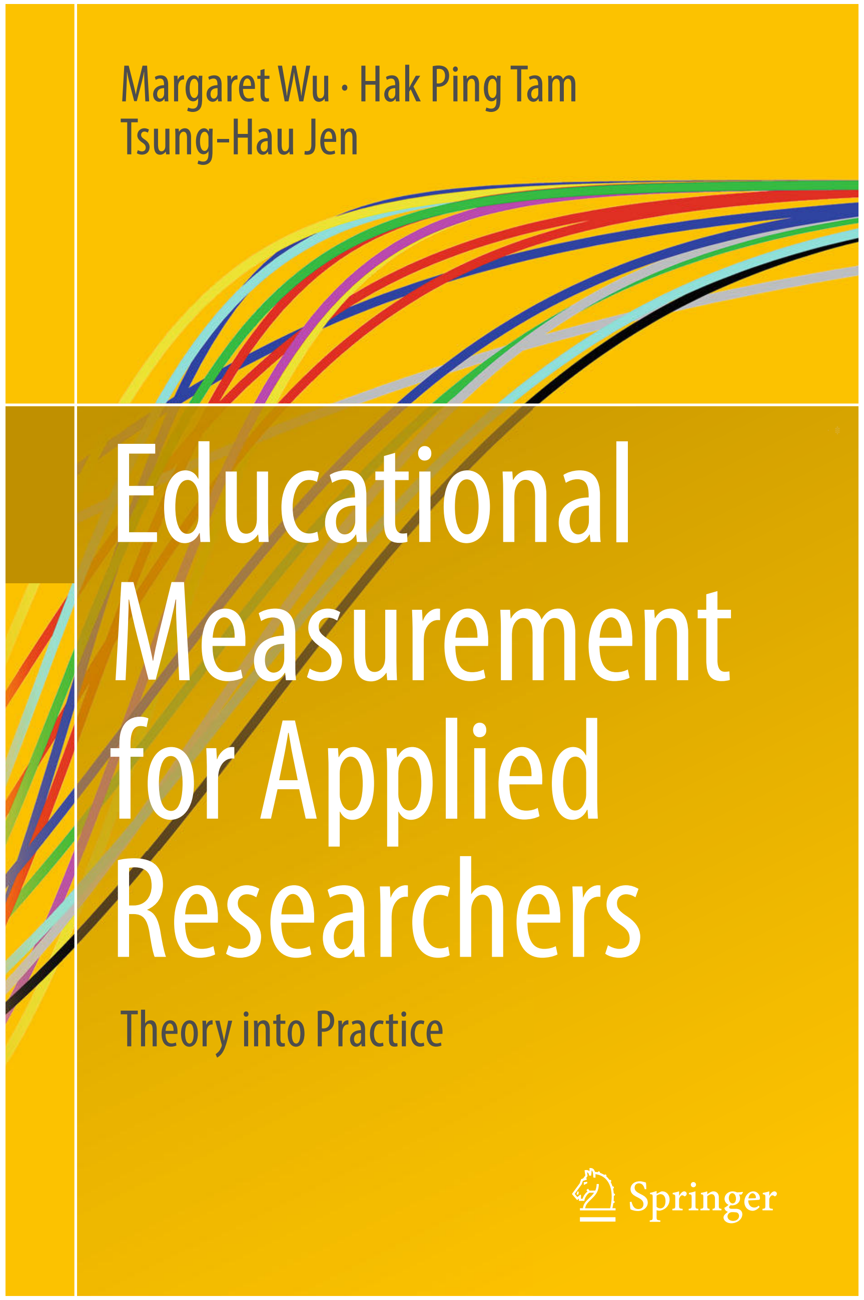
Modelo de Rasch-Master de Créditos parciais (Partial Credit Model)

Ricardo Primi
USF 2019
Programa de Pós Graduação Stricto Sensu em Psicologia

UNIVERSIDADE SÃO FRANCISCO

Capítulo 9


Margaret Wu · Hak Ping Tam
Tsung-Hau Jen
Educational Measurement for Applied Researchers
Theory into Practice
Springer

PSYCHOMETRIKA-VOL. 47, NO. 2.
JUNE, 1982

A RASCH MODEL FOR PARTIAL CREDIT SCORING

GEOFF N. MASTERS

UNIVERSITY OF CHICAGO

A unidimensional latent trait model for responses scored in two or more ordered categories is developed. This "Partial Credit" model is a member of the family of latent trait models which share the property of parameter separability and so permit "specifically objective" comparisons of persons and items. The model can be viewed as an extension of Andrich's Rating Scale model to situations in which ordered response alternatives are free to vary in number and structure from item to item. The difference between the parameters in this model and the "category boundaries" in Samejima's Graded Response model is demonstrated. An unconditional maximum likelihood procedure for estimating the model parameters is developed.

Key words: latent trait, Rasch model, ordered categories, partial credit.

PCM

- Dicotômico vs Politômico
- Rubricas
- Categorias ordenadas
- K categorias = K- limiares (thresholds)

O modelo de créditos parciais

The PCM specifies that, while conditioning on scoring a 0 or 1 (i.e., we know the score is either 0 or 1), the probability of a score of zero ($X = 0$) and the probability of a score of 1 ($X = 1$) are given by

$$\begin{aligned} p_{0/0,1} &= \Pr(X = 0/X = 0 \quad \text{or} \quad X = 1) = \frac{\Pr(X = 0)}{\Pr(X = 0) + \Pr(X = 1)} \\ &= \frac{1}{1 + \exp(\theta - \delta_1)} \end{aligned} \tag{9.1}$$

$$\begin{aligned} p_{1/0,1} &= \Pr(X = 1/X = 0 \quad \text{or} \quad X = 1) = \frac{\Pr(X = 1)}{\Pr(X = 0) + \Pr(X = 1)} \\ &= \frac{\exp(\theta - \delta_1)}{1 + \exp(\theta - \delta_1)} \end{aligned} \tag{9.2}$$

$$P(x_i = 1 \mid \theta, \delta_i) = \frac{P(x=1)}{P(x=0)+P(x=1)} = \frac{\exp(\theta - \delta_i)}{1+\exp(\theta - \delta_i)} \quad (1)$$

For this dichotomous case we have two probability equations:

$$P(x=0) = \frac{1}{1+\exp(\theta - \delta_1)} \text{; and}$$

$$P(x=1) = \frac{\exp(\theta - \delta_1)}{1+\exp(\theta - \delta_1)}.$$

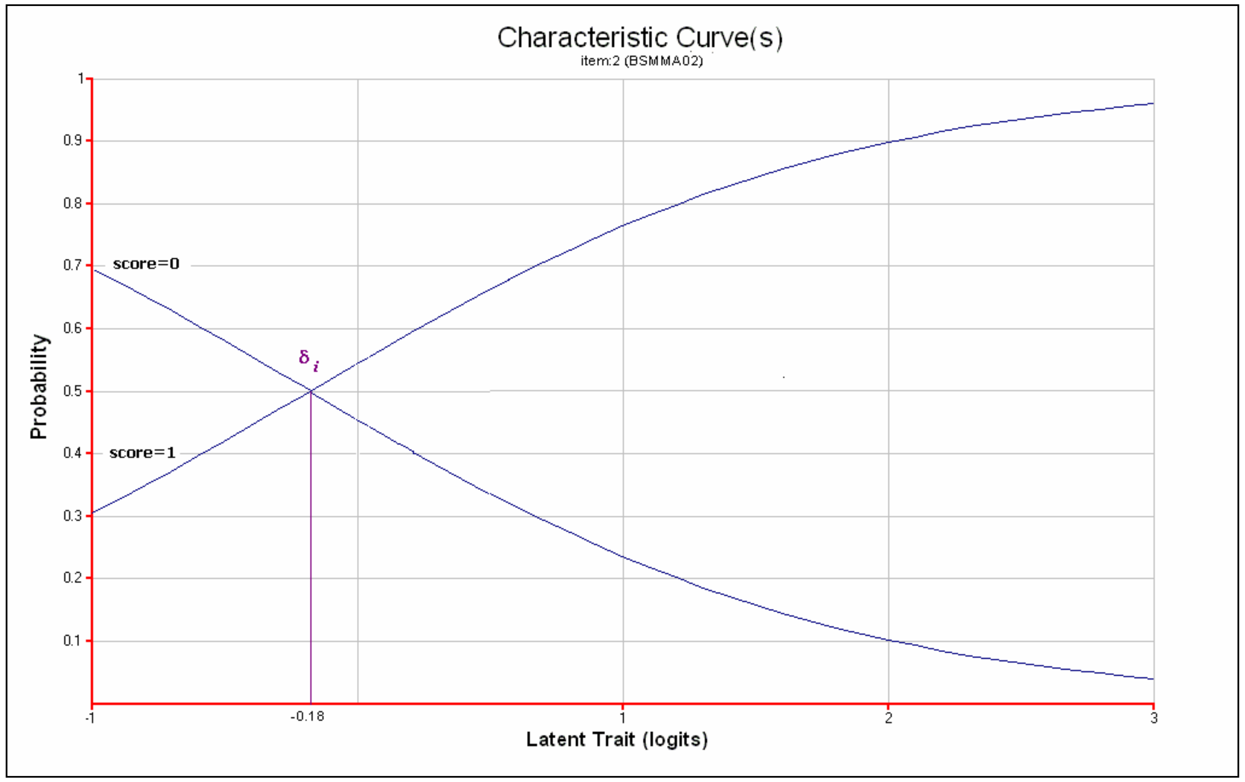


Figure 4. Item characteristic curve for a dichotomous (2-category) item.

Kennedy (2005). Constructing Measurement Models for MRCML Estimation: A Primer for Using the BEAR Scoring Engine. Berkeley: BEAR

PCM especifica a probabilidade condicional de duas categorias adjacentes

Equations (9.1) and (9.2) are in the form of the dichotomous Rasch probabilities.

Similarly, conditional on scoring a 1 or 2, the probability of X = 1 and the probability of X = 2 are given by

$$p_{1/1,2} = \Pr(X = 1/X = 1 \quad \text{or} \quad X = 2) = \frac{\Pr(X = 1)}{\Pr(X = 1) + \Pr(X = 2)}$$
$$= \frac{1}{1 + \exp(\theta - \delta_2)} \tag{9.3}$$

$$p_{2/1,2} = \Pr(X = 2/X = 1 \quad \text{or} \quad X = 2) = \frac{\Pr(X = 2)}{\Pr(X = 1) + \Pr(X = 2)}$$
$$= \frac{\exp(\theta - \delta_2)}{1 + \exp(\theta - \delta_2)} \tag{9.4}$$

Equations (9.3) and (9.4) are also in the form of the dichotomous Rasch probabilities.

PCM Probabilities for All Response Categories

While the derivation of the PCM is based on specifying probabilities for adjacent score categories, the probability for each score, when all score categories are considered collectively, can be derived. The following gives the probability of each score category for a 3-category (0, 1, 2) PCM.

$$p_0 = \Pr(X = 0) = \frac{1}{1 + \exp(\theta - \delta_1) + \exp(2\theta - (\delta_1 + \delta_2))} \tag{9.5}$$

$$p_1 = \Pr(X = 1) = \frac{\exp(\theta - \delta_1)}{1 + \exp(\theta - \delta_1) + \exp(2\theta - (\delta_1 + \delta_2))} \tag{9.6}$$

$$p_2 = \Pr(X = 2) = \frac{\exp(2\theta - (\delta_1 + \delta_2))}{1 + \exp(\theta - \delta_1) + \exp(2\theta - (\delta_1 + \delta_2))} \tag{9.7}$$

More generally, if item i is a polytomous item with score categories 0, 1, 2, …, m_i, the probability of person n scoring x on item i is given by

$$\Pr(X_{ni} = x) = \frac{\exp \sum_{k=0}^{x} (\theta_n - \delta_{ik})}{\sum_{h=0}^{m_i} \exp \sum_{k=0}^{h} (\theta_n - \delta_{ik})} \tag{9.8}$$

where we define $\exp \sum_{k=0}^{0} (\theta_n - \delta_{ik}) = 1$, and hence when the score is 0, the numerator of Eq. (9.8) is 1. The summation index k refers to score categories.

Note that the number of δ_k parameters is one less than the number of response categories. For example, if there are three response categories, 0, 1 and 2, then there are two δ parameters, δ_1 and δ_2. In the same way as for dichotomous items, when there are two response categories (e.g., correct and incorrect), there is one item difficulty parameter, δ.

Wu, M. & Adams, R. (2007). Applying the Rasch model to psycho-social measurement: A practical approach. Educational Measurement Solutions, Melbourne.

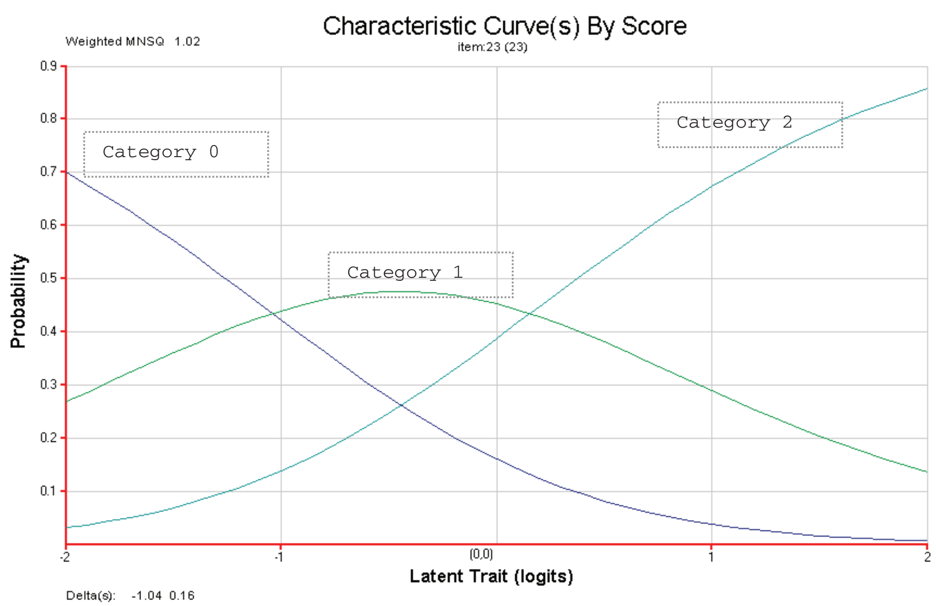


Figure 17 Theoretical Item Characteristic Curves for a 3-category Partial Credit Item

From Figure 17, it can be seen that as ability increases, the probability of being in a higher score category also increases.

Graphical interpretation of the delta (δ) parameters

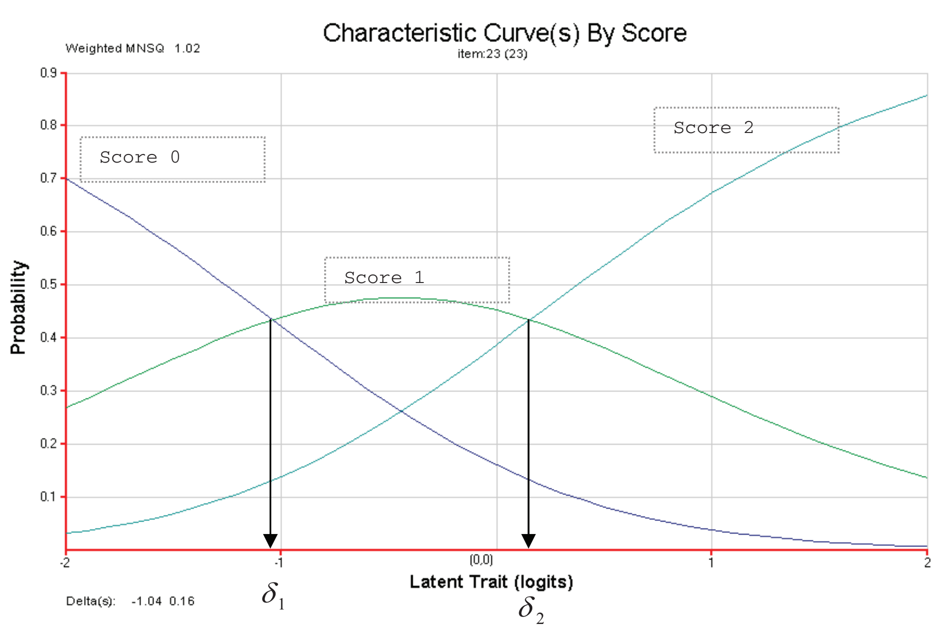


Figure 18 Graphical representations of the delta (δ) parameters

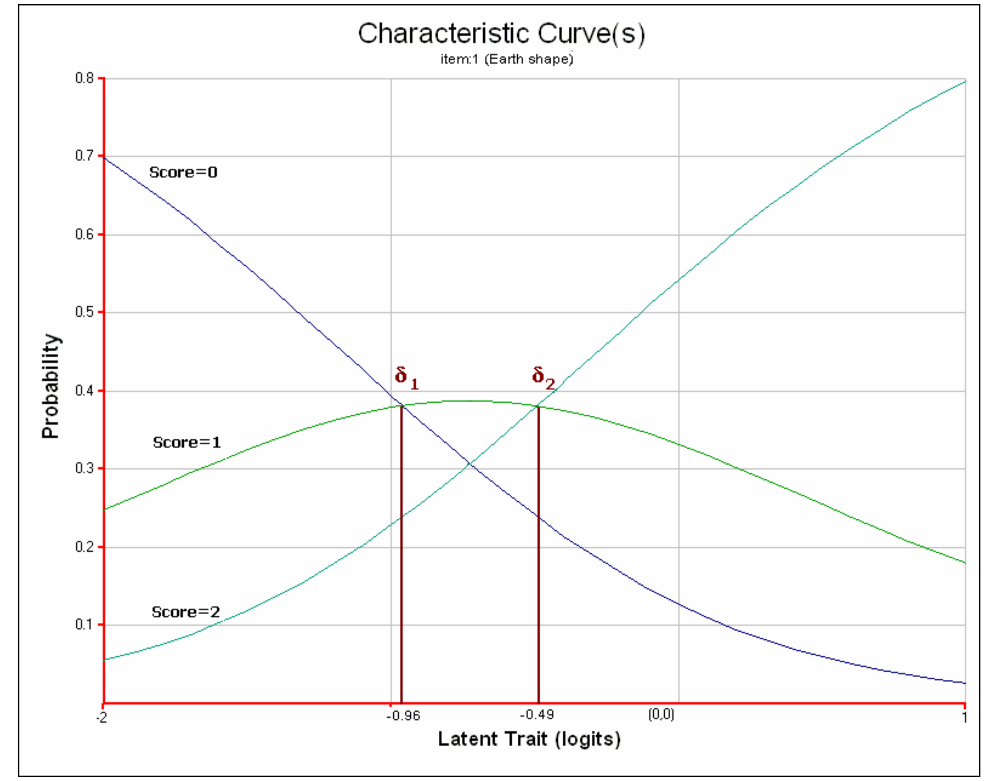


Figure 5. Category probability curves and δ_{ij} values for a 3-category polytomous item.

Most likely single score category
mas .. nao >.50

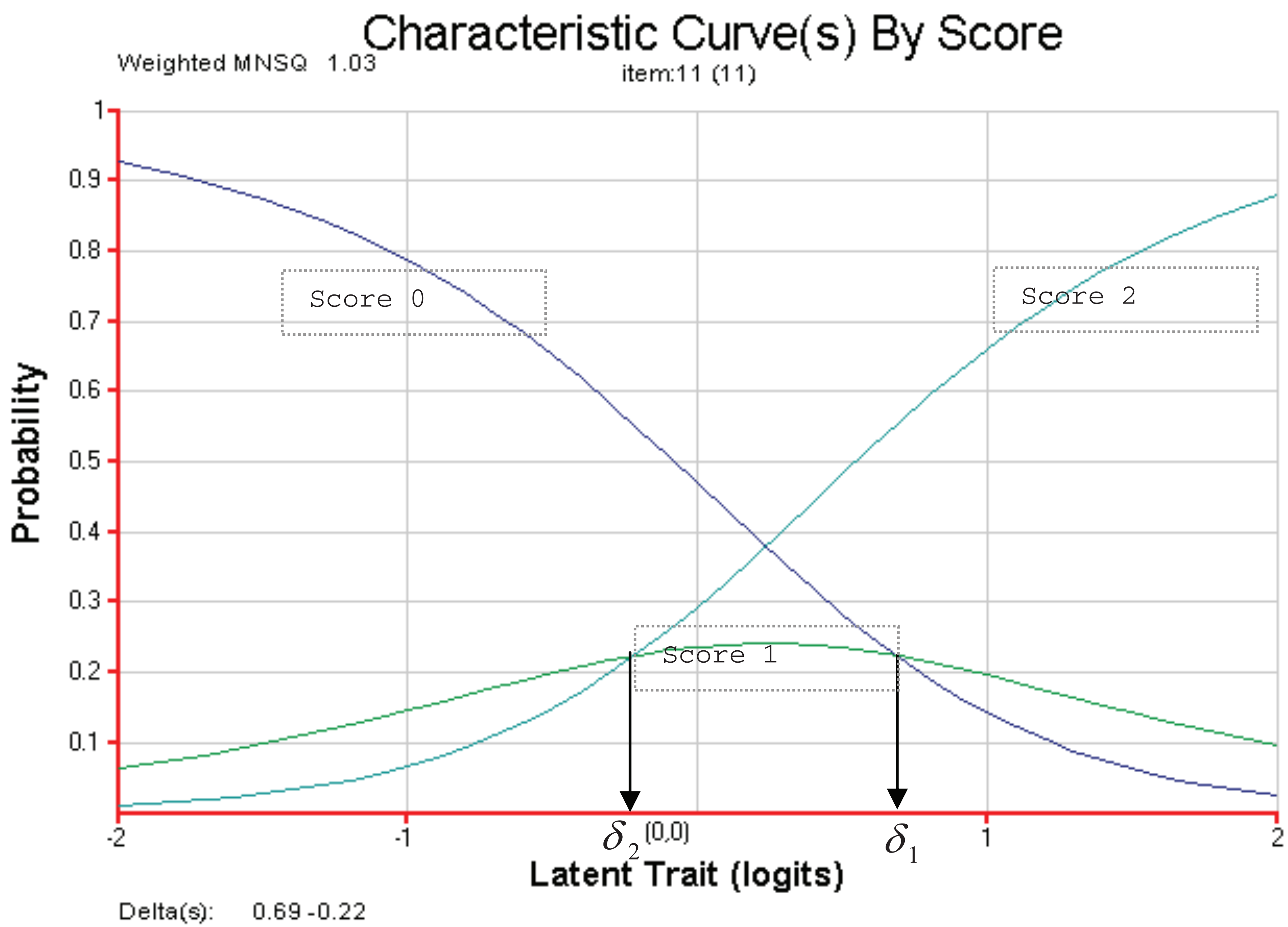


Figure 19 ICC for PCM where the delta parameters are dis-ordered

Item 5 - pharm

In the Pharmochem company, there are 57 employees. Each employee speaks either German or English, or both. 25 employees can speak German and 48 employees can speak English. How many employees can speak both German and English? Show how you found your answer.

Item analysis (Item 5 – pharm)

Response	Score	Count	% of tot	Pt Bis
16*	2	293	61.68	0.43
comp err	1	18	3.79	0.01
Other	0	117	24.63	-0.36

Scoring guide:

Fully correct answer was given a score of 2. For responses with correct method but incorrect computation, a score of 1 was awarded.

*Correct answer

Fig. 9.3 Item statistics for a partial credit scoring mathematics item

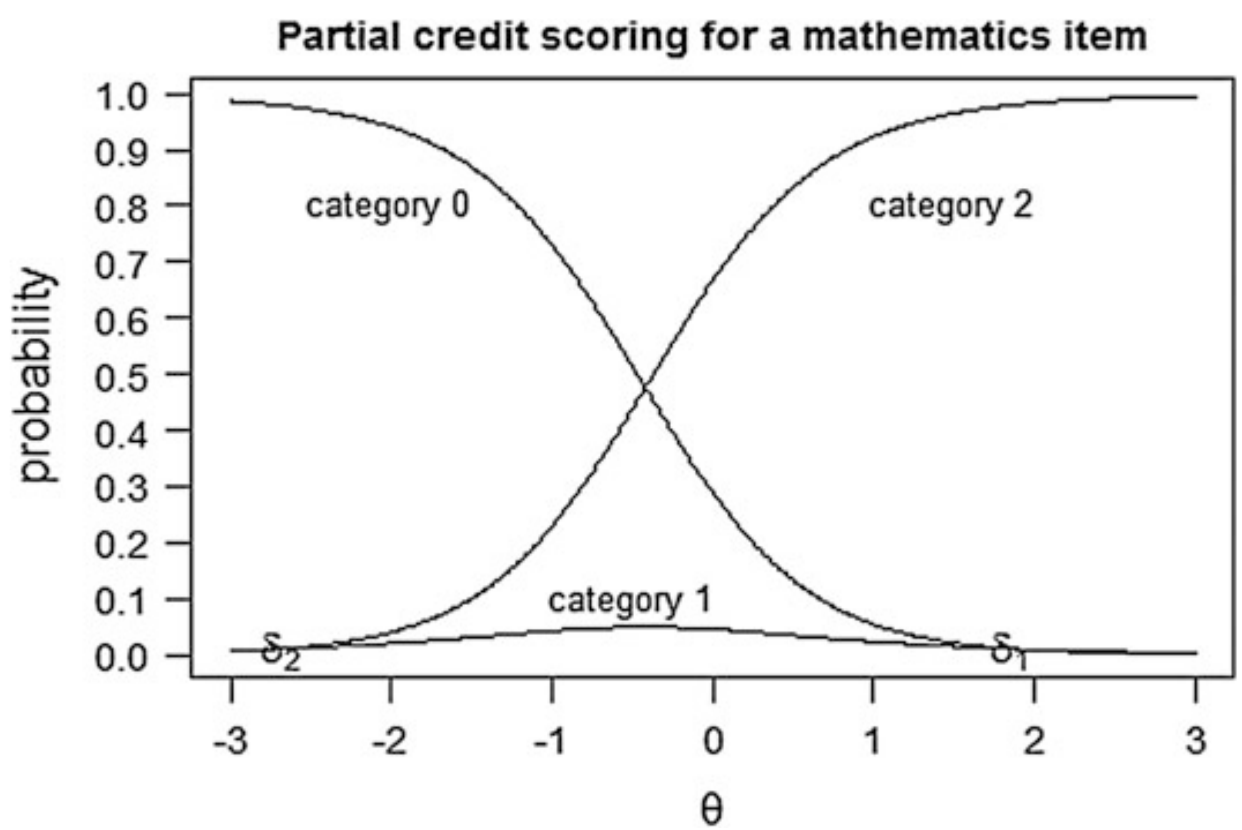


Fig. 9.4 ICC for a partial credit mathematics item with dis-ordered thresholds

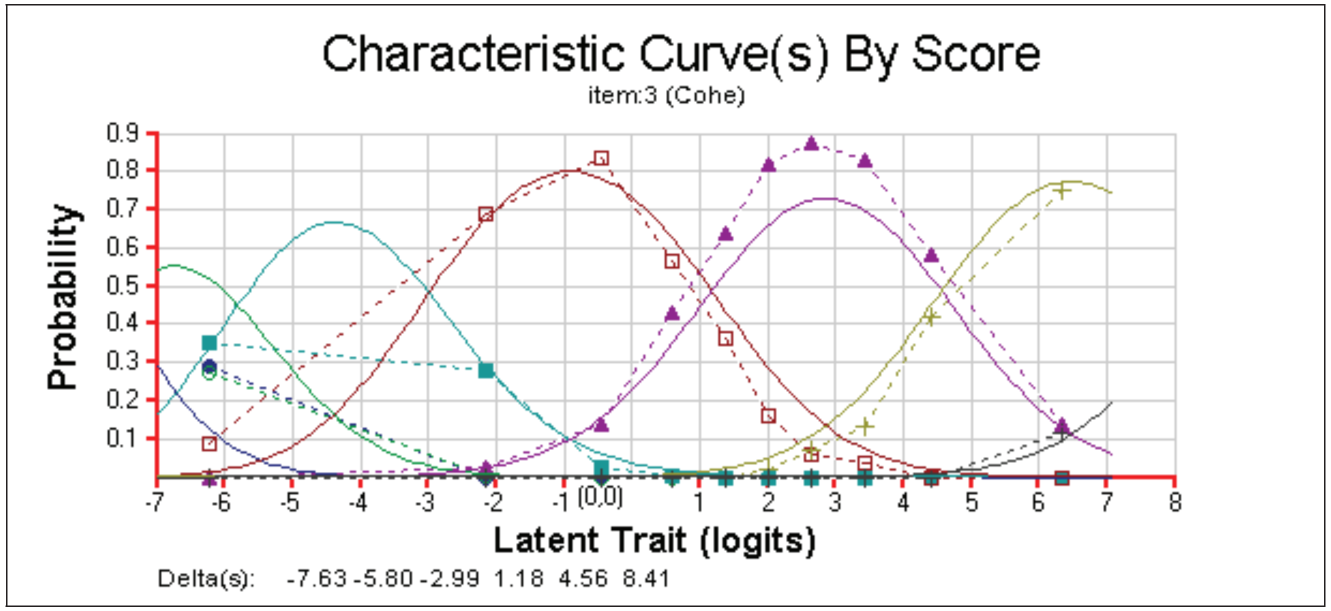


Figure 22 ICC for an essay marking criterion, "Cohesion", using PCM on a 6-point scale

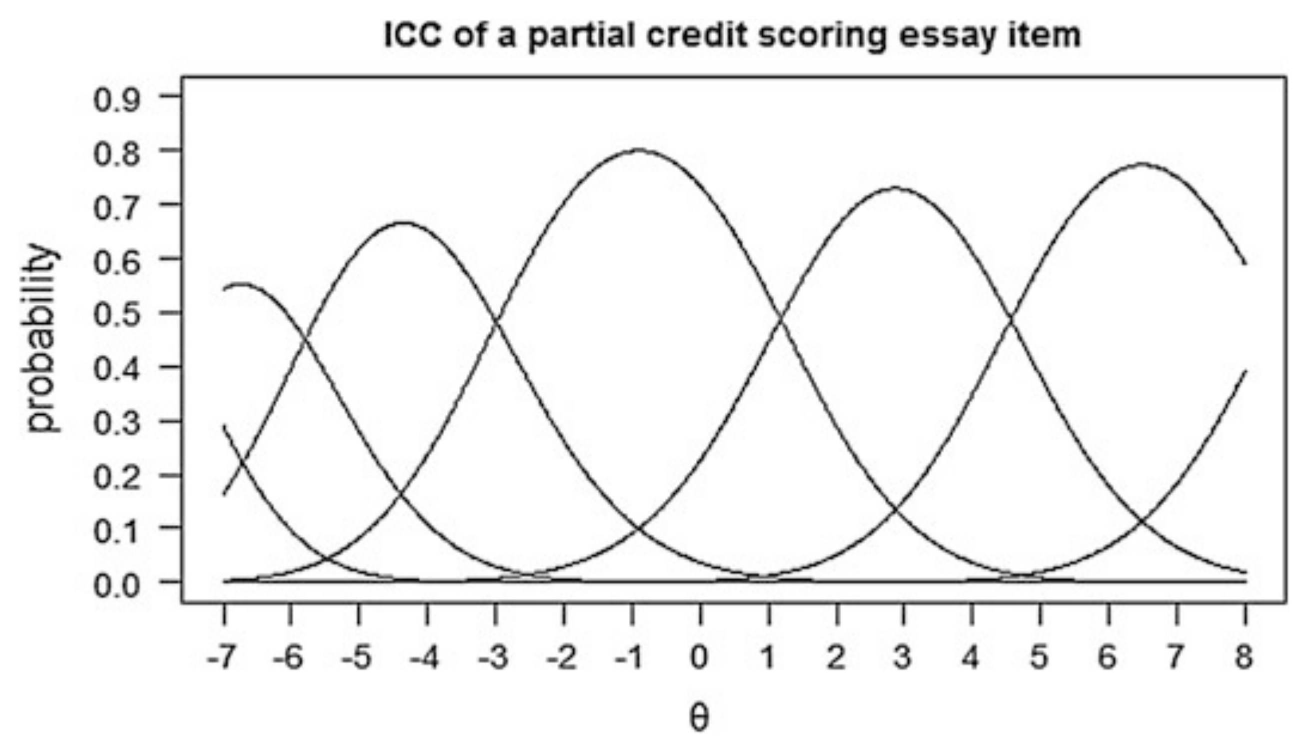


Fig. 9.5 ICC for an essay marking criterion, "Cohesion", using PCM on a 7-point scale

Tau e delta dot

A variation of the parameterisation of the PCM is the use of τ's (tau's) and $\delta_\bullet$ (delta dot). Mathematically, the delta (δ_{ik}) parameters in Eq. (9.8) can be re-written in the following way:

Using the notations as in Eq. (9.8) but dropping the index i for simplicity, let

$$\delta_\bullet = \sum_{k=1}^{m} \delta_k / m \tag{9.9}$$

where m is the maximum score. That is, the total number of response categories of an item is $m+1$.

Equation (9.9) shows that $\delta_\bullet$ is the average of the delta (δ_k) parameters for one item.

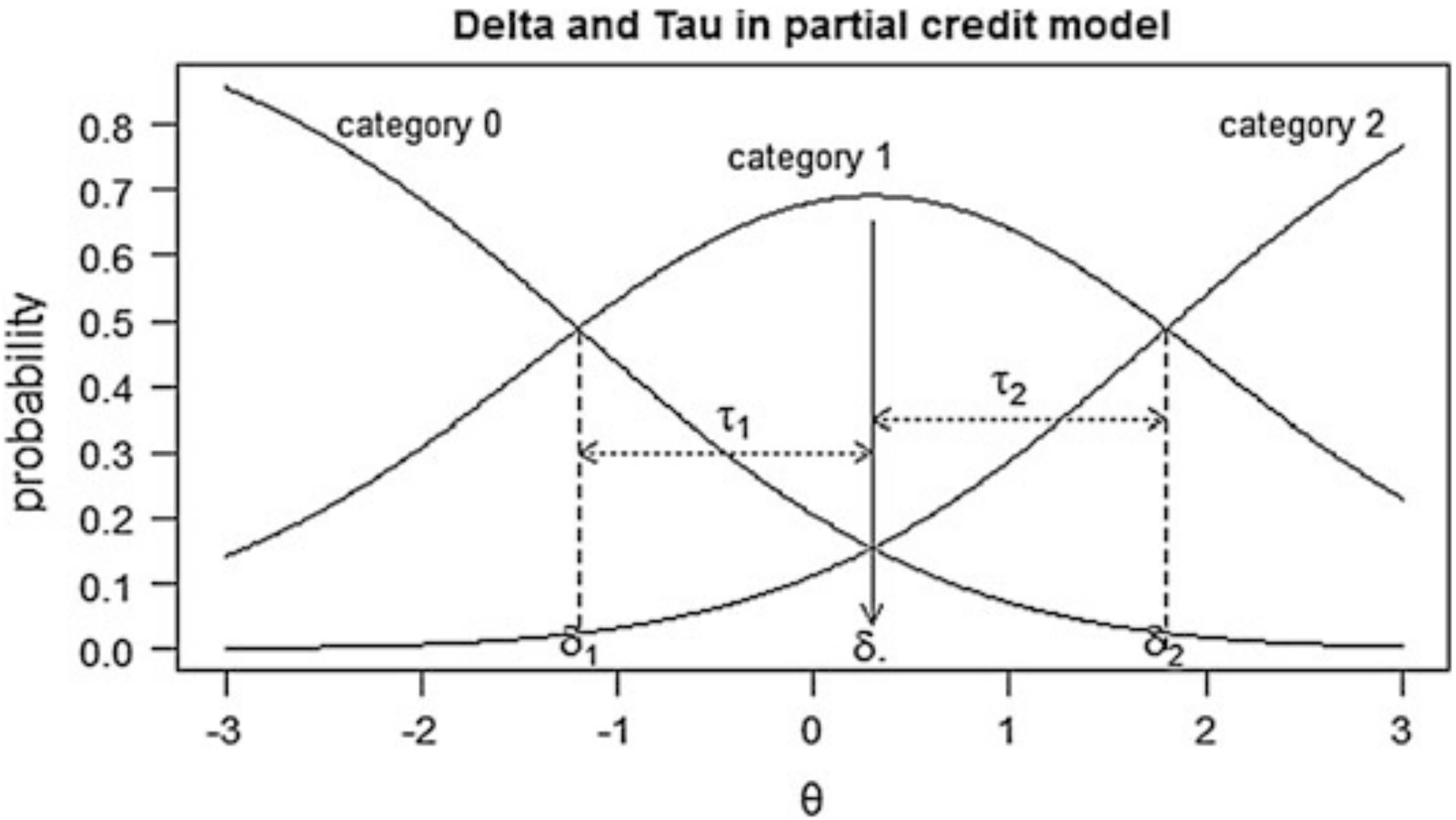
Next, let us define τ_k as the difference between $\delta_\bullet$ and δ_k. That is,

$$\tau_k = \delta_\bullet - \delta_k \tag{9.10}$$

Graphically, the relationships among τ_k, $\delta_\bullet$ and δ_k are illustrated in Fig. 9.6.

The parameterisation of the PCM using $\delta_\bullet$ and τ_k is mathematically equivalent to the parameterisation using δ_k. Using Eqs. (9.9) and (9.10), one can compute $\delta_\bullet$ and τ_k from δ_k. Conversely, given τ_k, and $\delta_\bullet$, one can compute δ_k as

$$\delta_k = \delta_\bullet - \tau_k \tag{9.11}$$



$$\sum_{k=1}^{m} \delta_k = \sum_{k=1}^{m} (\delta_\bullet - \tau_k)$$

$$\sum_{k=1}^{m} \delta_k = \sum_{k=1}^{m} \delta_\bullet - \sum_{k=1}^{m} \tau_k$$

$$\sum_{k=1}^{m} \delta_k = \sum_{k=1}^{m} \frac{\sum_{k=1}^{m} \delta_k}{m} - \sum_{k=1}^{m} \tau_k$$

$$\sum_{k=1}^{m} \delta_k = \sum_{k=1}^{m} \delta_k - \sum_{k=1}^{m} \tau_k$$

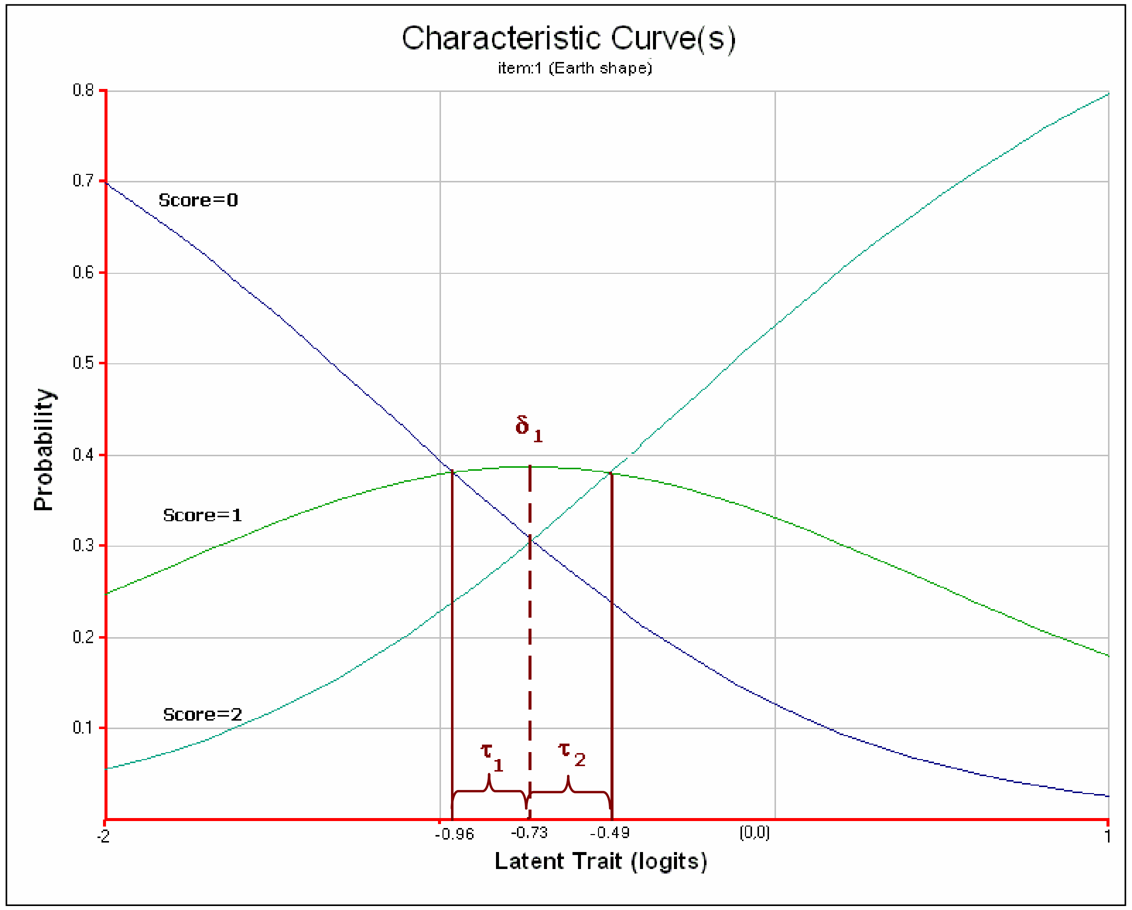
$$\sum_{k=1}^{m} \tau_k = 0$$

Rasch Andrich Rating Scale Model (Modelo de respostas Graduadas)

The rating scale model is a special case of the partial credit model in which the tau parameters for step j are the same for every item. That is, $\tau_{11}=\tau_{21}=\tau_{31}\ldots$, $\tau_{12}=\tau_{22}=\tau_{32}\ldots$, etc. In this formulation, our measurement model becomes

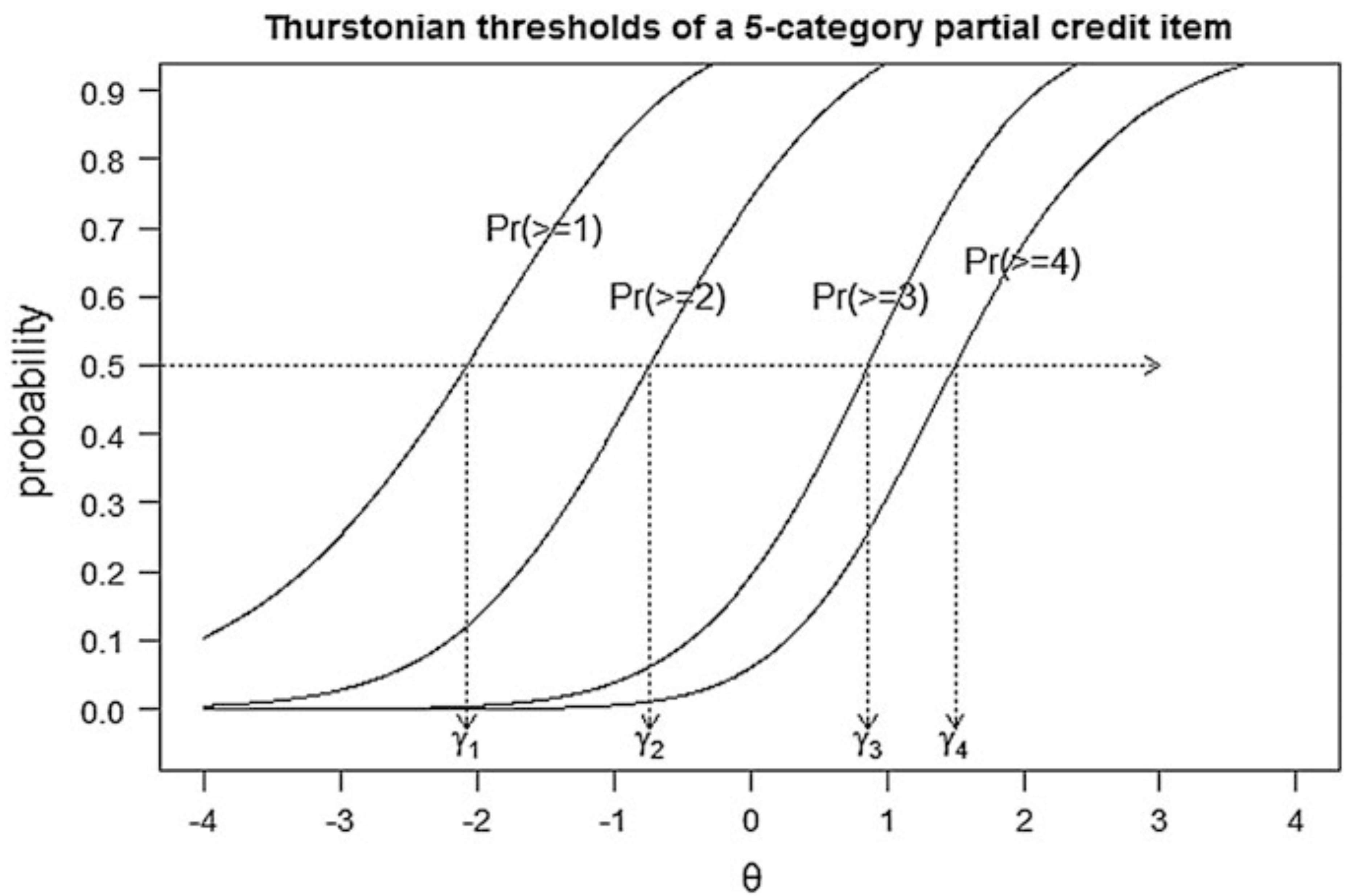
$$P(x_i = c \mid \xi_i, \theta) = \frac{\exp \sum_{j=0}^{c} [\theta - (\delta_i + \tau_j)]}{\sum_{k=0}^{m} \exp \sum_{j=0}^{k} [\theta - (\delta_i + \tau_j)]},$$

where $\xi_i=(\delta_i, \tau_1, \tau_2, \ldots \tau_{m-1})$. Again, the final tau parameter, τ_m, is not estimated because it is constrained to make the sum of all the tau parameters equal to zero.



Thurstonian Tresholds (limiares Thurstonianos): gama γ_k

- Baseado na probabilidade acumulada
- γ para uma categoria k é definido como a habilidade na qual a probabilidade da resposta x ou maior é igual a .50
- para um item de liker de cinco pontos teremos:
  - γ_1 ponto na escala de habilidade em que a p >=.50 de se ter escore 1 ou mais (2, 3, 4, 5)
  - γ_2 ponto na escala de habilidade em que a p >=.50 de se ter escore 2 ou mais (3, 4, 5)
  - γ_3 ponto na escala de habilidade em que a p >=.50 de se ter escore 3 ou mais (4, 5)
  - γ_4 ponto na escala de habilidade em que a p >=.50 de se ter escore de 4 ou mais (5)


Thurstonian thresholds of a 5-category partial credit item
probability
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Pr(>=1)
Pr(>=2)
Pr(>=3)
Pr(>=4)
γ1
γ2
γ3
γ4
-4
-3
-2
-1
0
1
2
3
4
θ

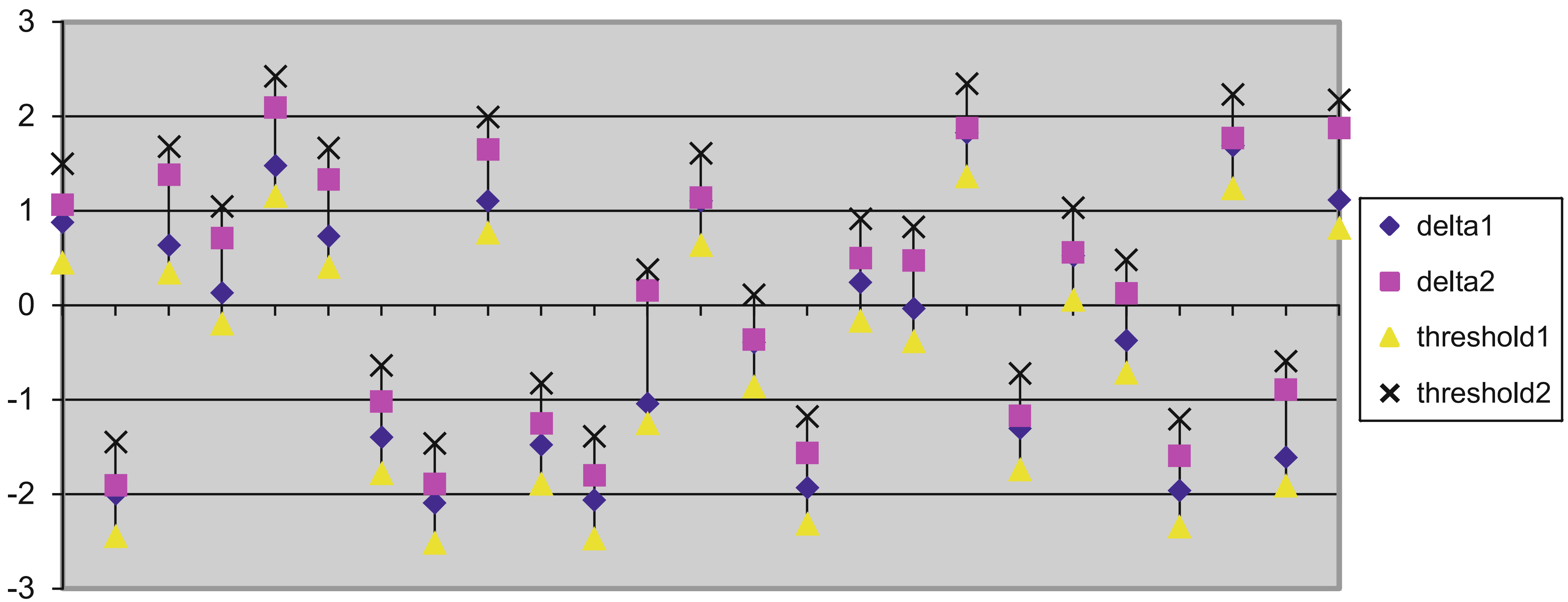


Fig. 9.8 Comparisons of threshold and delta values for 25 items

Escores esperados (expected escores)

Another measure of item difficulty for partial credit scoring items can be derived by computing the expected score on an item, as a function of ability. Consider an item with 3 score categories. The probabilities of scoring a 0, 1 or 2 are given by Eqs. (9.5)–(9.7). The expected score, E, on this item, as a function of the ability θ and delta parameters δ_1 and δ_2, is given by

$$E = 0 \times \Pr(X = 0) + 1 \times \Pr(X = 1) + 2 \times \Pr(X = 2) \tag{9.13}$$

Computing E as a function of θ, one can construct an Expected Scores Curve, similar to the item characteristic curve. Figure 9.9 shows an example.

Let E_1 be the ability at which the expected score on this item is 0.5. Let E_2 be the ability at which the expected score is 1.5. One might regard the region between E_1 and E_2 as the "score 1 region", and the ability continuum below E_1 as the "score 0 region", and the ability continuum above E_2 as the "score 2 region". In this way, E_1 could be regarded as an item difficulty parameter for score 1, and E_2 could be regarded as an item difficulty parameter for score 2.

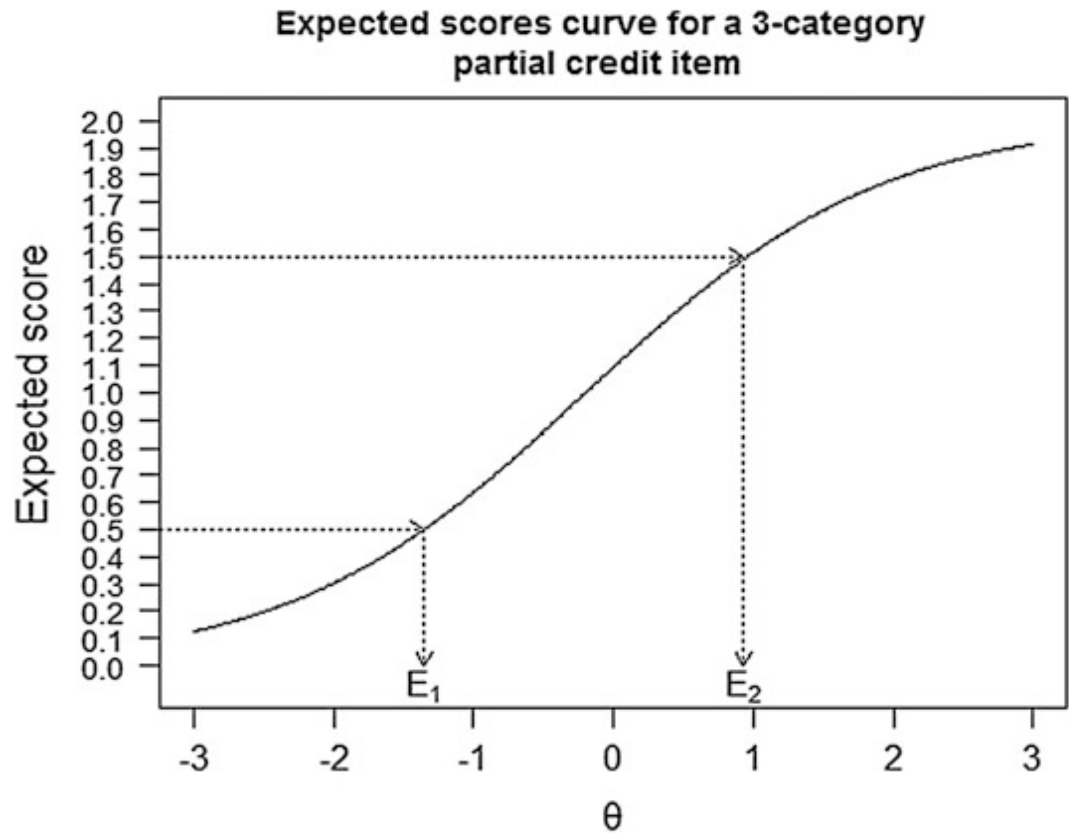


Fig. 9.9 Expected scores curve for a 3-category partial credit item

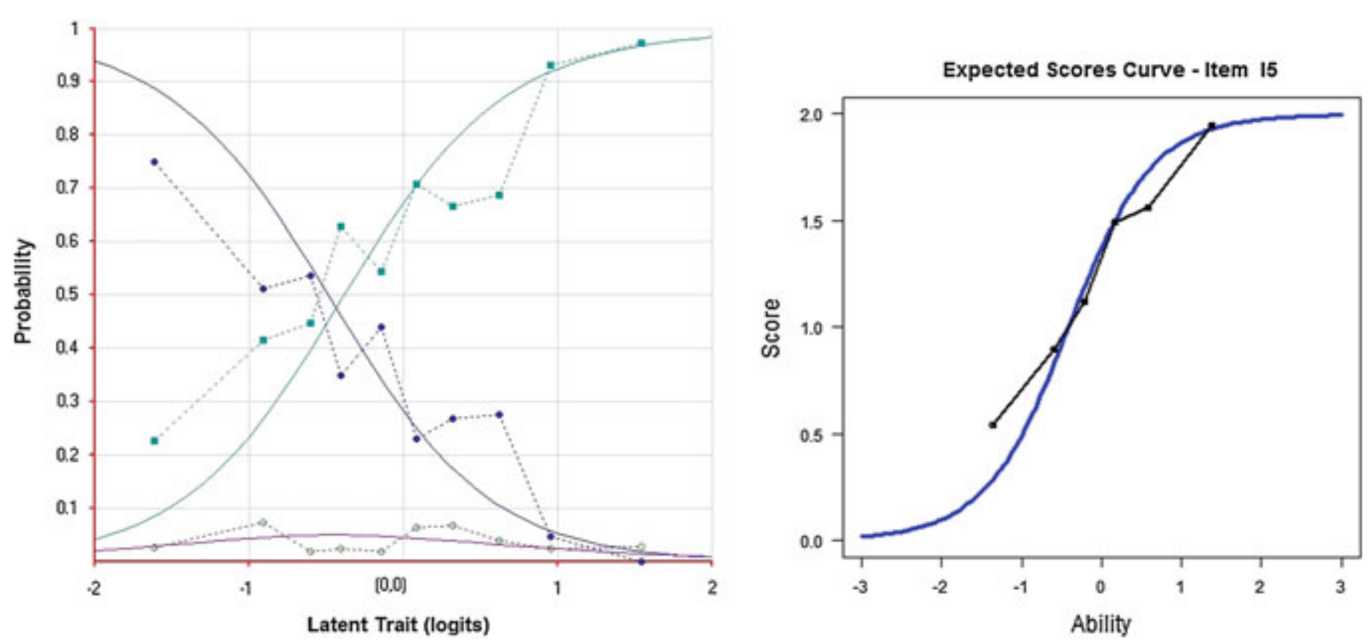


Fig. 9.10 ICCs (*left graph*) and expected scores curve (*right graph*) for a PCM item

Do manual do Winsteps (https://www.winsteps.com/winman/index.htm):Table 3.2+ Summary of dichotomous, rating scale or partial credit structures

---

Where does category 1 begin?

When describing a rating-scale to our audience, we may want to show the latent variable segmented into rating scale categories:

0----------------01------------------12----------------------2

There are 3 widely-used ways to do this:

1. "1" is the segment on the latent variable from where categories "0" and "1" are equally probable to where categories "1" and "2" are equally probably. These are the **Rasch-Andrich thresholds** (ANDRICH THRESHOLD) for categories 1 and 2.

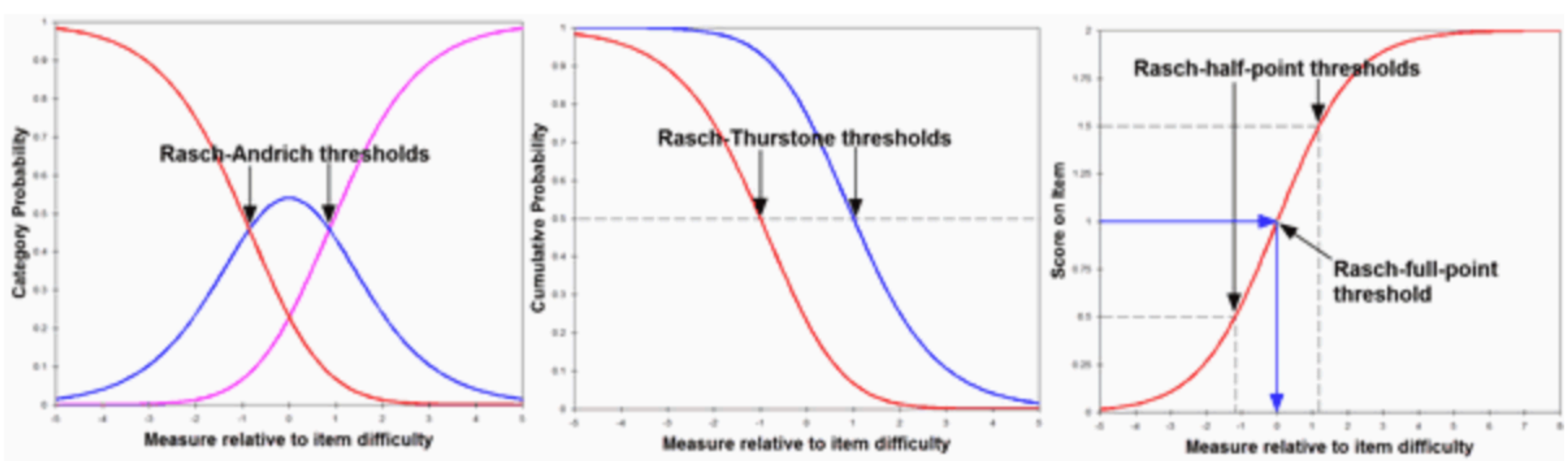
2. "1" is the segment on the latent variable from where categories "0" and "1+2" are equally probable to where categories "0+1" and "2" are equally probably. These are the **Rasch-Thurstone thresholds** (50% CUM. PROBABILITY) for categories 1 and 2.

3. "1" is the segment on the latent variable from where the expected score on the item is "0.5" to where the expected score is 1.5. These are the **Rasch-half-point thresholds** (ZONE) for category 1.

Alternatively, we may want a point on the latent variable correspond to the category:

----------0------------------1----------------------2-----------

4. "1" is the point on the latent variable where the expected average score is 1.0. This is the Rasch-Full-Point threshold (AT CAT.) for category 1.



Aplicações do modelo de créditos parciais

- Respostas construídas -> créditos parciais
- Escore máximo de um item equivale a seu peso !
    - itens ruins que não separam as pessoas em baixa e alta capacidade não devem receber um peso mais alto
    - itens que separam claramente os sujeitos devem receber um peso maior
    - = discriminação
- O que determina o número de pontos/categorias em um item: deveria ser seu índice de discriminação
- Quão forte um item está relacionado ao construto geral sendo medido / quanta informação o item traz para a medida ? -> mais peso !

An Example Item Analysis of Partial Credit Items

The data set for this example came from a mathematics problem solving test for grade 5 students. The test had 48 questions, arranged in 3 rotated test booklets. In total, 1086 students took part in the test, but each item had around 500 student responses. The test had a mix of dichotomously and polytomously scored items. IRT and CTT analyses were conducted on the data set.

The following shows one example item ("Average") from the test and corresponding initial proposed scoring scheme:

Item 4: Item "Average"
Megan obtained an average mark of 81 for her four science tests. The following shows her scores for Tests 1, 3 and 4? What was her test score for Test 2? Show how you found your answer.

Test 1	Test 2	Test 3	Test 4	Average mark of 4 tests
84	?	89	93	81

As students were requested to provide their working in solving the item, students' responses contained a variety of approaches and answers. These responses

Table 9.1 Scoring scheme for item "Average"

Response	Proposed score
Correct analytic method and correct answer of 58	4
Trial-and-error method, but still obtained the correct answer	3
Correct analytic method, but computation error, resulting in incorrect answer	2
Computed the average of the three scores, but unable to proceed to produce the correct answer	1
Other responses	0

Table 9.2 Item statistics for the item "Average"

Score category	Frequency	Percentage	Pt biserial correlation	Average ability
0	183	0.33	−0.60	−0.78
1	108	0.19	−0.12	−0.22
2	36	0.06	0.12	0.43
3	23	0.04	0.09	0.40
4	209	0.37	0.57	0.67

Table 9.3 Fit statistics for item "Average"

Parameter	Infit mean squares	Infit t
δ_1	0.99	−0.14
δ_2	0.96	−0.72
δ_3	1.18	2.85
δ_4	1.14	2.34

A few observations can be made from the item statistics in Table 9.2. First, very few students used the correct method but made a computational error leading to an incorrect answer (score category 2). Similarly, few students used the trial and error method and obtained the correct answer (score category 3). Second, the point biserial correlations for categories 2 and 3 are very similar. Third, the average abilities of students in categories 2 and 3 are very similar. These observations suggest that categories 2 and 3 can possibly be combined.

The fit statistics for this item are given in Table 9.3. The fact that the fit mean squares statistics are larger than 1 indicates that the item is not as discriminating as the model expects for an item with a maximum score of 4. This is further confirmed by the expected scores curve, as shown in Fig. 9.11. For high ability students, the observed score is lower than the expected score.

Based on these item statistics, a recoding of the score categories is made by collapsing categories 2 and 3 into a new category 2, and recoding the current category 4 as category 3. That is, the item has a maximum score of 3 after recoding.

Table 9.4 Item statistics for the item "Average", after recoding

Score Category	Frequency	Percentage	Pt biserial correlation	Average ability
0	183	0.33	−0.59	−0.78
1	108	0.19	−0.11	−0.20
2	59	0.11	0.17	0.47
3	209	0.37	0.54	0.66

Table 9.5 Fit statistics for item "Average", after recoding

Parameter	Infit mean squares	Infit t
δ_1	0.94	−0.76
δ_2	0.93	−1.55
δ_3	0.97	−0.52
δ_4	1.05	1.08

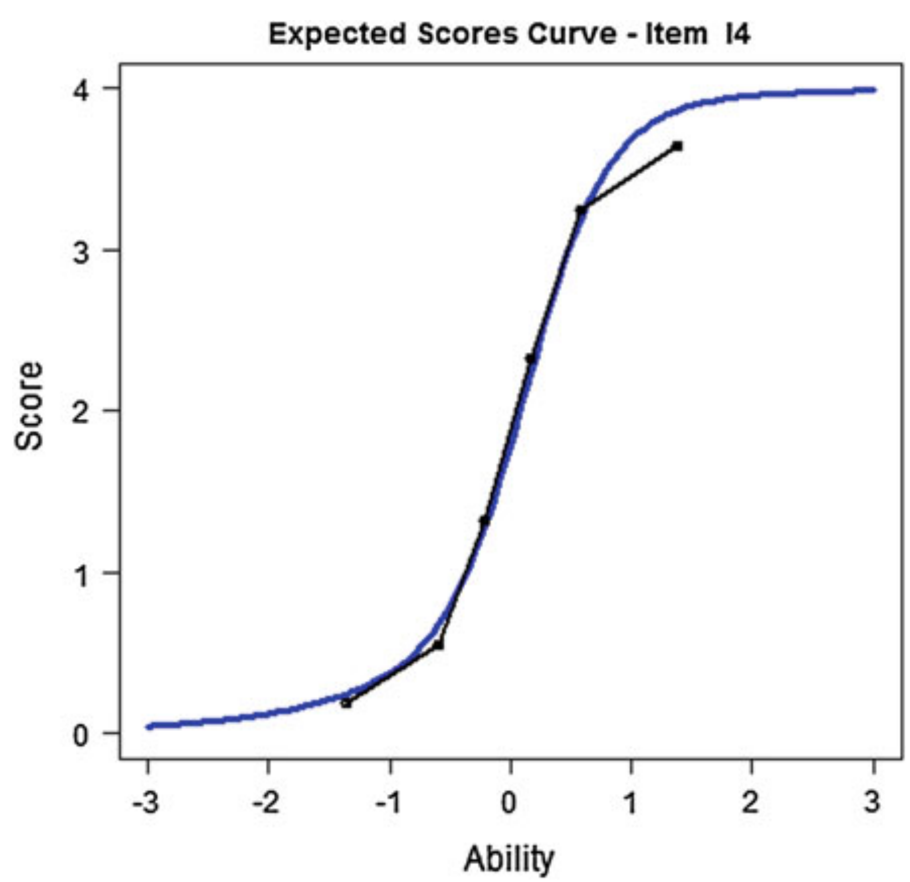


Fig. 9.11 Expected scores curve for item "Average"

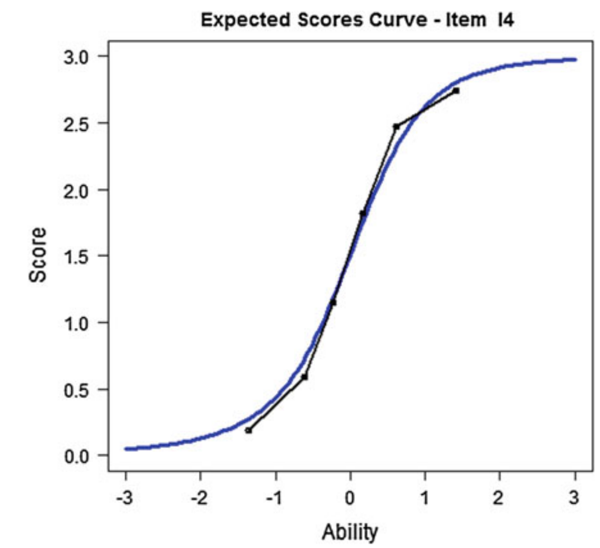


Fig. 9.12 Expected scores curve of item "Average", after recoding

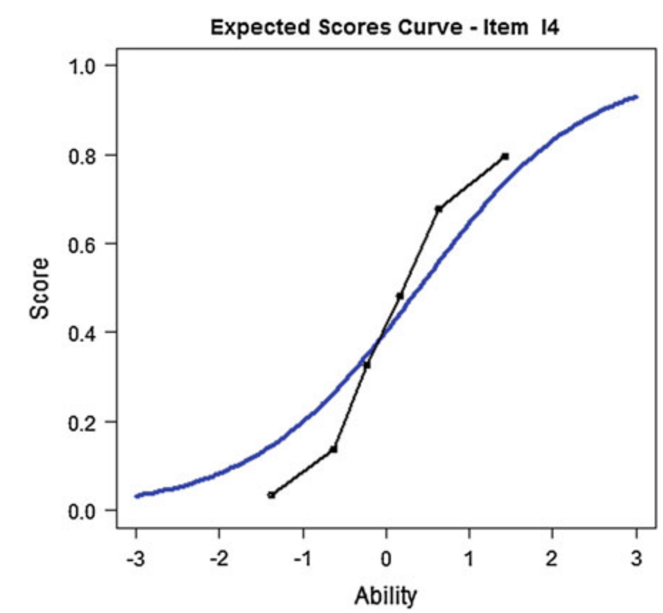


Fig. 9.13 Expected scores curve of item "Average", with "correct/incorrect" scoring

Modelo	Características (Parâmetros)
Rasch-Andrich Rating Scale Model (respostas graduais)	Generalização do modelo de Rasch para escalas likert Estima um índice de dificuldade por item e k-1 limiares (tresholds) gerais para as categorias (k = número de pontos na escala)
Rasch-Masters Partial Credit Model (créditos parciais)	Mais geral e flexível. Generalização do modelo de Rasch para itens politômicos Estima k-1 limires por item (k = número de pontos na escala)
Samejima's Graded Response Model	Generalização do modelo de 2 parâmetros para itens politômicos Estima um índice de discriminação e k-1 limiares (tresholds) as categorias (k = número de pontos na escala)
Generalized Rating Scale Model ou Muraki's Modified Graded Response Model	Generalização do modelo de 2 parâmetros para escalas likert Estima um índice de dificuldade por item e k-1 limiares (tresholds) gerais para as categorias (k = número de pontos na escala)
Muraki's Generalized Partial Credit Model	Generalização do modelo de 2 parâmetros para escalas likert Estima um índice de dificuldade por item e k-1 limiares (tresholds) gerais para as categorias (k = número de pontos na escala)

Educational Measurement
ISSUES AND PRACTICE

NCME national council on measurement in education

Educational Measurement: Issues and Practice
Spring 2014, Vol. 33, No. 1, pp. 36–48

An NCME Instructional Module on Polytomous Item Response Theory Models

Randall David Penfield, *The University of North Carolina at Greensboro*

A polytomous item is one for which the responses are scored according to three or more categories. Given the increasing use of polytomous items in assessment practices, item response theory (IRT) models specialized for polytomous items are becoming increasingly common. The purpose of this ITEMS module is to provide an accessible overview of polytomous IRT models. The module presents commonly encountered polytomous IRT models, describes their properties, and contrasts their defining principles and assumptions. After completing this module, the reader should have a sound understating of what a polytomous IRT model is, the manner in which the equations of the models are generated from the model's underlying step functions, how widely used polytomous IRT models differ with respect to their definitional properties, and how to interpret the parameters of polytomous IRT models.

Keywords: item response theory, polytomous items, partial credit model, graded response model, nominal response model

		Scored Categories for Y_i			
		$Y_i = 0$	$Y_i = 1$	$Y_i = 2$	$Y_i = 3$
Adjacent Category Approach	Step 1	F	S		
	Step 2		F	S	
	Step 3			F	S
Continuation Ratio Approach	Step 1	F	S	S	S
	Step 2		F	S	S
	Step 3			F	S
Cumulative Approach	Step 1	F	S	S	S
	Step 2	F	F	S	S
	Step 3	F	F	F	S
Nominal Approach	Step 1	S	F		
	Step 2	S		F	
	Step 3	S			F

Samejima's Graded Response Model equivalente ao modelo de dois parâmetros (veremos na aula de modelos de 2p)

FIGURE 4. Description of the step functions for a four-category polytomous item. Within each step, S represents success and F represents failure. Under the nominal approach for defining step functions, the outcome $Y_i = 0$ represents the correct option of a multiple-choice item and $Y_i = 1, 2, 3$ represent distractor options.

Exercício 3: Aplicando modelo de créditos parciais no SENNA