Chapter 20
The Polytomous Rasch Model I

Statistics Review 13: Distribution theory

In Chap. 3, it was shown that the analysis of items that are rated or given partial credit could be combined with items that are scored simply $x_{ni} = 0$ or $x_{ni} = 1$. In such items, the scores assigned are extended beyond 0 and 1 to give, for example $x_{ni} = 0$ or $x_{ni} = 1$ or $x_{ni} = 2$ or $x_{ni} = 3$. We have already used the term *dichotomous* for the case where items are scored 0 or 1 and the term *polytomous* when there are more than two graded categories. Sometimes you will see *polychotomous*. Psychometricians have discussed which is aetiologically correct and there seems to be a consensus that it is polytomous.

We stress here the ordering of the categories, such as when one awards the marks of 0 for poor performance, 1 for moderate performance and 2 for excellent performance, where these performances are defined operationally in some way. Even in the dichotomous case, however, the categories were ordered in the sense that there was a preferred outcome—a score of 1 for correct is considered better than a score of 0 for incorrect. Sometimes in attitude questionnaires, the direction of the ordering is genuinely arbitrary, but in any case, an order is implied and needs to be consistent among items.

The analysis of polytomous data generalizes readily from the dichotomous, but in order to see this, review the preliminary idea about average values from the dichotomous case to the polytomous one in *Statistics Review 13*.

The Model for Ordered Polytomous Responses

We first consider the case of an item with three ordered response categories. We have seen in *Statistics Review 13* how we can use the idea of a probability to obtain a theoretical mean, where in the dichotomous case the probability of success is the mean. We now set up the model in a form that gives the probability that each score

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_20

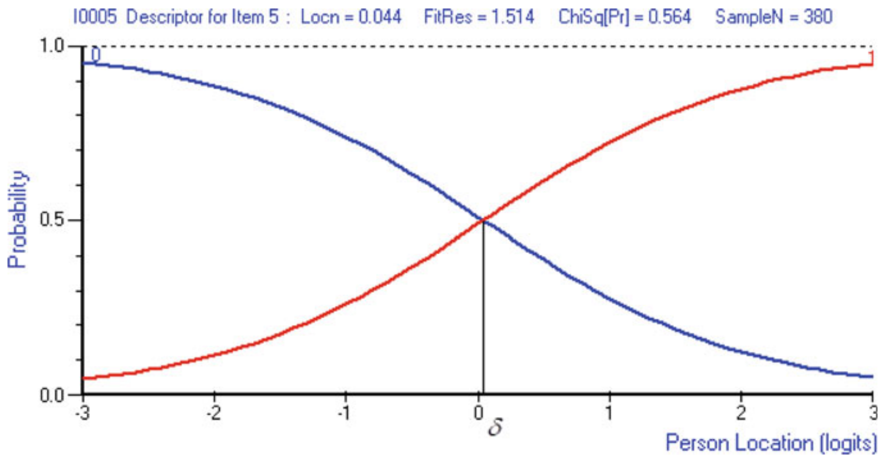

Fig. 20.1 ICCs for both the 0 response and the 1 response

will occur as a function of the proficiency of the person and the difficulty of the item. This is a generalization of the Rasch model for dichotomous items.

You will recall that, in full, the Rasch model for dichotomous responses is

$$\Pr\{x_{ni} = 1\} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}} \text{ and } \Pr\{x_{ni} = 0\} = \frac{1}{1 + e^{\beta_n - \delta_i}}. \tag{20.1}$$

So far we have drawn only the ICC for the correct response $x_{ni} = 1$. This is adequate in the dichotomous case because there are only two possible responses and the probability of a score of 0 is always a complement of a score of 1, $\Pr\{0\} + \Pr\{1\} = 1$.

Figure 20.1 shows the ICCs for both the 0 response and the 1 response for a simulated dichotomous item. It is clear from Fig. 20.1 that the probability of 0 decreases as the proficiency of the person increases, complementing the probability of 1.

Suppose, however, that we now have an item in which the possible scores are 0, 1 and 2. We might in advance consider the kind of probability curves these three responses should have. Figure 20.2 shows such response probabilities for a simulated polytomous item of difficulty $\delta = 0.067$. The item's thresholds τ_1 and τ_2, the points of equal probability for adjacent categories, are also shown.

Figure 20.2 shows that the response for the score $x_{ni} = 0$ is essentially the same as in the dichotomous case—as the proficiency increases, the probability of a score of 0 decreases. Also as the proficiency increases, the probability of a maximum score of 2 increases. Both as expected. However, between these curves is the curve which shows the probability of a score of 1. This curve shows that when a person is of moderate proficiency relative to the item's difficulty, then the most likely score is a 1. This structure is central to the model for graded, partial credit, or rating responses.

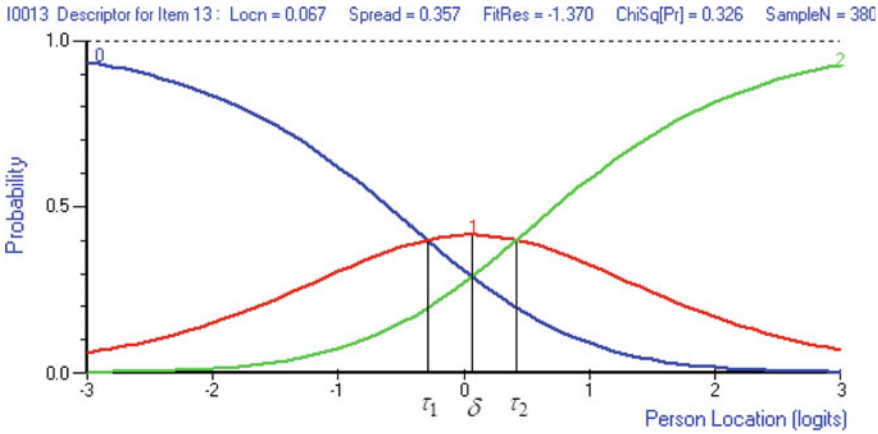

Fig. 20.2 ICCs for the 0, 1 and 2 responses in an item with three categories

In Fig. 20.2 there are two new parameters, τ_{1i} and τ_{2i}. These thresholds are the points where the probability of a response of either 0 or 1, and 1 or 2, respectively, are equally likely. In the case of a dichotomous response (with two categories), the only threshold is the difficulty which is the point where the probability of either 0 or 1 is equally likely. In the case of three categories there are two thresholds, each of which qualifies the average difficulty of the item, which is still denoted by δ_i and is the mean of the thresholds; $\delta_i = (\tau_{1i} + \tau_{2i})/2$ in the case of two thresholds.

The generalization of the Rasch model for dichotomous responses is now shown. You might be interested that the model is relatively recent as far as models are concerned. It was derived in two related papers, Andersen (1977) and Andrich (1978), and is based on the work of Rasch (1961).

First, we rewrite the case of the dichotomous model so that it is easier to generalize. The more complete and symmetric expressions for the parts of Eq. (20.1) are

$$\Pr\{x_{ni} = 0\} = \frac{e^{0(\beta_n - \delta_i)}}{e^{0(\beta_n - \delta_i)} + e^{1(\beta_n - \delta_i)}}; \tag{20.2a}$$

$$\Pr\{x_{ni} = 1\} = \frac{e^{1(\beta_n - \delta_i)}}{e^{0(\beta_n - \delta_i)} + e^{1(\beta_n - \delta_i)}}. \tag{20.2b}$$

Because any number to the power of 0 is 1, then $e^{0(\beta_n - \delta_i)} = e^0 = 1$ in Eq. (20.2a) giving $\Pr\{x_{ni} = 0\} = \frac{e^{0(\beta_n - \delta_i)}}{e^{0(\beta_n - \delta_i)} + e^{1(\beta_n - \delta_i)}} = \frac{1}{1 + e^{1(\beta_n - \delta_i)}}$, as required.

Notice that again the expressions, Eqs. (20.2a) and (20.2b), have the same denominator, which is the sum of the numerators. The numerators carry the essential form of the model, and the denominator simply ensures that the sum of the two probabilities is 1.

Notice also that the number multiplying $(\beta_n - \delta_i)$ is the score of the response—when the response is $x_{ni} = 0$, then $(\beta_n - \delta_i)$ is multiplied by 0 to give $0\,(\beta_n - \delta_i)$

in the exponent of the numerator; when the response is $x_{ni} = 1$, then $(\beta_n - \delta_i)$ is multiplied by 1 to give $1\ (\beta_n - \delta_i)$ in the exponent of the numerator. We might expect that when the item has three categories and the possible scores are 0, 1 and 2, that this feature will remain, and that when the response is $x_{ni} = 2$, then $(\beta_n - \delta_i)$ will be multiplied by 2. This is indeed the case.

With three categories, the model takes the form

$$\Pr\{x_{ni} = 0\} = \frac{e^{0(\beta_n - \delta_i)}}{e^{0(\beta_n - \delta_i)} + e^{-\tau_{1i} + 1(\beta_n - \delta_i)} + e^{-\tau_{1i} - \tau_{2i} + 2(\beta_n - \delta_i)}} \tag{20.3a}$$

$$\Pr\{x_{ni} = 1\} = \frac{e^{-\tau_{1i} + 1(\beta_n - \delta_i)}}{e^{0(\beta_n - \delta_i)} + e^{-\tau_{1i} + 1(\beta_n - \delta_i)} + e^{-\tau_{1i} - \tau_{2i} + 2(\beta_n - \delta_i)}} \tag{20.3b}$$

$$\Pr\{x_{ni} = 2\} = \frac{e^{-\tau_{1i} - \tau_{2i} + 2(\beta_n - \delta_i)}}{e^{0(\beta_n - \delta_i)} + e^{-\tau_{1i} + 1(\beta_n - \delta_i)} + e^{-\tau_{1i} - \tau_{2i} + 2(\beta_n - \delta_i)}} \tag{20.3c}$$

The equations involve the two thresholds. If the response is 0, and therefore no threshold has been exceeded, then no threshold appears in the numerator and the coefficient or multiplier of $(\beta_n - \delta_i)$ is 0. If the response is 1 and therefore only the first threshold has been exceeded and the rest have been failed, then the first threshold appears in the numerator and the coefficient or multiplier of $(\beta_n - \delta_i)$ is 1. If the response is 2 and therefore both the first and second thresholds have been exceeded, then both thresholds appear in the numerator and the coefficient or multiplier of $(\beta_n - \delta_i)$ is 2.

The denominator is once again the sum of all the numerators—in this case there are 3 numerators.

This kind of expression generalizes to any number of scores. It can be written for any score x_{ni} in the following form:

$$\Pr\{x_{ni} = x\} = \frac{e^{-\tau_{1i} - \tau_{2i} \ldots - \tau_{xi} + x(\beta_n - \delta_i)}}{\sum_{x'=0}^{m_i} e^{-\tau_{1i} - \tau_{2i} \ldots - \tau_{x'i} + x'(\beta_n - \delta_i)}}. \tag{20.4}$$

The denominator is just the sum of all of the numerators, and the numerator is a generalization of the case with three categories.

The above development of the model includes the hypothesis that the thresholds are ordered such that $\tau_{mi} > \tau_{m-1i} > \cdots \tau_{2i} > \tau_{1i}$. In estimates of the parameters, it is possible for them to show a reversed ordering. If there is a reversed ordering of the thresholds, then there is a problem with the way that the categories function. A fuller discussion of this feature is provided in subsequent chapters on the polytomous Rasch model.

Test of Fit Between the Data and the Model

A key aspect of checking fit between the model and the data is once again the comparison of the observed mean for a class interval and the theoretical mean or, formally, the expected value $E[X]$. Given the proficiency and the item parameter estimates, the probabilities of responding in each category for each item are estimated from Eq. (20.4).

The expected value (theoretical mean) is given by

Expected Value:

$$E[X_{ni}] = \sum_{x=0}^{m_i} P_{xi}(x_i) \tag{20.5}$$

where P_{xi} is the probability of a score of x determined from Eq. (20.4).

The observed mean is calculated by the same expression,

$$\text{Observed Mean} = \sum_{x=0}^{m_i} p_{xi}(x_i) \tag{20.6}$$

except that instead of the P_{xi} being a *probability* estimated according to the model, it is the *observed* proportion p_{xi} of the number of responses in category x. Each person's expected value is calculated, and then the expected values and observed scores of persons in each class interval are analyzed.

Interpretation from a Computer Output

Below is the RUMM2030 output from analysis according to the Rasch model for ordered response categories of the data in Chap. 3. Recall that in this analysis, some of the items scored 0 and 1 could be put naturally into sets. In particular, these were items 6, 9 and 10. The scores for each of these item sets were added together so that items 6, 9, and 10 are now polytomous items.

Proportions in Each Category

To get an orientation to the data and the analysis, Table 20.1 shows the distribution of responses of all persons in each of the categories for each of the items. This information is taken directly from the computer program used to analyze the data.

Table 20.1 Observed proportions of responses in each category for each item

Item number	Item label	Score				
		0	1	2	3	4
1	m001	0.08	0.92			
2	m002	0.04	0.96			
3	m003	0.12	0.88			
4	m004	0.14	0.86			
5	m005	0.08	0.92			
6	m006	0.02	0.02	0.08	0.39	0.49
7	m007	0.29	0.71			
8	m008	0.33	0.67			
9	m009	0.02	0.10	0.29	0.59	
10	m010	0.14	0.35	0.20	0.22	0.08

Table 20.2 Estimated thresholds for all items

Item number	Item label	Location estimate	Threshold estimates			
			1	2	3	4
1	m001	−0.997	0.000			
2	m002	−1.752	0.000			
3	m003	−0.378	0.000			
4	m004	−0.327	0.000			
5	m005	−0.868	0.000			
6	m006	0.452	0.326	−0.829	−0.614	1.118[a]
7	m007	0.745	0.000			
8	m008	0.960	0.000			
9	m009	0.259	−0.738	−0.018	0.757	
10	m010	1.910	−1.523	0.610	−0.183	1.096[a]

[a]These thresholds show disorder and this implies that there is a problem with the operation of the categories

Threshold Estimates for the Items

Table 20.2 shows the threshold estimates for all of the items. Note that there are no threshold estimates for the items that are scored simply 0 and 1. Also, there is a problem with two items because the thresholds are not correctly ordered. Later in this chapter, and the next chapter, the ordering of the thresholds is considered in more detail. Here we are simply providing an orientation to the analysis.

Table 20.3 Estimated difficulties for all items

Item number	Location estimate δ_i	Std. error
m001	−0.997	0.550
m002	−1.752	0.747
m003	−0.378	0.444
m004	−0.327	0.437
m005	−0.868	0.525
m006	0.452	0.192
m007	0.745	0.336
m008	0.960	0.325
m009	0.259	0.214
m010	1.910	0.147

Location (Difficulty) Estimates for the Items

Table 20.3 shows the difficulty δ_i estimates, headed LOCATION, and their standard errors for each of the items. Notice that the sum of the thresholds in Table 20.2 sum to 0.

The Test of Fit for a Dichotomous Item Scored 0 or 1

Table 20.4 shows the details for the test of fit for item m007, which was considered in detail in Chap. 13. It is one of the dichotomously scored items. The output is equivalent, though not identical, because the simultaneous estimates of the parameters with polytomous scoring with other items rearranged the values a little.

The item fits the model, as in the previous analysis when all items were treated as dichotomous, which is evident from the χ^2 value of 1.315 and significance of 0.505.

Figure 20.3 shows the item characteristic curve (ICC) for item m007 under the analysis where some items are scored as partial credit. In this case, the item does not discriminate as well as it did before when all the items were treated as dichotomous—the points showing the observed proportions are flatter than the theoretical curve.

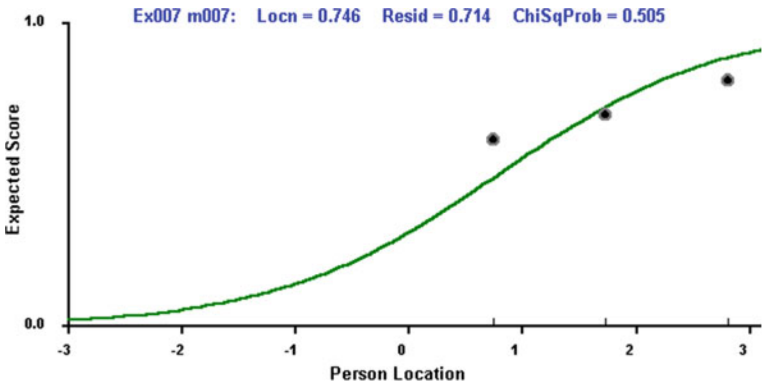

Fig. 20.3 Item characteristic curve for item m007

The Test of Fit for a Partial Credit Item m009 Scored 0, 1, 2 or 3

Table 20.5 shows the details for the test of fit for item m009 which has possible scores of 0, 1, 2 and 3. The structure of the output is the same as for dichotomously scored items.

According to the χ^2 statistic, item m009 shows poor fit. The item characteristic curve for this item is shown in Fig. 20.4. The observed means for each class interval

Table 20.4 Test of fit for item m007 with a location of 0.746

Group		Location		Component			Category responses	
No.	Size	Max	Mean	Residual	ChiSqu		0	1
1	13	1.186	0.736	0.858	0.737	OBS.P	0.38	0.62
						EST.P	0.50	0.50
OM = 0.62 EV = 0.50 OM-EV = 0.12 ES = 0.24						OBS.T		0.62
2	20	1.864	1.671	−0.146	0.021	OBS.P	0.30	0.70
						EST.P	0.28	0.72
OM = 0.70 EV = 0.71 OM-EV = −0.01 ES = −0.03						OBS.T		0.70
3	16	3.451	2.769	−0.746	0.557	OBS.P	0.19	0.81
						EST.P	0.12	0.88
OM = 0.81 EV = 0.87 OM-EV = −0.06 ES = −0.19						OBS.T		0.81
AVE = 0.78								
ITEM: df = 2.00 ChiSqu = 1.315 Significance = 0.505								

Table 20.5 Test of fit for item m009 with a location of 0.257

Group		Location		Component			Category responses			
No.	Size	Max	Mean	Residual	ChiSqu		0	1	2	3
1	13	1.186	0.736	1.414	2.001	OBS.P	0.00	0.23	0.31	0.46
						EST.P.	0.07	0.24	0.39	0.30
OM = 2.23 EV = 1.89 OM-EV = 0.34 ES = 0.39						OBS.T		1.00	0.57	0.60
2	20	1.864	1.671	−2.026	4.106	OBS.P	0.05	0.10	0.45	0.40
						EST.P	0.01	0.07	0.31	0.60
OM = 2.20 EV = 2.51 OM-EV = −0.31 ES = −0.45						OBS.T		0.67	0.82	0.47
3	16	3.451	2.769	1.184	1.401	OBS.P	0.00	0.00	0.06	0.94
						EST.P	0.00	0.01	0.15	0.84
OM = 2.94 EV = 2.81 OM-EV = 0.13 ES = 0.30						OBS.T		**	1.00	0.94
AVE = 2.46										
ITEM: df = 2.00 ChiSqu = 7.507 Significance = 0.000										

Note ** = undefined

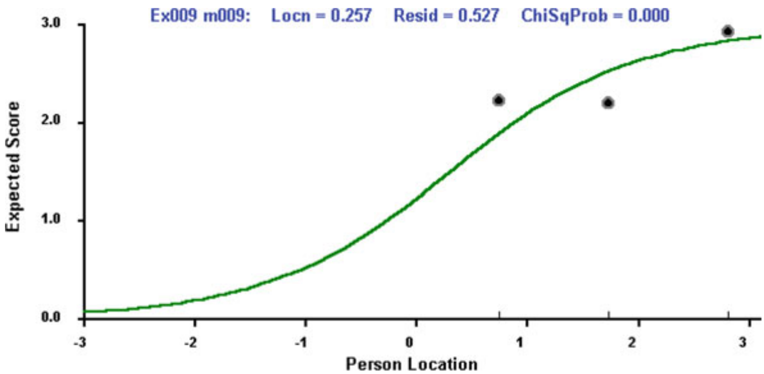

Fig. 20.4 Item characteristic curve for item m009

can again be compared to their expected values. It is evident that item m009 does not discriminate as well as it might across the first two groups.

Threshold Order for Item m009 Scored 0, 1, 2 and 3

In addition to the test of fit in terms of theoretical and observed means, the order of the thresholds which form the categories of the items is important. The thresholds are shown in Table 20.2, and for this item they are in the correct

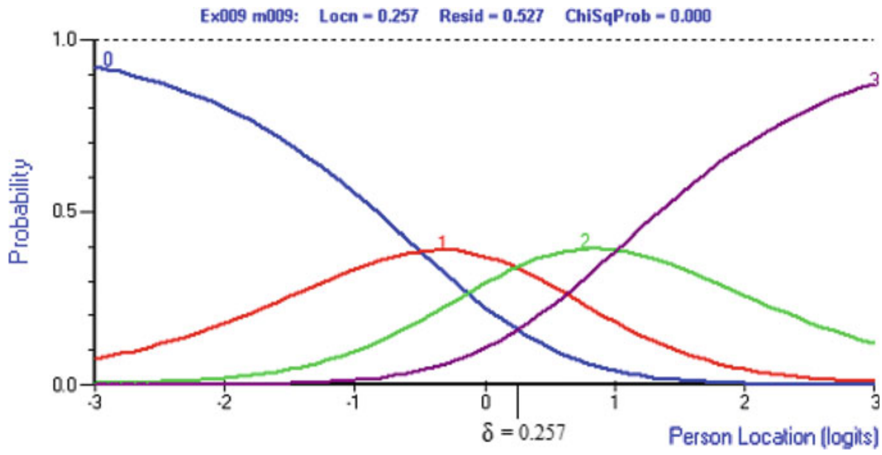

Fig. 20.5 Category characteristic curves for item m009

order: −0.738, −0.018, 0.757. By convention in the form of Eq. (20.4), these thresholds have a mean of 0. To locate the thresholds on the common scale, the item difficulty (0.257) has to be taken into account and located first—then the thresholds are located around the item difficulty. Figure 20.5 shows the category characteristic curves (CCCs) for this item. They show the probabilities of each response category as a function of person proficiency. Because the thresholds are ordered, the curves show the required relationship shown in Fig. 20.2.

Threshold Order for Item m010 Scored 0, 1, 2, 3 and 4

The thresholds for item m010 shown in Table 20.2 have the values −1.523, 0.610, −0.183 and 1.096. These are not in the correct order. Figure 20.6 shows the CCCs for this item. The curves show a relationship which is a mess. In particular, there is no region of the continuum in which a score of 2 is the most likely. That is, even in the region of proficiencies where the expected (mean) score is 2, people are more likely to obtain one of the other scores. This indicates that the categories are not working as intended; as the proficiency of persons increases, the probability of gaining a higher score does not increase systematically—a score of 2 is never the most likely. This means that the item should be studied to understand why categories are not working as intended.

Although the categories are not working as intended, according to the χ^2 test, the item fits the model. The ICC is shown in Fig. 20.7. It must be remembered that the

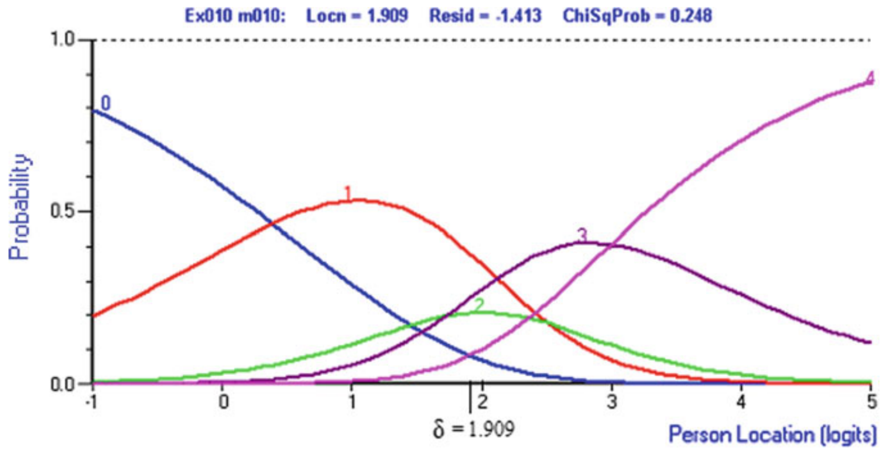

Fig. 20.6 Category characteristic curves for item m010

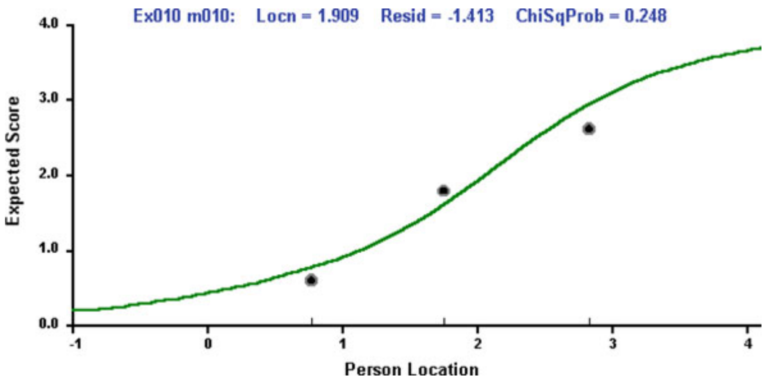

Fig. 20.7 Item characteristic curve for item m010

sample size is very small for detecting misfit this way, but it does show that the data may show a discrepancy from the model in one way and not in another way.

Estimates of the Proficiencies of the Persons

In the case with partial credit, the total score has the same property as it does with dichotomous items; if the data fit the Rasch model, then the total score contains all of the information for the person. Table 20.6 shows the proficiency estimates, and the standard errors, associated with each total score. Once again, different total scores show different standard errors, with the scores in the middle having smaller ones than those on the extreme.

Table 20.6 Proficiency estimates associated with each total score

Total score	Frequency	Proficiency estimate	Standard error
0	0	$-\infty$	∞
1	0	−2.690	1.076
2	0	−1.846	0.801
3	0	−1.308	0.675
4	0	−0.908	0.592
5	0	−0.591	0.537
6	1	−0.324	0.500
7	1	−0.084	0.481
8	1	0.144	0.476
9	0	0.374	0.485
10	2	0.619	0.505
11	3	0.888	0.532
12	5	1.186	0.560
13	11	1.513	0.582
14	0	1.864	0.603
15	6	2.247	0.641
16	5	2.714	0.738
17	5	3.451	1.024
18	1	$+\infty$	∞
Mean = 1.782 Std. deviation = 0.889			

Exercises

Exercise 2: Basic analysis of dichotomous and polytomous responses in Appendix C.
Exercise 4: Advanced analysis of polytomous responses in Appendix C.

References

Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika, 42,* 69–81.
Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 357–374.
Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), *Proceeding of the fourth Berkeley symposium on mathematical statistics and probability* (Vol. 4, pp. 321–333). Berkeley, California: University of California Press. Reprinted in Bartholomew, D. J. (Ed.). (2006) *Measurement: sage benchmarks in social research methods* (Vol. I, pp. 319–334). London: Sage Publications.