Georg Rasch and Benjamin Wright's Struggle With the Unidimensional Polytomous Model With Sufficient Statistics

Educational and Psychological

Measurement

2016, Vol. 76(5) 713–723

© The Author(s) 2016

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164416634790

epm.sagepub.com

David Andrich¹

Abstract

This article reproduces correspondence between Georg Rasch of The University of Copenhagen and Benjamin Wright of The University of Chicago in the period from January 1966 to July 1967. This correspondence reveals their struggle to operationalize a unidimensional measurement model with sufficient statistics for responses in a set of ordered categories. The article then explains the original approach taken by Rasch, Wright, and Andersen, and then how, from a different tack originating in 1961 and culminating in 1978, three distinct stages of development led to the current relatively simple and elegant form of the model. The article shows that over this period of almost two decades, the demand for sufficiency of a unidimensional parameter of the object of measurement, which enabled the separation of this parameter from the parameter of the instrument, drove the theoretical development of the model.

Keywords

sufficient statistics, polytomous Rasch model, rating scale, partial credit, ordered category formats

In his article “The Function of Measurement in Modern Physical Science,” Kuhn (1961) argues that, contrary to the common understanding that scientific theories are

¹The University of Western Australia, Perth, Western Australia, Australia

Corresponding Author:

David Andrich, The University of Western Australia, M428, 35 Stirling Highway, Crawley, Perth, Western Australia 6009, Australia.

Email: david.andrich@uwa.edu.au

derived from measurements, measurements are themselves derived from scientific theories. He summarizes this position as follows:

In textbooks the numbers that result from measurements usually appear as the archetypes of the ‘irreducible and stubborn facts’ to which the scientist must, by struggle, make his theories conform. But in scientific practice, as seen through the journal literature, the scientist often seems rather to be struggling with facts, trying to force them to conformity with a theory he does not doubt. Quantitative facts cease to seem simply the ‘given’. They must be fought for and with, and in this fight the theory with which they are to be compared proves the most potent weapon. Often scientists cannot get numbers that compare well with theory until they know what numbers they should be making nature yield (p. 193).

Because it is presented in text books and is programmed in software, we now take for granted the polytomous Rasch model (PRM) for ordered categories, both the rating scale and partial credit parameterizations. Therefore, the suggestion that, in the development of this model, there is an analogy to *struggling with facts* and having a theory as a *potent weapon* might come somewhat as a surprise.

Below are excerpts of correspondence between Georg Rasch and Ben Wright sharing their struggles to implement the model in the late 1960s. The correspondence refers to Erling Andersen, then a PhD student with Rasch, and subsequently his successor as Professor of Statistics as applied to the Social Sciences at The University of Copenhagen. The excerpts are from letters that deal with other matters, which are not included here. The letters were typed with a typewriter of course, and to retain the feel for the correspondence, I have retained the few typographical errors. Although Rasch’s first name was “Georg,” in this correspondence he signed “George”; Wright also sometimes addressed him as “George.” Following the excerpts, which give insights into their reasoning, inevitably not readily evident in text books, I summarize the approach taken by Rasch, Wright, and Andersen, and how their work was the basis for the final form of the model.

By 1960, and based on the requirement of invariant comparisons, Rasch had developed a *measurement theory* (Rasch, 1960, 1961) from which he explained and formulated models for measurement, whether in the physical or social sciences. He summarized this theory, which required the separation of object and instrument parameters, in Rasch (1977). To separate these parameters, any probabilistic model for measurement was required to have *sufficient statistics*. As shown below, the development of the PRM was a struggle, and in its development the requirement of statistical sufficiency proved the *potent weapon*—it made explicit the kind of model the measurement theory had to *yield*.

Georg Rasch and Ben Wright Correspondence

Dear Ben

19th January, 1966

The item analysis for more than two categories does seem to present considerable technical difficulties in spite of Erlings [*sic*] optimism when he was leaving. If I interpret your

remarks on it correctly you are going to try an a priori assumption of one -dimensionality of the parameters. In some cases it may be a way out and in principle it could be generalized as assuming $r = 2$, etc.

I am very anxious to know how it works, not least whether and how you get cheated when actually $r = 2$, while working on assumption that $r = 1$. Simulations may be illustrative.

Your sincerely, George.

Dear George: November 16, 1966

Erlings [*sic*] just spent two days with us here in Chicago and we did some good work together. Our topic was the case where a number of categories for an answer are more than two. After seeing the general maximum likelihood solution to this problem and realizing the difficulty in computing the more complicated symmetric functions involved, we concentrated on the more typical case, where the researcher expects of the categories to be ordered and is really only in doubt about whether the hypothesis of ordering fits the data and second what the weights opositions [*sic*] in order are. Solving this problem seems the simpler and we came to a possible solution requiring more less [*sic*?] computing than the general case. This is where we are at now.

Our plan is to get together again in a few months to try to create a computer program for doing the work and to test it out on some data. We hope to complete all this by next summer, and anticipate a third conference with Erling perhaps in July.

Sincerely, Benjamin D. Wright.

Dear George: January 25, 1967

Our second preoccupation continues to be how to deal with the case of more than two categories. We too decided that it was unrealistic to work on the hypothesis of one dimension underlying the various categories. Erling's full treatment of the problem is unmanageable even on a giant computer when there are more than five or six items. It is fantastic how the number of calculations and storage space necessary mount, as the number of items goes up to say, twenty, and the number of categories goes up to say, six. A little elementary arithmetic convinced us that we would never be able to solve any problems of that size with the algorithm Erling proposed.

As an alternative, we are sampling terms from the symmetric functions involved, and estimating the symmetric functions in this way. In principle, some of those symmetric functions had billions of terms in them, asking us to sample several thousands of terms at random [to] [*sic*] approximate the symmetric function in this way and make another round in the iteration of the item parameters. Of course we do not know whether this will work, but we like the idea and would benefit very much from your opinion of it.

Sincerely, Benjamin D. Wright.

Dear Ben, 9th February, 1967

A couple of years ago we considered the possibility of utilizing the analogue to Chapt. X.3 for $m > 2$. At that time we abandoned the idea as impractical. Recently I have reconsidered it and I think there is a point we missed by [*sic*] then. Of course the number of cases where category g in items i meets category $h \neq g$ in item j may not be large, even with

many observations at disposal, but for a fixed pair of items there is a way of pooling all of the pairs (g, h) leading to a fairly simple estimation of the distance between the item parameters. The method is equivalent to using $k = 2$ the conditional of the a_{+1} vectors given row-marginals, having broken the data up into groups according to those, i.e. according to $a_{vi} + a_{vj}$ - vectors, i.e. according to combinations of categories g and h . Each such distribution is binomial and the distributions are independent, the latter fact giving the possibility of pooling.

Yours truly, George

Dear Georg: March 14, 1967

The news is better than good. It is marvellous. We are having surprisingly good success with the $M > 2$ model. The pair-wise algorithm that you reminded us about is marvelously quick and surprisingly efficient. It will certainly serve as a most excellent starting point for any iterations to meet the maximum likelihood criterion. Maybe in some case the pair-wise approach will be as good as maximum likelihood approach unless one is willing to spend quite a bit of computer time improving the estimates.

Summing up our current work:

We have coded and tested two different programs for estimating the parameters in the $M > 2$ case. We also have a program which simulates data according to a 2 or 3 factor model, adds random noise and provides us with as many cases as we want of natural looking data but of a known structure.

We also have a principal component routine for factoring rectangular matrices so that when a matrix of estimated item category parameters is obtained, we can factor it into item and category component (sic) and evaluate its rank.

The first fastest and surprisingly accurate algorithm is the pair-wise approach. In this algorithm we cross tabulate the category responses for each of items, take the log of the ratio of symmetric cells, and average these logs over categories and items. The resulting average when normalised so that all margins means are zero, forms a very good estimate of the generating parameters.

Our second algorithm uses the maximum likelihood equation for the $M > 2$ case but instead of computing the symmetric function recursively, we compute them for each score vector sampling of its term rather than compute the whole function.

This method is quite a bit slower than the pair-wise method, taking perhaps ten times longer for the same size problem. It depends a lot on how large a sample of the terms we take.

Sincerely, Benjamin D Wright

Dear Georg: July 6, 1967

Our interest in this has been increased because computing symmetric functions for the multinomial case where $M > 2$ has turned out to be quite difficult. The recursive methods so far devised accumulate sizeable round off errors so that there is a definite limit on how far the estimation iterations can go.

We have a practical working program for getting estimates in the multinomial case, but it produces rather coarse estimates because of this accumulation of round off error.

This correspondence was given to the author here by Ben Wright some time in 1977, and there is no more on the model in this correspondence. There is a letter dated July 17, 1967 with no reference to the model. Wright did write a paper with Dorothy Vogt (Vogt and Wright, n.d.) which must have been written in the late 1960s, showing estimation equations along the lines captured in the above excerpts. It had no example. It seems the paper was never published and in my studies as a student with Wright (October 1971 to September 1973), the polytomous model did not come up.

The Approach Taken by Georg Rasch, Ben Wright, and Erling Andersen

In the summary below, the notation that is now more familiar than that used by Rasch is shown. The response vector for a dichotomous response is simply $(X) = (x_0, x_1) = (0, 1)$ where in general, $x_1 = 1$ is the positive response. In current standard form, the Rasch model for such a response is given by

$$\begin{aligned} \Pr\{X_{ni} = x_1 = 1; \beta_n, \delta_i\} &= \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)}, \\ \Pr\{X_{ni} = x_0 = 0; \beta_n, \delta_i\} &= \frac{1}{1 + \exp(\beta_n - \delta_i)}, \end{aligned} \quad (1)$$

where β_n , δ_i are respectively, the scalar locations of object n and instrument i on a single dimension. For purposes of exposition, the object and instrument are, respectively, specialized to be a *person* and an *item*.

The generalization that Rasch first made in the case of more than two ordered categories was to specify the response vector of $m+1$ categories in the form $(X) = (x_0, x_1, \dots, x_k, \dots, x_m) = (0, 0, 0, \dots, 1, \dots, 0, 0)$ where 1 is the score of the category of the response $x_k = 1$, $k \neq 0$; $x_j = 0$, $j \neq k$; and $x_0 = 0$. Then the generalization of the model of Equation (1) takes the form

$$\begin{aligned} \Pr\{X_{ni} = x_k = 1; (\beta_n), (\delta_i)\} &= \frac{\exp(\beta_{nk} - \delta_{ik})}{1 + \sum_{k=1}^m \exp(\beta_{nk} - \delta_{ik})}, \quad k \neq 0 \\ \Pr\{X_{ni} = x_0 = 0; (\beta_n), (\delta_i)\} &= \frac{1}{1 + \sum_{k=1}^m \exp(\beta_{nk} - \delta_{ik})}, \end{aligned} \quad (2)$$

where $(\beta_n) = (\beta_{n1}, \beta_{n2}, \dots, \beta_{nx}, \dots, \beta_{nm})$, $(\delta_i) = (\delta_{i1}, \delta_{i2}, \dots, \delta_{ix}, \dots, \delta_{im})$ are vectors of order m for person n and item i , respectively.

Vector Sufficiency and the Polytomous Rasch Model

Taking complete data in the sense that every person responded to every item $i = 1, 2, 3, \dots, I$, in the dichotomous model where $m = 1$, the total score of the responses $x_{ni} = 1$ across the items, $r_n = \sum_{i=1}^I x_{ni}$, is the sufficient statistic for the scalar β_n . Then conditional on this total score, the resultant equation is a function only

of the vector of I item parameters $(\delta) = (\delta_1, \delta_2, \dots, \delta_i, \dots, \delta_I)$ and independent of the scalar parameter β_n .

In his work with polytomous items, Rasch took the categories to be the same in each item, there being $m + 1$ of them. Because the parameter for each person n in Eq. (2) is a vector, its sufficient statistic is also a vector, specifically $(r_n) = (r_{n1}, r_{n2}, \dots, r_{nm}) = (\sum_{i=1}^m x_{ni1}, \sum_{i=1}^m x_{ni2}, \dots, \sum_{i=1}^m x_{nim})$ of total scores of categories $k = 1, 2, 3, \dots, m$ across the items, a generalization of the sufficient statistic in the dichotomous case. Then, conditional on this vector, the resultant matrix of responses is a function only of the I by m matrix of item parameters:

$$[\delta] = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \cdots & \delta_{1m} \\ \delta_{21} & \delta_{22} & \delta_{23} & \cdots & \delta_{2m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \delta_{I1} & \delta_{I2} & \delta_{I3} & \cdots & \delta_{Im} \end{bmatrix}. \quad (3)$$

Rasch and Wright, with Andersen, worked on estimating this matrix, and their correspondence focuses on the various issues that arise from that. As is evident from their correspondence, there are many complications in this approach and the most successful algorithm involved pairwise conditioning, taking two items at a time.

Having estimated the matrix of item parameters, and on the assumption that the categories in some sense are of increasing intensity on a single dimension or variable, and that the item parameters are on this dimension, Rasch proposed that this multi-parameter matrix be decomposed into the unidimensional form

$$\delta_{ix} = \varphi_x \delta_i, \quad (4)$$

where φ_x , $x = 1, 2, \dots, m$ were referred to as scoring functions and were expected to reflect increasing values on the variable, that is, $\varphi_{x+1} > \varphi_x$, $0 < x < m$; and δ_i without a category subscript, was a unidimensional scalar characterizing item i . This is the decomposition referred to in the correspondence. They evidently also made the unidimensionality assumption of Equation (4) and attempted to estimate the item and category parameters directly.

Although the approach of estimating the matrix of item parameters was challenging, Andersen (1972), whose theoretical work in advancing Rasch's measurement theory is sometimes underestimated, provided and operationalized a set of equations that essentially provided a solution to this problem. However, because the solution was not easy to implement and the results not always easy to understand, it would not have become a standard, practical approach. Andersen himself then took a different tack based on Rasch's (1961) earlier formulation. His was one of three sequential but distinct stages, each involving struggles to make the model conform to the *potent weapon* of sufficiency.

Stage 1: A Unidimensional Expression of the Polytomous Rasch Model

Although it was not followed up initially, Rasch (1961) had specialized the multidimensional model of Equation (2) algebraically to a unidimensional form. Thus, rather than maintaining a vector of parameters for each item to the point of estimation, and then reducing the whole matrix of item parameter estimates to a unidimensional form, the vector of parameters for each item was reduced to a unidimensional form algebraically with separate category parameters, and then its consequences studied.

From a sequence of derivations requiring invariance of comparisons of parameters through sufficient statistics, Rasch successively simplified the model of Equation (2) to the form

$$\Pr\{X_{ni}=x\} = \frac{\exp(\kappa_x + \varphi_x(\beta_n - \delta_i))}{\sum_{k=0}^m \exp(\kappa_k + \varphi_k(\beta_n - \delta_i))}, \quad (5)$$

where β_n , δ_i are scalars on the same single dimension as in the dichotomous model, φ_x , $x=0, 1, 2, \dots, m$ are the same kind of scoring functions as in Equation (4), and κ_x , $x=0, 1, 2, \dots, m$ are category coefficients. The terms φ_x and κ_x , which arose from solving a differential equation requiring sufficiency, were *not interpreted*, with only the expectation that successive values of φ_x should increase.

Stage 2: Scalar Sufficiency and the Polytomous Rasch Model

Abandoning the simultaneous estimation of a matrix of item parameters, Andersen (1977) focused on the further consequences of sufficiency for the scalar parameter β_n in the model in the form of Equation (5). He derived the following condition that had to be met by the scoring function φ_x :

$$\varphi_{x+1} - \varphi_x = \varphi_x - \varphi_{x-1}. \quad (6)$$

He further showed that only if $\varphi_{x+1} = \varphi_x$ was it possible to combine the categories $x+1$ and x while retaining the model of Equation (5).

Clearly, this condition for combining categories was not obvious in advance. Indeed, Jansen and Roskam (1986) considered this property so counterintuitive that it rendered the model unsuitable for the analysis of ordered categorical data! Nevertheless, driven by the requirement of sufficiency, it was an insightful derivation by Andersen, one which proved critical. Although he derived the constraint of Equation (6) for the scoring function φ_x , like Rasch, Andersen did not explain it in any way, and furthermore, made no reference to the category coefficient κ_x , $x=0, 1, 2, \dots, m$. This was clearly an unfinished position. However, without *doubting* the requirements of sufficiency, any further development of the model had to, not only be consistent with Equation (5), but had to satisfy Andersen's condition in Equation (6), and if possible, also explain the constraint $\varphi_{x+1} = \varphi_x$ for combining categories.

Stage 3: A Response Process That Is Characterized by the Polytomous Rasch Model

That third stage, beginning independently of Equation (2) but with a focus on it and Equation (5), was to hypothesize a *latent* dichotomous response process at each of the $m > 1$ thresholds where in the case of just two categories there is only one such threshold. Each threshold τ_k , $k = 1, 2, 3, \dots, m$ was taken to qualify a single item location parameter δ_i additively to give $\delta_i + \tau_k$ as the parameter between two adjacent categories at which the probabilities of the two categories are equal. Because the category structure was assumed to be the same across items, the parameter τ_k was not subscripted by i . Each of the latent response processes was characterised, not immediately by the model of Equation (1) for dichotomous responses, but by the model of Birnbaum (1968) to give in which the qualified parameter $\delta_i + \tau_k$ of the item characterizes the difficulty at threshold k , α_k is the discrimination at this threshold, and all thresholds involve the same, single person parameter β_n .

$$\begin{aligned} \Pr\{X_{ni} = x_k = 1; (\beta_n), (\delta_i), (\alpha_k)\} &= \frac{\exp \alpha_k (\beta_n - (\delta_i + \tau_k))}{1 + \sum_{k=1}^m \exp \alpha_k (\beta_n - (\delta_i + \tau_k))}, \quad k \neq x_0 \\ \Pr\{X_{ni} = x_0 = 0; (\beta_n), (\delta_i), (\alpha_k)\} &= \frac{1}{1 + \sum_{k=1}^m \exp \alpha_k (\beta_n - (\delta_i + \tau_k))}, \end{aligned} \quad (7)$$

Then to account for the ordering of the categories, the m latent responses were constrained to one *manifest* response by further hypothesizing an ordering of the thresholds. Although hypothesizing a process at each threshold was different from specifying sufficiency for a scalar person parameter and considering the consequences, it was consistent with the approach taken by Rasch, Wright, and Andersen in the sense that the process began with a vector of item parameters, and then before any estimation, the model was constrained to have a single person parameter and the response process constrained to take account of the ordering of the categories. This stage was described in detail in Andrich (1978) in which the resultant model was identified with Equation (5). In particular, the article showed that the scoring function was given by the recursive relationship

$$\varphi_x = \sum_{k=1}^x \alpha_k, \quad x = 1, 2, 3, \dots, m; \quad \varphi_0 = 0, \quad (8)$$

where α_k was the discrimination at threshold τ_k , $k = 1, 2, 3, \dots, m$ and that if the discrimination was made the same across thresholds within an item to apply the model of Equation (1), for example, $\alpha_k = 1$, $k = 1, 2, 3, \dots, m$, then

$$\varphi_x = x, \quad x = 0, 1, 2, \dots, x. \quad (9)$$

Equation (9) clearly met Andersen's condition in Equation (6)—the thresholds needed to have equal discrimination, in the same way that dichotomous items must have equal discrimination, to provide scalar sufficiency. Equation (8), which can be expressed in the recursive form $\varphi_{x+1} = \varphi_x + \alpha_{k+1}$; $k = x$, also explained the condition

that two categories x and $x+1$ could be combined into the one category x only if $\varphi_x = \varphi_{x+1}$ —the equation shows that they could be combined only if $\alpha_{k+1} = 0$, or in general terms, that the discrimination between the two categories was 0. This explanation seemed eminently sensible. Moreover, Equation (8) explained why the scoring function in the general Equation (5) would increase with successive categories, that is, $\varphi_{x+1} > \varphi_x$, $0 < x < m$. It resulted simply from φ_x being the sum of the discriminations at successive thresholds up to and including threshold $k = x$, which if the thresholds were discriminating in the same direction as required, $\alpha_k > 0$, $k = 1, 2, 3, \dots, m$, would mean that the scoring function $\varphi_{x+1} = \varphi_x + \alpha_{k+1}$; $k = x$ was necessarily increasing.

In addition, after specifying $\alpha_k = 1$, $k = 1, 2, 3, \dots, m$, the category coefficients, κ_x , $x = 0, 1, 2, \dots, m$, were interpreted in terms of the thresholds:

$$\kappa_x = - \sum_{k=1}^x \tau_k; \quad \kappa_0 = 0, \quad \kappa_m = 0, \quad (10)$$

which gives the simplified form of Equation (5) with κ_x identified by Equation (10):

$$\Pr\{X_{ni} = x\} = \frac{\exp(\kappa_x + x(\beta_n - \delta_i))}{\sum_{k=0}^m \exp(\kappa_k + k(\beta_n - \delta_i))}, \quad x = 0, 1, 2, \dots, m. \quad (11)$$

Because many rating scales have the same categories across items, this is referred to as a *rating scale* parameterization. If the thresholds are allowed to vary in number and value among items, then the model takes the form equivalent to

$$\Pr\{X_{ni} = x\} = \frac{\exp(\kappa_{xi} + x(\beta_n - \delta_i))}{\sum_{k=0}^{m_i} \exp(\kappa_{ki} + k(\beta_n - \delta_i))} \quad (12)$$

where $\kappa_{xi} = - \sum_{k=1}^x \tau_{ik}$, $\kappa_{i0} = 0$; $\kappa_{im} = 0$ in which the thresholds are now subscripted by item i and m_i is the maximum score of item i . Because such items are common in proficiency assessment, Equation (12) is referred to as a *partial credit* parameterization (Wright and Masters, 1982). At the level of a response of a single person to a single item, the parameterizations of Equations (11) and (12) are identical both algebraically and in the interpretation of their parameters.

Thus rather than being estimated, the scoring function φ_x , $x = 0, 1, 2, 3, \dots, m$ became simply the successive integers beginning with 0. Together with being a generalization of the φ_x , $x = 0, 1$ scoring of a dichotomous response, this was a considerable simplification. Rather elegantly, the integer count $\varphi_x = x$ not only characterized the ordinal category of response descriptively but also characterized the number of successive thresholds deemed to have been exceeded (where $m - x$ characterized the number not exceeded), where $\varphi_0 = 0$ implied that none had been exceeded. Then the total score $r_n = \sum_{i=1}^I x_{ni}$ of a person on a set of items was the simple sufficient statistic for the person scalar parameter β_n , which, although with its own idiosyncrasies, was also a direct generalization of the dichotomous case and was much simpler to implement in estimation than was the vector of sufficient statistics.

Finally, the category coefficients κ_x could also be interpreted in terms of the unique item parameter δ_i and familiar threshold parameters τ_k , $k = 1, 2, 3, \dots, m$, which enter additively in the model. The estimation of these thresholds was a direct generalization of the estimation of a single threshold for each item in the dichotomous case, with the scores of the categories being jointly sufficient statistics for the thresholds. (The *joint* sufficiency needs to be emphasized because the constraint on order of the categories has the consequence that the estimates of the thresholds are not in one-to-one correspondence with the magnitudes of the scores.)

Reflections on the Role of Rasch's Measurement Theory and the Demand for Sufficiency

It is difficult to imagine that the recursive relationships in Equations (8) and (10), which involve the familiar concepts of thresholds (points of equal probability between two adjacent categories), and discrimination at these thresholds, could have been identified simply by deducing algebraically the consequences of sufficiency from Equation (2) or even Equation (5). Thus, it is not surprising to read about the struggles Georg Rasch and Ben Wright, with the help of Erling Andersen, had with their initial approach to the model. However, their work in operationalizing a model for ordered categories which revealed complications, and the theoretical work which tried to get around these complications, was necessary to set up the equations that had to be satisfied for sufficiency to hold. Then when the category coefficients and scoring function were identified with the model of Equation (5), and the latter explained the conditions formulated by Andersen, it meant the model of Equation (11), where φ_x and κ_x were defined by Equations (9) and (10), respectively, had to be the one that sufficiency was supposed to *yield*.

It is relevant to note that the articles Rasch (1961), Andersen (1977), and Andrich (1978), were entirely theoretical, had no examples meaning that the model was not derived to account for any data set, and focused on the response of one person to one polytomous item. It is also salutary to note, in this day of outcomes, publication demands of academics, and so on, that the publications described above spanned some 18 years. Thus, work over more than 18 years initially gave rise to just these three, theoretical, publications. The publications and applications they have generated since have been in the thousands.

From 1978 (a decade after the last piece of correspondence between Ben Wright and Georg Rasch on the model), Ben Wright again became interested in this model in the form of Equations (11) and (12), and introduced it to his many students. On seeing an application of the simple form of the model in the contingency table context (Andrich, 1979) in the journal *Biometrics*, of which Rasch was a founding member, Rasch seemed beguiled and showed great interest. Unfortunately, Rasch died in 1980 and could neither fully appreciate how his measurement theory led to the final, elegant form of the model, nor witness its many applications.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Andersen, E. B. (1972). The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistical Society Series B*, 34, 42-54.
- 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, 561-574.
- Andrich, D. (1979). A model for contingency tables having an ordered response classification. *Biometrics*, 35, 403-415.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick. *Statistical theories of mental test scores* (pp. 397-545). Reading, MA: Addison-Wesley.
- Jansen, P. G. W., & Roskam, E. E. (1986). Latent trait models and dichotomization of graded responses. *Psychometrika*, 51, 69-91.
- Kuhn, T. S. (1961). The function of measurement in modern physical science. *Isis*, 52, 161-193.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* (Copenhagen, Danish Institute for Educational Research). Expanded edition (1980) with foreword and afterword by B. D. Wright (1980). Chicago, IL: The University of Chicago Press. Reprinted (1993). Chicago, IL: MESA Press. Available from <http://www.rasch.org/books.htm>.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 321-334). Berkeley: University of California Press.
- Rasch, G. (1977). On specific objectivity: an attempt at formalising the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy*, 14, 58-94.
- Vogt, D. K. & Wright, B. D. (n.d.). *Parameter estimation for the polychotomous Rasch model*. Unpublished manuscript. Chicago, IL: The University of Chicago, School of Education. (Note. *polychotomous* was a common term used for ordered categories which has since been replaced by polytomous.)
- Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago, IL: MESA Press.