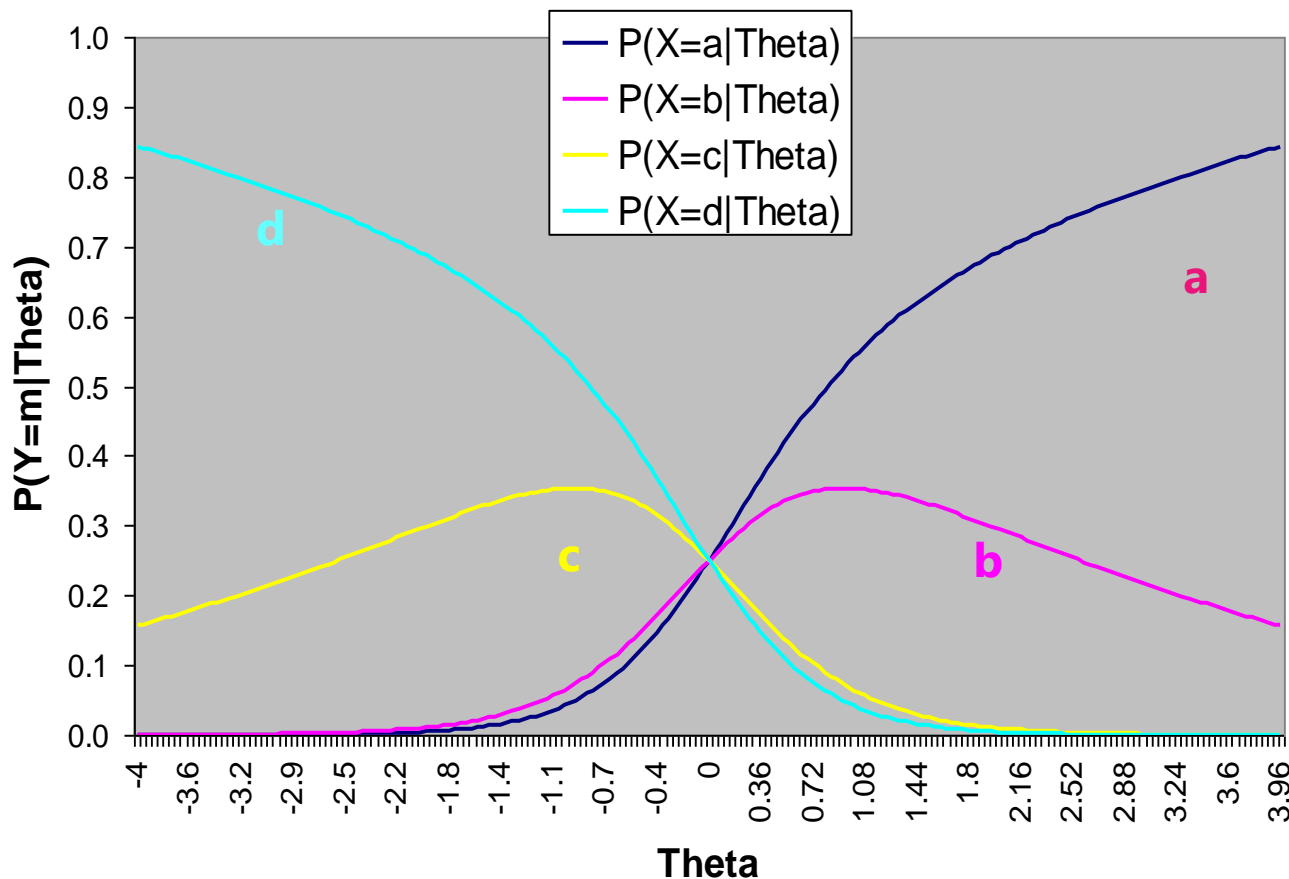
**IFA version of the GRM—what is estimated in Mplus (with intercepts):**

- **If 3** or **0**:  $\text{Logit}[p(y_{is} = 0)] = -\tau_{1i} + \lambda_{1i}F_s \rightarrow p1(y_{is} = 0) = \frac{\exp(\mu_{1i} + \lambda_{1i}F_s)}{1 + \exp(\mu_{1i} + \lambda_{1i}F_s)}$
- **If 3** or **1**:  $\text{Logit}[p(y_{is} = 1)] = -\tau_{2i} + \lambda_{2i}F_s \rightarrow p2(y_{is} = 1) = \frac{\exp(\mu_{2i} + \lambda_{2i}F_s)}{1 + \exp(\mu_{2i} + \lambda_{2i}F_s)}$
- **If 3** or **2**:  $\text{Logit}[p(y_{is} = 2)] = -\tau_{3i} + \lambda_{3i}F_s \rightarrow p3(y_{is} = 2) = \frac{\exp(\mu_{3i} + \lambda_{3i}F_s)}{1 + \exp(\mu_{3i} + \lambda_{3i}F_s)}$

**IRT  $b_{ki}$**  = “difficulty” = trait needed for a 50% probability (logit=0) of the **non-reference binary category**; **IFA  $\mu_{ki}$**  = “intercept” = logit of the probability of the **non-reference binary category** when Factor=0

# Category Response Curves (NRM for 5-Category Item)

## Nominal Response Item Response Function



Example Analysis  
of Multiple Choice  
Distractors:

People low in Theta  
are most likely to  
pick **d**, but **c** is their  
second choice

People high in Theta  
are most likely to  
pick **a**, but **b** is their  
second choice

# Summary: Polytomous Models

Many kinds of polytomous IRT/IFA models...

- Some **assume order** of response options...
  - **Graded Response Model Family** → “cumulative logit model”
    - Models cumulative change in categories using all data for each submodel
- Some allow you to **test order** of sequential response options...
  - **(Generalized) Partial Credit Model Family** → “adjacent category logit model”
    - Models adjacent category thresholds only, so they allow you to see reversals (empirical mis-ordering of your response options with respect to Theta)
    - (G)PCM useful for identifying separability and adequacy of categories
- Some estimate order of response options...
  - **Nominal Model** → “baseline category logit model”
    - Useful to examine probability of each response option or semi-ordered options
    - Can be hard to estimate all parameters (especially separate  $\alpha$  or  $\lambda$  per submodel)



# The Bigger Picture of Latent Trait Models

- When fitting latent factor models (or when just predicting observed outcomes from observed predictors instead), you have many options to fit non-normal response outcome distributions...
  - **CFA:** Continuous outcomes with normal residuals,  $X \rightarrow Y$  is linear
    - If residuals may not be normal *but a linear  $X \rightarrow Y$  relationship is still plausible*, you can use MLR estimation instead of ML to control for that
  - **IRT and IFA:** Categorical or ordinal outcomes with Bernoulli/multinomial residuals,  $X \rightarrow$  transformed  $Y$  is linear;  $X \rightarrow$  original  $Y$  is nonlinear
    - Full information MML traditionally paired with IRT version of model; limited information WLSMV traditionally paired with IFA version of model instead
    - See example 6b from [CLDP 948](#) for how to address “not applicable” responses
  - **Count family:** Non-negative integer outcomes,  $X \rightarrow \text{Log}(Y)$  is linear
    - Residuals can be Poisson (where mean = variance) or negative binomial (where variance > mean); either can be zero-inflated or zero-truncated
    - Hurdle or two-part may be more direct way to predict/interpret excess zeros (predict zero or not and how much, rather than two kinds of zeros)
    - See example 6c from [CLDP 948](#) for factor models for count-type outcomes