

Measuring Attitudes With a Threshold Model Drawing on a Traditional Scaling Concept

Jürgen Rost
University of Kiel

This paper presents a generalized Rasch model for measuring attitudes which is based on the concepts of Thurstone's method of successive intervals. The model combines the rating scale and the dispersion model proposed by Andrich and a submodel of the partial credit model proposed by Masters. An estimation procedure for unconditional maximum likelihood (ML) estimates is outlined. A recursion formula for the symmetric functions, which is needed for conditional ML

procedures, is given. The benefits of the model are illustrated with a study on students' interest in physics. The fit of different threshold models can be compared using conditional likelihood values and conditional likelihood ratio tests. *Index terms: attitude measurement, conditional likelihood ratio test, partial credit model, Rasch model, rating scales, successive intervals, threshold model.*

Guilford's (1954) statement about rating scales in his textbook on psychometric methods remains valid: "Of the psychological measurement methods that depend upon human judgment, rating-scale procedures exceed them all for popularity and use. . . . Generally speaking, rating methods belong logically under the heading of the method of successive intervals" (p. 263).

A basic difference exists, however, between using rating scales to measure attitudes and the method of successive intervals. The latter placed objects on a scale by using judgments of a number of persons on a graded response format. The measurement of persons' attitudes took place in a second step once the scale values of the items were known.

In contrast, the goal of Likert's technique of attitude scaling (Likert, 1932) was to scale persons. He said that it is sufficient to weight response categories with successive integers in order to get scores for the attitudes. A scaling of items and categories did not take place.

The family of generalized Rasch models, to which the model presented here belongs, scales persons and objects simultaneously. Nevertheless, scaling techniques in the Thurstonian tradition are based on a typical scaling concept, which can be fruitfully developed in the framework of threshold models and generalized Rasch models.

Thurstone's Scaling Tradition

The method of successive intervals, introduced by Thurstone (Edwards & Thurstone, 1952), attempted to solve a twofold scaling problem. First, the particular rating scale of a questionnaire—the response

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 12, No. 4, December 1988, pp. 397–409
© Copyright 1988 Applied Psychological Measurement Inc.
0146-6216/88/040397-13\$1.90

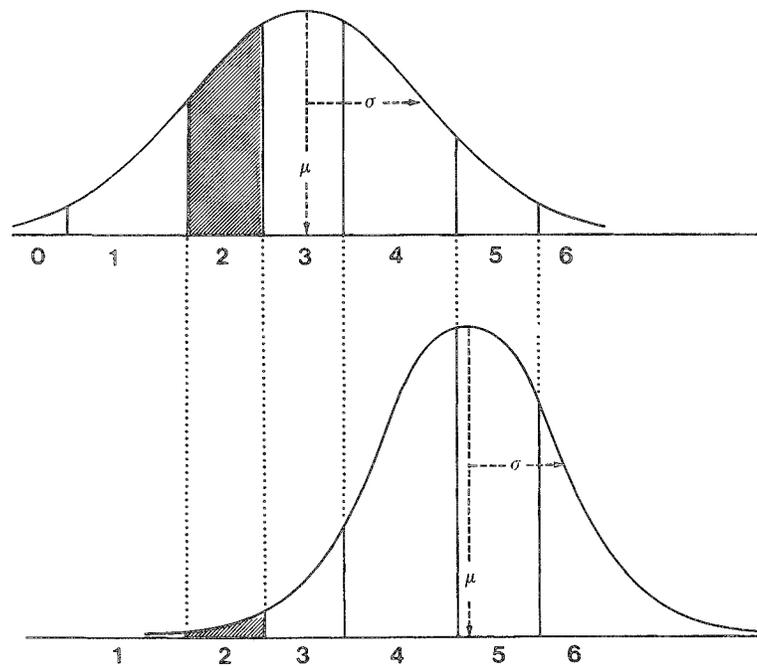
categories used to judge a series of stimuli—is scaled. Categories are ordered on a continuum, but no assumption is made about their “distances” or “intervals” on that continuum. The first scaling problem is estimating the widths of the successive intervals as a characteristic of the response categories, independent of the stimuli (test items) and persons.

The second scaling problem is scaling the test items, or the stimuli or statements to be judged. The original formulation of the method of successive intervals (Edwards & Thurstone, 1952) only allowed the stimuli to vary according to their location on the continuum spanned by successive categories. The widths of the intervals on the continuum, as well as the location of the stimuli on this continuum, were estimated under the assumption of normally distributed judgments, as shown in Figure 1. The response probabilities correspond to the areas under a normal distribution and are determined by the widths of intervals and the parameters μ and σ .

Figure 1, however, goes further by allowing different dispersion parameters for each stimulus. This generalization of the method of successive intervals was introduced by Burros (1955) and Rimoldi and Hormaeche (1955). It accounts for the typical characteristic of Thurstone’s scaling models (the Law of Categorical Judgment) by stating that “there are at least two important kinds of parameters, represented respectively by S_j , the scale value of the j th stimulus, and σ_j , the corresponding discriminial dispersion” (Burros, 1955, p. 300).

Thus, the general case of the successive interval technique covers three types of parameters defining the widths of intervals, the scale values of stimuli, and the dispersions of judgments for each stimulus. When transposing this approach to psychometric modeling, a fourth type of parameter, the person parameter, must be considered.

Figure 1
 The Assumption of Normally Distributed Ratings in the Method of Successive Intervals



Psychometric Models

Within the past 10 years, powerful psychometric models have been developed for rating-scale data based on Rasch's (1960/1980) theory of measurement (e.g., Andrich, 1978a, 1978b, 1982; Wright & Masters, 1982). Psychometric models differ in two fundamental aspects from the method of successive intervals. First, they do not require any distributional assumption about the judgments of individuals. Second, they simultaneously solve an additional scaling problem, namely the scaling of persons. However, none of these models solves the original problem addressed by the method of successive intervals, which was scaling the widths of intervals under the assumptions of different dispersions for the stimuli.

Andrich (1978a, 1978b) introduced the "rating-scale model" as a direct generalization of Rasch's (1960/1980) dichotomous test model. In the rating-scale model

$$p(r_{vi} = x) = \frac{\exp(x\xi_v + x\lambda_i + \psi_x)}{d_{vi}} \quad (1)$$

where

$$d_{vi} = \sum_{x=0}^m \exp(x\xi_v + x\lambda_i + \psi_x) \quad (2)$$

and

$$\psi_0 = \psi_m = 0 \quad (3)$$

the probability of a judgment in category x , given by person v on item i with response categories from 0 to m , equals the logistic function of a person parameter ξ_v , an item parameter λ_i , and a category parameter ψ_x . Andrich derived this model based on a threshold definition and showed that the category parameters ψ_x can be interpreted as cumulative threshold parameters τ_x defined by

$$\psi_x = \sum_{s=0}^x \tau_s \quad (4)$$

In the resulting version of the model in Equation 1, τ_x parameterizes the location of threshold x (the threshold between response category $x - 1$ and x) on the latent continuum (Andrich, 1978a).

Although some consider the threshold approach unconvincing (e.g., Jansen & Roskam, 1986), the same model follows from what Rasch (1965) and Andersen (1977a) have shown concerning specific objective generalizations of the dichotomous Rasch model. Rasch (1965) showed that the following model structure is necessary for specific objective comparisons on the basis of multicategorical data, when both item and person parameters are scalars:

$$p(r_{vi} = x) = \frac{\exp[\phi_x(\xi_v + \lambda_i) + \psi_x]}{d_{vi}} \quad (5)$$

The denominator in Equation 5, d_{vi} , is the sum of the numerator over all categories. Into this so-called "unidimensional Rasch model" (Andersen, 1977b), a multiplicative parameter ϕ_x is introduced which, like ψ_x , only depends on the categories. Andersen (1977a) has shown that the parameters ϕ_x must be equidistant in order to obtain a specific objective model. Moreover, the ϕ_x parameters are subject to two norming conditions (usually $\phi_0 = 1$ and $\phi_m = 0$). These restrictions taken together lead directly to the rating-scale model in Equation 1, because the parameter ϕ_x can be replaced by x if the norming conditions $\phi_0 = 0$ and $\phi_m = m$ are chosen.

In the rating-scale model, each item i is parameterized by only one parameter λ_i , which quantifies the mean of all threshold locations on the latent continuum (Andrich, 1978a). The idea inherent in the method of successive intervals—to allow for varying dispersions of item responses—is not realized in

this model. It is met, however, in another threshold model which has also been proposed by Andrich (1982). In his "dispersion model,"

$$p(r_{vi} = x) = \frac{\exp[x\xi_{vi} + x\lambda_i + x(m-x)\delta_i]}{d_{vi}}, \quad (6)$$

two parameters, λ_i and δ_i , describe each item. Whereas λ_i is the mean location of all thresholds, δ_i quantifies half the distance between two neighboring thresholds. Hence, the model assumes equidistant thresholds within each item, but varying distances among the items. Again, only one part of the method of successive intervals is realized.

This paper combines both the rating-scale and the dispersion models as a straightforward threshold formalization of what the developers of the method of successive intervals considered an adequate way of analyzing rating-scale data. All of these models are submodels of the partial-credit model developed by Masters (1982), in which no restrictions are imposed on the threshold parameters. Figure 2 illustrates the assumptions made in the different models.

A Threshold Model Drawing on Thurstone's Tradition

The model derived here is based on a parameterization of threshold probabilities. The probability of passing a threshold x is defined as the conditional probability of responding in category x under the condition that the response is either in $x - 1$ or in x :

Figure 2
 Illustration of Different Restrictions on Threshold Distances

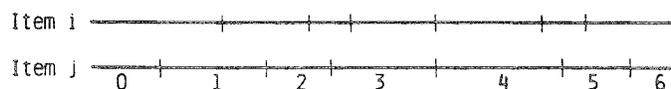
Rating Scale Model: Constant Distances for All Items



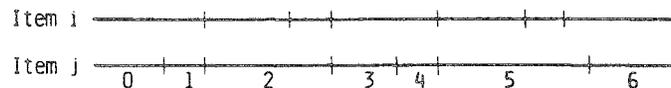
Dispersion Model: Equidistant Thresholds for Each Item



Successive Interval Model: A Combination of the Rating Scale and Dispersion Model



Partial Credit Model: No Restriction on Threshold Intervals



$$\pi_x = p(r_{vi} = x | r_{vi} = x-1 \text{ or } r_{vi} = x) = \frac{p(r_{vi} = x)}{p(r_{vi} = x-1) + p(r_{vi} = x)} \quad (7)$$

The relevant property of this definition is that it relies only on the probabilities of two neighboring response categories; this contrasts, for example, with defining the probability of reaching a particular "response level," $p(r_{vi} \geq x)$, as in the model of Samejima (1969). As Wright and Masters (1982) pointed out, this property is the reason that a simple model structure with separability of person and item parameters can be obtained.

The model follows from the assumption that the threshold probability is a logistic function of a person parameter ξ_v , two item parameters λ_i and δ_i , and a threshold parameter τ_x :

$$\pi_x = \frac{\exp[\xi_v + \lambda_i + (m+1-2x)\delta_i + \tau_x]}{1 + \exp[\xi_v + \lambda_i + (m+1-2x)\delta_i + \tau_x]} \quad (x > 0) \quad (8)$$

The interpretation of ξ_v and λ_i is straightforward. They represent the person's level on the latent variable, and the item easiness or its tendency to facilitate passing a threshold. However, in combination with the remaining two parameters, δ_i and τ_x , the interpretation of λ_i must be refined. Their interpretation is shown by means of a fictitious example (Figure 3).

Figure 3
 Threshold Characteristic Curves for Two Items With
 Five Response Categories According to Equation 8

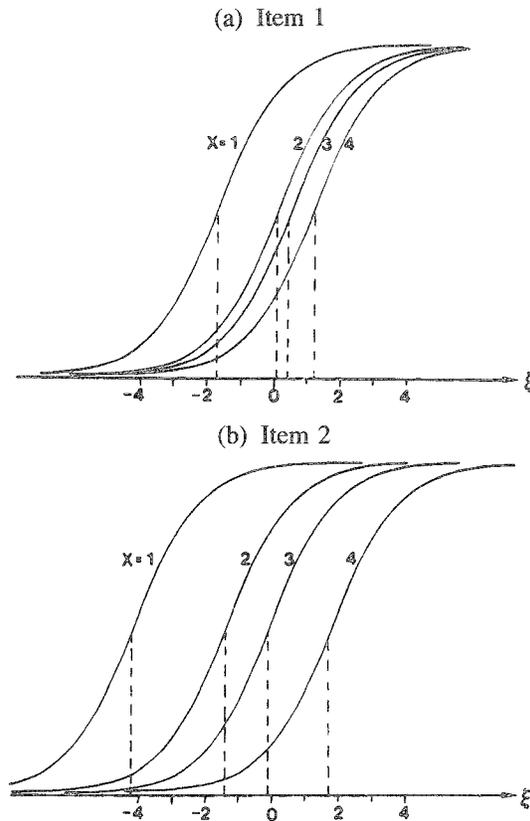


Figure 3 shows the "threshold characteristic curves" (TCCs), or the threshold probability as a function of the latent variable, for two items and five response categories. The parameter values of these items are $\lambda_1 = 1$, $\lambda_2 = 0$, $\delta_1 = .4$, and $\delta_2 = -.1$, and the threshold parameters are $\tau_1 = 2$, $\tau_2 = 0$, $\tau_3 = -.5$, and $\tau_4 = -1.5$.

As in the rating-scale and dispersion models (Andrich, 1978a, 1982), λ_i is a location parameter and parameterizes the mean location of all thresholds of an item, if the norming condition for the other parameters is chosen appropriately. The interpretations of the threshold parameters τ_x and the second item parameters δ_i are also identical to those of the rating scale and the dispersion model, but must be related to each other. The threshold parameters τ_x no longer parameterize threshold distances that are constant across the items, because the item parameters δ_i also contribute to these distances. Instead, they quantify the *mean* threshold distances for all items. Each item has threshold distances which differ by a certain amount from the distances defined by τ_x . This item-specific contribution to the threshold distances is parameterized by δ_i .

In Figure 3, the "basic threshold distances" defined by τ_x are $\Delta_{12} = 2.0$ between thresholds 1 and 2, $\Delta_{23} = .5$ between thresholds 2 and 3, and $\Delta_{34} = 1.0$ between thresholds 3 and 4. Because δ_i parameterizes only half the deviation of a threshold distance within item i from these mean distances, the final distances of both items are $\Delta_{12} = 2.8$, $\Delta_{23} = 1.3$, and $\Delta_{34} = 1.8$ for item 1, and $\Delta_{12} = 1.8$, $\Delta_{23} = .3$, and $\Delta_{34} = .8$ for item 2 (see Figure 3).

This example illustrates that the proposed model is nothing more than a combination of the rating-scale and the dispersion models. In order to avoid overparameterization, both category parameters and dispersion parameters must be normed. The norming conditions are

$$\sum_{x=1}^m \tau_x = \sum_{i=1}^k \delta_i = \sum_{i=1}^k \lambda_i = 0 \quad (9)$$

The model thus has $2(k - 1)$ independent item parameters, $m - 1$ independent threshold parameters and, of course, n person parameters.

The model equation is easily obtained by following the same derivation as in other threshold models (e.g., Masters & Wright, 1984, appendix), so that the response probabilities are

$$p(r_{vi} = x) = \frac{\exp \left[\sum_{x=1}^m \xi_v + \lambda_i + (m + 1 - 2x)\delta_i + \tau_x \right]}{d_{vi}} = \frac{\exp[x\xi_v + x\lambda_i + x(m - x)\delta_i + \psi_x]}{d_{vi}} \quad (10)$$

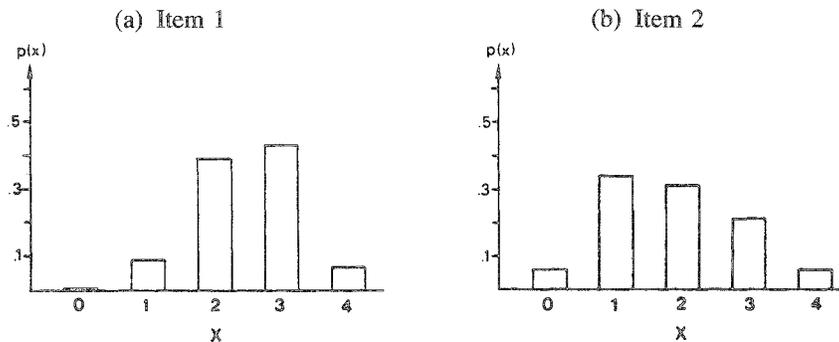
where

$$\psi_x = \sum_{x=1}^m \tau_x \quad (11)$$

and d_{vi} , as usual, is the sum of the numerator across all categories. If $\psi_0 = 0$ is defined, Equations 10 and 11 also hold for $x = 0$.

To show the analogy to the method of successive intervals, Figure 4 gives the response probabilities for the two fictitious items of Figure 3, when all persons have the same level on the latent variable ($\xi_v = 0$). Item 2 has a higher dispersion than item 1. Thus, the distribution of item 1 is steeper due to higher threshold distances of item 1, because of its positive dispersion parameter ($\delta_i = +.4$). In addition, item 2 also shows that the response probability of category 1 is somewhat higher than those of categories 2 and 3. Again this is caused by the higher distance between thresholds 1 and 2 ($\Delta_{12} = 1.8$) as compared with the two following distances ($\Delta_{23} = .3$ and $\Delta_{34} = .8$). As a rule, the response probability of a category increases as the distance between its adjacent thresholds increases.

Figure 4
 Response Probabilities for the Fictitious Items of Figure 3
 and a Mean Degree on the Latent Variable ($\xi_v = 0$)



Thus, a strong analogy exists between the method of successive intervals based on category boundaries under a normal distribution and the present model based on a threshold definition. In both cases a monotone relation exists between the widths of the threshold or boundary intervals and the size of the response probabilities. However, it is only an analogy. By their definitions, thresholds and category boundaries have quite different interpretations (e.g., Andrich, 1982; Wright & Masters, 1982). Despite these differences, the model in Equations 10 and 11 parallels the method of successive intervals inasmuch as the threshold distances are decomposed into a category-specific part (ψ_x) and an item-specific part (δ_i) that causes the dispersions of item responses to vary.

Parameter Estimation

Parameter estimates can be obtained by means of conditional and unconditional maximum likelihood (ML) procedures. The latter is outlined here, and a recursive formula for computing the symmetric functions needed for conditional parameter estimation is provided following an example.

Unconditional parameter estimation can be done using the modified Newton-Raphson algorithm as described by Andrich (1978b) for the rating-scale model or by Wright and Masters (1982) for the partial-credit model. According to this iterative algorithm, new parameter estimates $\hat{\epsilon}^*$ of a parameter ϵ are computed on the basis of preliminary estimates $\hat{\epsilon}$, the first-order partial derivatives of the log-likelihood function for this parameter, $\ln L'(\epsilon)$, and the second-order derivatives for the same parameter, $\ln L''(\epsilon)$,

$$\hat{\epsilon}^* = \frac{\hat{\epsilon} - \ln L'(\epsilon)}{\ln L''(\epsilon)} \quad (12)$$

Note that the complete matrix of second-order derivatives is not required in this algorithm, but only its diagonal elements. The likelihood function is given by

$$L = \prod_v \prod_i \frac{\exp[x\xi_v + x\lambda_i + x(m-x)\delta_i + \psi_x]}{d_{vi}} \quad (13)$$

and the partial derivatives of its logarithm are

$$\ln L'(\xi_v) = r - \sum_i \sum_x x p_{rix} \quad (14)$$

$$\ln L'(\lambda_i) = r_i - \sum_r \sum_x n_r x p_{rix} \quad (15)$$

$$\ln L'(\delta_i) = q_i - \sum_r n_r \sum_x x(m-x)p_{rix} \quad , \quad (16)$$

and

$$\ln L'(\psi_x) = r_x - \sum_r n_r \sum_i p_{rix} \quad , \quad (17)$$

where r is the raw score of a person,

$r_i = \sum_v r_{vi}$ is the score of item i ,

r_x is the number of responses in category x , and

$q_i = \sum_v r_{vi}(m - r_{vi})$ is the sum of the scoring coefficients of the dispersion parameter.

p_{rix} is the probability of a person with score r answering item i in category x , or the response probability given by Equations 10 and 11, which is identical for all persons with the same score r . The second-order derivatives then are

$$\ln L''(\xi_r) = \sum_i \left[\left(\sum_x x p_{rix} \right)^2 - \sum_x x^2 p_{rix} \right] \quad , \quad (18)$$

$$\ln L''(\lambda_i) = \sum_r n_r \left[\left(\sum_x x p_{rix} \right)^2 - \sum_x x^2 p_{rix} \right] \quad , \quad (19)$$

$$\ln L''(\delta_i) = \sum_r n_r \left\{ \left[\sum_x x(m-x)p_{rix} \right]^2 - \sum_x [x(m-x)]^2 p_{rix} \right\} \quad , \quad (20)$$

and

$$\ln L''(\psi_x) = \sum_r n_r \sum_i (p_{rix}^2 - p_{rix}) \quad (21)$$

(Andrich, 1978b). The algorithm may be started by using 0s as initial estimates for all parameters. A proven sequence of reestimating parameters in each iteration is λ_i , ψ_x , δ_i , and ξ_r . The new estimates must be normed in each iteration according to Equation 9. The procedure is quick; the estimation process for the example described below took about 30 iterations to obtain stable estimates.

As usual, the unconditional ML estimates are biased for small numbers of items. The well-known correction term $(k - 1)/k$ (Masters, 1982) must be applied here. Every parameter estimate should be multiplied by this factor to correct for estimation bias.

Example

To illustrate the application of the proposed model, part of a questionnaire on students' interest in physics and technology (Hoffmann, Lehrke, & Todt, 1985) was analyzed with the successive interval model (Equations 10 and 11). In this study, 4,035 students from Grades 5–9 answered the questionnaire. Part of the questionnaire concerned students' interest in eight areas of physics and technology in different contexts (Häußler, 1987). For example, interest in learning about acoustics was assumed to be different when presented in a pure intellectual context rather than in an everyday context. An item in an intellectual context was "Learning more about what the pitch and intensity of a tone depend on," and for an everyday context "Measuring the intensity of different sources of noise (e.g., traffic, music) by means of appliances."

The interest in each of these items was rated on a five-point scale of "very high," "high," "medium," "low," and "very low." For the present purpose, the items for all eight topics were analyzed in two different contexts ("intellectual" and "everyday life"). Hence, the data consisted of two sets of items, one set measuring the interest in physics presented in an intellectual context, and one set in an everyday-life context.

Figure 5 shows the parameter estimates of both sets of items after being transformed into threshold locations on the latent continuum. The symbols in Figure 5, therefore, mark the position of a threshold x , under four different models (see Equations 1, 6, 10, and 23). These locations follow from the definition of threshold probabilities under a particular model (see Equation 8). In the case of the successive interval model, they are given by

$$\text{loc}(x) = \lambda_i + (m+1-2x)\delta_i + \psi_x - \psi_{x-1} \quad (22)$$

Figure 5 shows that for these data all thresholds are ordered, so that each subsequent threshold is more difficult than the preceding one. This is not necessarily the case, but it should be expected from the rationale of the threshold approach (Wright & Masters, 1982). Put differently, this result indicates that the five response categories were used as a rating scale by the respondents (Rost, in press).

The results also show that the middle threshold distance is the largest for all items. This may indicate a response set toward the mean. Obviously, the middle response category has been marked with a higher frequency than should be expected for equidistant thresholds.

A comparison of the items concerned with two different contexts of learning physics (Figure 5a vs. Figure 5b) clearly shows that they vary in difficulties, mean threshold distances, and item dispersions. The threshold distances of the items with an everyday-life context (Figure 5b) are generally smaller than for the intellectual context (Figure 5a). Because the size of threshold intervals is related to the discrimination of the item (the smaller the distance, the higher the discrimination; Andrich, 1983), it appears that everyday-life items have greater discriminative power for measuring students' interests than do intellectual items.

When evaluating the successive interval model, it may be more informative to compare the estimates of this model with those of the rating-scale, the dispersion, and the partial-credit models. The threshold locations of all these models are depicted in Figure 5.

Figure 5a clearly shows that the successive interval model simply combines the rating-scale and the dispersion models. The dispersion model, with its equidistance assumption, fails to model the enlarged distance between the two middle thresholds (compare the crosses in Figure 5a). However, the rating-scale model, with a constant item dispersion, underestimates the distances of the first four items and overestimates the distances of the last four items (the arrowheads in Figure 5a). The partial-credit parameters (the dots in Figure 5) show the position of thresholds if no assumption on threshold distances is made. They are close to the successive interval parameters (circles) and naturally a bit farther from the rating-scale and dispersion parameters.

Model Comparisons by Means of Conditional Likelihoods

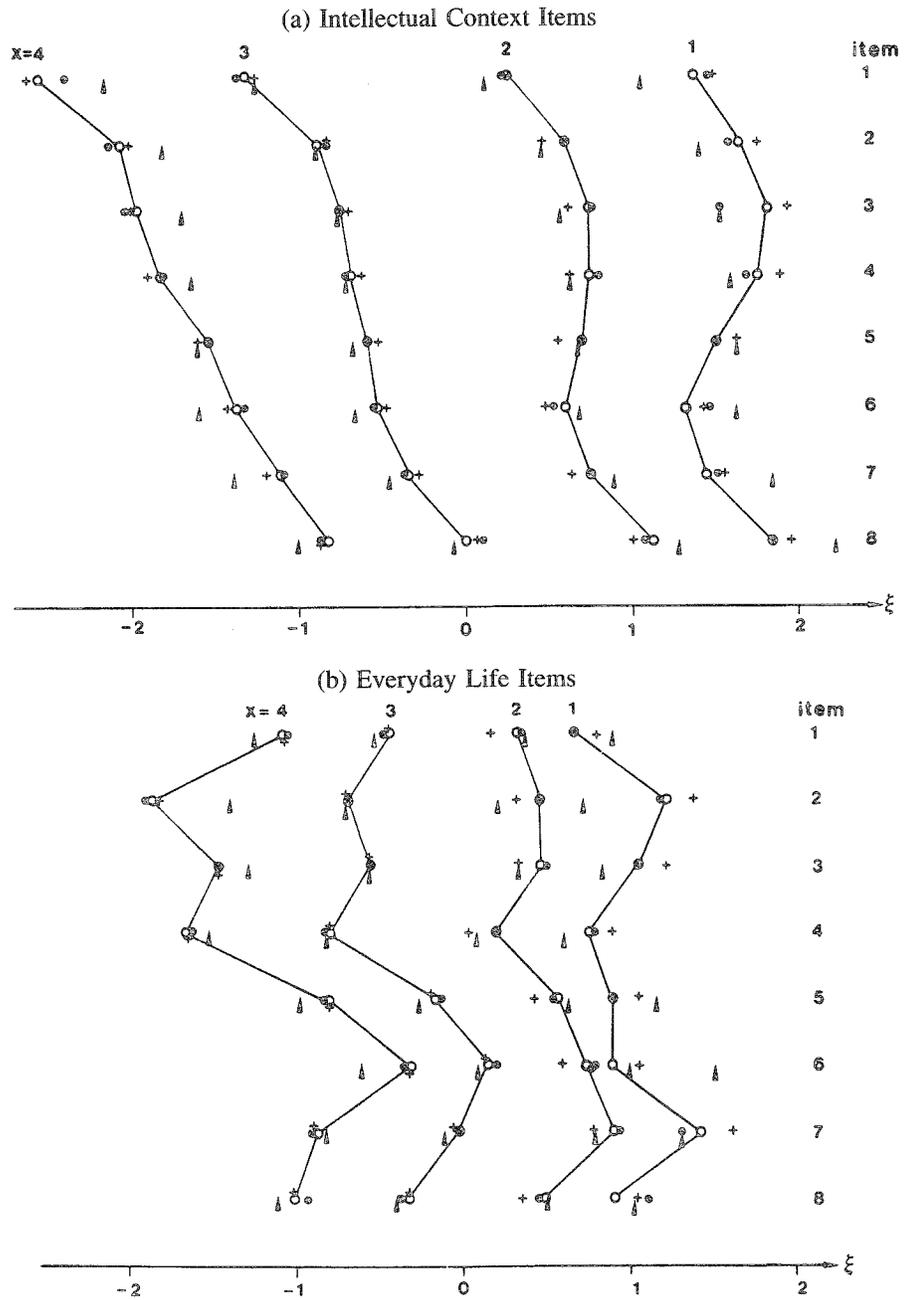
The fit of different threshold models applied to the same set of data can be compared by means of the conditional likelihood (CL) of the data. Statistical inference is possible by using conditional likelihood ratio (CLR) tests. The CL depends only on the structural parameters of a model (item and category parameters), and incidental parameters (the person parameters) are "conditioned out." The CL function of the partial-credit model

$$p(r_{vi} = x) = \frac{\exp(x\xi_v + \varepsilon_{ix})}{d_{vi}} \quad (23)$$

is

$$\text{CL} = \prod_x \left[\frac{\prod_i \eta_{ix}}{\gamma_x(\eta)} \right]^{n(x)}, \quad (24)$$

Figure 5
 Threshold Locations on the Latent Continuum for
 Two Sets of Items Under Four Different Models
 (○-○ Successive Interval, ● Partial Credit, ▲ Rating Scale, + Dispersion)



where $\gamma_r(\boldsymbol{\eta})$ is the symmetric function of order r ,

$$\gamma_r(\boldsymbol{\eta}) = \sum_{\mathbf{x}} \prod_i \eta_{ix} \quad , \quad (25)$$

and

$$\eta_{ix} = \exp(\varepsilon_{ix}) \quad , \quad (26)$$

where \mathbf{x} denotes the response vector and $n(\mathbf{x})$ its observed frequency. The same equation holds for the other threshold models discussed here, if the threshold parameters ε_{ix} are substituted by the linear decomposition assumed in a particular model, for example,

$$\varepsilon_{ix} = x\lambda_i + x(m-x)\delta_i + \psi_x \quad (27)$$

in the case of the successive interval model. For larger numbers of items, the symmetric functions can only be computed with the aid of a recursive formula. As a generalization of the recursive formula for dichotomous items (Andersen, 1972), the following equations have been derived (Rost, 1988):

$$\gamma_r(\boldsymbol{\eta}) = \frac{1}{r} \sum_i \sum_x x \eta_{ix} \gamma_{r-x}^{(i)}(\boldsymbol{\eta}) \quad (28)$$

and

$$\gamma_r^{(i)}(\boldsymbol{\eta}) = \gamma_r(\boldsymbol{\eta}) - \sum_x \eta_{ix} \gamma_{r-x}^{(i)}(\boldsymbol{\eta}) \quad . \quad (29)$$

They enable rapid calculation of the conditional likelihoods and computation of the CLR statistic

$$Z = -2 \ln \frac{CL_1}{CL_2} \quad , \quad (30)$$

where CL_1 and CL_2 denote the likelihood of the data under two different models, with Model 1 a submodel of Model 2. Z is asymptotically chi-square distributed and the number of degrees of freedom equals the number of independent parameters under Model 2 minus those of Model 1.

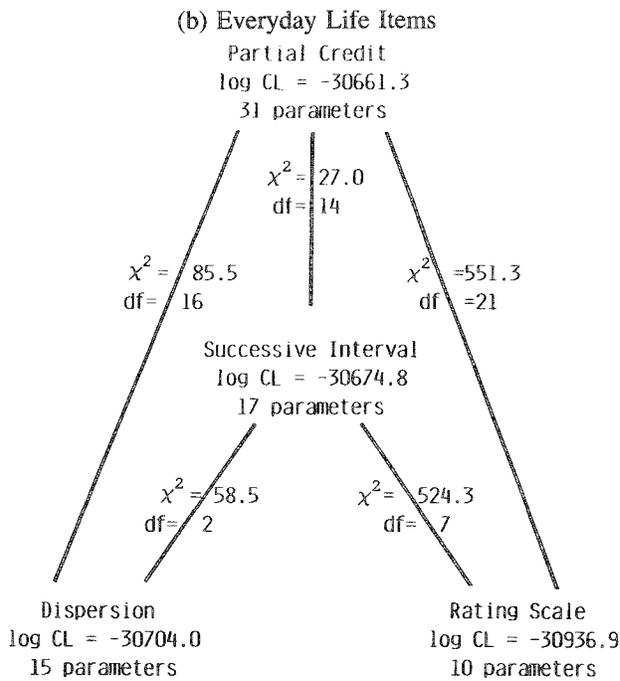
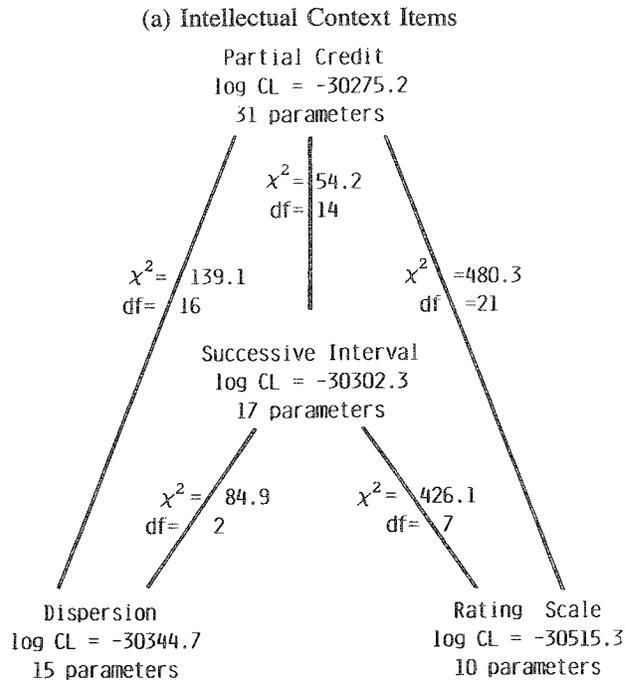
Figure 6 shows the logarithms of CL values under four different models, as well as the results of CLR tests between all pairs of these hierarchically related models for the students' interest data.

For both sets of items, both the rating-scale and dispersion models provide a poorer fit (smaller log likelihood) than the partial-credit model. As compared with the former two, the difference between the successive interval model and the partial-credit model is relatively small and non-significant at the .01 level in the case of the everyday-life items. Comparison of the successive interval model with the dispersion and the rating-scale model shows that only a few additional parameters—two compared with the dispersion model and seven compared with the rating-scale model—cause a drastic increment of the conditional likelihood.

Discussion

In its middle position between the partial-credit model and the rating-scale and dispersion models, the successive interval model combines advantages of both. Because it neither assumes equidistance nor postulates constant threshold distances for all items, it easily accommodates varying threshold locations. On the other hand, it is not completely unrestricted like the partial-credit model, but takes account of the prominent feature of rating-scale data, namely the fact that the same response categories are used for all items. For those data, the model is not only parsimonious with regard to the number of parameters, but the parameters are also easier to interpret than under unrestricted threshold locations.

Figure 6
 Conditional Likelihoods (Logarithms) and CLR Tests
 Between Four Hierarchically Related Models



The idea of combining category-specific threshold distances with item-specific influences on these distances can also be extended to models with skewness or kurtosis parameters as proposed by Andrich (1985). This, however, may result in an uncountable number of models having very specialized applications. In contrast to these extensions, the model of successive intervals is general and is simply a formalization of procedures for rating data that have existed from the beginning of psychometric theory.

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. (1977a). Sufficient statistics and latent trait models. *Psychometrika*, 42, 69–81.
- Andersen, E. B. (1977b). The logistic model for m answer categories. In W. F. Kempf & B. H. Repp (Eds.), *Mathematical models for social psychology* (pp. 59–80). Bern: Huber.
- Andrich, D. (1978a). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, 4, 581–594.
- Andrich, D. (1978b). A rating formulation for ordered response categories. *Psychometrika*, 43, 561–573.
- Andrich, D. (1982). An extension of the Rasch model for ratings providing both location and dispersion parameters. *Psychometrika*, 47, 105–113.
- Andrich, D. (1983). A latent trait model which takes account of dependencies among responses: Implications for test construction and analysis. In S. E. Embretson (Ed.), *Test design: Developments in psychology and psychometrics*. Orlando FL: Academic Press.
- Andrich, D. (1985). An elaboration of Guttman scaling with Rasch models for measurement. In N. B. Tuma (Ed.), *Sociological methodology* (pp. 33–80). San Francisco: Jossey-Bass.
- Burros, R. H. (1955). The estimation of the discriminial dispersion in the method of successive intervals. *Psychometrika*, 20, 299–305.
- Edwards, A. L., & Thurstone, L. L. (1952). An internal consistency check for scale values determined by the method of successive intervals. *Psychometrika*, 17, 169–180.
- Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.
- Häubler, P. (1987). Measuring students' interest in physics—design and results of a cross-sectional study in the Federal Republic of Germany. *International Journal of Science Education*, 9, 79–92.
- Hoffmann, L., Lehrke, M., & Todt, E. (1985). Development and change in pupils' interests in physics (grade 5 to 10): Design of a longitudinal study. In M. Lehrke, L. Hoffmann, & P. L. Gardner (Eds.), *Interests in science and technology education. Proceedings of the 12th IPN Symposium 1984* (pp. 71–80). Kiel: IPN.
- Jansen, P. G. W., & Roskam, E. E. (1986). Latent trait models and dichotomization of graded responses. *Psychometrika*, 51, 69–91.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, No. 140.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- Masters, G. N., & Wright, B. D. (1984). The essential process in a family of measurement models. *Psychometrika*, 49, 529–544.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research. [Expanded edition, University of Chicago Press, 1980.]
- Rasch, G. (1965). Kolloquium über Meßmodelle. (Manuscript of a lecture by G. Rasch at the Statistical Institute of the University of Copenhagen, Spring 1965, transcribed by Jon Stene; translated by Bruno H. Repp.)
- Rimoldi, H. J. A., & Hormaeche, M. (1955). The law of comparative judgment in the successive intervals and graphic rating scale methods. *Psychometrika*, 20, 307–318.
- Rost, J. (1988). Test theory with qualitative and quantitative latent variables. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 147–171). New York: Plenum.
- Rost, J. (in press). Rating scale analysis with latent class models. *Psychometrika*.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika, Monograph Supplement* No. 17.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis. Rasch measurement*. Chicago: MESA Press.

Author's Address

Send requests for reprints or further information to Jürgen Rost, IPN-Institute for Science Education at the University of Kiel, Olshausenstraße 62, D-2300 Kiel 1, Federal Republic of Germany.