Chapter 21
The Polytomous Rasch Model II

This chapter revises and goes beyond Chap. 20 in understanding the polytomous Rasch model. The chapter revises the model for more than two ordered categories as a direct, generalization of the simple logistic model of Rasch for dichotomous responses. Examples of such a response format include the familiar Likert-style attitude or personality items in questionnaires, as well as some partial credit structures in assessing proficiency.

Key features of the model are that (i) the successive categories are scored with successive integers as is done in Classical Test Theory (CTT); (ii) nevertheless, distances between thresholds which define the categories are estimated and no assumptions about equal sizes of the categories need be made; (iii) the model retains the distinctive properties of Rasch's model in terms of the separation of person and item parameters; (iv) the threshold estimates do not have to be in the natural order. The ordering of threshold estimates is a property of the data, and if the estimates are not in their natural order, it shows that the categories are not working as required.

Andrich (1978) shows the original derivation of the model in which the scoring of the categories by integers and the interpretation of the more general parameters derived by Rasch and Andersen were clarified. Andrich (2010a, b) shows recent summaries of the model.

Rated Data in the Social Sciences

(a) According to Dawes (1972) 60% of studies have only rated dependent variables. Since then, because of an increase in performance assessment and applications in health outcomes, it is likely to have increased.

(b) Rating is often used in place of measurement, but the measurement analogy is retained, e.g. grading of essays as in the structure below.

D. Andrich and I. Marais, *A Course in Rasch Measurement Theory*,
Springer Texts in Education, https://doi.org/10.1007/978-981-13-7496-8_21

0	1	2	3	4	Scores or ratings

τ_1 τ_2 τ_3 τ_4

E	D	C	B	A
Fail	Pass	Credit	Distinction	High distinction

The continuum is partitioned into hypothetical regions by thresholds, τ_1, τ_2, τ_3, τ_4, and the scores in the successive categories follow the measurement analogy. In the case of the dichotomous model, there is only the one threshold.

(c) Often more than one sub-criterion is used, though in the end the scores on these criteria are aggregated, e.g. for the grading of essays:

Criterion	Rating	x_{ni} for person n
(i) Organization	0–3	x_{n1}
(ii) Content	0–4	x_{n2}
(iii) Grammar	0–5	x_{n3}
	Total rating	$\sum_{i=1}^{3} x_{ni} = r_n$

Successive categories are scored with successive integers, irrespective of threshold distances between categories. Note that different criteria can have different numbers of categories.

The Partial Credit and Rating Scale Specifications

Often when the number of categories is the same for all items, and the format for all items is the same, it is possible to estimate only one set of thresholds for all items. In that case, the model has been called the *rating scale model*. When the items have different numbers of categories, or where all the items have the same number of categories but the thresholds are estimated for each of the items, the model has been called the *partial credit model* (Masters, 2016). However, these are modifications only to the number of parameters estimated—the response structure for the response of a person to an item is identical. Therefore, unless specialized, we refer to the model as the Polytomous Rasch Model (PRM), and refer to the former as the rating scale parameterization and the latter as the partial credit parameterization. We write this difference formally in a subsequent section.

The Generalization to Three Ordered Categories

Figure 21.1 includes the probability curve of the third category, which is in the middle of the two other categories scored 0 and 2. First, with three categories there are two thresholds. Even without knowledge of the Rasch model, if we were to draw the probability of such a category as a function of the proficiency of a person, we saw in the last chapter that we would draw something like the graph in Fig. 21.1. This graph shows a region where, in between the two thresholds, the category has a higher probability than either of the two other categories.

The explicit Eqs. (20.3a, 20.3b, 20.3c) from the last chapter are repeated below and interpreted further. Notice how the thresholds appear successively as sums of the thresholds up to the threshold corresponding to the score. However, do not be deceived into thinking that the process is somehow successive—the response is simply a *classification* into one of three *ordered* categories.

$$\Pr\{0; \beta, \delta, \tau_1, \tau_2\} = \frac{1}{\gamma}$$
$$\Pr\{1; \beta, \delta, \tau_1, \tau_2\} = \frac{1}{\gamma} \exp[-\tau_1 + 1(\beta - \delta)]$$
$$\Pr\{2; \beta, \delta, \tau_1, \tau_2\} = \frac{1}{\gamma} \exp[-(\tau_1 + \tau_2) + 2(\beta - \delta)]$$

where $\gamma = 1 + \exp[-\tau_1 + 1(\beta - \delta)] + \exp[-(\tau_1 + \tau_2) + 2(\beta - \delta)]$.

Notice that the denominator is still the sum of all the terms of the numerator.

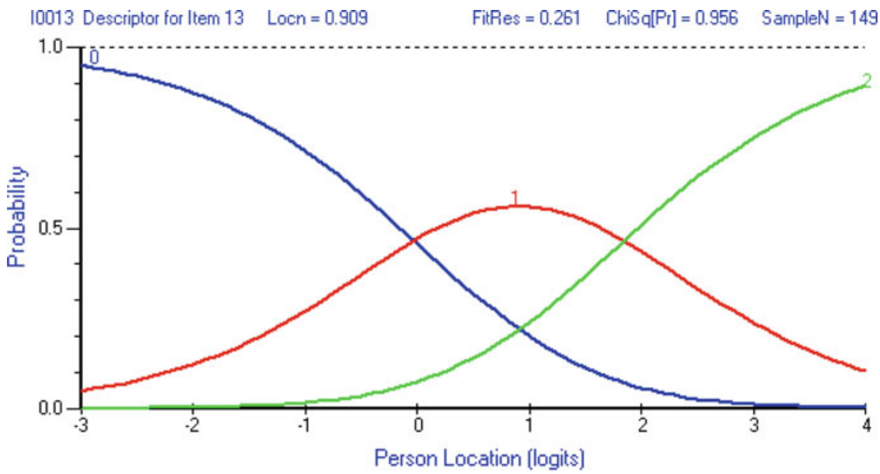

Fig. 21.1 Probability curves for three ordered categories

For completeness and symmetry of the expressions, we can introduce a threshold τ_0, $\tau_0 \equiv 0$. It does not change any values but we can write the first term in the same format as the other terms. This gives

$$\Pr\{0; \beta, \delta, (\underline{\tau})\} = \frac{1}{\gamma}\exp[-\tau_0 + 0(\beta - \delta)];\ \tau_0 \equiv 0$$

$$\Pr\{1; \beta, \delta, (\underline{\tau})\} = \frac{1}{\gamma}\exp[-(\tau_0 + \tau_1) + 1(\beta - \delta)]$$

$$\Pr\{2; \beta, \delta, (\underline{\tau})\} = \frac{1}{\gamma}\exp[-(\tau_0 + \tau_1 + \tau_2) + 2(\beta - \delta)];\ \tau_1 + \tau_2 = 0$$

where $x \in \{0, 1, 2\}$, $\gamma = \sum_{k=0}^{2}\exp\left[-\sum_{x'=0}^{k}\tau_{x'} + k(\beta - \delta)\right]$ and $\underline{\tau}$ is the vector of thresholds of the item.

Notice that in the above formulation we make the thresholds sum to zero: $\tau_1 + \tau_2 = 0$.

This means that the parameter δ is the difficulty of the item and that the thresholds are located around the item, with the item's difficulty in the middle of the thresholds.

Note that the successive categories are scored by successive integers beginning with 0, where 0, 1, 2 is just an extension of the 0, 1 scoring for two ordered categories. In addition, even though the successive categories are scored with successive integers, the thresholds are estimated. They do not have to be equidistant.

In early discussions of more formal models for ordered categories there was a belief that integer scoring required equal distances. You can read in the literature the reason the integer scoring appears, but it is not because the distances between categories are somehow equal. In any case, with three categories, there are just two thresholds and there is only one distance as such between them, not three.

The Expected Value Curve

We saw in *Statistics Review 14* that the expected value $E[X]$ and the probability $\pi_1 = \pi$ of the response of 1, in the dichotomous Bernoulli variable, was the same: $E[X] = \pi_1 = \pi$.

However, in the case of more than two ordered categories, $m + 1$, with scores 0, 1, 2, …, m,

$$E[X] = \sum_{x=0}^{m} x\pi_x$$

where

π_x is the probability of a response with score x.

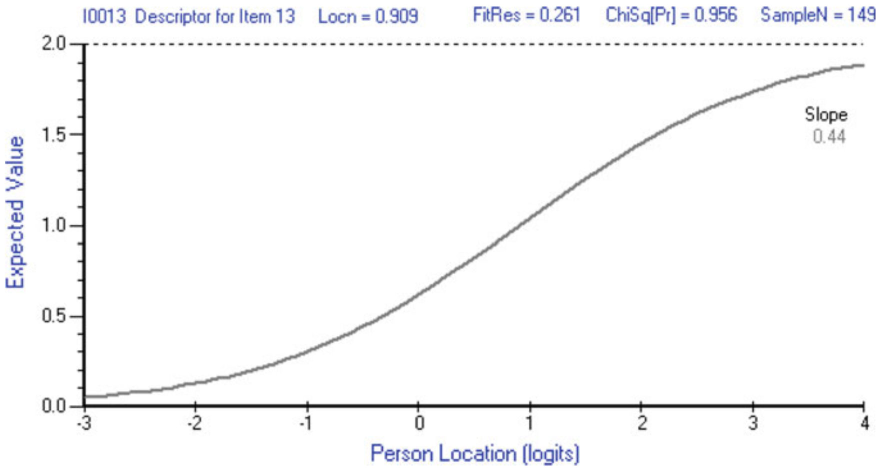

Fig. 21.2 $E[X]$ for an item with three ordered categories

Figure 21.2 shows the $E[X]$ for item 13 where $m = 2$.

We notice that there is a slope parameter with this item. It is the rate of change of the expected value at the location of the item $\delta = 0.909$. The slope is considered further later in the chapter.

The Structure of the PRM

The structure of the PRM and its relation to the dichotomous Rasch model at the thresholds, can be seen by forming the probability of a response in the higher of two adjacent categories, *given* that the response is in one of these two categories.

For example, consider

$$\begin{aligned}
&\frac{\Pr\{2;\beta,\delta,\underline{\tau}\}}{\Pr\{1;\beta,\delta,\underline{\tau}\}+\Pr\{2;\beta,\delta,\underline{\tau}\}}\\
&\quad=\frac{\{\exp[-(\tau_0+\tau_1+\tau_2)+2(\beta-\delta)]\}/\gamma}{\{\exp[-(\tau_0+\tau_1)+1(\beta-\delta)]\}/\gamma+\exp[-(\tau_0+\tau_1+\tau_2)+2(\beta-\delta)]\}/\gamma}\\
&\quad=\frac{\exp[-\tau_2+(\beta-\delta)]}{1+\exp[-\tau_2+(\beta-\delta)]}\\
&\quad=\frac{\exp[\beta+(-\delta-\tau_2)]}{1+\exp[\beta+(-\delta-\tau_2)]}\\
&\quad=\frac{\exp[\beta-(\delta+\tau_2)]}{1+\exp[\beta-(\delta+\tau_2)]}
\end{aligned}$$

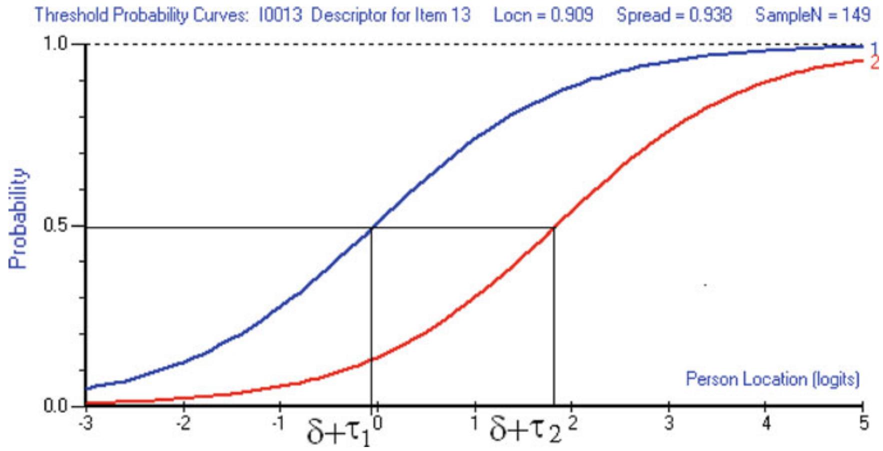

Fig. 21.3 Conditional probability of a dichotomous response in the higher of two adjacent categories

where $\delta_2 = \delta + \tau_2$. We can see that this is just the dichotomous Rasch model with the difficulty of the item in the dichotomous model replaced by the difficulty of threshold 2, where the difficulty τ_2 of the threshold, relative to the item's difficulty δ, is added to this item's difficulty.

Figure 21.3 shows the two curves: $\frac{\Pr\{1\}}{\Pr\{0\}+\Pr\{1\}}$ and $\frac{\Pr\{2\}}{\Pr\{1\}+\Pr\{2\}}$ for item 13. The spread parameter is defined in Chap. 22.

Thus, the structure of the polytomous item is that the thresholds are simply characterized by the dichotomous Rasch model. These curves are parallel as in the usual case of dichotomous items which form a single set of items.

The Generalization to Any Number of Categories $m + 1$

$$\Pr\{0; \beta, \delta, \underline{\tau}\} = \frac{1}{\gamma} \exp[-\tau_0 + 0(\beta - \delta)]$$

$$\Pr\{1; \beta, \delta, \underline{\tau}\} = \frac{1}{\gamma} \exp[-(\tau_0 + \tau_1) + 1(\beta - \delta)]$$

$$\Pr\{2; \beta, \delta, \underline{\tau}\} = \frac{1}{\gamma} \exp[-(\tau_0 + \tau_1 + \tau_2) + 2(\beta - \delta)]$$

$$\Pr\{3; \beta, \delta, \underline{\tau}\} = \frac{1}{\gamma} \exp[-(\tau_0 + \tau_1 + \tau_2 + \tau_3) + 3(\beta - \delta)]$$

...

$$\Pr\{x; \beta, \delta, \underline{\tau}\} = \frac{1}{\gamma} \exp[-(\tau_0 + \tau_1 + \tau_2 + \tau_3 + \tau_x) + x(\beta - \delta)]$$

...

$$\Pr\{m; \beta, \delta, \underline{\tau}\} = \frac{1}{\gamma} \exp[-(\tau_0 + \tau_1 + \tau_2 + \tau_3 + \cdots + \tau_x + \cdots + \tau_m) + m(\beta - \delta)] \tag{21.1}$$

where $\sum_{x'=0}^{m} \tau = 0$ and $\gamma = \sum_{k=0}^{m} \exp\left[-\left(\sum_{x'=0}^{k} \tau_{x'}\right) + k(\beta - \delta)\right]$.

Define $\kappa_0 = \kappa_m = 0$,

$$\kappa_1 = -\tau_1$$
$$\kappa_2 = -\tau_1 - \tau_2$$
$$\kappa_m = -\tau_1 - \tau_2 - \cdots - \tau_m = 0$$

Then,

$$\Pr\{k; \beta, \delta, \underline{\tau}\} = \frac{1}{\gamma} \exp[\kappa_x + x(\beta - \delta)] \tag{21.2}$$

where $\gamma = \sum_{k=0}^{m} \exp[\kappa_k + k(\beta - \delta)]$ and the κs are known as category coefficients.

In the simple specialization of Eq. (21.1) to the dichotomous case when $m = 1$, $\kappa_0 = \kappa_m \equiv 0$.

Figure 21.4 shows category probability or characteristic curves for an item with unequally spaced thresholds and 4 categories.

Figure 21.5 shows the expected value curve for the item in Fig. 21.4. Notice that now the maximum value of $E[X]$ is 3.

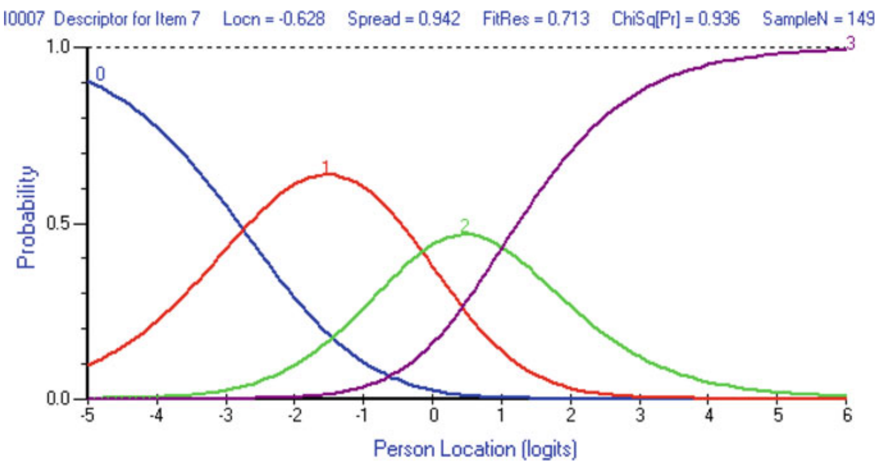

Fig. 21.4 Category characteristic curves for an item with unequally spaced thresholds and 4 categories

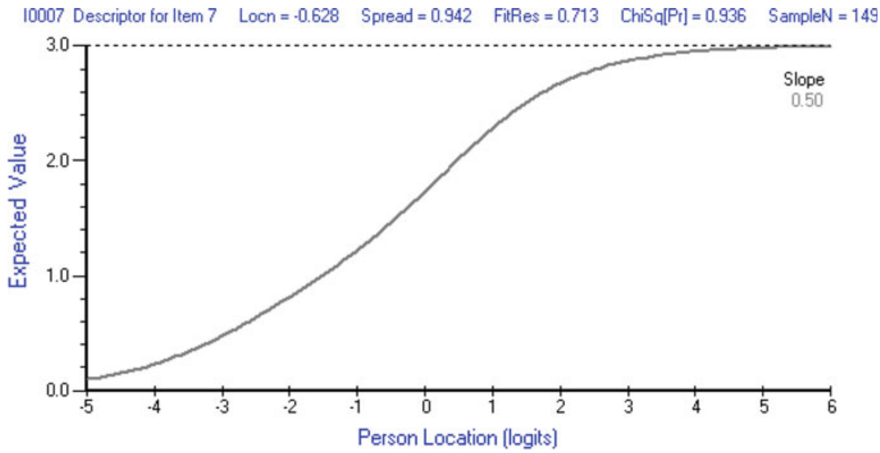

Fig. 21.5 The expected value curve for an item with unequally spaced thresholds and 4 categories

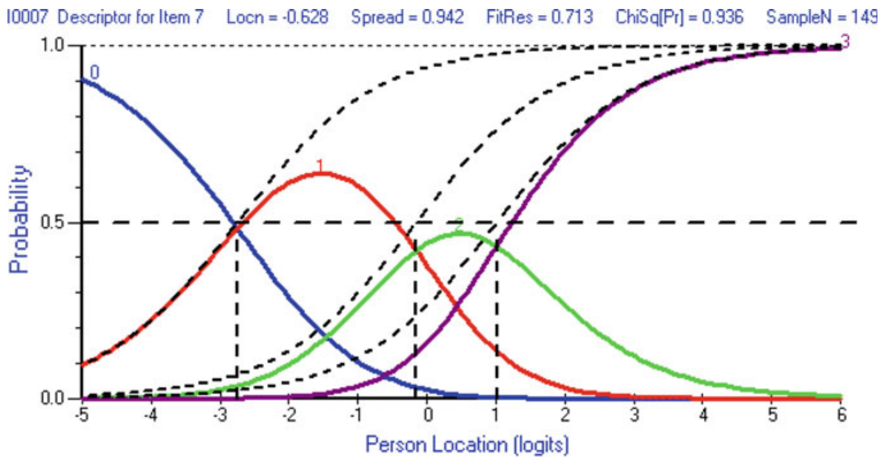

Fig. 21.6 Conditional dichotomous responses for three thresholds

Figure 21.6 shows the threshold probability curves, as in Fig. 21.3, but now superimposed on the category characteristic curves for the same item as in Fig. 21.4. These are shown as dotted curves.

The Slope of $E[X]$

The slope of $E[X]$ is a function of the distance between the thresholds, the closer the thresholds, the steeper the slope. Of course, the slope of the curve changes at each point, but we summarise the slope at the value of the location of the item δ_i. This is

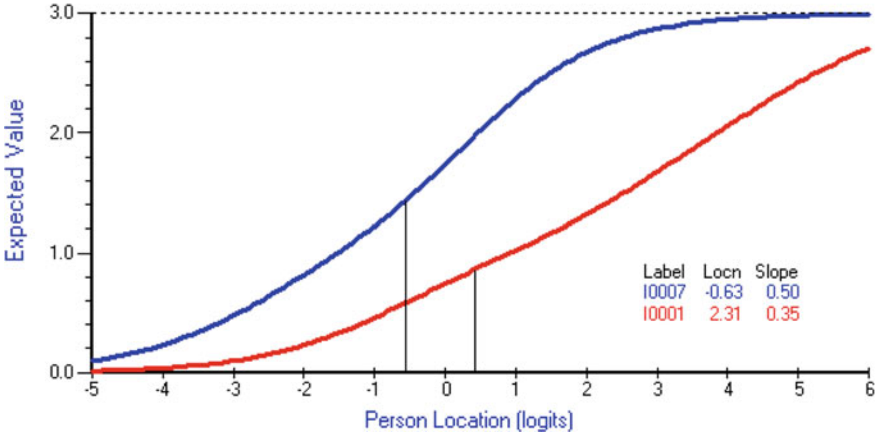

Fig. 21.7 Two items with different slopes of $E[X]$

the mean of the thresholds on the continuum. Figure 21.7 shows two items, where one has a steeper slope than the other.

Latent Threshold Curves

Figure 21.8 shows the category characteristic curves for the two items in Fig. 21.7. Although the slopes of the $E[X]$ curves are different, the latent dichotomous thresholds characteristic curves in Fig. 21.8 for both items, and all of their thresholds, are parallel.

The threshold curves are shown in dotted lines because they are not observed. They are part of the latent structure of the PRM. The threshold characteristic curve, the conditional probability of the higher of two categories given that the response

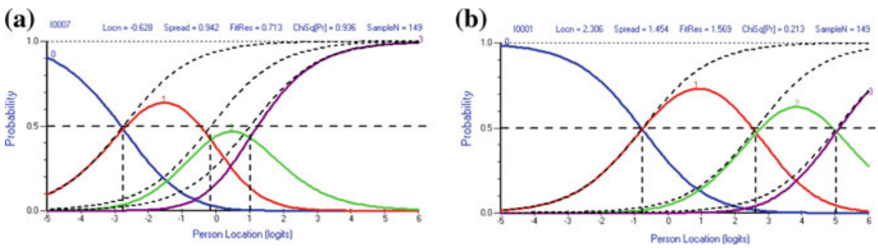

Fig. 21.8 **a** An item with a smaller average distance between thresholds. **b** An item with a larger average distance between thresholds

is in one of the two categories, is inferred—there is no dichotomous response at a threshold.

Diagnosing Problems with the Functioning of the Categories

The Rasch model for ordered categories has two unusual properties. First, in the model, and this was known to Rasch (1966), it is not an arbitrary matter to collapse and combine categories. If one has three categories functioning well, then if two categories are combined, they will not fit the data as well. Second, the thresholds that define the categories on the continuum do not need to be correctly ordered. If the estimates from the data are not correctly ordered, then this means that the categories are not functioning as intended.

Figure 21.9 shows the category curves for item 10 from the same example as the items shown above. This example has a number of attitude items which are graded in four categories: strongly disagree, disagree, agree, and strongly agree. Consider the shape and the relationship amongst the category response functions for this item, and in particular, consider the response in the category scored 2. There is a region in which the third category (score 2) never has a maximum probability. This indicates a problem with the operation of the category.

Although there is a problem with the category, the evidence does not explain why there is a problem. There may be many reasons why there is a problem. We consider this example briefly by considering some complementary evidence. One of the most important things to consider is the distribution of persons in relation to the thresholds. It may be that there are just not enough people in the relevant

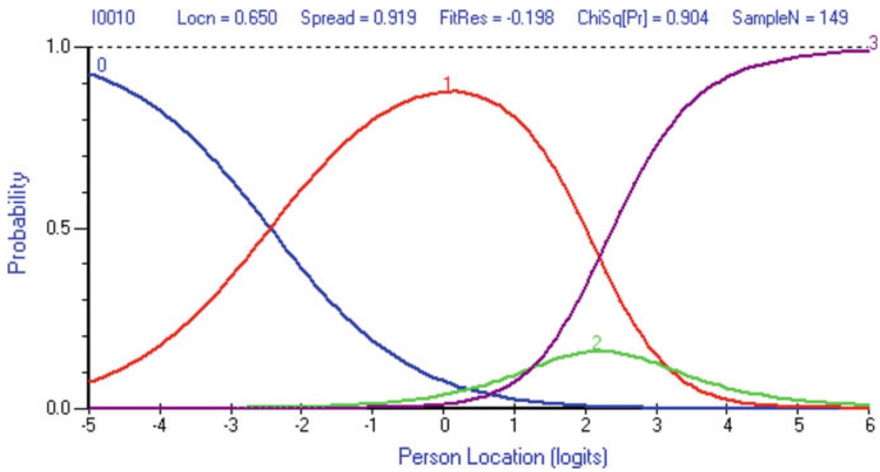

Fig. 21.9 An example where a category is not functioning as intended

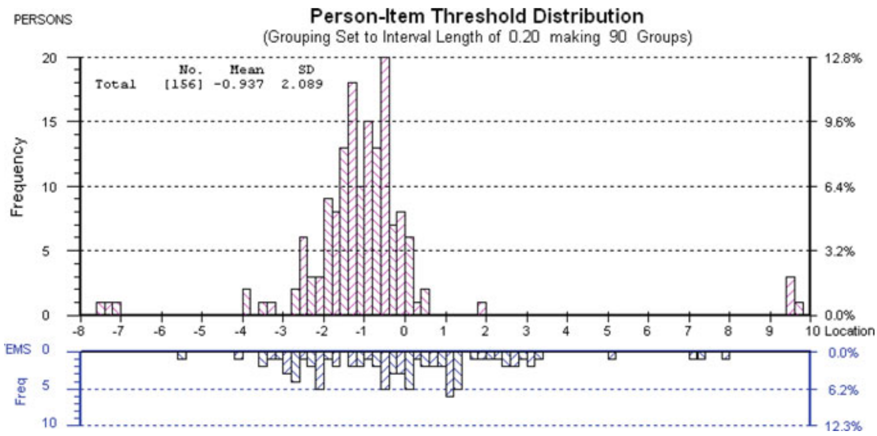

Fig. 21.10 Distribution of persons and thresholds

categories to give sound estimates. This appears to be the case for these items. First, Fig. 21.10 shows the person distribution relative to the thresholds. It is evident that most people are between −4 and 2 logits with some people at the extremes. However, from Fig. 21.9, the thresholds of the item 10 go beyond 3 logits. This suggests there might be a lack of data in the region of the reversed thresholds of the item.

Table 21.1 below shows the frequencies of responses in each category for each item. It is evident that there are indeed very few cases in the categories for scores of 2 and 3, and therefore not much can be read from the reversed thresholds in this case.

However, it is possible to get reversed thresholds even if there are many people in the relevant categories. Andrich (2011) shows such an example. Also, examples where there are structural explanations of reversed thresholds in the assessment of educational proficiency is given in van Wyke (2003). He considers the implication of evidence of reversed thresholds and shows the application of the model to construct a continuum of educational proficiency in mathematics.

The Partial Credit and Rating Parameterizations of the PRM

We begin with Eq. (21.1) where the thresholds are explicit.

$$\Pr\{x; \beta, \delta, \underline{\tau}\} = \frac{1}{\gamma} \exp[-(\tau_0 + \tau_1 + \tau_2 + \tau_3 + \cdots \tau_x) + x(\beta - \delta)] \qquad (21.3)$$

Table 21.1 Frequencies of responses in each category highlighting item 10

Seq	Code	Cat 1	Cat 2	Cat 3	Cat 4
1	I0001	82	65	2	0
2	I0002	8	34	87	20
3	I0003	19	85	44	1
4	I0004	12	36	61	39
5	I0005	39	89	17	4
6	I0006	2	57	73	17
7	I0007	22	81	39	7
8	I0008	26	93	26	4
9	I0009	46	87	12	4
10	**I0010**	**33**	**113**	**2**	**1**
11	I0011	30	70	42	7
12	I0012	11	133	5	0
13	I0013	104	36	6	3
14	I0014	27	119	3	0
15	I0015	22	59	66	2
16	I0016	15	75	42	17
17	I0017	76	69	2	2
18	I0018	64	75	10	0
19	I0019	36	106	6	1
20	I0020	52	81	14	2
21	I0021	19	71	39	20
22	I0022	13	94	36	6
23	I0023	35	86	26	2
24	I0024	20	59	49	21
25	I0025	13	82	51	3
26	I0026	62	60	23	4
27	I0027	31	75	36	7

The Rating Scale Parameterization

If the items have a different overall location parameter δ, we notate it δ_i. Then if all the items have the same number of categories, and if they have the same descriptors, we might hypothesize that the thresholds for the different items are all the same value. In that case, we do *not* subscript the thresholds with different items. An example where the same descriptors might be present is in attitude or opinion surveys where a number of questions have the same categories labels such as

Strongly Disagree (SD)	Disagree (D)	Agree (A)	Strongly Agree (SA)

Then the model takes the form

$$\Pr\{x; \beta, \delta_i, \underline{\tau}\} = \frac{1}{\gamma}\exp[-(\tau_0 + \tau_1 + \tau_2 + \tau_3 + \cdots \tau_x) + x(\beta - \delta_i)] \tag{21.4}$$

where the thresholds are *not* subscripted by the item parameter, which the overall location is so subscripted.

Of course, it is possible that the threshold estimates are not equidistant across items. In that case, there is an interaction between the distances between the thresholds and the items.

Sometimes there are two kinds of items, for example those worded positively and negatively. In that case, it might be that the negatively worded items have similar threshold distances and the positively worded items have similar threshold distances, but that the positively and negatively worded items have different threshold distances. We may write the model differently, with the derivation below.

Expanding the numerator in Eq. (21.4), we have

$$\begin{aligned}
\Pr\{x; \beta, \delta_i, \underline{\tau}\} &= \frac{1}{\gamma}\exp[-(\tau_0 + \tau_1 + \tau_2 + \tau_3 + \cdots \tau_x) + x(\beta - \delta_i)] \\
&= \frac{1}{\gamma}\exp[-(\tau_0 + \tau_1 + \tau_2 + \tau_3 + \cdots \tau_x) + x\beta - x\delta_i] \\
&= \frac{1}{\gamma}\exp[-(\tau_0 + \tau_1 + \tau_2 + \tau_3 + \cdots \tau_x) - \underbrace{(\delta_i + \delta_i + \delta_i + \cdots \delta_i)}_{x} + x\beta] \\
&= \frac{1}{\gamma}\exp[-(\tau_0 + \tau_1 + \delta_i + \tau_2 + \delta_i + \tau_3 + \delta_i \cdots \tau_x + \delta_i) + x\beta] \\
&= \frac{1}{\gamma}\exp[-(\tau_{i0} + \delta_{i1} + \delta_{i2} + \delta_{i3} + \cdots + \delta_{ix}) + x\beta] \\
&= \frac{1}{\gamma}\exp\left[-\sum_{k=0}^{x}\delta_{ik} + x\beta\right]
\end{aligned}$$

where now $\delta_{ix} = \delta_i + \tau_x$ and $\delta_{i0} \equiv 0$.

The thresholds δ_{ix} for all items have the same origin and can be compared across items. The thresholds $\tau_x,\ x = 1, 2, \ldots, m$ are mean deviated from each item's difficulty. Both forms can be instructive depending on the context of interpretation. In RUMM2030 $\tau_x,\ x = 1, 2, \ldots, m$ are called *centralized* thresholds because they are centred about the item's location δ_i, and δ_{ix} are called *uncentralized* thresholds.

The Partial Credit Parameterization

In tests of proficiency, different items may have different numbers of categories and therefore the thresholds cannot be expected to be the same distance apart. In that case, the location of the item δ and the thresholds $\tau_1, \tau_2, \tau_3, \ldots, \tau_x, \ldots \tau_m$ are subscripted with i. Furthermore, the maximum score m is subscripted by i. This gives the parameterization

$$\Pr\{x; \beta, \delta_i, \underline{\tau}\} = \frac{1}{\gamma} \exp[-(\tau_{1i} + \tau_{2i} + \tau_{3i} + \cdots \tau_{xi}) + x(\beta - \delta_i)] \tag{21.5}$$

where $\gamma = \sum_{k=0}^{m_i} \exp\left[-\left(\sum_{x'=0}^{k} \tau_{x'}\right) + k(\beta - \delta_i)\right]$.

This parameterization can also be cast into the form

$$\Pr\{x; \beta, \underline{\delta}\} = \frac{1}{\gamma} \exp\left[-\sum_{k=0}^{x} \delta_{ik} + x\beta\right] \text{where } \gamma = \sum_{k=0}^{m_i} \exp\left[-\left(\sum_{x'=0}^{k} \delta_{ix'}\right) + k\beta\right] \tag{21.6}$$

In summary, the only difference between the rating and partial credit parametrizations is that the former has all items having the same number of thresholds and the thresholds, which are deviations from the location of the items, are all the same, while in the latter the mean deviated thresholds are not the same. The rating and partial credit parameterizations can both be used when all items have the same number of categories. In the case that there are different numbers of categories among items, then the partial credit parameterization needs to be used. When the item location is added to the thresholds, then the thresholds are referenced to the same origin and can be compared.

Exercises

Exercise 2: Basic analysis of dichotomous and polytomous responses in Appendix C.

Exercise 4: Advanced analysis of polytomous responses in Appendix C.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 357–374.

Andrich, D. (2010a). Educational measurement: Rasch models. In P. Peterson, E. L. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed., Vol. 4, pp. 111–122). Elsevier.

Andrich, D. (2010b). Understanding the response structure and process in the polytomous Rasch model. In M. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models: Developments and applications* (pp. 123–152). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Andrich, D. (2011). Rating scales and Rasch measurement. *Expert Review of Pharmacoeconomics & Outcomes Research, 11*(5), 571–585.

Dawes, R. M. (1972). *Fundamentals of attitude measurement*. New York: Wiley.

Masters, G. N. (2016). Partial credit model. In W. J. van der Linden (Ed.), *Handbook of item response theory: Models* (Vol. 1, Chap. 7, pp. 109–126). Boca Raton, Florida: Taylor and Francis.

Rasch, G. (1966). An individualistic approach to item analysis. In P. F. Lazarsfeld & N. W. Henry (Eds.), *Readings in mathematical social science* (pp. 89–108). Chicago: Science Research Associates.

van Wyke, J. F. (2003). Constructing and interpreting achievement scales using polytomously scored items: A comparison between the Rasch and Thurstone models. Professional Doctorate thesis, Murdoch University, Western Australia.