Application of a Psychometric Rating Model to Ordered Categories Which Are Scored with Successive Integers

David Andrich
The University of Western Australia

A latent trait measurement model in which ordered response categories are both parameterized and scored with successive integers is investigated and applied to a summated rating or Likert questionnaire. In addition to each category, each item of the questionnaire and each subject are parameterized in the model; and maximum likelihood estimates for these parameters are derived. Among the features of the model which make it attractive for applications to Likert questionnaires is that the total score is a sufficient statistic for a subject's attitude measure. Thus, the model provides a formalization of a familiar and practical procedure for measuring attitudes.

One of the most persistent features in the scoring of response categories which are ordered is the assignment of successive integers to the successive categories. This procedure is particularly evident in the scoring of items of attitude questionnaires constructed in the popular Likert tradition in which people rate their degree of agreement or disagreement to statements or items. In addition, the summary statistic indicating a person's attitude with respect to a set of items on such a questionnaire is commonly taken to be the sum of the integer scores assigned to the items. Although this is the common approach to dealing with such ordered response categories, certain assumptions of this scoring procedure, particularly that with respect to the implied assumption of "equal distances" between categories, continues to be questioned.

The general approach for overcoming objections to the integer-scoring procedure is to use a response model which keeps track of the category in which a person responds. Two basic types of models with such a property have been studied extensively. In the first type, exemplified by Samejima (1969) and Kolakowski and Bock (1972), a unidimensional latent variable with distances between response categories is postulated. The boundary positions of the distances are considered threshold points of the continuous latent variable, and the response is supposed to be determined by the interval in which the value of the latent variable falls. In the second type of model, considered by Rasch (1961, 1968) and discussed in some detail by Andersen (1973), a multidimensional parametric structure is postulated with respect to the response categories; and after the parameter estimation is carried out, checks on the possible reduction of the dimensionality of this structure are made.

Through a special case of this model, in which a single dimension is postulated immediately, Andersen (1977) and Andrich (in press–b) have revived interest in the "equidistant" scoring of the categories. First, Andersen has shown that if the sum of scores assigned to categories is to provide a sufficient statistic for a person's attitude and if certain plausible ordering properties among response patterns are to hold, then the categories must be scored in an "equidistant," or equivalently, integer-scoring fashion. Secondly, Andrich has interpreted the model in terms of thresholds on a latent continuum. This interpretation is set in a similar context to that of the traditional threshold formulation but is distinctively different from it. Thus, when it holds, the model formalizes in terms of a latent trait approach the intuitive procedure of the Likert-type questionnaires. (For full details of the statistical rationale and the interpretation of the model, the reader is referred to the above two papers. However, since they contain no applications, an operationalization and an application of this model are illustrated in the present paper.)

The Rating Response Model

The rating response model takes the form

$$p\{X=x \mid \beta_v, \delta_i, \underset{\sim}{\kappa}, m\} = \exp\{\kappa_x + x(\beta_v - \delta_i)\}/\gamma , \qquad [1]$$

where

$$\gamma = \gamma(\beta_v, \delta_i, \underset{\sim}{\kappa}, m) = \sum_{k=0}^{m} \exp\{\kappa_k + k(\beta_v - \delta_i)\} \qquad [1a]$$

is a normalizing factor.

In Model 1 the various variables are interpreted as follows. Firstly, it is proposed that when a person responds to an item in one of $m+1$ ordered response categories, the categories are separated by m ordered thresholds on a latent continuum. Secondly, the value x of random variable X, where $x \epsilon \{0, 1, \ldots, m\}$, designates the number of thresholds passed. Specifically, $x=0$ indicates a response in the first (or lowest) category, in which case no threshold is passed, while $x=m$ indicates a response in the last (or highest) category, in which case all thresholds are passed. Thirdly, and most importantly, the category coefficients κ_x are defined in terms of the thresholds τ_k, $k=1, m$ as follows:

$$\kappa_x = 0 \text{ for } x = 0, \qquad [2]$$

$$\kappa_x = -\sum_{k=1}^{x} \tau_k, \ x=1, m.$$

Finally, β and δ, respectively, represent the attitude of a person and the affective value of an item.

Two related points are stressed with respect to this model—points developed in Andrich (in press–b). Firstly, the scoring of successive categories with successive integers depends on the equal discrimination at the thresholds and not on equal distances between thresholds. This is in contrast to traditional formulations in which integer scoring is considered to depend on equal distances between thresholds. Secondly, the values of thresholds are estimated. The relationship of each category coefficient to the thresholds is that except for the first coefficient (which is always zero), it is the sum of the thresholds up to that category. The characteristic curves for the response categories in relation to the thresholds are shown in Figure 1 in the context of the illustrative example.

Before proceeding with the estimation, it is noted that when $m=1$—that is, when the response is dichotomous—Model 1 reduces to Rasch's simple logistic model (SLM). In the case $exp \ \kappa_x = \begin{bmatrix} m \\ x \end{bmatrix}$—that is, when the coefficients are specified in advance to be binomial—it reduces to the binomial logistic model (BLM) described in Andrich (1978, in press–a).

Estimation

Because Model 1 is a straightforward generalization of the SLM and the BLM, it is possible to approach its parameter estimation from the two main approaches which have been studied with respect to these special cases. They are the so-called *conditional* and *unconditional* approaches. In the conditional approach, sufficient statistics are identified for the person and category parameters; and then an equation, conditional on these statistics and containing only the item parameters, is obtained. With respect to a sample of data, a conditional likelihood equation may be derived which can be used to obtain maximum likelihood estimates of the item parameters. On the basis of the item parameter estimates so obtained, which are then treated as fixed and known quantities, maximum likelihood person and threshold parameters may be estimated.

Although this approach, exemplified by Andersen (1973) and Wright and Douglas (1977), is ideal theoretically, implementation problems for routine computer analyses do seem to arise in the solution algorithms. In particular, round-off errors in associated symmetric functions seem to mitigate against successful estimation with large numbers of items. Douglas (in press) has studied this approach with respect to the BLM; the problem is exaggerated as the number of categories is also increased. In the model of Equation 1, these same drawbacks seem to exist.

In the unconditional approach, used by Wright and Panchapakesan (1969), Wright and Douglas (1977), and Andrich (1978), a direct likelihood function is obtained; and the item and person parameters are effectively estimated simultaneously. While this approach is not hampered by large numbers of items, it has the weakness that the item parameters are not consistent. This weakness, discussed by Andersen (1973) for the SLM, is due to the retention of the "incidental" person parameters (which vary from person to person) in the estimation. Fortunately, a simple correction factor, which gives estimates identical to those obtained in the conditional approach in the case of two items and two response categories, does exist. Wright and Douglas provide a rationale for this factor in the SLM; the factor will be discussed later with respect to the model studied in this paper.

Unconditional Estimation

If N subjects respond to I items in one of $m+1$ ordered categories which are successively scored from 0 to m, then the joint probability of the data matrix $[x_{vi}]$, given the parameter vectors $\underset{\sim}{\kappa}, \underset{\sim}{\beta}, \underset{\sim}{\delta}$, is given by

$$p\{[x_{vi}] | \underset{\sim}{\kappa}, \underset{\sim}{\beta}, \underset{\sim}{\delta}\} = \prod_{v=1}^{N} \prod_{i=1}^{I} \left[\frac{\exp(\kappa_{x_{vi}} + x_{vi}(\beta_v - \delta_i))}{\sum\limits_{k=0}^{m} \exp(\kappa_k + k(\beta_v - \delta_i))} \right]. \qquad [3]$$

Expanding, simplifying, and taking the logarithm of Equation 3 gives

$$
\log L = \sum_x T_x \kappa_x + \sum_v r_v \beta_v - \sum_i S_i \delta_i - \sum_{vi} \log \left[\sum_k \exp(\kappa_k + k(\beta_v - \delta_i)) \right], \quad [4]
$$

where T_x is the total number of responses with respect to all subjects and all items which are in the category x;

$r_v = \sum_i x_{vi}$ is the total score of subject v;

$S_i = \sum_v x_{vi}$ is the total score of item i; and

L is the likelihood.

It can now be seen readily from the logarithmic form of Equation 4, and according to the factorization theorem of sufficient statistics, that the total person score r_v is a sufficient statistic for β_v. Similarly, it can be seen that T_k and S_i are respectively sufficient statistics for κ_x and δ_i. These are key features of the model, which not only make it relatively tractable, but also provide very clear connections with more intuitive approaches to the analysis of questionnaires with ordered response categories. This is especially the case with respect to finding an index for a person's attitude.

Reparameterization

By differentiating Equation 4 with respect to each of κ_x, β_v, and δ_i and equating the resultant expressions to zero, maximum likelihood equations may be obtained directly. Before doing so, however, it proves expedient to anticipate two dependencies in the resultant equations: one among the item parameters and one among the category parameters. These can readily be seen from Equation 3, since the probability statement is unchanged if a constant is added to each δ_i, providing it is also added to each β_v; and it is unchanged if a constant is added to each κ_x.

Adding the equation $\sum_i \hat{\delta_i} = 0$ provides an origin and a simple method for placing a constraint on the items. The most efficient and useful method for handling the indeterminacy on the categories is to reparameterize them in a way analogous to the procedure used by Bock (1972) for a similar model in which the categories are nominal rather than ordered.

Reparameterizing according to

$$
\underset{(m+1)\times 1}{\kappa} = \underset{(m+1)\times m}{B} \ \underset{m\times 1}{\tau} \ ,
$$

[5]

where the matrix **B** (called the basis matrix) takes the form

$$
\underset{(m+1)\,\times\,m}{B} =
\begin{bmatrix}
0 & 0 & 0 & \cdots & 0 \\
-1 & 0 & 0 & \cdots & 0 \\
-1 & -1 & 0 & \cdots & 0 \\
-1 & -1 & -1 & \cdots & 0 \\
\cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & & \cdot \\
-1 & -1 & -1 & & -1
\end{bmatrix}
\qquad [6]
$$

reduces by 1 the number of parameters to be estimated. This overcomes the indeterminacy and provides immediately an estimate of the thresholds rather than the category coefficients. This effect is ideal, since it is the thresholds and not the category coefficients which can be interpreted directly.

The Solution Equations

With the above general reparameterization, the log likelihood takes the form

$$
\log L = \sum_x T_x \sum_k b_{xk}\tau_k + \sum_v r_v\beta_v - \sum_i S_i\delta_i \qquad [7]
$$

$$
- \sum_{vi} \log \left[\sum_x \exp(\sum_k b_{xk}\tau_k + x\beta_v - x\delta_i) \right],
$$

where b_{xk} is the (x, k)th element of **B**. In the estimation, the following relationships, which can be shown readily, help simplify the resulting expression considerably:

$$
\partial\{\log \sum_{x=0}^{m} \exp(\sum_k b_{xk}\tau_k + x\beta_v - x\delta_i)\}/\partial\delta_i = \sum_{x=0}^{m} xp_{xvi} , \qquad [8]
$$

$$
\partial\{\log \sum_{x=0}^{m} \exp(\sum_k b_{xk}\tau_k + x\beta_v - x\delta_i)\}/\partial\tau_k = \sum_{x=0}^{m} b_{xk}p_{xvi} , \qquad [9]
$$

$$\partial\{\log \sum_{x=0}^{m} \exp(\sum_{k} b_{xk}\tau_{k} + x\beta_{v} - x\delta_{i})\}/\partial\beta_{v} = \sum_{x=0}^{m} xp_{xvi} \quad . \tag{10}$$

Using the above results, differentiating Equation 7 with respect to δ_i, τ_k, and β_v, and equating the resultant expressions to 0 provides the following solution equations:

$$\frac{\partial \log L}{\partial \delta_i} = -S_i + \sum_{vx} xp_{xvi} = 0, \quad i=1, I, \tag{11}$$

$$\frac{\partial \log L}{\partial \tau_k} = \sum_{x} (T_x - \sum_{vi} p_{xvi})b_{xk} = 0, \quad k=1, m, \tag{12}$$

$$\frac{\partial \log L}{\partial \beta_v} = r_v - \sum_{ix} xp_{xvi} = 0, \quad v=1, N. \tag{13}$$

A Solution Algorithm

Because there are usually many more persons than items, more than one person will have the same total score. Then, because all people with the same total score will have the same parameter estimate—a consequence of the sufficient statistic for β_v and clearly evident from Equation 13—it is necessary to estimate only a parameter with respect to each total score. With these observations, the parameter estimation procedure is divided into three stages and is analogous to that used by Wright and Panchapakesan (1969) and Wright and Douglas (1977) for the SLM and that used by Lord (1968) and Bock (1972) for other similar latent trait models. Because the procedure has become familiar in the literature, these stages will now be described only in brief and general terms.

In the first stage of the estimation, person and threshold parameters are provisionally fixed, and Equation 11 is solved for the item parameters. These implicit equations (as are the others) are solved iteratively using the Newton-Raphson algorithm. Before proceeding to the second stage, the constraint $\sum_i \hat{\delta}_i = 0$ is imposed on these estimates. In the second stage, these new item parameters are considered fixed, and Equation 12 is solved for the threshold parameters. In the third stage, with both item and threshold parameters fixed, Equation 13 is solved for $\hat{\beta}_r$, the attitude estimate of any person with a total score of r. Although the possible total scores for a person range from 0 to Im, the β estimates corresponding to the two extreme scores are not finite; consequently, Equation 13 is solved only for $r= 1, Im - 1$, and any person with either of these extreme scores is eliminated in the estimation. With these new person parameters taken as fixed, Stage 1 and Stage 2 are repeated. When two successive sets of solutions for the items and threshold parameters differ by less than some arbitrary small value, the iterative process is terminated.

Inconsistency of Estimates

As was indicated earlier, the item parameter estimates obtained for the SLM from the above unconditional maximum likelihood approach are not consistent. That is, for a fixed number of items, and with a sample size increasing without limit, the parameter estimates do not tend to the actual values of the parameters. The correction factor for the SLM, for which Wright and Douglas (1977) provide a rationale, involves multiplying the obtained item parameter estimates, $\hat{\delta}_i$, by $(I-1)/I$. In the extended model described above, the same correction factor appears to be appropriate for the items, but a correction factor of $(m-1)/m$ also appears necessary for the thresholds. The details of these observations, which have been made from a series of simulations, are not reported in this paper. However, a full discussion on the nature of the inconsistency of estimates and the effectiveness of appropriate correction factors with respect to this model would be instructive. With the corrected item and threshold parameters, a revised set of person parameter estimates can be obtained by re-solving Equation 13.

Standard Errors of Estimates

In maximum likelihood estimation, the negative inverse of the second derivatives provides the limiting variance of the parameter estimates. Their square roots therefore provide large sample standard errors. With respect to each set of parameters, the second derivatives are used in the Newton-Raphson algorithm and therefore are routinely available. However, the appropriate standard errors for the corrected item and threshold values are estimated by evaluating them after correction for bias.

These second derivatives take the following simple forms:

$$\frac{\partial^2 \log L}{\partial \delta_i^2} = \sum_v \left\{ \left(\sum_x x p_{xvi} \right)^2 - \left(\sum_x x^2 p_{xvi} \right) \right\} , \; i=1, \; I, \quad [14]$$

$$\frac{\partial^2 \log L}{\partial \tau_k^2} = \sum_{vi} \left\{ \left(\sum_x b_{xk} p_{xvi} \right)^2 - \left(\sum_x b_{xk}^2 p_{xvi} \right) \right\} , \; k=1, \; m, \quad [15]$$

$$\frac{\partial^2 \log L}{\partial \beta_v^2} = \sum_i \left\{ \left(\sum_x x p_{xvi} \right)^2 - \left(\sum_x x^2 p_{xvi} \right) \right\} , \; v=1, \; N. \quad [16]$$

An Example

The set of data analyzed for illustrative purposes involves the responses of 273 Asian and Australian students from the University of Western Australia. These students answered 20 items pertain-

ing to an attitude to learning. The items and results, provided by Mastuhu (1977), were designed to provide information about the relative independence or dependence from authority that these students showed with respect to their learning in a University environment. These responses were used to compare the attitudes of Australian and Asian students. As examples, three of the items were

> I like to learn things which are definite.
> I enjoy teasing ideas out in an argument.
> I am confident that the lecturer always gives correct information in lectures.

The students responded to these questions across a continuum of strongly agree (SA), agree (A), disagree (D), and strongly disagree (SD), in the Likert tradition. However, no neutral category was provided because of the apparent general consensus in the literature that this category can attract responses other than those representing simple neutrality, such as indifference, misunderstanding, and the like (cf. Dubois & Burns, 1976). Four of the items indicated dependence on authority, while the others indicated independence. The scoring of the former were reversed, making a high score indicate greater independence. The analysis described below pertains to checking the internal consistency or unidimensionality of the items.

Results

The goodness of fit of the responses to the model was evaluated by a Pearsonian chi-square statistic generalized from the one described by Wright and Panchapakesan (1969) for the SLM and that described in Andrich (in press–a) for the BLM. Because the number of persons in each possible score group was small, people were grouped into four adjacent class intervals of approximately equal size.

Table 1 provides the values of the test statistic together with the associated probabilities under the hypothesis of the fit of the data to the model. Two analyses are shown, one containing the entire set of 20 items and the other containing a subset of 16 items, the 4 items showing misfit to the model having been deleted. The chi-square for each item was considered to be distributed with approximately 3 degrees of freedom—one less than the number of class intervals G. For the joint test of fit the number of degrees of freedom was taken to be $I(G-1)-(G-1)-m = (I-1)(G-1)-m$, where $G-1$ degrees of freedom correspond to the number of constraints across class intervals and m degrees of freedom correspond to the number of category parameters estimated. These figures are also shown in Table 1.

It is evident from this table that the entire set of 20 items does not accord with the model, although the individual item fit suggests that this is due primarily to the poor fit of three or four items. Because the individual item fit test is conservative, four rather than only three of the least well-fitting items—Items 3, 5, 11, and 12—were deleted and the analyses repeated. With the reduced set of items, the fit according to the χ^2 criterion is very satisfactory, neither too good nor too bad.

The above test of fit, however, is a very general one and therefore does not disclose certain specific violations of the model. In particular, scores are amalgamated among categories so that the test does not indicate specifically whether the responses *among* categories accord with expectation. In order to get an orientation to this aspect of the data, the average attitude $\hat{\beta}_g$ of the people in each class interval g was calculated. Then, the probability that a person of attitude $\hat{\beta}_g$ responded in category x to item i was obtained by entering $\hat{\beta}_g$ and $\hat{\delta}_i$ into the probability statement of the model—Equation 1. If the responses accord with the model, then these probabilities should be similar to the proportion of persons in each class interval who respond in each category to each item. The patterns of observed proportions and model probabilities are illustrated in Figure 1 with Items 6, 15, and 7. These items were chosen because they represent a spread across the continuum.

Table 1
Goodness of Fit of 20-Item and 16-Item Sets

Item Number[a]	20-item set (χ^2 with approx. 3df)		16-item set (χ^2 with approx. 3df)	
19	0.82	$p < .80$	0.88	$p < .80$
20	1.47	.50	1.24	.70
6	1.83	.50	0.35	.95
13	1.99	.50	6.46	.05
10	2.23	.50	2.61	.30
14	2.23	.50	3.99	.20
15	2.52	.30	1.61	.50
17	2.85	.30	3.56	.30
7	3.51	.30	0.63	.80
8	3.81	.20	3.46	.30
1	4.33	.20	4.39	.20
9	5.03	.10	4.78	.10
4	5.42	.10	5.53	.10
16	6.10	.10	1.36	.70
18	6.45	.10	7.75	.05
2	6.56	.05	5.24	.10
11	8.96	.02	--	--
3	10.31	.02	--	--
5	11.32	.01	--	--
12	24.37	.01	--	--

Total χ^2 111.35 with approx.   53.83 with approx.
54 df, $p < .001$ 42 df, $p < 0.20$

[a]For ease of visual inspection, items have been ordered according to the fit on the first analysis.

It is apparent from Figure 1 that the corresponding proportions and probabilities do tend to follow each other (with the probability of a high score increasing as either the person attitude increases or the item affective value decreases), although, as expected in a probabilistic model, there were discrepancies. If the model does account for the data, however, then these discrepancies, with respect to all class intervals and all items, should not show any systematic pattern.

To investigate the pattern of these discrepancies, each observed proportion which was greater than the corresponding model probability was classed as a positive residual. The number of positive residuals with respect to all class intervals and all items are shown in Table 2. If the items do accord with the model, then about half of the residuals in each category (shown as expected frequencies in Table 2) should be positive. From this Table it is evident that the number of positive residuals, in both the 20-item and 16-item sets, was not untenable under a random sampling hypothesis.

With no evidence that the 16-item set did not accord with the model, the various parameter estimates are shown in Tables 3 and 4. For comparison purposes the figures for the 20-item set as well as

Figure 1
**Observed proportions (+) and model probabilities of responses in each category for
3 items and 4 class intervals**

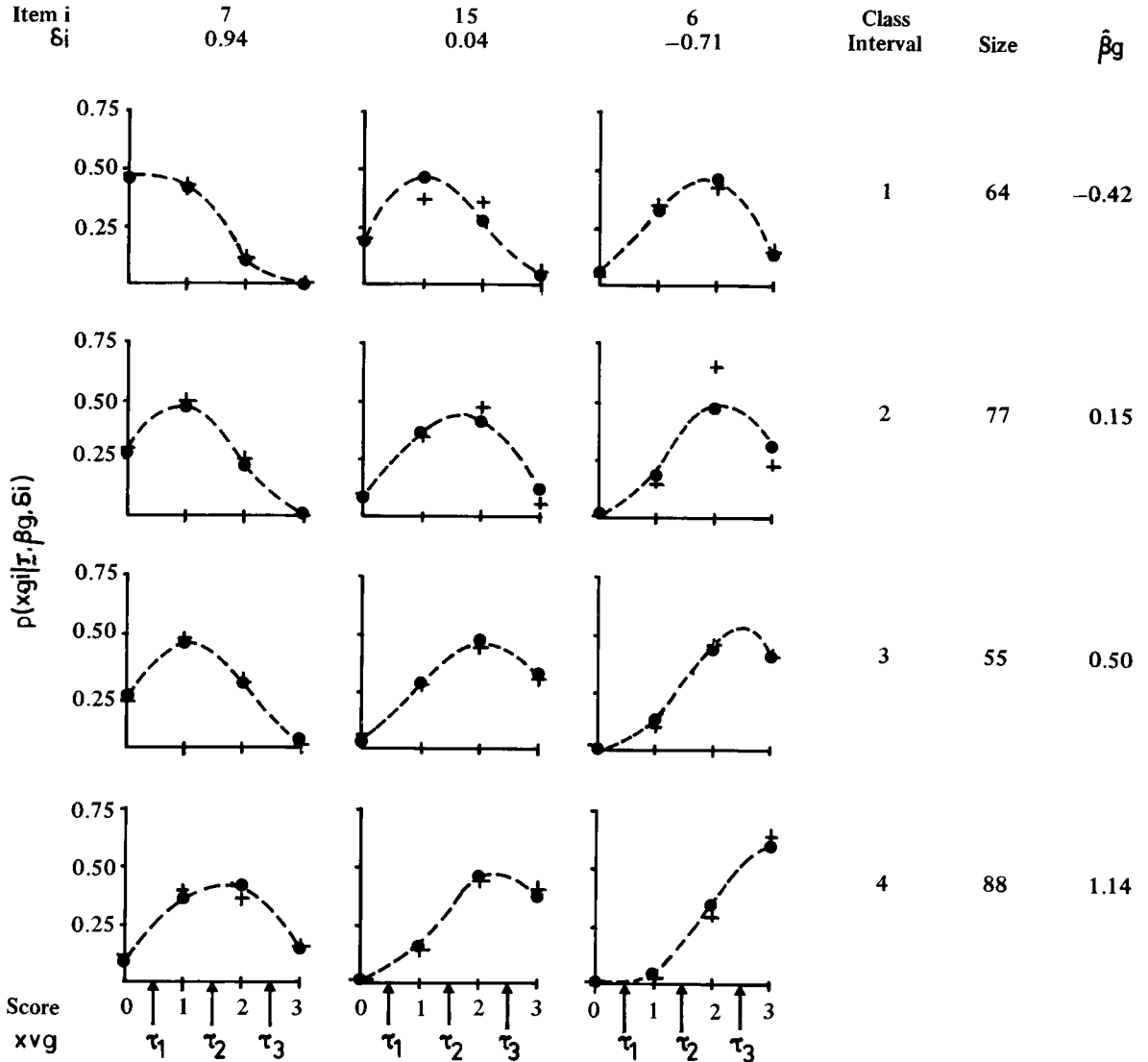For ease of visual inspection, adjacent model probabilities have been joined by curves.

Table 2
Number of Positive Residuals with
Respect to Each Response Category

	20-Item Set				
Category	SD	D	A	SA	
Score	0	1	2	3	Total
Observed	46	34	38	39	157
Expected	40	40	40	40	160
	16-Item Set				
Observed	32	27	30	33	122
Expected	32	32	32	32	128

the 16-item set are shown. One of the most interesting results with respect to the parameter estimates is that of the threshold estimates shown in Table 3. In both the 20- and 16-item sets, and more closely with the latter set with only the conformable items, not only were the distances between thresholds statistically equivalent but these distances were one unit. That is, within 95% confidence limits for the estimates, the respective threshold estimates were −1.0, 0.0, and 1.0.

Three points should perhaps be stressed in relation to this result. Firstly, although the threshold estimates are constrained to sum to zero, there are no further constraints placed upon them. Secondly, these values follow after a correction for bias, where the original estimates were multiplied by a factor of $(m-1)/m$; the appropriateness of this factor, as mentioned earlier, requires further study. Finally, it is reiterated that the equality of distances between successive thresholds is not what governs the appropriateness of the integer scoring; it is instead an equality of discriminations at the thresholds.

With unit distances between threshold values centered at 0.0, Figure 2 displays the category characteristic curves as a function of the difference between a person's attitude, β, and an item's affective value, δ. Clearly, as the attitude increases relative to an item's affective value, the probability of a high score increases.

Discussion

Unlike the simple logistic specialization of the general multidimensional Rasch model which has been studied extensively with respect to its implementation and application to both achievement and attitude tests, other specializations have received relatively little attention. This paper has dealt with one of these alternate specializations, which should prove useful in educational and psychological scaling applications.

The special feature of the model is the integer scoring of categories, even though the thresholds between them are parameterized separately, and its relatively successful application in the illustrative example. Perhaps the most desirable aspect of the integer-scoring function, an aspect which has been mentioned before, is that a person's attitude is a function of his/her total score with respect to items. Thus, in the first step of obtaining an attitude estimate, one proceeds as in the usual summated rating approach. For the second step, which requires a non-linear transformation of the total score, the con-

Table 3
Item and Threshold Estimates for
20-Item and 16-Item Sets

| Item i	20-Item Set		16-Item Set	
	$\hat{\delta}_i$	$SE(\hat{\delta}_i)$	$\hat{\delta}_i$	$SE(\hat{\delta}_i)$
1	.13	.07	.04	.07
2	.61	.07	.50	.07
3	-.71	.08	--	--
4	-.34	.08	-.43	.07
5	-1.20	.09	--	--
6	0.63	.08	-.71	.08
7	1.03	.07	.94	.07
8	-.44	.08	-.52	.08
9	.33	.07	.23	.07
10	-.06	.07	-.14	.07
11	-.59	.08	--	--
12	1.02	.07	--	--
13	-.18	.07	-.27	.07
14	.34	.07	.24	.07
15	.13	.07	.04	.07
16	-.54	.08	-.65	.08
17	.29	.07	.18	.07
18	1.59	.08	1.46	.08
19	-.72	.08	-.79	.08
20	-.04	.07	-.11	.07
Threshold k	$\hat{\tau}_k$	$SE(\hat{\tau}_k)$	$\hat{\tau}_k$	$SE(\hat{\tau}_k)$
1	-1.06	.05	-1.00	.05
2	.05	.03	-0.02	.04
3	1.01	.03	1.02	.04

version table can be made readily available from the corresponding item analysis. The familiarity of these steps with the SLM should facilitate the application of the model described to Likert questionnaires.

In addition to these advantageous properties for person measurement, the model has the same advantageous properties for item analysis as those discussed in Andrich (1978, in press–a) with respect to the model where the thresholds are not estimated, but the category coefficients are specified to be binomial. First, checks on the internal consistency of items can be made with explicit probability statements. Secondly, and unlike traditional analyses of Likert questionnaires, estimates of scale values which reflect affective characteristics of items are provided.

Table 4
Attitude Estimates and Standard Errors
Corresponding to Every Third Possible
Score for 20-Item and 16-Item Sets

Score r	20-Item Set		16-Item Set	
	$\hat{\beta}_r$	$SE(\hat{\beta}_r)$	$\hat{\beta}_r$	$SE(\hat{\beta}_r)$
1	-4.24	1.01	-3.90	1.01
4	-2.78	0.52	-2.45	0.52
7	-2.15	0.41	-1.82	0.41
10	-1.72	0.35	-1.39	0.36
13	-1.38	0.32	-1.04	0.33
16	-1.09	0.31	-0.73	0.31
19	-0.83	0.29	-0.46	0.30
22	-0.59	0.28	-0.19	0.30
25	-0.37	0.27	0.07	0.30
28	-0.15	0.27	0.34	0.30
31	0.07	0.27	0.61	0.31
34	0.29	0.27	0.91	0.33
37	0.51	0.28	1.26	0.35
40	0.75	0.28	1.67	0.40
43	1.00	0.30	2.24	0.48
46	1.28	0.32	3.24	0.73
49	1.60	0.34	--	--
52	2.00	0.39	--	--
55	2.54	0.47	--	--
58	3.53	0.72	--	--

Figure 2
**Obtained Characteristic Curves for Four Categories as a Function of the Difference
Between Person Attitude and Item Affective Value**
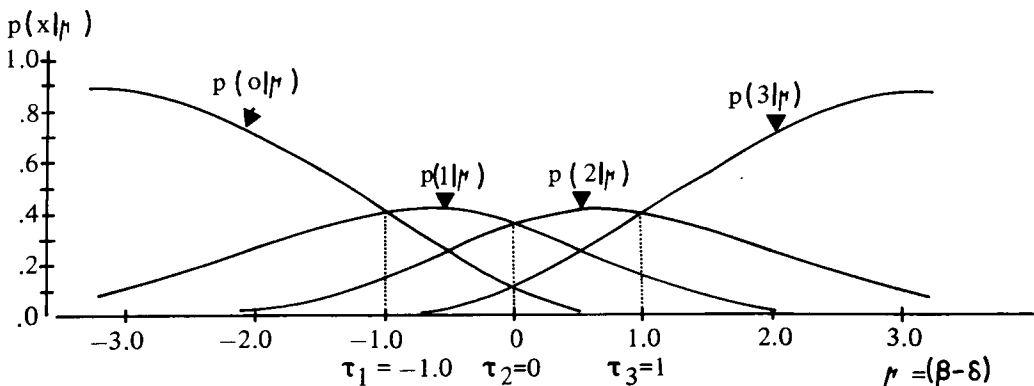
References

Andersen, E. B. Conditional inference for multiple-choice questionnaires. *British Journal of Mathematical and Statistical Psychology,* 1973, *26,* 31–44.

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

Andrich, D. A binomial latent trait model for the study of Likert style attitude questionnaires. *British Journal of Mathematical and Statistical Psychology,* 1978, *31,* 84–98.

Andrich, D. Scaling attitude items constructed and scored in the Likert tradition. *Educational and Psychological Measurement,* in press. (a)

Andrich, D. A rating formulation for ordered response categories. *Psychometrika,* in press. (b)

Bock, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika,* 1972, *37,* 29–51.

Douglas, G. A. Conditional maximum likelihood estimation for a multiplicative binomial response model. *British Journal of Mathematical and Statistical Psychology,* in press.

Dubois, B., & Burns, J. A. An analysis of meaning of the question mark response category in attitude scales. *Educational and Psychological Measurement,* 1975, *35,* 869–884.

Kolakowski, D., & Bock, R. D. *A fortran IV program for maximum likelihood item analysis and test scoring: Logistic model for multiple item responses* (Research Memorandum No. 13). University of Chicago, Department of Education, Statistical Laboratory, 1972.

Lord, F. M. An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement,* 1968, *28,* 989–1200.

Mastuhu. *A comparison of the attitudes to learning of Asian, Australian, and European students at the University of Western Australia.* Unpublished Master's thesis, University of Western Australia, 1977.

Rasch, G. On general laws and the meaning of measurement in psychology. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. IV.* Berkeley, CA: University of California Press, 1961, 321–34.

Rasch, G. *A mathematical theory of objectivity and its consequences for model construction.* European Meeting on Statistics, Econometrics, and Management Science, Amsterdam, September 1968.

Samejima, F. Estimation of latent ability using a response pattern for graded scores. *Psychometrika,* Monograph Supplement No. 17, 1969.

Wright, B. D., & Panchapakesan, N. A. A procedure for sample-free item analysis. *Educational and Psychological Measurement,* 1969, *29,* 23–48.

Wright, B. D., & Douglas, G. A. Best procedures for sample-free item analysis. *Applied Psychological Measurement,* 1977, *1,* 281–295.

Author's Address

David Andrich, Department of Education, The University of Western Australia, Nedlands, Western Australia, 6009.