

Essay 1

Rasch's incomplete form of the polytomous model

David Andrich (The University of Western Australia)

The thresholds form of the polytomous Rasch model for ordered categories continues to be debated for various reasons. It seems to have been debated for 40 years since its publication in 1978. This is a series of essays that describe the development of this form of the model focussing on comments excerpted from various pieces of correspondence or commentary. However, although specific wording may be specific to a person, in general the substance of the commentary is not been unique – I have been debating it for 40 years, either in person, in papers, and on websites.

These essays use these comments as a springboard for considering, in a combination of formal and informal forms (i) the history, (ii) the structure and (iii) the function of the model. This model, with the rating or partial credit parameterizations, is now mainstreamed and some of the relevant history, structure and function can easily be lost in academically didactic papers.

The first essay is in response to the comment that

Rasch himself never mentioned or dealt with the threshold form of the model, he never required threshold ordering, and therefore by implication the ordering of the thresholds, presumably in any form, is irrelevant.

My concern here is that this might appear as a valid argument against the kind of ordering of the threshold estimates from data, developed in Andrich (1978) and first exemplified in Andrich (1979), to researchers engaged in constructing instruments, It cannot be a valid argument in principle – it is simply appealing to authority.

However, beyond this point, in this essay I suggest that Rasch had not developed the model to the point where, even on his own terms, it had reached theoretical and practical closure as realised through sufficiency. This closure was subsequently reached in two further stages with the threshold form of the model. Therefore, Rasch could not have dealt with the threshold form of the model, nor argued for or against the implications of threshold ordering.

The state of Rasch's work till 1978

My suggestion that Rasch had left the model *undeveloped*, both in theory and for practice as realised through sufficiency, is not a criticism of course – Rasch did remarkable work and wrote many important things that we have not picked upon in social measurement and beyond.

In the Rasch (1961) (titled *On general laws and the meaning of measurement in psychology*), Rasch developed a multidimensional generalisation of the dichotomous model for the data in multiple categories. For each category relative to the first, there was a person and category parameter. Then in Andersen (1977), Erling Andersen who was

Rasch's PhD student and successor as Professor of Statistics as Applied to the Social Sciences, at the University of Copenhagen, took the case further. He had already programmed the approach of first estimating a multidimensional vector of item parameters for items with the same number and type of categories, and then decomposing this matrix into a product of an item parameter and category parameters. I discussed this approach in some detail in Andrich (2016). Ben Wright, with Rasch and Andersen worked on this approach as well.

Although it requires some summarising from Rasch's 1961 paper and Andersen's 1977 paper, the model for ordered categories simplifies, in Latin letters for convenience here, to

$$\text{Eq. (1)} \quad P_{nix} = \Pr\{X_{ni} = x; \beta_n, (\delta_i)\} = \exp\{(\kappa_{xi}) + \varphi_{xi}(\beta_n - \delta_i)\} / \gamma_{ni}.$$

I will not repeat what the parameters stand for as they are well known. This form is also known in Rasch's other work. The κ_{xi} parameters Rasch called *category coefficients* and the φ_{xi} parameters the *scoring functions*.

Rasch, Wright and Andersen worked on estimating the parameters of Eq. (1) in different ways.

The first piece of evidence that Rasch's work was not complete is from the structure and location of φ_{xi} in Eq. (1). It multiplies $(\beta_n - \delta_i)$; therefore it is in the form of a *discrimination* parameter. It specialises to it in the dichotomous case. It may seem surprising, therefore, that Rasch and Andersen, and then both with Wright tried to estimate this parameter when Wright considered, even in the late 1960s, that it could not be estimated in the dichotomous case. I summarised the correspondence on this work between Rasch and Wright in Andrich (2016).

The second, related sense in which Rasch's work can be seen as incomplete comes from Erling Andersen's work. In a brilliant, difficult paper, in 1977 Andersen showed that if the scoring function φ_{xi} was indeed an estimated real number, then there is no data reduction and that there is *no sufficiency* for the person parameter. This further confirms that Rasch had *not* completed the derivation of his model for ordered categories and a unidimensional parameterisation to the point where it had clear mutual sufficiency of the person and item parameters, something he required.

In the threshold form, if the discrimination parameters are equal at the thresholds, then we have exactly the scoring function that Andersen required that gives mutual sufficiency of the person and item parameters. The model takes the form

$$\begin{aligned}
 P_{nix} &= \Pr\{X_{ni} = x > 0; \beta_n, (\delta_i)\} = \exp\{(-\delta_{i0} - \delta_{i1} - \delta_{i2} - \dots - \delta_{ix}) + x\beta_n\} / \gamma_{ni} \\
 \text{Eq. (2)} \quad &= \Pr\{X_{ni} = x > 0; \beta_n, (\delta_i)\} = \exp\{(-\tau_{i0} - \tau_{i1} - \tau_{i2} - \dots - \tau_{ix}) + x(\beta_n - \delta_i)\} / \gamma_{ni}, \\
 P_{ni0} &= 1 / \gamma_{ni},
 \end{aligned}$$

where the total integer score of a person on the items is the sufficient statistic for the person parameter. The model specialises to the dichotomous case.

In summary, I suggest that Rasch said nothing about thresholds because he did not have this form of the model, and that this is not a reason in itself to argue that what I (or anyone else for that matter) says about thresholds is wrong. It seems to me that had no one developed the threshold form of the model, it would only be used in esoteric research as it was up to 1977, and not applied as widely, and routinely, as it is now. This perspective is further confirmed by the fact that the so called nominal response model by Darrell Bock (1972) is exactly of the form of Eq. (1) and from time to time this has been researched. However, estimating two parameters for each category, with one constraint on each set, proves to be very difficult. If it were not difficult, the IRT people would be using the nominal response model all the time. This example is further justification for the point that had the Rasch model, with sufficiency, not been developed *beyond* what Rasch had, it would not be used. It would be like the nominal response model, occasionally researched.

Although Erling Andersen had what he called an equidistant scoring function for F above, he scored it back to front from what we do now.

What is remarkable as an insight about this model, as early as 1967 Rasch was aware that his model did not permit the arbitrary collapsing of categories post hoc, and that the structure of the model was sensitive to the parameterisation of the categories.

However, neither Rasch nor Andersen explained the scoring function and the category coefficients in any substantive terms, though they did expect the scoring function to increase across successive categories. How and why I approached Eq. (1) to develop the threshold form of the model I describe in Essay 2.