LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Gerhard Tutz

On the Construction of Adjacent Categories Latent Trait Models from Binary Variables, Motivating Processes and the Interpretation of Parameters

Technical Report Number 218, 2018
Department of Statistics
University of Munich

http://www.statistik.uni-muenchen.de

On the Construction of Adjacent Categories Latent Trait Models from Binary Variables, Motivating Processes and the Interpretation of Parameters

Gerhard Tutz

Ludwig-Maximilians-Universität München

Akademiestraße 1, 80799 München

December 14, 2018

Abstract

Latent trait models have a mathematical representation that provides a link between person and item parameters and the probability of a response in categories. The usefulness of specific models is mainly determined by the motivation of models, the interpretation of parameters and the narratives around models. The focus is on the partial credit model, for which differing and contradicting motivations, interpretations and narratives have been given over time. It is shown that the model can be derived by assuming that binary Rasch models hold for binary variables that are always present in multi-categorical response models. An alternative derivation is based on binary Rasch models for latent variables that compare adjacent categories. It is shown that the PCM can generally be characterized as a model that conditionally compares two categories from the set of response categories. The representation as an adjacent categories model can be seen as just a specific parameterization. It is demonstrated that the confusion of these alternative binary variables in the PCM can be misleading and generate inappropriate interpretation of parameters.

Keywords: Ordered responses, latent trait models, item response theory, partial credit model, polytomous Rasch model, Rasch model

1 Introduction

Ordered latent trait models have a long tradition in psychometrics. Comprehensive overviews were given by Nering and Ostini (2011), Van der Linden (2016), see also von Davier and Carstensen (2007). The dominating models for ordered responses are Samejima's graded response model (Samejima, 1997), the polytomous Rasch model Andersen (1977), the partial credit model (Masters, 1982) and the rating scale model

(Andrich, 1978). The models can be grouped into two classes of models. The graded response model is a *cumulative response model*, which is distinctly different from the other models, which are from the class of *adjacent categories models*. Moreover, the polytomous Rasch model is equivalent to the partial credit model, the latter is just a reparameterization. The rating scale model is a simplified version of the polytomous Rasch model, which uses a restricted parameterizaion.

We focus on adjacent categories models and mostly use the partial credit parameterization. We also consider briefly the graded response model. This is necessary to clarify the differences between model classes, which are sometimes blurred in the literature. We do not consider explicitly the sequential model (Tutz, 1990; Verhelst et al., 1997), which builds a third class of ordered response models.

The starting point is the construction of ordered models from *binary response models*, which have been thoroughly investigated in the literature. They play a crucial role in all of the models but are most important in the partial credit model since ignoring them might yield an insufficient and misleading interpretation of parameters. One objective is to show how easy ordered latent trait models that are in common use can be derived from assuming that binary Rasch models hold for specific dichotomizations of the response categories. The main tool that is used is that *any* polytomous response can be uniquely represented by a sequence of binary variables for which realizations have the simple structure (1,...,1,0,...,0), that is, a sequence of ones is followed by a sequence of zeros. The structure has been investigated before, among others by Andrich (1978), Andrich (2013), but it has not been used that any latent trait model has this structure.

In the partial credit model a second type of binary variables is present, they represent the choice between adjacent categories. It is shown that the focus on adjacent categories is arbitrary. A general characterization is given that shows that the model is a structured collection of binary conditional models, in which the binary variables do not determine the response in adjacent categories but in a selected set of pairs of categories. The use of adjacent categories just yields a specific parametrization. In addition we consider the interpretation of parameters and possible processes behind models. In particular processes behind the partial credit model seem not to have been always distinguished clearly from processes behind the graded response model and the sequential model.

2 Models for Ordered Responses obtained from Split Variables

Let $Y_{pi} \in \{0, 1, \ldots, k\}$, $p = 1, \ldots, P$, $i = 1, \ldots, I$, denote the ordinal response of person p on item i. It is assumed that the categories $\{0, 1, \ldots, k\}$ are ordered and that there is a response model $P(Y_{pi} = r)$, $r = 0, \ldots, k$.

Various motivations for ordered models have been given. For example, Samejima's graded response model (Samejima, 1997) can be derived from an underlying continuous trait. Rather than referring to motivations of this sort we start with binary variables that are contained in the response Y_{pi} and aim at constructing ordinal models from them.

TABLE 1: *Binary variables that represent an ordinal response $Y_{pi} \in \{0,1,2,3\}$, Left: dummy coding, right: split variables.*

Dummy Coding		Split Coding	
Y_{pi}	$(Y_{pi}^{(1)}, Y_{pi}^{(2)}, , Y_{pi}^{(3)})$	Y_{pi}	$(Y_{pi}^{(1)}, Y_{pi}^{(2)}, , Y_{pi}^{(3)})$
0	(0,0,0)	0	(0,0,0)
1	(1,0,0)	1	(1,0,0)
2	(0,1,0)	2	(1,1,0)
3	(0,0,1)	3	(1,1,1)

2.1 Split Variables

Since $Y_{pi} \in \{0, 1, \ldots, k\}$ one has a polytomous response, which is not unidimensional but multidimensional. There are several ways to define (univariate) binary variable that represent Y_{pi} in multidimensional form. A common version uses variables $Y_{pi}^{(1)}, \ldots, Y_{pi}^{(k)}$, with $Y_{pi}^{(r)}$ defined by $Y_{pi}^{(r)} = 1$ if $Y_{pi} = r$, and $Y_{pi}^{(r)} = 0$, otherwise. Then one obtains

$$Y_{pi} = r \quad \Leftrightarrow \quad (Y_{pi}^{(1)} \ldots, Y_{pi}^{(k)}) = (0, \ldots, 0, 1, 0, \ldots, 0).$$

Thus, if $Y_{pi} = r$ only the variable $Y_{pi}^{(r)}$ has value one, all others have the value zero, see Table 1, left column. The downside of this set of variables is that it does does not use the ordering of categories.

Therefore, we will consider an alternative set of variables. Let the whole set of variables be defined by

$$(Y_{pi}^{(1)} = 1, \ldots, Y_{pi}^{(r)} = 1, Y_{pi}^{(r+1)} = 0, \ldots, Y_{pi}^{(k)} = 0) \text{ if } Y_{pi} = r. \tag{1}$$

The definition implies that only outcomes of the form $(1, \ldots, 1, 0, \ldots, 0)$ can occur, see Table 1, right column. Vectors of this form, which are given by a sequence of ones followed by a sequence of zeros, form a *Guttman space*, the corresponding variables are called *Guttman variables*, see, for example, Andrich (2013).

It is easy to derive the marginal distributions of the variables, which are given by

$$Y_{pi}^{(r)} = \begin{cases} 1 & Y_{pi} \in \{r, \ldots, k\} \\ 0 & Y_{pi} \in \{0, \ldots, r-1\}. \end{cases} \tag{2}$$

Therefore, the variable $Y_{pi}^{(r)}$ represents the *dichotomization* of the response categories into the subsets $\{0, \ldots, r-1\}$ and $\{r, \ldots, k\}$. $Y_{pi}^{(r)} = 1$ indicates that the response is in category r or in a higher one. Since the variables code the splitting of the set of categories into two groups of adjacent categories, we will also refer to them as *split variables*.

One could also have started with the definition of the split variables (2) and derive that (1) holds. Our starting point was the definition of the binary variables given in (1) since it makes clear that the representation of the ordinal responses Y_{pi} as a vector of Guttman variables entails that the variables are split variables.

In summary, split variables form a Guttman space, and if one represents the response by a Guttman space the Guttman variables are split variables. For the link between the split variables and the response in $k+1$ categories one has the equivalence

$$Y_{pi} = r \quad \Leftrightarrow \quad (Y_{pi}^{(1)} \dots, Y_{pi}^{(k)}) = (1, \dots, 1, 0, \dots, 0),$$

where $Y_{pi}^{(r)}$ is the last of the sequence of binary variables with a value 1. It should be emphasized that the Guttman space is a general concept. Although it has been used in the literature to describe adjacent categories models it is not linked to a specific model. It is just one way to represent ordered responses in multivariate form. In the following we will always use the coding of polytomous variables through the split variables, which form the Guttman space.

It is obvious that split variables are not independent. The covariance between two split variables is given by $\text{cov}(Y_{pi}^{(r)}, Y_{pi}^{(s)}) = P(Y_{pi} \geq s) - P(Y_{pi} \geq r)P(Y_{pi} \geq s) = P(Y_{pi} \geq s)(1 - P(Y_{pi} \geq r))$ for $r < s$. It is seen that $\text{cov}(Y_{pi}^{(r)}, Y_{pi}^{(s)}) = 0$ only if $P(Y_{pi} \geq r) = 1$.

The split variables defined in (2) refer to the ordering of categories because they use adjacent categories. In the following it is first shown how ordered latent trait model may be derived by assuming binary Rasch models to hold for the split variables.

2.2 Assuming Rasch models for the Split Variables

The binary split variables (2) distinguish between low and high categories, more concrete, $Y_{pi}^{(r)}$ distinguishes between $Y_{pi} < r$ and $Y_{pi} \geq r$. The simplest and most widely used latent trait model is the binary Rasch model. Therefore, one might assume that the binary variables $Y_{pi}^{(r)}$, which build the ordinal response, are determined by binary Rasch models. One implicitly accounts for the ordering of categories since $Y_{pi}^{(r)}$ uses the ordering. In concrete parameterization, one assumes

$$P(Y_{pi}^{(r)} = 1) = F(\theta_p - \delta_{ir}), \tag{3}$$

where $F(\eta) = \exp(\eta)/(1 + \exp(\eta))$ is the logistic distribution function, θ_p is a person parameter and δ_{ir} the item difficulty connected to the dichtomization $Y_{pi} < r$ and $Y_{pi} \geq r$. The assumption that in all of the binary models the same person parameters is present serves to obtain a simple one-dimensional model. In fact, it is easily derived that assuming (3) yields for the ordinal response Samejima's *graded response model*,

$$P(Y_{pi} \geq r) = F(\theta_p - \delta_{ir}), \quad r = 1, \dots, k,$$

see Samejima (1997, 2016). Consequently, the item difficulties have to fulfill $\delta_{i1} \leq \dots \leq \delta_{ik}$, since $\delta_{ir} > \delta_{i,r+1}$ for any r would yield negative probabilities. This conceptualization of the model differs from the usual motivation from a latent continuous response that is observed only in categories. It also shows that Guttman variables play a role in models that differ from the partial credit model.

2.3 Assuming Rasch Models Conditionally

An alternative way to obtain an ordered response model is to assume that Rasch models hold for the split variables given other splits have specific values. Let us consider the assumption

$$P(Y_{pi}^{(r)} = 1 | Y_{pi}^{(1)} = 1, \dots, Y_{pi}^{(r-1)} = 1, Y_{pi}^{(r+1)} = 0, \dots, Y_{pi}^{(k)} = 0) = F(\theta_p - \delta_{ir}). \tag{4}$$

It means that the Rasch model holds for the split into categories $Y_{pi} < r$ and $Y_{pi} \geq r$ *given* all splits for lower categories (into $Y_{pi} < s$ and $Y_{pi} \geq s$, $s < r$) are in favor of the higher category and all splits for higher categories (into $Y_{pi} < s$ and $Y_{pi} \geq s$, $s > r$) are in favor of the lower category.

Although the assumption seems less natural than the assumption that split variables themselves follow a Rasch model it is an assumption that yields a widely used model. It is not hard to show that if (4) holds one obtains the *polytomous Rasch model* (PRM) or *partial credit model* (PCM),

$$P(Y_{pi} = r) = \frac{\exp(r\theta_p - \sum_{l=1}^{r} \delta_{il})}{\sum_{s=0}^{k} \exp(\sum_{l=1}^{s}(\theta_p - \delta_{il}))}, \tag{5}$$

where $\sum_{k=1}^{0}(\theta_p - \delta_{ik}) = 0$. A more accessible form of the model is given by

$$\log\left(\frac{P(Y_{pi} = r)}{P(Y_{pi} = r-1)}\right) = \theta_p - \delta_{ir}, \quad r = 1, \ldots, k. \tag{6}$$

The representations (5) and (6) are known as the partial credit parameterizations of the model. An equivalent model is obtained by using the reparameterization $\delta_{il} = \delta_i + \tau_{il}$, $l = 1, \ldots, k$, where $\sum_{l=1}^{k} \tau_{il} = 0$. The representation

$$P(Y_{pi} = r) = \frac{1}{\gamma_{pi}} \exp(\sum_{l=1}^{r}\{\theta_p - \delta_i - \tau_{il}\}) = \frac{1}{\gamma_{pi}} \exp(r(\theta_p - \delta_i) - \sum_{l=1}^{r} \tau_{il}) \quad r = 1, \ldots, k,$$

with $\gamma_{pi} = \sum_{s=0}^{k} \exp(\sum_{l=1}^{s} \theta_p - \delta_i - \tau_{il})$ is usually referred to as the polytomous Rasch model, see also Andrich (1978).

One may also use used the weaker assumption

$$P(Y_{pi}^{(r)} = 1 | Y_{pi}^{(r-1)} = 1, Y_{pi}^{(r+1)} = 0) = F(\theta_p - \delta_{ir}), \tag{7}$$

and assume merely that the variables $(Y_{pi}^{(1)}, \ldots, Y_{pi}^{(k)})$ form a Guttman space. The weaker condition $Y_{pi}^{(r-1)} = 1, Y_{pi}^{(r+1)} = 0$ in (7) can be used since in a Guttman space the condition is equivalent to the more general condition $Y_{pi}^{(1)} = 1, \ldots, Y_{pi}^{(r-1)} = 1, Y_{pi}^{(r+1)} = 0, \ldots, Y_{pi}^{(k)} = 1$ used in (4). For the interpretation of parameters it does not matter which condition is used, the stronger or the weaker one. However, one can and should use that the variables are split variables because, as shown in Section 2.1, the representation of ordinal responses in a Guttman space means that the Guttman variables are split-variables (given in (2)). One consequence is that both conditions $Y_{pi}^{(r-1)} = 1, Y_{pi}^{(r+1)} = 0$ and $Y_{pi}^{(1)} = 1, \ldots, Y_{pi}^{(r-1)} = 1, Y_{pi}^{(r+1)} = 0, \ldots, Y_{pi}^{(k)} = 0$ mean that all the dichotomized decisions with the exception of $Y_{pi}^{(r)}$ are already fixed. Thus, the condition $Y_{pi}^{(r-1)} = 1, Y_{pi}^{(r+1)} = 0$ is equivalent to $Y_{pi} \in \{r-1, r\}$. The same holds for the stronger restriction. With all the dichotomized decisions with the exception of $Y_{pi}^{(r)}$ being fixed by the condition, the split variable $Y_{pi}^{(r)}$ distinguishes only between the categories $r-1$ and r. More formally one obtains

$$P(Y_{pi}^{(r)} = 1 | Y_{pi}^{(r-1)} = 1, Y_{pi}^{(r+1)} = 0) = P(Y_{pi} = r | Y_{pi} \in \{r-1, r\}). \tag{8}$$

That means assuming a Rasch model model for the conditional variables $Y_{pi}^{(r)}|Y_{pi}^{(r-1)} = 1, Y_{pi}^{(r+1)} = 0$ is equivalent to assuming a Rasch model for $Y_{pi}|Y_{pi} \in \{r-1, r\}$.

Derivations of the polytomous Rasch model from Guttman variables were also given by Andrich (1978), Luo (2005), Andrich (2013). However, they use a different representation, in particular, they do not use that Guttman variables are split variables and try to *infer the existence* of Guttman spaces by investigating various spaces. They do not use that Guttman spaces are just a recoding of ordered responses as shown in Section 2.1. Nevertheless, since the Guttman variable in an orderd response model are unique they must have implicitly considered the same variables.

An important difference of the present derivation and the derivation given by Andrich (1978, 2013) is that in the latter it is argued that the thresholds $\delta_{i1}, \dots, \delta_{ik}$ should be ordered to obtain a sensible model. In the derivation given here there is no necessity that thresholds should be ordered.

3 Alternative Constructions of the Partial Credit Model

The model form (6), which directly compares the probabilities of obtaining categories $r-1$ and r, seems to suggest an alternative and simple way to construct the PCM from the comparison of two categories and therefore a binary choice. In the following it is shown that the PCM can indeed be constructed rather generally from binary choices that refer to pairs of categories. It makes the PCM a general model for pairs of categories.

3.1 The Partial Credit Model as a Model for Pairs of Categories

Let us again assume that one wants to construct an ordered latent trait model from binary models. One might indeed think of adjacent categories and model the preference for the higher category as a function of the person parameter, which should be the same for all pairs of adjacent categories in order to obtain a one dimensional model. However, one might also think of alternative pairs of categories. Let us, for example, consider the pairs $(0,1), (0,2), \dots, (0,k)$ and postulate

$$\log\left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=0)}\right) = r\tilde{\theta}_p - \tilde{\delta}_{ir}, \quad r = 1, \dots, k,$$

where the ability is scaled by r yielding the ability $r\tilde{\theta}_p$ for the comparison of categories r and 0. The scaling reflects the difference between categories, $r = r - 0$. The model simply compares the preference of category r over category 0. As in the PCM it is a conditional preference as seen from the equivalent representation that makes the condition obvious,

$$P(Y_{pi} = r|Y_{pi} \in \{0, r\}) = F(r\tilde{\theta}_p - \tilde{\delta}_{ir}). \tag{9}$$

The model is a collection of *conditional* binary models for the pairs $(0,1), (0,2), \dots, (0,k)$. The interesting point is that the model (9) is equivalent to the PCM, it is just an alternative representation, however, a representation that does not use adjacent categories.

The construction principle to obtain an ordinal model from the comparison of a pair of categories applies much more generally. One can choose among quite different sets of pairs of categories.

Let $P = \{(s_1, r_1), (s_k, r_k)\}$ denote a set of k pairs of categories, for which $s_i < r_i$, $i = 1, \dots, k$, and $s_i, r_i \in \{0, \dots, k\}$. Let the model for proportions of probabilities be defined by

$$\log(\frac{P(Y_{pi} = r)}{P(Y_{pi} = s)}) = (r - s)\tilde{\theta}_p - \tilde{\delta}_{irs}, \quad (s, r) \in P, \tag{10}$$

or equivalently by

$$P(Y_{pi} = r | Y_{pi} \in \{s, r\}) = F((r - s)\tilde{\theta}_p - \tilde{\delta}_{irs}) \quad (s, r) \in P. \tag{11}$$

The predictor on the right hand includes a scaling that reflects the difference between categories and the item parameters depend on the considered pair. Since P contains k pairs the number of item parameters is the same as in the PCM, namely k.

It is easy to see that a further condition is needed to make (11) a sensible model. For example, if a category, say 0, is not present in the considered pairs one will not obtain a probability for category 0. Therefore, only non-redundant sets of category pairs are considered admissible. *Non-redundant* sets of pairs are those for which the restrictions on probabilities specified in (11) are independent. For a concise definition let the model be given in matrix form

$$\boldsymbol{D} \log(\boldsymbol{\pi}) = \boldsymbol{\theta},$$

where $\boldsymbol{\pi}^T = (\pi_0, \dots, \pi_k)$, $\pi_r = P(Y_{pi} = r)$, is the vector of probabilities, $\log(\boldsymbol{\pi})^T = (\log(\pi_1), \dots, \log(\pi_k))$ are the logarithms of probabilities, and $\boldsymbol{\theta}$ is a vector that contains the right hand terms in (10). The matrix $\boldsymbol{D}$ generates the differences between the logarithms of probabilities. One row in the matrix $\boldsymbol{D}$ has the form $(0, \dots, 0, -1, 0, \dots, 1, \dots, 0)$, where the -1 is in column s and 1 is in column r (starting with 0 since a row vector has length $k + 1$), which generates the difference $\log(\pi_r)) - \log(\pi_s)$. The model is based on a non-redundant set of pairs if the matrix $\boldsymbol{D}$ has k independent rows. As shown in the Appendix the following proposition holds.

Proposition 3.1 *Let P be a set of non-redundant pairs of categories. Then the model for proportions of probabilities given in (10) is equivalent to the partial credit model.*

That means any set of non-redundant pairs can be used to define the partial credit model. The partial credit model in the adjacent categories formulation (6) is just one specific parameterization that uses adjacent categories.

It should be noted that Proposition 3.1 goes beyond a simple reparameterization of the partial credit model. It shows that the comparison of pairs of categories is a *necessary and sufficient* condition for the PCM. It is straightforward to derive that it is a necessary condition since (10) follows after some simple reformulation if one assumes that the PCM holds. This has been used before, for example by Wilson (1992), Wilson and Masters (1993). However, Proposition 3.1 shows that comparison of pairs of categories is also a sufficient condition. If one specifies that the comparison of pairs for a sufficiently large set of pairs follows a Rasch model one obtains that the PCM holds. Thus the PCM can be *constructed* from comparing a non-redundant pair of categories. It makes the PCM a structured collection of conditional binary Rasch models for any non-redundant pairs of categories.

3.2 Interpreting Parameters of the Partial Credit Model

The interpretation of the parameters of the PCM has undergone changes over the years. In early versions of the PCM the model was interpreted as a model for consecutive steps (Masters, 1982) and parameters were interpreted as difficulties in these steps. However, the conditioning on adjacent categories shows that it cannot be a model for *consecutive* steps. A model for consecutive steps is the sequential or step model, which is a genuine sequential processing model and has a quite different form (Molenaar, 1983; Tutz, 1990; Verhelst et al., 1997). Thus, interpretation of parameter should not refer to consecutive steps.

There are a least two alternative ways to obtain an interpretation. One starts from the model as representing probabilities for response categories or one refers to the hidden variables that are determined by Rasch models. We consider both approaches, but for simplicity, confine ourselves to the adjacent categories parameterization, which is the most widely used one.

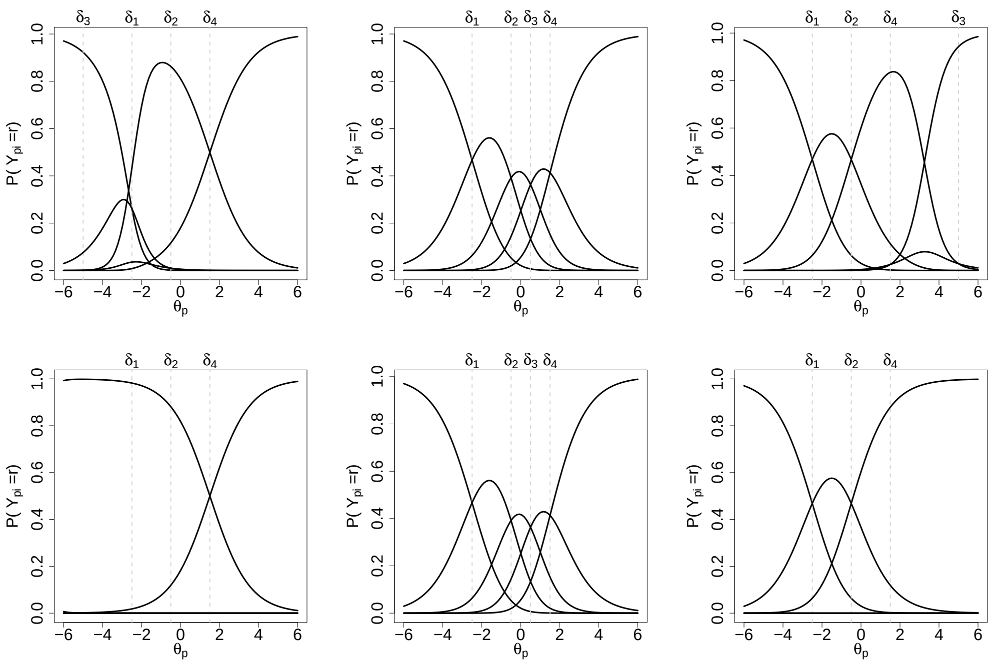

FIGURE 1: *Probability response curves for items with five categories; left:* $\delta_{i3} = -5$ *(upper panel) and* $\delta_{i3} = -10$ *(lower panel); right:* $\delta_{i3} = 5$ *(upper panel) and* $\delta_{i3} = 10$ *(lower panel)*

Let us consider the representation (6), which can also be given in the form

$$\frac{P(Y_{pi} = r)}{P(Y_{pi} = r-1)} = e^{\theta_p - \delta_{ir}}, \quad r = 1, \dots, k. \tag{12}$$

It directly parameterizes the comparison of probabilities of adjacent categories. In achievement tests θ_p represents the ability and it is obvious that an increase in θ_p increases the probability of the higher category of the pair of categories. Since θ_p is the same in all comparisons this holds for all pairs of adjacent categories. Consequently, larger person parameters yield higher scores. The item parameters can be seen as thresholds because they are the intersection points of the successive pairs of category probability curves as seen from the examples in Figure 1. They have also be described as the points which have a 50% probability of a response in either category, which holds only if consideration is restricted to two categories. The interpretation as thresholds will be unproblematic if it refers only to the property that $P(Y_{pi} = r-1) = P(Y_{pi} = r)$ for $\theta_p = \delta_{ir}$ but does not mean that steps or transitions are involved, a point that was emphasized by Andrich (2015) and will also be discussed later.

The crucial point in the interpretation of the item parameter is that in (12) just two categories are compared at a time, although the comparison has effects on all the other probabilities. The basic tendency is: if δ_{ir} is small the probability of the lower category, $P(Y_{pi} = r - 1)$, will be small, if δ_{ir} is high the probability of the higher category will be small. This is illustrated in the first row of Figure 1. The panel in the middle is the reference with sensibly ordered item parameters. In the right panel the threshold δ_{i3} has been increased to $\delta_{i3} = 5$. Consequently the probability of category 3 is small. In the left panel the threshold δ_{i3} has been decreased to $\delta_{ir} = -5$ yielding that the probability of observing category 2 is very small, but also category 1 has smaller probabilities than in the middle panel. One can also look at extreme cases. If all other parameters are fixed one obtains $P(Y_{pi} = r) = 0$ if $\delta_{ir} \to \infty$, and $P(Y_{pi} = r - 1) = 0$ if $\delta_{ir} \to -\infty$. However, it has not only consequences for the adjacent categories as illustrated by the second row in Figure 1 ($\delta_{i3} = -10$ in the left panel, $\delta_{i3} = 10$ in the right panel). It is seen that if $P(Y_{pi} = 2)$ tends to zero also the probabilities $P(Y_{pi} = 0), P(Y_{pi} = 1)$ will tend to zero (left panel). If $P(Y_{pi} = 3)$ tends to zero also $P(Y_{pi} = 4)$ will tend to zero (right panel). Thus, although the thresholds determine only the proportion between adjacent categories they have a strong impact on the other probabilities. Therefore, the common interpretation that δ_{ir} is a threshold that determines the preference of category r over $r - 1$ is rather misleading since it does not make it explicit that the comparison holds only for these categories. Actually it is a *conditional* comparison, *given the response is in categories categories* $r - 1$ *or* r.

For the clarification of the interpretation of parameters it is useful to consider the nature of the binary responses that are modeled as Rasch models more closely. At first sight it seems that they are simple binary variables but since they use just two categories they are actually conditional variables given by

$$Y_{pi}^{(r)} = \begin{cases} 1 & Y_{pi} \geq r \text{ given } Y_{pi} \in \{r-1, r\} \\ 0 & Y_{pi} < r \text{ given } Y_{pi} \in \{r-1, r\}. \end{cases} \tag{13}$$

For these variables the binary Rasch models are assumed to hold. Important features of these variables are that they are *latent* and *conditional*. Since with ordinal responses there is one observed response in one category only the condition is never observed and consequently variables are latent. Of course, latent models may be used to interpret parameters and this has been done in the literature. Andrich (2015) argued that the item parameters can be interpreted exactly as they are in the binary case, namely as *difficulty parameters*. The interpretation is based on the hidden Rasch models and uses the different conceptualization of the PCM that was outlined before in several papers (Andrich, 1978, 2013). A crucial element in these derivations and the interpretation of parameters is that the latent binary variables form a Guttman space. However, as has been shown in Section 2.1 the Guttman space in latent trait model is *observable*. One should not confuse the conditional binary variables given in (13) with the variables that show the Guttman structure. The former are latent and are never observed whereas the latter are observable.

The main problem with the interpretation of model parameters obtained from the binary models for adjacent categories is due to the presence of the condition in the variables (13). The model representations (6) and (12) are indeed seductive, they suggest that the condition can be ignored. This is the reason why often the item parameter (given the person parameter) is seen as determining the "preference of category r over category $r-1$". The restriction to adjacent categories is simply dropped when interpreting parameters. However, the condition is present in the building block of the PCM and must not be ignored if one uses the Rasch model to interpret parameters.

The crucial condition behind the PCM is

$$P(Y_{pi} = r | Y_{pi} \in \{r-1, r\}) = F(\theta_p - \delta_{ir}), \tag{14}$$

Taking it into account yields a *conditional* interpretation of parameters, one obtains

> θ_p is the person parameter and δ_{ir} the item parameter in a binary Rasch model that distinguishes between categories $r-1$ and r *given the response is in category* $r-1$ *or* r.

The real problem lies in the nature of the condition. The PCM assumes that the models hold *simultaneously* for all pairs of adjacent categories. Thus, when characterizing the probability of a response in category r one conditions on the *outcome* of Y_{pi}. The distribution of Y_{pi} as determined by (14) uses in the condition the outcome of the variable whose distribution is to be determined.

Let us assume that we have an achievement test with k ordered categories. Taking (14) seriously means, for example, that δ_{ir} is the difficulty that a person shows the highest possible performance (category k) given the person is a high performer, where high performer is defined by $Y_{pi} \in \{k-1, k\}$. In the same way, the model specifies the probability of observing category 1 given the person is a low performer, where low performer is defined by $Y_{pi} \in \{0, 1\}$. The interpretation of the difficulty parameter in the binary Rasch model implies that one already knows that the person is a poor, medium or strong performer. Somewhat loosely one can describe the threshold parameter as the difficulty that determines the probability of observing high categories given the location on the response categories is fixed.

Let us comment briefly on the step metaphor. Andrich (2015) gives various reasons why the step metaphor is incompatible with the PCM and writes that the model "does not characterize the process by which the person being assessed reaches a location on the continuum". One reason he gives is that the probability of a response in any category is a function of all the thresholds, which led to the characterization as a *divide by total model* (Thissen and Steinberg, 1986). The dependence of the probabilities on all the thresholds is obvious from denominator in (5) and has been illustrated in Figure 1. However, the reasoning is not compulsory if one allows for *conditional* steps. Andrich also outlined that in an achievement test with binary variables and response categories *incorrect* and *correct* one does not assume a transition or step from incorrect to correct. One only observes a correct or incorrect response. Thus there seems to be no reason why the step or transition concept should be useful in polytomous items with more than two levels of performance. Nevertheless, sometimes the response to items can be seen as *solving a problem* or *not solving a problem*. Then it is not totally inappropriate that solving a problem corresponds to a transition from the status unsolved to the status solved. In particular an item with no response can be seen as incorrect or can be seen as an unsolved problem, it simply remains unsolved. It takes a transition to obtain the status solved.

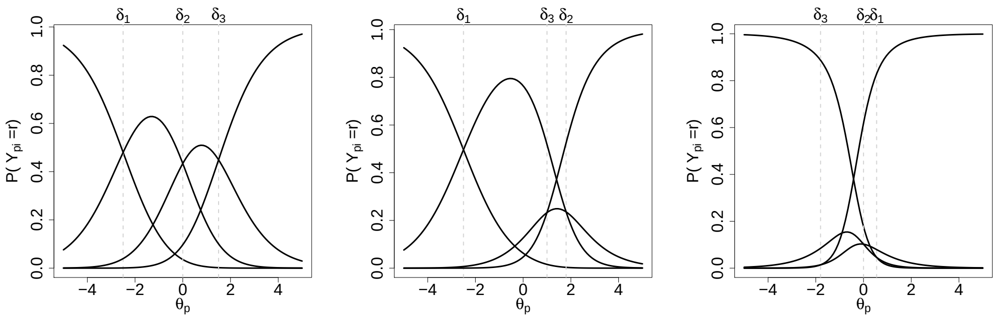

FIGURE 2: *Probability response curves for items with four categories; left: ordered item parameters, middle: $\delta_3 < \delta_2$, right: item parameters inversed, $\delta_3 < \delta_2 < \delta_1$*

3.3 The Ordering of Parameters

There has been some dispute on the question if item parameters in the PCM should be ordered or not, see, for example, Adams et al. (2012), Andrich (2013). Therefore, when investigating the properties of the PCM it seems warranted to consider the problem briefly.

Nothing in the construction of the model from binary variables calls for an ordering of categories. Since the variables (13) are conditional and the condition for each variable is different there is no need that item parameters are ordered. Nevertheless, Andrich (2015) argues that they should be ordered referring to an "intended order" and writes "that successive categories are intended to be ordered in the sense that they successively reflect more of the trait, the implication is that the thresholds which define the boundaries of the categories are not only expected, but *required* to increase in difficulty." (p. 11). The concept of intended order seems sensible but is also vague and hardly warrants the conclusion that thresholds are required to be ordered.

Figure 2 shows three items, one with ordered difficulties, $\delta_{i1} < \delta_{i2} < \delta_{i3}$, one where the order is reversed for two difficulties, $\delta_{i2} > \delta_{i3}$, and one with reversed order, $\delta_{i1} > \delta_{i2} > \delta_{i3}$. In the case of ordered difficulties for each category there are person parameters, for which the probability to observe the category is larger than for all other categories. If $\delta_{i2} > \delta_{i3}$ the probability of observing category 2 is strongly reduced. In an achievement test item with these thresholds a person that has at least performance level 2 has a very large probability to show performance level 3. It is important that the modes of the probability curves remain ordered, which might be interpreted in the sense that categories successively reflect more of the trait. In the last item with reversed order in all of the difficulties the dominating categories are category 0 and 3, which makes it an almost dichotomous item. Certainly as an item for ordered categories it is not the best choice since not much information is gained from the other categories. Nevertheless, probability curves are ordered and therefore the ordering of categories is exploited, although information gain could be better. Items with ordered thresholds are better at exploiting the information provided by ordered categories, however, that does not mean that thresholds have to be ordered to make it an item for ordered categories.

4 More on the Interpretation of Parameters: Comparing Models

In the derivation of latent trait models and the interpretation of parameters it is often helpful to specify a process that is behind the model. In general a process can be seen as the probabilistic mechanism that generates the response. In some cases it is an attempt to describe the mental process that determines an individual's behavior. However, it is usually unknown how a respondent eventually finds the category he/she prefers. Quite different processes might be behind the decision process. There have been several attempts to use binary models as elements in an underlying process. One example are item response trees, which more recently have been proposed for Likert items. They are consistent with a "sequential process, according to which a response is constructed based on a respondent's answers to a series of mental queries." (Böckenholt, 2017). The mental queries correspond to binary decisions, which sum up to a polytomous response model that are able to include response styles, for more on item response trees see also De Boeck and Partchev (2012), Böckenholt and Meiser (2017).

Although some caution is warranted the construction of the latent trait models from binary models considered here might also be seen as potential processes that describe how responses could have been generated. It seems appropriate to distinguish between measurement of attitudes or personalities, in which decisions are involved, and proficiency tests, which are linked to success rather than decisions.

Let us first consider rating scales that refer to *attitudes or personalities*. Both models considered here use the dichotomizations into groups of categories $\{0, \ldots, r-1\}$ and $\{r, \ldots, k\}$ represented by the split variables $Y_{pi}^{(r)}$.

In the *graded response model* one might assume that respondents evaluate all the possible dichotomizations, which are determined by binary Rasch models with the same person parameter θ_p. Moreover, the dichotomizations have to be compatible, that means, if a person decides that $Y_{pi} \geq 5$ he/she should also decide $Y_{pi} \geq 3$. More general,

> if $Y_{pi} \geq r$ is preferred over $Y_{pi} < r$ also $Y_{pi} \geq s$ should be preferred over $Y_{pi} < s$ for all $s < r$, or equivalently, if $Y_{pi}^{(r)} = 1$ then one has $Y_{pi}^{(s)} = 1$ for all $s < r$ (which is the Guttman property).

That means the person shows a specific sort of consistency when evaluating all the dichotomizations. The property warrants the conclusion that the process behind the graded response model is the one consistent with assessment in ordered categories. A quite different view was propagated by Andrich (2015) (p.13), who writes that the process behind the *adjacent categories models* is the one consistent with assessment in ordered categories.

In the *polytomous Rasch model* or PCM the Rasch model does not determine the dichotomizations into subgroups of categories. Instead pairs of categories are at the heart of the model, therefore possible process models should also consider pairs of categories. The conditional choice between adjacent categories comes into play after the person has narrowed down the decision to categories $r-1, r$. Thus, given $Y_{pi}^{(s)} = 1$ for $s = 1, \ldots, r-1$ and $Y_{pi}^{(s)} = 0$ for $s = r+1, \ldots, k$ the preference of r over $r-1$ is determined by the binary Rasch model with the same person parameter θ_p but thresholds that vary over the categories. The model also implies some consistency in the conditioning part. When choosing between categories $r-1$ and r it is assumed that

> $Y_{pi} \geq r$ is preferred over $Y_{pi} < r$ for all $s < r$ and $Y_{pi} \geq s$ is not preferred over $Y_{pi} < s$ for all $s > r$, or equivalently, one has $Y_{pi}^{(r)} = 1$ for all $s < r$ and $Y_{pi}^{(s)} = 0$ or all $s > r$.

Unfortunately, the described process model is complicated and not fully convincing. The main problem is that there is no conceptualization of the process that narrows down the choice to the alternatives $r-1$ and r. It certainly has to be assumed that this process depends on the trait, for example, if the person parameter is large one expects pairs of categories between which the choice is made to be close to k. Even if one postulates a process for the choice of adjacent categories it has no impact on the model structure since the model is already defined when the response probabilities given adjacent categories are fixed.

The search for a process is hindered by the structure of the model. As shown in previous sections the condition involves that the response or outcome is already partially fixed, that is, *given* $Y_{pi} \in \{r-1, r\}$. The conditional structure seems to make it impossible to find a simple process that generates the model.

In *proficiency tests* categories are not determined by decisions but by the performance in solving problems. Then, θ_p is the ability of a person and the item parameters are item difficulties or thresholds. The dichotomization into the groups of categories represented by the split variables $Y_{pi}^{(r)}$ can be seen as representing a performance level. Let $Y_{pi}^{(r)} = 1$ indicate that at least *performance level* r is obtained, or in dichotomizations, that $Y_{pi} \geq r$.

In the *graded response model* one postulates consistency of dichotomizations that refer to the performance, the performance levels are strictly ordered, that is,

> if a person has at least performance level r ($Y_{pi}^{(r)} = 1$) he/she also has at least performance level $s, s < r$.

It should be noted that no history is involved. The model does not assume that performance levels are reached successively as does the sequential model. Moreover, it is sufficient to postulate the consistency downwards, that is, nothing is assumed for the performance levels larger than r. In the Guttman structure $Y_{pi}^{(r)} = 1$ means only that the variables $Y_{pi}^{(s)} = 1, s < r$ are restricted, there is no restriction on the variables $Y_{pi}^{(s)} = 1, s > r$.

In the *polytomous Rasch model* or PCM the performance level r is modeled under the condition that

> a person has at least performance level $r-1$ ($Y_{pi}^{(r-1)} = 1$) and also $Y_{pi}^{(s)} = 1$ for $s < r-1$) but does not have performance level $r+1$ ($Y_{pi}^{(r+1)} = 0$ and also $Y_{pi}^{(s)} = 0$ for $s > r+1$).

Again the conditioning makes it a problematic process. Andrich (2015) explicitly considers the assessment in an item with 3 categories. When interpreting the parameters in the polytomous Rasch models he considers the dichotomous responses at the thresholds of an item with four categories and writes

"If the assessment ... is in category 2 (a score of 1), the implication is having exceeded threshold 1 but neither threshold 2 nor threshold 3; likewise, if an assessment is in category 3 (a score of 2), the implication is having exceeded both thresholds 1 and 2, but not threshold 3. And elegantly, if the response is in the first category (a score of 0), the number of thresholds exceeded is 0; if the response is in category 4 (score of 3), all three thresholds have been exceeded. This characterization reflects the severe constraint that order places on the latent, dichotomous, responses at the thresholds."

This can be seen as a fitting description of the *graded response model* referring to the thresholds in the Guttman variables $Y_{pi}^{(1)}, \dots, Y_{pi}^{(k)}$. It is indeed the case that in the graded response model the response is exactly the number of thresholds that are exceeded though it also the case in the polytomous Rasch model. It is also obvious that in the graded response model thresholds have to be ordered since otherwise the model would not be well defined. However, the description was meant to refer to the *polytomous Rasch model* and the latent Guttman variables. But the Guttman variables are observable and not latent. It can also not refer to the conditional variables $Y_{pi}^{(r)}|Y_{pi} \in \{r-1, r\}$ considered in Section 3. They are indeed latent but for these variables it does not make any sense to consider which thresholds have been exceeded given the assessment is in a specific category. Given the assessment is in category r one can infer that $Y_{pi}^{(r)}|Y_{pi} \in \{r-1, r\} = 1$ and $Y_{pi}^{(r+1)}|Y_{pi} \in \{r-1, r\} = 0$. But given the assessment is in category 5 it does not make sense to consider $Y_{pi}^{(3)}|Y_{pi} \in \{2, 3\}$ because the condition excludes that category 5 can be the resulting category. The severe constraint in Andrich's description refers to the Guttman structure, but it ignores the more important constraint that binary responses are conditional.

The description given by Andrich (2015) can be seen as a description of a possible process behind the graded response model but not behind the polytomous Rasch model. It might not have been intended as a description of a process model although the wording "having exceeded thresholds" hints at some sort of process. It also seems to imply some transition or step. The important point here is that it does not provide a process model for the PCM, which seems hard to obtain.

5 Concluding Remarks

It has been demonstrated that the interpretation of the parameters and the process behind the PCM and polytomous Rasch model are sometimes misleading. In the early times of the PCM it seems to have been confused with the sequential model. In the more recent literature there is a tendency to confuse it with the graded response model. However, the model is distinctly different from these two models and one should be careful when interpreting parameters.

When interpreted appropriately the PCM has various strengths. It allows for sufficient statistics of item and person parameters so that a conditional maximum likelihood procedure for the estimation of parameters is readily available (Andersen, 1977; Andrich, 2010). It exploits the ordering of categories as seen, for example, by the ordering of modes of probability curves. It does not impose an ordering of item parameters, which is an advantage over the graded response model because estimation problems are avoided and the model can be extended, for example to include response styles (Tutz et al., 2018) without having to account for clumsy restrictions. Moreover, as has been shown, the model can be seen as a general model that parameterizes pairs of categories by Rasch models. Here we do not investigate the implications of this property but it can be used, for example, to obtain extended models when the categories are divided into disagreement and agreement categories. One can parameterize pairs of categories within the groups of categories and link the categories by specifying a model for two categories, one from the disagreement group, one from the agreement group. The models within groups can be used to include response style parameters that reflect the tendency to middle and extreme categories.

Appendix: Proof of Proposition 3.1

We again considered the specific model that uses the pairs $(0,1),(0,2),\dots,(0,k)$ and postulate that $P(Y_{pi}=r|Y_{pi}\in\{0,r\})=F(r\tilde{\theta}_p-\tilde{\delta}_{ir})$ holds for $r=1,\dots,k$. For simplicity it is called the *category zero model*. Simple reparameterization shows that it is equivalent to the PCM.

In matrix form the model is given by $\boldsymbol{D}^{(0)}\log(\boldsymbol{\pi})=\boldsymbol{\theta}^{(0)}$; $\boldsymbol{D}^{(0)}$ is a $(k\times(k+1))$ matrix that can be partitioned into $\boldsymbol{D}^{(0)}=(-\mathbf{1}|\boldsymbol{I})$, where $\mathbf{1}^T=(1,\dots,1)$ is a vector of length k and $\boldsymbol{I}$ is a $(k\times k)$ unit matrix.

Let us now consider a general model $\boldsymbol{D}\log(\boldsymbol{\pi})=\boldsymbol{\theta}$, where $\boldsymbol{D}$ has k independent rows. Let $\tilde{\boldsymbol{D}}$ denote the $(k\times k)$-matrix obtained by deleting the first column in $\tilde{\boldsymbol{D}}$. Then the model can alternatively be given by $\tilde{\boldsymbol{D}}\log(\boldsymbol{\pi}^{(0)})=\boldsymbol{\theta}$, where $\boldsymbol{\pi}^{(0)}=(\pi_1/\pi_0,\dots,\pi_k/\pi_0)$, $\pi_r=P(Y_{pi}=r)$. This follows from the representation $\log(\pi_r)-\log(\pi_s)=\log(\pi_r/\pi_0)-\log(\pi_s/\pi_0)$. The vector $\boldsymbol{\pi}^{(0)}$ can also be used to define the category zero model for the pairs $(0,1),(0,2),\dots,(0,k)$ and be given in simpler form by $\log(\boldsymbol{\pi}^{(0)})=\boldsymbol{\theta}^{(0)}$.

Therefore, one has

$$\boldsymbol{\theta}=\boldsymbol{D}\log(\boldsymbol{\pi})=\tilde{\boldsymbol{D}}\log(\boldsymbol{\pi}^{(0)})=\tilde{\boldsymbol{D}}\boldsymbol{\theta}^{(0)},$$

which links the parameter of the general model to the parameters of the category zero model.

Let $\operatorname{span}\{\boldsymbol{r}_1,\dots,\boldsymbol{r}_k\}$ denote the space that is spanned by the row vectors $\boldsymbol{r}_1,\dots,\boldsymbol{r}_k$ of $\boldsymbol{D}$, and $\operatorname{span}\{\boldsymbol{r}_1^{(0)},\dots,\boldsymbol{r}_k^{(0)}\}$ denote the space that is spanned by the row vectors $\boldsymbol{r}_1^{(0)},\dots,\boldsymbol{r}_k^{(0)}$ of $\boldsymbol{D}^{(0)}$. Since each vector $\boldsymbol{r}_l$ is contained in $\operatorname{span}\{\boldsymbol{r}_1^{(0)},\dots,\boldsymbol{r}_k^{(0)}\}$, and both spaces have the dimension k the spaces are identical. Let us now consider the k-dimensional vectors $\bar{\boldsymbol{r}}_1,\dots,\bar{\boldsymbol{r}}_k$ that are obtained by deleting the first component of the vectors. In the same way $\bar{\boldsymbol{r}}_1^{(0)},\dots,\bar{\boldsymbol{r}}_k^{(0)}$ are obtained from $\boldsymbol{r}_1^{(0)},\dots,\boldsymbol{r}_k^{(0)}$ by deleting the first component. Since $\operatorname{span}\{\boldsymbol{r}_1,\dots,\boldsymbol{r}_k\}=\operatorname{span}\{\boldsymbol{r}_1^{(0)},\dots,\boldsymbol{r}_k^{(0)}\}$ also $\operatorname{span}\{\bar{\boldsymbol{r}}_1,\dots,\bar{\boldsymbol{r}}_k\}=\operatorname{span}\{\bar{\boldsymbol{r}}_1^{(0)},\dots,\bar{\boldsymbol{r}}_k^{(0)}\}$ holds. Since $\operatorname{span}\{\bar{\boldsymbol{r}}_1^{(0)},\dots,\bar{\boldsymbol{r}}_k^{(0)}\}=R^k$ and $\operatorname{span}\{\bar{\boldsymbol{r}}_1,\dots,\bar{\boldsymbol{r}}_k\}$ are the rows of $\tilde{\boldsymbol{D}}$ the $k\times k$ matrix $\tilde{\boldsymbol{D}}$ has full rank. Therefore, it is invertible and $\boldsymbol{\theta}^{(0)}=\tilde{\boldsymbol{D}}^{-1}\boldsymbol{\theta}$.

That means the parameters $\boldsymbol{\theta}^{(0)}$ and $\boldsymbol{\theta}$ can be transformed into each other and the general model and the category zero model are equivalent. Since the category zero model is equivalent to the PCM the same holds for the general model.

References

Adams, R. J., M. L. Wu, and M. Wilson (2012). The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement 72*(4), 547–573.

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*(4), 561–573.

Andrich, D. (2010). Sufficiency and conditional estimation of person parameters in the polytomous Rasch model. *Psychometrika 75*(2), 292–308.

Andrich, D. (2013). An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any 'threshold disorder controversy'. *Educational and Psychological Measurement 73*(1), 78–124.

Andrich, D. (2015). The problem with the step metaphor for polytomous models for ordinal assessments. *Educational Measurement: Issues and Practice 34*(2), 8–14.

Böckenholt, U. (2017). Measuring response styles in Likert items. *Psychological Methods* (22), 69–83.

Böckenholt, U. and T. Meiser (2017). Response style analysis with threshold and multi-process irt models: A review and tutorial. *British Journal of Mathematical and Statistical Psychology 70*(1), 159–181.

De Boeck, P. and I. Partchev (2012). Irtrees: Tree-based item response models of the glmm family. *Journal of Statistical Software 48*(1), 1–28.

Luo, G. (2005). The relationship between the rating scale and partial credit models and the implication of disordered thresholds of the Rasch models for polytomous responses. *Journal of Applied Measurement 6*(4), 443–455.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika 47*, 149–174.

Molenaar, I. (1983). Item steps (Heymans Bulletin 83-630-ex). *Groningen, The Netherlands: University of Groningen, Department of Statistics and Measurement Theory*.

Nering, M. L. and R. Ostini (2011). *Handbook of polytomous item response theory models*. Taylor & Francis.

Samejima, F. (1997). Graded response model. *Handbook of Modern Item Response Theory*, 85–100.

Samejima, F. (2016). Graded response model. In W. Van der Linden (Ed.), *Handbook of Item Response Theory*, pp. 95–108.

Thissen, D. and L. Steinberg (1986). A taxonomy of item response models. *Psychometrika 51*(4), 567–577.

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Statistical and Mathematical Psychology 43*, 39–55.

Tutz, G., G. Schauberger, and M. Berger (2018). Response styles in the partial credit model. *Applied Psychological Measurement 42*, 407–427.

Van der Linden, W. (2016). *Handbook of Item Response Theory*. Springer: New York.

Verhelst, N. D., C. Glas, and H. De Vries (1997). A steps model to analyze partial credit. In W. Van der Linden (Ed.), *Handbook of Modern Item Response Theory*, pp. 123–138. Springer.

von Davier, M. and C. H. Carstensen (2007). *Multivariate and mixture distribution Rasch models*. Springer Science+ Business Media, LLC New York, NY.

Wilson, M. (1992). The ordered artition model: An extension of the partial credit model. *Applied Psychological Measurement 16*(4), 309–325.

Wilson, M. and G. N. Masters (1993). The partial credit model and null categories. *Psychometrika 58*(1), 87–99.