JOURNAL OF APPLIED MEASUREMENT, *6*(4), 443–455

The Relationship between the Rating Scale and Partial Credit Models and the Implication of Disordered Thresholds of the Rasch Models for Polytomous Responses

Guanzhong Luo
Hong Kong Examinations and Assessment Authority

There is a perception in the literature that the Rating Scale Model (RSM) and Partial Credit Model (PCM) are two different types of Rasch models. This paper clarifies the relationship between the RSM and PCM from the perspectives of literature history and mathematical logic. It is shown that not only are the RSM and the PCM identical, but the two approaches used to introduce them are statistically equivalent. Then the implication of disordered thresholds is discussed. In addition, the difference between the structural thresholds and the Thurstone thresholds are clarified.

Requests for reprints should be sent to Guanzhong Luo, Hong Kong Examinations and Assessment Authority, 13/F., Southorn Centre, 133 Hennessy Road, Wan Chai, Hong Kong, e-mail: gluo@hkeaa.edu.hk.

Introduction

In the simple Rasch model, a dichotomous response X_{ni} is governed by only one item location parameter δ_i and one person location parameter β_n:

$$\Pr\{X_{ni} = 1 \mid \beta_n, \delta_i\} = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)}. \qquad (1)$$

Since late 1970s, an important development in educational measurement was the establishment and application of Rasch models for polytomous responses. Among the works frequently referenced in the literature on Rasch models, Andersen (1977), Andrich (1978), and Masters (1982) are widely considered as the most influential ones. In particular, Andrich (1978), which is based on the equidistant assumption in Andersen (1977), is often referred to as the Rating Scale Model (RSM) and Masters' (1982) expression of the Rasch model referred to as the Partial Credit Model (PCM).

Calling these two forms of Rasch models by different names reflects the perception in the literature that they are two different models. This perception seems to be getting more and more prevalent in recent publications. For example, there were two separate chapters in *the handbook of modern item response theory* (edited by van der Linden and Hambleton, 1997), one with the title *The rating scale model* and another with the title *The partial credit model*. Another example is the multi-volume *Encyclopedia of Social Measurement*, published by Academic Press in November 2004. In this book, the rating scale model and the partial credit model are two separate entries in the subject area of measurement models.

As explained in the rest of this paper, this perception was deeply rooted in the history of the development of the Rasch model for polytomous responses. First, it is possibly due to the complexity of the model introduced by Andersen (1977), in which the conceptualization about the rating model was first suggested in terms of scoring functions. Based on this assumption and a focus on the response process of a person to an item, Andrich (1978) used the *rating formulation* to introduce his expression of the Rasch model for polytomous responses in which the thresholds of an item were explicitly parameterised. Later, Masters (1982) introduced his PCM, in which the thresholds of an item were envisaged as *step difficulties*. In that paper, Masters also considered a special case of the PCM in which all items have the same set of step difficulties relative to the item scale values. He termed it a rating scale model. Wright and Masters (1982) had some more detailed discussion on the model and referred to Masters (1982) as in press. It is perhaps in Masters (1982) that first mentioned the term "rating scale model." However, it gave the impression that the approach in Andrich (1978) derived only this special case. This impression was further reinforced by a debate on whether items with disordered thresholds should be considered acceptable, or should be taken as a symptom of the ill-structuring of the data, or at least an opportunity to further examine the structure of the item.

Revisiting the development of the Rasch model for polytomous responses, this paper starts with a clarification on the definition of a rating scale model. It is then pointed out that the works by Andrich (1978) and Masters (1982) lead to an **identical** model from different approaches. Furthermore, this paper demonstrates that the two approaches, used by Andrich (1978) and Masters (1982) are necessary and sufficient for each other. That is, the two different approaches are mathematically equivalent. Therefore, it is suggested to simply call the model the Rasch model for polytomous responses, or in short the *Polytomous Rasch Model* (PRM).

The paper continues with a discussion of whether the order of the thresholds is an essential requirement of the PRM and whether an item with disordered thresholds is really problematic. It is argued that although the derivation of the PRM does not rely on the order of the thresholds, and an item with disordered thresholds does not violate the basic assumption for fundamental measurement, it causes a problem in using a restricted sample space to represent the original sample space. Some discussions on related issues are also provided. For the readers who are less interested in the mathematical derivations, the proofs of the theorems in this paper may be skipped.

What is the rating scale model?

This section traces the origin of the term "rating scale model." Before proceeding to the derivation of the PCM, Masters (1982) considered some observation formats. Under a subtitle *rating scales*, he wrote:

> A third type of ordered category data comes from rating scales in which a fixed set of ordered response alternatives is used with every item. When performances are being rated, judges may be provided with response alternatives like
>
> POOR FAIR GOOD
>
> Or, when self-reports on attitude are sought, alternatives like
>
> STRONGLY DISAGREE DISAGREE AGREE STRONGLY AGREE
>
> may be provided. The distinguishing feature of this third observation format is the use of a single set of response categories with every item in the scale. (p. 150)

Then after the PCM is introduced (see the later section in this paper for the expression of the PCM), Masters (1982) presented *rating scales* as a special case of the PCM:

> In a rating scale, however, the ordered response levels are not defined by a series of item subtasks, but by the fixed set of ordered rating points used with the items. As the same set of rating points is used with every item, the relative difficulties of the steps within each item should not vary greatly from item to item.
>
> This expectation can be incorporated into the Partial Credit Model by resolving each item step into two components so that
>
> $\delta_{ij} = \delta_i + \tau_j$
>
> where δ_i is the location or "scale value" of item *i* on the variable and τ_j is the location of the j^{th} step in each item relative to that item's scale value. Under this simplification, the only difference remaining among items is the difference in location on the variable. (pp. 162-163)

Further on, Masters and Wright (1997) considered the model in Andrich (1978) as a special case of the PCM:

> Andrich's (1978) model for rating scales, for example, can be thought of as a version of the PCM with the added expectation that the response categories are defined and function in the same way for each item in an instrument. With this added function, rather than modeling a set of m_i parameters for each item, a single parameter δ_i is modelled for item *i*, and a set of *m* parameters $(\tau_1, \tau_2, \ldots \tau_m)$ are proposed for the common response categories. To obtain the rating scale version of the PCM, each item parameter in the model is redefined as $\delta_{ix} = \delta_i + \tau_x$. (p. 101)

However, a search of the literature finds discrepancies on the definition of the rating scale model. For example, in an article titled "*The rating scale model,*" Andersen (1997) explained the concept of the rating scale model very differently:

> The rating scale model is a latent structure model for polytomous responses to a set of items.... It is called a rating scale model because the response categories are scored such that the total score for all items constitute a rating of the respondents on a latent scale. It is assumed that the category scores are equally spaced, giving the highest score to the first or the last category. Thus the phrasing of the response categories must reflect a scaling of the responses, e. g., "very good," "good," "not so good," "bad." (p. 67)

In that paper, Andersen re-introduced his expression of the Rasch model for polytomous responses in this way:

Under the rating scale model, the response functions have the form

$$P_{ih}(\theta) = \frac{e^{w_h - a_{ih}}}{\sum_{h=1}^{m} e^{w_h - a_{ih}}}.$$

In this expression $w_1,\ldots,w_m$ are the category scores, which prescribe how the m response categories are scored, while a_{ih} are item parameters connected with the items and categories. (p. 68)

Furthermore, Andersen (1997) wrote:

The rating scale model is based on the assumption that the category scores $w_1,\ldots,w_h$ are equidistant, i.e., the differences

$d_h = w_h - w_{h-1}$, $h = 2,\ldots, m$,

are all equal. This assumption was first suggested in Andersen (1977) based on certain properties connected with the sufficient statistic for θ, where θ is regarded as an unknown parameter to be estimated. When Andrich (1978) introduced the rating scale models, he showed that $w_h = h$ could be interpreted as the numbers of the categories in which the response occurs, or as thresholds being exceeded. (p. 69)

In the quotation above, θ is the person parameter and is denoted as β_n in the current paper. We can see that Andersen's original assumption about a rating scale is that the category scores within an item are equally spaced rather than that relative difficulties of "steps" to the item scale values are invariant across items (Masters, 1982). It also should be highlighted here that the category scores are not item parameters. Instead, they are the coefficients in the model that have to be specified rather than estimated. The basic assumption for the rating scale models above is also termed "equidistant scoring" (Andersen, 1977). As can be seen in the next section, this specification does not require that the item thresholds be equally spaced, nor the relative locations of the thresholds be invariant across the items involved.

Andrich's approach: The rating formulation, thresholds and the rating model

Andrich (1978) constructed the Rasch model for ordered categories by introducing a series of dichotomous Rasch response variables $(Z_1,Z_2,\ldots,Z_m)$ with locations $\{\delta_1,\delta_2,\ldots,\delta_m\}$. These dichotomous response variables are termed latent variables as they are not direct observed responses. The approach used in Andrich (1978) is termed the *rating formulation*.

Before introducing the rating formulation, Andrich (1978) summarised the work of Rasch (1968) and Andersen (1977) for the case of three categories and made the expression in Andersen (1977) more explicit:

In a specification … where θ and σ are considered functions of uni-dimensional parameters β and δ respectively, and still with the requirements of sufficient statistics for the subject and item parameters, the response model can be expressed as

$$p\{X = 0 \mid \beta,\delta,\kappa,\phi\} = \frac{1}{\gamma},$$

$$p\{X = x \mid \beta,\delta,\kappa,\phi\} = \frac{\exp[\kappa_x + \phi_x(\beta-\delta)]}{\gamma}, \quad x = 1,\ 2,$$

where

$$\gamma = \gamma(\beta,\delta,\kappa,\phi) = 1 + \sum_{k=1}^{2} \exp[\kappa_k + \phi_k(\beta-\delta)].$$

(p. 564)

The aims of Andrich (1978) are twofold. One is to specify the coefficients $\{\phi_k\}$, and another to parameterise the parameters $\{\kappa_k\}$ in terms of the locations of the latent variables (Z_1,Z_2). The *rating formulation* is summarised as follows.

When (Z_1,Z_2) are independent, the entire sample space Ω includes $2^2 = 4$ response patterns $\{(0,0), (1,0), (0,1), (1,1)\}$. Let Ω' be the collection of the Guttman response patterns in Ω: $\Omega' = \{(0,0), (1,0), (1,1)\}$; and define

$$p_1 = \Pr\{Z_1 = 1\};\ q_1 = 1 - p_1 = \Pr\{Z_1 = 0\};$$
$$p_2 = \Pr\{Z_2 = 1\};\ q_2 = 1 - p_2 = \Pr\{Z_2 = 0\}. \quad (2)$$

It is evident that

$$\Pr\{\Omega'\} = \Pr\{Z_1 = 0\}\Pr\{Z_2 = 0\} + \Pr\{Z_1 = 1\}\Pr\{Z_2 = 0\} + \Pr\{Z_1 = 1\}\Pr\{Z_2 = 1\} = q_1 q_2 + p_1 q_2 + p_1 p_2.$$

For a polytomous response variable $X{:}x \in \{(0,1,2\}$, the one-to-one correspondence between the observations of X and the elements of Ω' was defined as

$$X = 0 \quad \Leftrightarrow (0,\ 0),$$
$$X = 1 \quad \Leftrightarrow (1,\ 0),$$
$$X = 2 \quad \Leftrightarrow (1,\ 1\ ). \quad (3)$$

Then the probability $\Pr\{X = k\}$ is defined as the conditional probability of the corresponding pattern on the constrained space Ω', $k = 0,1,2$. That is,

$$\begin{aligned}
&\Pr\{X = 0\} = \Pr\{Z_1 = 0\}\Pr\{Z_2 = 0\}/\Pr\{\Omega'\} \\
&= q_1 q_2 /(q_1 q_2 + p_1 q_2 + p_1 p_2); \\
&\Pr\{X = 1\} = \Pr\{Z_1 = 1\}\Pr\{Z_2 = 0\}/\Pr\{\Omega'\} \\
&= p_1 q_2 /(q_1 q_2 + p_1 q_2 + p_1 p_2); \\
&\Pr\{X = 2\} = \Pr\{Z_1 = 1\}\Pr\{Z_2 = 1\}/\Pr\{\Omega'\} \\
&= p_1 p_2 /(q_1 q_2 + p_1 q_2 + p_1 p_2).
\end{aligned} \quad (4)$$

Or in general,

$$\Pr\{X = x\} = \frac{\prod_{k=1}^{x} p_k \prod_{k=x+1}^{2} q_k}{\Pr\{\Omega'\}} = \frac{\prod_{k=1}^{x} p_k \prod_{k=x+1}^{2} q_k}{\sum_{l=0}^{2}(\prod_{k=1}^{l} p_k \prod_{k=l+1}^{2} q_k)};\quad x = 0,\ 1,\ 2; \quad (5)$$

where for notational convenience,

$$\prod_{k=1}^{0} p_k = 1.$$

When the dichotomous response variables (Z_1,Z_2) follow the simple Rasch model of Equation 1 with locations $\{\delta_1,\delta_2\}$, that is,

$$p_1 = \Pr\{Z_1 = 1 \mid \beta_n, \delta_i\} = \frac{\exp(\beta_n - \delta_1)}{1 + \exp(\beta_n - \delta_1)};$$
$$p_2 = \Pr\{Z_2 = 1 \mid \beta_n, \delta_i\} = \frac{\exp(\beta_n - \delta_2)}{1 + \exp(\beta_n - \delta_2)}; \quad (6)$$

the polytomous variable X has the probabilities for the legitimate responses as:

$$\Pr\{X = x \mid \beta_n, (\delta_k)\} = \frac{\exp\{\sum_{k=0}^{x}(\beta_n - \delta_k)\}}{\sum_{l=0}^{2}\exp\{\sum_{k=0}^{l}(\beta_n - \delta_k)\}},\ x = 0,1,2; \quad (7)$$

where for notational convenience,

$$\sum_{k=0}^{0}(\beta_n - \delta_k) \equiv 0$$

for any β. Equation 7 can be further simplified by parameterising the mean of $\{\delta_1,\delta_2\}$ as the *location* of the polytomous variable X:

$$\delta = \frac{1}{2}(\delta_1 + \delta_2); \quad (8)$$

and the deviations $\{\tau_k = \delta_k - \delta, k = 1,...,2\}$ as the *centralised* thresholds of the polytomous response model. The locations (δ_1,δ_2) of the dichotomous variables (Z_1,Z_2) are termed *uncentralised* thresholds. Then the model has the form

$$\Pr\{X = x \mid \beta_n, \delta, (\tau_k)\} = \frac{\exp\{x(\beta_n - \delta) - \sum_{k=0}^{x}\tau_k\}}{\sum_{l=0}^{2}\exp\{l(\beta_n - \delta) - \sum_{k=0}^{l}\tau_k\}};$$
$$x = 0,1,2; \quad (9)$$

where for notational convenience, $\tau_0 \equiv 0$.

Figure 1 shows the probabilistic functions of Equation 9 and those of the corresponding dichotomous Rasch variables (Z_1,Z_2). It can be seen graphically that the crossing point of the probabilistic functions for adjacent categories are the locations of the corresponding dichotomous variable.

In the derivation above, X is a response variable on a particular item. In a general situation, to identify the response X with respect to item i with the maximum score of m_i $(i = 1,\ldots, I)$, the model of Equation 5 can be written as

$$\Pr\{X_{ni} = x\} = \frac{(\prod_{k=1}^{x} p_k)(\prod_{k=x+1}^{m} q_k)}{\Pr\{\Omega'\}}$$
$$= \frac{\exp[\sum_{k=0}^{x}(\beta_n - \delta_{ik})]}{\sum_{l=0}^{m_i}\exp[\sum_{k=0}^{l}(\beta_n - \delta_{ik})]}$$
$$= \frac{\exp\{x(\beta_n - \delta) - \sum_{k=0}^{x}\tau_k\}}{\sum_{l=0}^{m}\exp\{l(\beta_n - \delta) - \sum_{k=0}^{l}\tau_k\}},$$
$$x = 0,1,...,m_i; \qquad (10)$$

where for notational convenience,

$$\sum_{k=0}^{0}(\beta_n - \delta_{ik}) \equiv 0$$

for any β_n. $\{\tau_k\}$ are now termed *centralized thresholds* and

$$\sum_{k=0}^{0}\tau_k = 0. \qquad (11)$$

Historically, in Andrich (1978), as the focus was on the response process of a person to an item, the thresholds of an item were generally denoted as $\{\tau_k; k = 1,\ldots, m\}$ without the subscripts to the item. However, as explained above, the rating formulation does not require that the $\{\tau_k\}$ be the same for all items. Andrich (1982) gives a more comprehensive clarification on the structure of the model and claimed "The Rasch model was distinguished from the traditional one by being termed a *rating model.*" Furthermore, Andrich (1985) provides a detailed elaboration on this model. To emphasize that the centralized thresholds may be different for different items, they are denoted there as $\{\tau_{ki}; k = 1,\ldots, m_i; i = 1,\ldots,1\}$. Luo (2001) revisits the rating formulation and extends it into a general context. In fact, a general form of unfolding models, which follows a single-peaked response process rather than the cumulative response process that the Rasch model follows, is constructed using the principle of rating formulation in Luo (2001).

Masters' approach: The step difficulty and the partial credit model

Masters (1982) takes partial credit as an observation format that is different from, but more general than, the rating scale form:

> The fourth general type of data comes from a observation format which requires the prior identification of several ordered levels of performance on each item and thereby awards partial credit for partial success on items. …Under

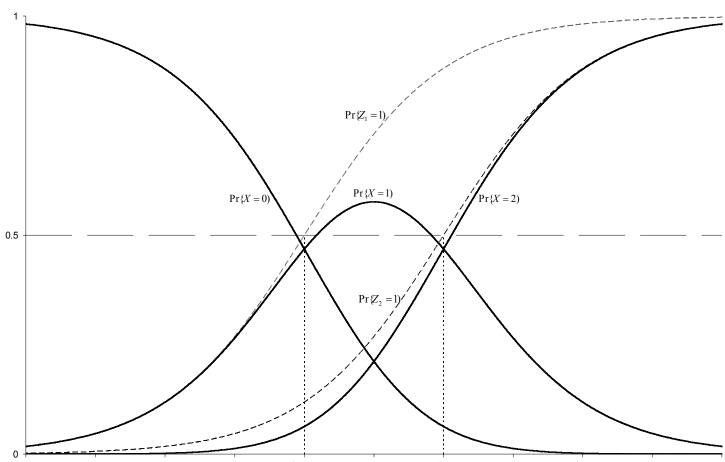

Figure 1. Probabilistic functions of the Rasch model for three categories and the corresponding dichotomous Rasch variables.

> this format the number of steps into which an item is divided and the relative difficulties of these steps can vary from item to item. …The model developed in this paper for the analysis of partial credit data is an extension of Andrich's rating scale model to situations in which response alternatives are free to vary in number and structure from item to item. (p. 150)

Before proceeding to the mathematical derivation of the model, Masters (1982) introduces the concept of steps:

> When performances on an item are recorded in the $m + 1$ ordered levels 0, 1,…, m, it is convenient to think in terms of m "steps" which have to be taken to complete the item. (p. 155)

To argue that the step interpretation also can be applied to a rating scale, Masters (1982) wrote:

> For an item on an attitude questionnaire, "completing the k^{th} response alternative over the $(k–1)^{\text{th}}$ in response to the item. Thus a person who chose to AGREE with a statement on an attitude questionnaire … can be considered to have chosen DISAGREE over STRONGLY DISAGREE (first step taken) and also AGREE over DISAGREE (second step taken), but to have failed to chose STRONGLY AGREE over AGREE (third step rejected). (p. 156)

After giving an example of three step mathematics item, he further stresses the importance of the order of the steps:

> These three steps must be taken in order. The second step in the item can be taken only if the first has been completed, and the third step can be taken only if the first two steps have been completed. (p. 155)

Then the difficulty of a step is defined as a conditional probability on the adjacent response levels:

> The third step in item i, for example, is from level 2 to level 3. The difficulty of this third step governs how likely it is that a person who has already reached level 2 will complete the third step to level 3. Another way of saying is that the difficulty of the third step governs how likely it is that a person will make a 3 rather than a 2 on the item. A simple model for this third step in item i is
>
> $$\Phi_{3ni} = \frac{\pi_{3ni}}{\pi_{2ni} + \pi_{3ni}} = \frac{\exp(\beta_n - \delta_{i3})}{1 + \exp(\beta_n - \delta_{i3})}.$$
>
> In this expression $\pi_{2ni}+\pi_{3ni}$ is person n's probability of scoring either 2 or 3 on item i, and $\pi_{3ni}/(\pi_{2ni}+\pi_{3ni})$ is the probability of person n completing the third step in item i to score 3 rather than 2 . …now δ_{i3} is defined as the difficulty of the third "step" in item i. (pp. 157-158)

In general, for person n and item i with the maximum score of m, the conditional probability for any pair of adjacent response categories is required to take the form of the dichotomous Rasch model. That is,

$$\Phi_{kni} = \frac{\Pr\{X_{ni} = K\}}{\Pr\{X_{ni} = K - 1\} + \Pr\{X_{ni} = K\}} = \frac{\exp\{\beta_n - \delta_{ik}\}}{1 + \exp\{\beta_n - \delta_{ik}\}}, \quad k = 1, ..., m_i, \tag{12}$$

with the required constraint

$$\sum_{k=0}^{m} \Pr\{X_{ni} = k\} = 1. \tag{13}$$

Equation 12 for $k = 1,...,m_i$ are solved to obtain the probabilities of the polytomous response variable X_{ni}:

$$\Pr\{X_{ni} = x \mid \beta_n, \delta_{ki}\} = \frac{\exp\{\sum_{k=0}^{x}(\beta_n - \delta_{ki})\}}{\sum_{l=0}^{m_i}\exp\{\sum_{k=0}^{l}(\beta_n - \delta_{ki})\}}, \quad x = 1, ..., m_i. \tag{14}$$

Now we can see the model above is **identical** to that of Equation 7. The only difference is that the item parameters in Equation 7 are not given subscripts with respect to the item number, as the focus in Andrich (1978) is on the response process of a person to an item. This difference is considered only a difference in the use of the notations. The step difficulties $\{\delta_{ki}\}$ are in fact the uncentralized thresholds in Equation 7.

The equivalence of the approaches by Andrich and Masters

In the discussions above, it is seen that Andrich (1978) and Masters (1982) lead to the identical model from different approaches. This is not coincidental. Luo (2001) demonstrates that these two approaches are in fact equivalent in an even more general context regardless of the response process employed. In Luo (2001), the main result on the equivalence of these two approaches is summarised in the following theorem.

Theorem 1. Let $Z = (Z_1, Z_2, \ldots, Z_m)$ be a set of dichotomous response variables,

$$\begin{aligned} &p_k \equiv \Pr\{Z_k = 1\}, \\ &q_k \equiv 1 - p_k = \Pr\{Z_k = 0\}, \\ &k = 1,\ \ldots,\ m. \end{aligned} \tag{15}$$

X is a polytomous response variable with probabilities $\Pr\{X = k\}$, $k = 1,\ldots,m$. Then the sufficient and necessary condition for

$$\begin{aligned} &\Pr\{X = k\} = \frac{(\prod_{l=1}^{k} p_l)(\prod_{l=k+1}^{m} q_l)}{\gamma}; \\ &\gamma = \sum_{j=0}^{m} (\prod_{l=1}^{j} p_l)(\prod_{l=j+1}^{m} q_l); \\ &k = 0,\ldots,m; \end{aligned} \tag{16}$$

is

$$\begin{aligned} &p_k = \frac{\Pr\{X = k\}}{\Pr\{X = k-1\} + \Pr\{X = k\}},\quad k = 1,\ \ldots, m; \\ &\sum_{k=0}^{m} \Pr\{X = k\} = 1. \end{aligned} \tag{17}$$

Proof:

Necessity. From (16), it is evident that

$$\sum_{k=0}^{m} \Pr\{X = k\} = 1.$$

For any k, $0 < k \le m$, and direct substitution leads to

$$\begin{aligned} &\frac{\Pr\{X = k\}}{\Pr\{X = k-1\} + \Pr\{X = k\}} \\ &= \frac{(\prod_{l=1}^{k} p_l)(\prod_{l=k+1}^{m} q_l)}{\gamma} \Big/ \frac{(\prod_{l=1}^{k-1} p_l)(\prod_{l=k}^{m} q_l) + (\prod_{l=1}^{k} p_l)(\prod_{l=k+1}^{m} q_l)}{\gamma} \\ &\frac{(\prod_{l=1}^{k} p_l)(\prod_{l=k+1}^{m} q_l)}{(\prod_{l=1}^{k-1} p_l)(\prod_{l=k}^{m} q_l) + (\prod_{l=1}^{k} p_l)(\prod_{l=k+1}^{m} q_l)} \\ &= \frac{p_k}{q_k + p_k} = p_k. \end{aligned} \tag{18}$$

Sufficiency. From (17), for any k, $0 < k \le m$

$$\begin{aligned} &\prod_{l=1}^{k} p_l \prod_{l=k+1}^{m} q_l \\ &= (\prod_{l=1}^{k} \frac{\Pr\{X = l\}}{\Pr\{X = l-1\} + \Pr\{X = l\}})(\prod_{l=k+1}^{m} \frac{\Pr\{X = l-1\}}{\Pr\{X = l-1\} + \Pr\{X = l\}}) \\ &= \frac{(\prod_{l=1}^{k} \Pr\{X = l\})(\prod_{l=k+1}^{m} \Pr\{X = l-1\})}{\prod_{l=1}^{m} [\Pr\{X = l-1\} + \Pr\{X = l\}]} \\ &= \Pr\{X = k\} \cdot \frac{\prod_{l=1}^{m-1} \Pr\{X = l\}}{\prod_{l=1}^{m} [\Pr\{X = l-1\} + \Pr\{X = l\}]}. \end{aligned}$$

That is

$$\Pr\{X = k\} = \prod_{l=1}^{k} p_l \prod_{l=k+1}^{m} q_l \Bigg/ \frac{\prod_{l=1}^{m-1} \Pr\{X = l\}}{\prod_{l=1}^{m} [\Pr\{X = l-1\} + \Pr\{X = l\}]}. \tag{19}$$

Because

$$\sum_{k=0}^{m} \Pr\{X = k\} = 1,$$

by the definition of γ as a normalisation factor, it follows that

$$\frac{\prod_{l=1}^{m-1}\Pr\{X=l\}}{\prod_{l=1}^{m}[\Pr\{X=l-1\}+\Pr\{X=l\}]}=\gamma$$
$$=\sum_{j=0}^{m}(\prod_{l=1}^{j}p_l)(\prod_{l=j+1}^{m}q_l). \quad (20)$$

Thus, it is not surprising that the mathematical equivalence of the two approaches used in Andrich (1978) and Masters (1982) lead to the identical model. In the rest of this paper, therefore, the model is simply called the Rasch model for polytomous responses, or in short the *Polytomous Rasch model* (PRM).

However, it is noted that although the approaches of Andrich and Masters are equivalent, the explicit focus of each is different. Andrich's approach focuses on the overall structure of the latent dichotomous responses while Masters' approach focuses on the pairs of adjacent response categories. This difference in focus leads to different understandings of the implication of the order of the thresholds, as clarified in the next section.

Importance of the order of the thresholds

In the discussions so far, the order of the dichotomous response variables $(Z_1,Z_2,\ldots,Z_m)$ in terms of their location values is not explicitly imposed for the development of the PRM. Andrich (1996) explained this phenomenon as follows:

> …it turns out that there is nothing in the structure of the parameters or in the way the summary statistics appear in any of the estimation equations (irrespective of method of estimation) that constrains the thresholds estimates to be ordered… A common reaction to this evidence that the model permits the **estimates** to be disordered is that the disordered estimates should simply be accepted—either there is something wrong with the model or the parameters should be interpreted as showing something other than the data fail to satisfy the ordering requirement. This perspective is supported by the feature that the usual statistical tests of fit based on the chi-square distribution can be constructed and that the data can be shown to fit the model according to such a test of fit, which operate with different powers, are **necessary and sufficient** to conclude that the data fit the model. (p. 20)

However, in Andrich (1978) and his follow-up works, the ordering of the dichotomous variables is considered the central requirement in the structure of the PRM. The following two facts contribute to the main rationale for this requirement.

First, recall the derivation of the PRM with the rating formulation: For a given order of the dichotomous latent variables $(Z_1,Z_2,\ldots,Z_m)$, the collection of Guttman type responses Ω' is used for mapping the polytomouse responses. It can be seen that the difference of the probabilities $\Pr\{\Omega\} - \Pr\{\Omega'\}$ has the minimum value when $(Z_1,Z_2,\ldots,Z_m)$ are in the order of their locations:

$$\delta_1 < \delta_2 < \ldots < \delta_m. \quad (21)$$

For a simple situation when $m = 2$, as discussed earlier in this paper, $\Omega = \{(0,0), (1,0), (0,1), (1,1)\}$, $\Omega' = \{(0,0), (1,0), (1,1)\}$. In this case,

$$\Pr\{\Omega'\} = \Pr\{(0,0)\} + \Pr\{(1,0)\} + \Pr\{(1,1)\} \quad (22)$$

When $\delta_1 < \delta_2$, for any person location β, we have $\exp(\beta - \delta_1) > \exp(\beta - \delta_2)$. As a result,

$$\begin{aligned}
&\Pr\{(1,0)\} = \Pr\{Z_1 = 1\}\Pr\{Z_2 = 0\} \\
&= \frac{\exp(\beta-\delta_1)}{1+\exp(\beta-\delta_1)}\frac{1}{1+\exp(\beta-\delta_2)} \\
&> \frac{\exp(\beta-\delta_2)}{1+\exp(\beta-\delta_2)}\frac{1}{1+\exp(\beta-\delta_1)} \\
&= \Pr\{Z_1 = 0\}\Pr\{Z_2 = 1\} = \Pr\{(0,1)\}.
\end{aligned} \tag{23}$$

That is, (1,0) has the greatest probability among the response patterns with a total score of 1. On the other hand, if (Z_1,Z_2) is in the reversed order of their locations: $\delta_1 > \delta_2$, then the mapping $X = 1 \Leftrightarrow (1,0)$ does not allow the response pattern for the greatest probability. Similar situations can be observed for $m > 2$. In general, since the essence of the rating formulation is to substitute Ω with the constrained subspace Ω', the property of (21) ensures that Ω' is close to Ω as much as possible. In this situation, the structure of Ω is mostly allowed for by Ω'. Second, when the thresholds are ordered in their values, then for any $0 < k < m$, in the intervals $[\delta_{ki},\delta_{(k+1),i}]$, the probability $\Pr\{X = k\}$ has the greatest value among all categories. That is, within this interval, $\{X = k\}$ is the most likely.

Therefore, when a polytomous response is mapped into a set of dichotomous responses, the relationship between these dichotomous responses cannot be arbitrary. Though the ordering requirement is not the mathematical property of the probabilistic function of the PRM, this requirement ensures a proper structure of the measurement.

Nevertheless, the ordering requirement is not always considered as necessary by all researchers and practitioners. In particular, Masters (1988) argues that the thresholds can have any order:

> There is no connection between the order of the logically defined item parameters and the intended order of the response categories. There may sometimes be reasons for desiring particular patterns of item parameters in the Rasch PCM, but there is nothing in the logic of their definition or in the requirements of objective measurement that would lead us to expect or require these locally defined item parameters to have any particular order. (p. 287)

To respond to Masters' view above, Andrich (1988) argues that the locations of the thresholds are in fact all related through the denominator in Equation 14 which contains all thresholds of the item. Therefore, "*the division of the whole into its constituent parts cannot be arbitrary when the objectivity is required*" (Jansen and Roskam, 1984). In practice, the order of the thresholds estimated is the property of the data, and the appearance of disordered thresholds in the estimation should be taken as diagnostic opportunities (Andrich, de Jong, and Sheridan, 1997):

> Because the very construction of the model requires an ordering of thresholds, it is argues here that whenever the threshold estimates are reversed, it provides evidence that the ordering is not operating as intended. In addition, it is emphasized that a perspective that does not take the reversal of threshold estimates to be a real problem of ordering, would provide no basis for examining the statements in the first place with a view to explaining and understanding in incorrect ordering. (p. 68)

Masters (1988) also pointed out that when the item thresholds are not in strictly increasing order, the "regions of most probable response" cannot be identified. Instead of looking into the situation more closely, however, Masters (1988) and Adams (1988) advocated an alternative procedure, termed *zone maps* or *response spaces*, to describe the performance of the items with disordered thresholds. This procedure is summarised as follows.

With the PRM of Equation 14, for any $k = 0, 1, \ldots, m_i - 1$, consider the sum of probabilities for the first $k + 1$ categories:

$$P_k = \sum_{x=0}^{k} \Pr\{X_{ni} = x \mid \beta_n, (\delta_{ki})\}$$
$$= \sum_{x=0}^{k} \frac{\exp\{\sum_{k=0}^{x} (\beta_n - \delta_{ki})\}}{\sum_{l=0}^{m_i} \exp\{\sum_{k=0}^{l} (\beta_n - \delta_{ki})\}}. \qquad (24)$$

It is evident that $0 \le P_k \le 1$ for any k. And more significantly, for any $k < t < m_i - 1$,

$$P_k < P_t. \qquad (25)$$

Figure 2 draws the values of $\{P_k\}$ where $m_i = 3$ and the item thresholds are –0.94, 0.59 and –0.07 respectively as specified in Adams (1988). In Figure 2, rather than being a vertical axis as in Adams (1988) and Masters (1988), the latent trait continuum is drawn on the horizontal axis so that it is easily compared with the probabilistic curves of individual categories, which are shown in Figure 3.

The k^{th} Thurstone thresholds γ_k are defined as the points on the latent continuum on which $P_{k-1} = 0.5$, $0 < k \le m_i$. It is evident that for any $0 < k < m_i$, $\gamma_k < \gamma_{k-1}$. However, the concern on the use of the Thurstone thresholds is that the Thurstone

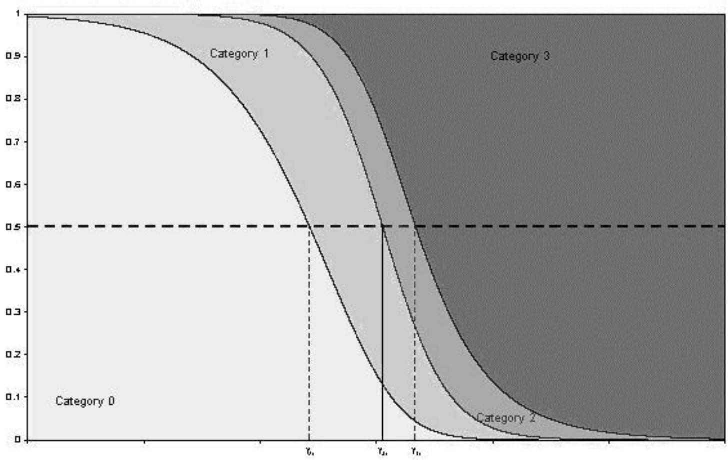

Figure 2. An item response map of the Rasch model for four categories with disordered thresholds.

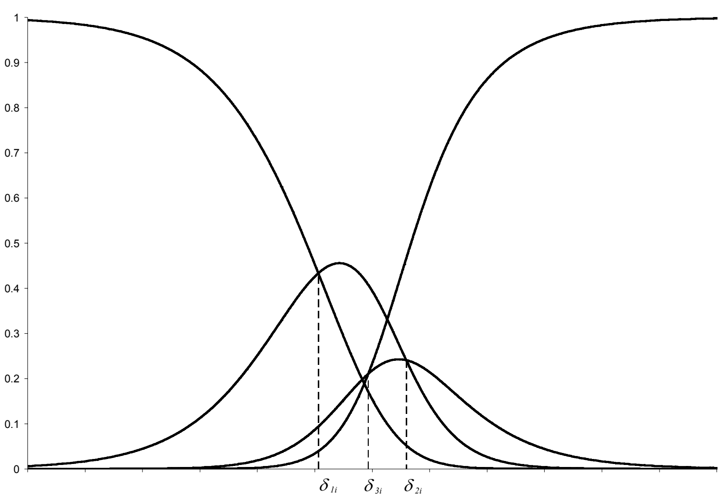

Figure 3. Probabilistic functions of the Rasch model for four categories with disordered thresholds.

thresholds are not structural parameters of the Rasch model in forms of various expressions discussed in this paper. It seems very hard, if not impossible, to re-parameterize the PRM into a form in which the Thurstone thresholds are explicitly (structurally) involved.

Summary and Discussions

Though the focus in Andrich (1978) is on the person-item response process, the principle of the rating formulation of Andrich (1978) is valid for the case that different items have different thresholds. In fact, the model derived with the rating formulation is mathematically identical to that derived in Masters (1982), except for the difference in the notation systems used in their works. Furthermore, the two approaches used to introduce them are statistically equivalent. Therefore, to keep calling an identical model derived with two different names according to which work is referred to reflects a misunderstanding of the model, or a misunderstanding of the approaches used to derive the model. It is suggested in this paper that the model be generally termed the Polytomous Rasch Model (PRM).

For those who are convinced that items with disordered thresholds estimated should be examined within the framework of the PRM, the two questions to be answered are a) what are the causes of the disordered thresholds and b) how can the performance of the item with disordered thresholds can be improved.

References

Adams, R. J. (1988). Applying the partial credit model to educational diagnosis. *Applied Measurement in Education, 4,* 347-361.

Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika, 42,* 69-81.

Andersen, E. B. (1997). The rating scale model. In W. J. van der Linden, and R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 67-84). New York: Springer.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561-574.

Andrich, D. (1982). An extension of the Rasch model for ratings providing both location and dispersion parameters. *Psychometrika*, *47,* 105-113.

Andrich, D. (l985). An elaboration of Guttman scaling with Rasch models for measurement. In N. Brandon-Tuma (Ed.), *Sociological methodology* (pp. 33-80). San Francisco: Jossey-Bass.

Andrich, D. (1988). A general form of Rasch's extended logistic model for partial credit scoring. *Applied Measurement in Education, 4,* 363-378.

Andrich, D. (1996). Measurement criteria for choosing among models for graded responses. In A. von Eye and C.C. Clogg (Eds.), *Analysis of categorical variables in developmental research* (pp. 3-35). Orlando, FL: Academic Press.

Andrich, D., de Jong, J., and Sheridan, B. S. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. In J. Rost and R. Langeheine (Eds.), *Application of latent trait and latent class models in the social sciences* (pp. 58-68). Munster, Germany: Waxmann.

Jansen, P. G. W., and Roskam, E. E. (1984). The polytomous Rasch model and dichotomization of graded responses. In E. Degreef and J. van Buggenhaut (Eds.), *Trends in mathematical psychology* (pp. 413-430). Amsterdam: North Holland.

Luo, G. (2001). A class of probabilistic unfolding models for polytomous responses. *Journal of Mathematical Psychology, 45*, 224-248.

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

Masters, G. N. (1988).The analysis of partial credit scoring. *Applied Measurement in Education, 4,* 279-298.

Masters, G. N., and Wright, B. D. (1997). The partial credit model. In W. J. van der Linden and R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 101-121). New York: Springer.

Rasch, G. (1968). *A mathematical theory of objectivity and its consequences for model construction*. Paper presented at the European Meeting on Statistics, Econometrics and Management Sciences. Amsterdam.

van der Linden, W. J., and Hambleton, R. K. (Eds.) (1997). *Handbook of modern item response theory.* New York: Springer.

Wright, B. D., and Masters, G. N. (1982). *Rating scale analysis.* Chicago: MESA Press.