

# Fitting a Polytomous Item Response Model to Likert-Type Data

Eiji Muraki  
Educational Testing Service

This study examined the application of the MML-EM algorithm to the parameter estimation problems of the normal ogive and logistic polytomous response models for Likert-type items. A rating-scale model was developed based on Samejima's (1969) graded response model. The graded response model includes a separate slope parameter for each item and an item response parameter. In the rating-scale model, the item response parameter is resolved into two parameters: the item location parameter, and the category threshold parameter characterizing the boundary between response categories. For a Likert-type questionnaire,

where a single scale is employed to elicit different responses to the items, this item response model is expected to be more useful for analysis because the item parameters can be estimated separately from the threshold parameters associated with the points on a single Likert scale. The advantages of this type of model are shown by analyzing simulated data and data from the General Social Surveys. *Index terms:* EM algorithm, General Social Surveys, graded response model, item response model, Likert scale, marginal maximum likelihood, polytomous item response model, rating-scale model.

An item response model expresses a probabilistic relationship between an examinee's performance on a test item and the examinee's latent trait. Several types of item response models have been proposed. The applicability of dichotomous item response models (Birnbaum, 1968; Lord, 1980; Lord & Novick, 1968; Rasch, 1960/1980) to cognitive item response data has been extensively studied.

Several polytomous item response models have been formulated based on these dichotomous models (Andrich, 1978; Bock, 1972; Masters, 1982; Rost, 1988; Samejima, 1969). Bock's (1972) nominal response model is applicable to a test item in which the response options are not necessarily ordered. If the options on a rating scale are successively ordered, as in a Likert scale (Likert, 1932), then a graded response model (Samejima, 1969) or a Rasch family of polytomous item response models—including the partial credit model (Masters, 1982), the rating-scale model (Andrich, 1978), and the successive interval model (Rost, 1988)—are more appropriate.

Bock and Lieberman (1970) proposed the marginal maximum likelihood (MML) method to estimate the parameters of item response models. Although this method is applicable to virtually any type of item response model, their procedure is not computationally practical for large numbers of items. Influenced by Dempster, Laird, and Rubin's (1977) EM algorithm, Bock and Aitkin (1981) reformulated the Bock-

Lieberman MML method. This revised MML method, the so-called MML-EM algorithm, is computationally feasible for both small and large numbers of items. It has been applied to the estimation problems of the two-parameter normal ogive model (Bock & Aitkin, 1981), the one-parameter logistic model (Thissen, 1982), and Thurstone's multiple-factor model (Bock & Aitkin, 1981; Bock, Gibbons, & Muraki, 1988; Thurstone, 1947). Other applications include a time-dependent model for the maintenance of parallel test forms (Bock, Muraki, & Pfeiffenberger, 1988) and the polytomous item response models of Samejima (1969) and Bock (1972), for which Thissen (1986) has developed a computer program.

In this study, Samejima's (1969) polytomous item response model is modified and called a rating-scale model, and the MML-EM algorithm is applied to estimation problems involving this special case of her model. The main purpose of this study is to show that the MML-EM algorithm can be used to obtain estimates of this model's parameters and that it facilitates the analysis of Likert-type response data.

### The Categorical Item Response Model

The operating characteristic (Samejima, 1972) is the essential feature of polytomous item response models. It expresses how the probability of a specific categorical response is formulated according to the laws of probability and the psychological assumptions regarding item response behavior.

In Samejima's model, a person's probability of responding in category  $k$  to a specific item  $j$ ,  $P_{jk}(\theta)$ , is obtained by subtracting the probability of responding in or below category  $k - 1$  from the probability of responding in or below category  $k$ . The operating characteristic of the graded item response for the latent trait variable  $\theta$  can therefore be expressed as

$$P_{jk}(\theta) = P_{jk}^+(\theta) - P_{j,k-1}^+(\theta) \quad , \quad (1)$$

where  $P_{jk}^+(\theta)$  and  $P_{j,k-1}^+(\theta)$  constitute the response model of the binary item for all response categories less than  $k$  and  $k - 1$ .

In the normal ogive model (Samejima, 1969), the formula for  $P_{jk}^+(\theta)$  is given by

$$P_{jk}^+(\theta) = \int_{-\infty}^{a_j(\theta - b_j)} \phi(x) dx \quad , \quad (2)$$

where  $\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ ,

$a_j$  is a discrimination parameter for item  $j$ , and

$b_{jk}$  is an item response parameter for item  $j$  and category  $k$ .

For each  $m$ -category item, there are  $m - 1$  category boundaries. Because the probability of the categorical response is determined by the lower boundary of the category  $k$  that the respondent endorses and the boundary that the respondent fails to endorse, the values of the category threshold parameter increase successively.

The rating-scale model is a special case of Samejima's graded response model, that is,

$$P_{jk}(\theta) = \int_{a_j(\theta - b_j + c_{k-1})}^{a_j(\theta - b_j + c_k)} \phi(x) dx \quad . \quad (3)$$

The logistic form of the rating-scale model is

$$P_{jk}(\theta) = \frac{\exp[Da_j(\theta - b_j + c_k)]}{1 + \exp[Da_j(\theta - b_j + c_k)]} - \frac{\exp[Da_j(\theta - b_j + c_{k-1})]}{1 + \exp[Da_j(\theta - b_j + c_{k-1})]} \quad , \quad (4)$$

where  $D$  is a scaling constant that puts the  $\theta$  scale in the same metric as the normal ogive model. In the discussion that follows, the slope parameter, the item location parameter, and the category threshold parameter are respectively denoted  $a_j$ ,  $b_j$ , and  $c_k$ .

When the items in a questionnaire differ in their number of response categories,  $m_j$  ( $j = 1, 2, \dots, n$ ), Samejima's model in Equation 2 may be more appropriate than the polytomous item response model in Equations 3 or 4. However, for a Likert-type questionnaire, where only a single scale is employed to evoke different responses for test items, the rating-scale model is expected to be more useful because the parameters associated with the items and the  $c_k$  parameters associated with the points on a single Likert scale can be estimated separately. The items can be ordered by their  $b_j$  estimates, and the psychological distance between the points on the Likert scale can be estimated independently of the item parameters.

The scaling of the widths between the category boundaries independent of the  $b_j$ s was introduced by Thurstone (Edwards & Thurstone, 1952) as the method of successive categories, and its parameter estimation was discussed by Bock and Jones (1968). The rating-scale model in Equation 3, therefore, can be thought of as a latent trait model derived from Thurstone's scaling tradition. In contrast, Andrich's rating-scale model belongs to a Rasch family of polytomous item response models and it was formulated using a different operating characteristic (Andrich, 1978). Although both rating-scale models are suitable for the analysis of Likert-type response data and their item threshold parameters are functionally equivalent, their parameterizations of categorical components are quite distinctive (Muraki, 1983).

### Parameter Estimation

Let  $U_{jki}$  represent an element in the matrix of the observed response pattern  $i$ .  $U_{jki} = 1$  if item  $j$  is rated by the  $k$ th category of a Likert scale, otherwise  $U_{jki} = 0$ . By the principle of local independence (Birnbau, 1968), the conditional probability of a response pattern  $i$ , given  $\theta$ , for  $m$  response categories and  $n$  items, as denoted by a response matrix  $(U_{jk})$ , is the joint probability:

$$P_i[(U_{jk})|\theta] = \prod_{j=1}^n \prod_{k=1}^m [P_{jk}(\theta)]^{U_{jki}} \quad (5)$$

For examinees randomly sampled from a population with a normal distribution of the latent trait variable,  $\phi(\theta)$ , the unconditional probability of the observed response pattern  $i$  is

$$P_i(U_{jk}) = \int_{-\infty}^{\infty} P_i[(U_{jk})|\theta]\phi(\theta) d\theta \quad (6)$$

If an examinee responds to  $n$  items with  $m$  categories, his/her response pattern  $i$  can then be assigned to one of  $m^n$  mutually exclusive patterns. Let  $r_i$  represent the number of examinees observed in such a pattern  $i$ , and let  $N$  be the total number of examinees sampled from the population. Then  $r_i$  is multinomially distributed with parameters  $N$  and  $P_i(U_{jk})$ . This probability can be interpreted as the likelihood function of the parameters  $a_j$ ,  $b_j$ , and  $c_k$ :

$$L = \frac{N!}{\prod_{i=1}^{m^n} r_i!} \prod_{i=1}^{m^n} [P_i(U_{jk})]^{r_i} \quad (7)$$

Taking the natural logarithm of Equation 7 yields

$$\ln L = \ln N! - \ln \sum_{i=1}^{m^n} r_i! + \sum_{i=1}^{m^n} r_i \ln P_i(U_{jk}) \quad (8)$$

The likelihood equations for  $\hat{a}_j$ ,  $\hat{b}_j$ , and  $\hat{c}_k$  can be derived from the first derivatives of Equation 8 with respect to each parameter, and respectively set to 0.

### Item Parameter Estimation

Let  $v_j$  represent the parameters  $a_j$  and  $b_j$ . With respect to  $v_h$ , which is the parameter  $v_j$  for the specific item  $j = h$ , the likelihood in Equation 8 can be differentiated as

$$\frac{\partial \ln L}{\partial v_h} = \sum_{i=1}^{m^n} \frac{r_i}{P_i(U_{jk})} \int_{-\infty}^{\infty} P_i[(U_{jk})|\theta] \sum_{k=1}^m \frac{\partial [P_{hk}(\theta)]^{U_{hki}}}{\partial v_h} \frac{\phi(\theta) d\theta}{[P_{hk}(\theta)]^{U_{hki}}} \quad (9)$$

Now let the observed score patterns be indexed by  $\ell = 1, 2, \dots, S$ , where  $S \leq \min(N, m^n)$ . If the number of examinees with response pattern  $\ell$  is denoted by  $r_\ell$ , then

$$\sum_{\ell=1}^S r_\ell = N \quad (10)$$

The first derivative of the likelihood function in Equation 9 can be approximated by using the Gauss-Hermite quadrature, such that

$$\frac{\partial \ln L}{\partial v_h} \approx \sum_{\ell=1}^S \sum_{f=1}^F \frac{r_\ell L_\ell A(X_f)}{\bar{P}_\ell} \sum_{k=1}^m \frac{\partial [P_{hk}(X_f)]^{U_{hkf}}}{\partial v_h} \frac{1}{[P_{hk}(X_f)]^{U_{hkf}}} \quad (11)$$

where

$$\bar{P}_\ell = \sum_{f=1}^F \prod_{j=1}^n \prod_{k=1}^m [P_{jk}(X_f)]^{U_{jkt}} A(X_f) \quad (12)$$

and

$$L_\ell = \prod_{j=1}^n \prod_{k=1}^m [P_{jk}(X_f)]^{U_{jkt}} \quad (13)$$

In Equation 11,  $A(X_f)$  is the weight of the Gauss-Hermite quadrature, and  $X_f$  is the quadrature point (Stroud & Secrest, 1966). The quadrature weight  $A(X_f)$  is approximately the standard normal probability density at the point  $X_f$ , such that  $\sum_{f=1}^F A(X_f) = 1$ . Because  $U_{hkf}$  can take on only two possible values, 1 and 0, Equation 11 can be rewritten as

$$\sum_{f=1}^F \sum_{k=1}^m \frac{\bar{r}_{hkf}}{P_{hk}(X_f)} \frac{\partial P_{hk}(X_f)}{\partial v_h} \quad (14)$$

where

$$\bar{r}_{hkf} = \frac{r_\ell L_\ell A(X_f) U_{hkf}}{\bar{P}_\ell} \quad (15)$$

and  $\bar{r}_{hkf}$  is the provisional expected frequency of the  $k$ th categorical response of item  $h$  at the  $f$ th quadrature point.

Bock and Aitkin (1981) applied the EM algorithm (Dempster et al., 1977) to estimate the parameters for each item individually, and then repeated the iteration process over  $n$  items until the estimates of all the items became stable to the required number of decimal places. This is in contrast to the Fisher-scoring procedure of Bock and Lieberman (1970). The  $q$ th cycle of the iterative process can be expressed as

$$\mathbf{v}_q = \mathbf{v}_{q-1} + \mathbf{V}^{-1} \mathbf{t} \quad (16)$$

where  $\mathbf{v}_q$  and  $\mathbf{v}_{q-1}$  are the parameter estimates of the  $q$ th and  $(q - 1)$ th cycles respectively,  $\mathbf{V}^{-1}$  is the inverse of the information matrix, and  $\mathbf{t}$  is the gradient vector. For the item parameter estimation, the elements of  $\mathbf{t}$  and  $\mathbf{V}$  are

$$t_{v_h} = \sum_{f=1}^F \sum_{k=1}^m \frac{\bar{r}_{hkf}}{P_{hk}(X_f)} \frac{\partial P_{hk}(X_f)}{\partial v_h} \quad (17)$$

$$V_{v_h \omega_h} = \sum_{f=1}^F \bar{N}_f \sum_{k=1}^m \frac{1}{P_{hk}(X_f)} \frac{\partial P_{hk}(X_f)}{\partial v_h} \frac{\partial P_{hk}(X_f)}{\partial \omega_h}, \quad (18)$$

where  $v_h = a_h$  or  $b_h$  and  $\omega_h = a_h$  or  $b_h$ .

In Equation 18,  $\bar{N}_f$  is called the provisional expected sample size at quadrature point  $f$ , and is computed by

$$\bar{N}_f = \sum_{\ell=1}^S \left[ \frac{r_\ell L_\ell A(X_f)}{\bar{P}_\ell} \right]. \quad (19)$$

The rigorous proof of the approximation of the second derivatives in Equation 18 by the product of the first derivatives is given by Kendall and Stuart (1973).

### Category Threshold Parameter Estimation

Because the category threshold parameter  $c_g$ , which is the parameter  $c_k$  for the specific category  $k = g$ , is contained in both  $P_{jg}(\theta)$  and  $P_{j,g+1}(\theta)$  as shown in Equation 1, the first derivative of the likelihood function in Equation 13 with respect to  $c_g$  can be derived by the following process:

$$\frac{\partial \prod_{j=1}^n \prod_{k=1}^m [P_{jk}(\theta)]^{U_{jki}}}{\partial c_g} = \prod_{j=1}^n \prod_{k=1}^m [P_{jk}(\theta)]^{U_{jki}} \sum_{j=1}^n \left[ \frac{U_{jgi}}{P_{jg}(\theta)} - \frac{U_{j,g+1,i}}{P_{j,g+1}(\theta)} \right] \frac{\partial P_{jg}^+(\theta)}{\partial c_g}. \quad (20)$$

Because  $U_{jgi} = 1$  or  $0$ ,  $U_{j,g+1,i} = 1$  or  $0$ ,

$$\frac{\partial P_{jg}(\theta)}{\partial c_g} = \frac{\partial P_{jg}^+(\theta)}{\partial c_g} \quad (21)$$

and

$$\frac{\partial P_{j,g+1}(\theta)}{\partial c_g} = - \frac{\partial P_{j,g+1}^+(\theta)}{\partial c_g}, \quad (22)$$

then

$$\frac{\partial \ln L}{\partial c_g} = \sum_{i=1}^{m^n} \frac{r_i}{P_i(U_{jki})} \int_{-\infty}^{\infty} \prod_{j=1}^n \prod_{k=1}^m [P_i(U_{jk}|\theta)]^{U_{jki}} \sum_{j=1}^n \left[ \frac{U_{jgi}}{P_{jg}(\theta)} - \frac{U_{j,g+1,i}}{P_{j,g+1}(\theta)} \right] \frac{\partial P_{jg}^+(\theta)}{\partial c_g} \phi(\theta) d\theta. \quad (23)$$

The element of the gradient vector for the  $c_k$  parameter estimation can be numerically computed as

$$t_{c_g} = \sum_{f=1}^F \sum_{j=1}^n \left[ \frac{\bar{r}_{jgf}}{P_{jg}(X_f)} - \frac{\bar{r}_{j,g+1,f}}{P_{j,g+1}(X_f)} \right] \frac{\partial P_{jg}^+(X_f)}{\partial c_g}. \quad (24)$$

Likert scales employing more than 10 categories are not commonly used in psychological or sociological research. If the number of categories is small, say  $m \leq 12$ , the Fisher-scoring method can be used to estimate the parameters without difficulty. Furthermore, because the covariance between any categories that differ by more than two points is simply 0, the information matrix becomes an  $(m - 1) \times (m - 1)$  tridiagonal symmetric matrix. The generation and inversion of this matrix is relatively simple. The elements of the information matrix are

$$V_{c_g c_g} = \sum_{f=1}^F \bar{N}_f \sum_{j=1}^n \left[ \frac{1}{P_{jg}(X_f)} + \frac{1}{P_{j,g+1}(X_f)} \right] \left[ \frac{\partial P_{jg}^+(X_f)}{\partial c_g} \right], \quad (25)$$

$$V_{c_g c_{g-1}} = \sum_{f=1}^F \bar{N}_f \sum_{j=1}^n \frac{1}{P_{jg}(X_f)} \frac{\partial P_{jg}^+(X_f)}{\partial c_g} \frac{\partial P_{j,g-1}^+(X_f)}{\partial c_{g-1}}, \quad (26)$$

and finally if  $|g - g'| \geq 2$ , then

$$V_{c_g c_{g'}} = 0. \quad (27)$$

### The MML-EM Algorithm

The MML-EM algorithm consists of two steps. The first is the *expectation* step (the E-step) where the provisional expected frequency and the provisional expected sample size are computed by using Equations 15 and 19, respectively. With these expected values, the information matrix and the gradient vector are computed. Then, in the *maximization* step (the M-step), the Newton-Raphson method of Equation 16 is performed to obtain the MML estimates. Both the E-step and the M-step are repeated (the EM cycle) until all estimates become stable.

Each EM cycle consists of two estimation processes. First, the item parameters,  $a_j$  and  $b_j$ , of item  $j$  are estimated one item at a time, and then the same iteration process is repeated for  $j = 1, 2, \dots, n$ . Second, after obtaining stable item parameter estimates for all items, the estimation process for the  $m - 1$   $c_k$  parameters is repeated until their values become stable at a specified level of precision.

The indeterminacy of the  $c_k$  parameters and their relationship to the  $a_j$  and  $b_j$  parameters and  $\theta$  can be expressed as

$$a_j(\theta - b_j + c_k) = \frac{a_j}{s} [s\theta - (sb_j - t) + (sc_k - t)] \quad (28)$$

where  $s$  is a scaling factor and  $t$  is a location factor. Equation 28 shows that shifting the center of the category metric results in a shift of  $b_j$  in the same direction by the same units. If the intervals of the category scale are expanded by the factor  $s$  and  $\theta$  is integrated out by  $N(0, s^2)$ , the  $b_j s$  will expand and the  $a_j s$  will contract by the same factor. Therefore, in the one-parameter model, the center of the  $c_k$  parameters must be set at an arbitrary point. The two-parameter model requires the additional constraint of scaling to obtain converged values of  $a_j$  and  $b_j$ . This relationship among parameters will be demonstrated in the analysis of the simulated data.

The algorithm described above is implemented in the PARSCALE program (Muraki & Bock, 1988). In the program, the initial values of the  $c_k$  estimates are obtained by the inverse normal transformation of the sample categorical responses. During the process of  $c_k$  estimation, if  $a_j$  is constant, one end of the category boundaries is fixed and is not estimated. In the case where  $a_j$  varies for items, both ends are fixed because of indeterminacy in the origin and scale of the  $c_k$  parameter.

## Data and Results

### Simulated Data

Five thousand Likert-type response vectors for five items and five categories were generated with a normal distribution of  $\theta$ ,  $N(0, 1)$ . Table 1 presents the original values of  $a_j$ ,  $b_j$ , and  $c_k$  parameters. These simulated data were analyzed three times under different conditions of  $c_k$  values. Ten quadrature points were used and the precision level .001 was set for all estimations.

In Analysis 1, the simulated data were analyzed with the two endpoints of the category scale fixed. The high degree of similarity of the recovered middle  $a_j$ ,  $b_j$ , and  $c_k$  parameters to the original values

Table 1  
Values of Category, Slope, and Location, and Parameters  
for Original Simulated Data and for Three Analyses

Parameter	Original	Analysis		
		1	2	3
<b>Category</b>				
1	-2.000*	-2.000*	-1.500*	-3.000*
2	-1.000*	-1.009	-0.509*	-1.514*
3	0.000*	-.007	.493*	-.011*
4	1.000*	1.000*	1.500*	1.500*
<b>Slope</b>				
1	.800*	.792	.792	.528
2	1.000*	.990	.990	.660
3	1.200*	1.188	1.188	.792
4	1.000*	.998	.998	.665
5	.800*	.789	.789	.526
<b>Location</b>				
1	-2.000*	-2.025	-1.524	-3.037
2	-1.000*	-.975	-.474	-1.462
3	0.000*	-.014	.487	-.020
4	1.000*	1.006	1.506	1.509
5	2.000*	1.992	2.493	2.990
Prior	N(0,1)	N(0,1)	N(0,1)	N(0,1.5 <sup>2</sup> )

\*Constants supplied to the estimation program.

indicates that the MML-EM algorithm developed here works for estimating parameters if observed Likert-type responses are generated based on the item response model in Equation 4.

As expressed by Equation 28, shifting the origin of the category scale affects the  $b_j$  parameters by the same amount, but does not affect the  $a_j$  parameters. In Analysis 2, the  $c_k$  estimates from Analysis 1 were increased by .5 and only the  $a_j$  and  $b_j$  parameters were estimated. As shown in Table 1, the  $a_j$  estimates were unchanged after the iteration was converged, but all the  $b_j$  estimates were shifted by .5.

In Analysis 3, a normal prior distribution for  $\theta$ ,  $N(0,1.5^2)$ , was used for the integration. All  $c_k$  parameters were multiplied by 1.5 and supplied to the estimation process as fixed points. When the estimation process was converged, the  $a_j$  estimates contracted and the  $b_j$  estimates expanded by the same amount. These results demonstrate that direct comparisons of  $a_j$  and  $b_j$  estimates are not possible unless their corresponding sets of  $c_k$  estimates are identical or their scaling and location factors are adjusted by some equating method.

The main feature of the rating-scale model is that the item parameters can be separated from the  $c_k$  parameters. Although the origins and scales of the item  $a_j$  and  $b_j$  metrics are arbitrary and depend on those of the  $c_k$  estimates after the distribution of  $\theta$  is controlled, they are theoretically invariant. To demonstrate this feature, the simulated dataset described above was analyzed with collapsed response categories. In Analysis 4, the second and third categorical responses were collapsed into a single category. Then, in Analysis 5, the three middle categorical responses were collapsed into a single categorical response. Table 2 shows the  $a_j$ ,  $b_j$ , and  $c_k$  estimates, along with their standard errors. The standard errors of the item parameter estimates increase slightly along with the further collapsing of response categories; however, their estimated values never differ by more than .02 among the three simulated data analyses. They are quite consistent regardless of the number of response categories.

Table 2  
 Parameter Estimates (Est.) and Their  
 Standard Errors (SE) for Analyses 4 and 5 of  
 Simulated Data, Collapsing Number of Categories

Estimate	Analysis 4		Analysis 5	
	Est.	SE	Est.	SE
Category				
1	-2.000*	0.000	-2.000*	0.000
2				
3	-.007	.006		
4	1.000*	0.000	1.000*	0.000
Slope				
1	.790	.011	.787	.012
2	.987	.009	.970	.011
3	1.202	.010	1.213	.012
4	.985	.011	.997	.013
5	.794	.014	.816	.016
Location				
1	-2.028	.017	-2.034	.018
2	-.976	.010	-.967	.012
3	-.012	.007	-.021	.009
4	1.007	.011	1.009	.012
5	1.990	.023	1.972	.022
Prior	N(0,1)		N(0,1)	

\*Constants supplied to the estimation program.

### The General Social Surveys

This dataset was extracted from the General Social Surveys file, archived by the National Opinion Research Center (Davis, 1988). The General Social Surveys have been conducted almost yearly since 1972. The questionnaire form used over the years has been quite consistent, although a few questions have been added or deleted in certain years. Two sets of Likert-type questions were drawn from the questionnaire. The first set of items, the so-called Satisfaction items, has been used since 1973; the second set of items, the so-called Country items, has been used since 1974. For this study, the datasets of 1974, 1977, and 1983 were analyzed.

The directions for the Satisfaction items are: "For each area of life I am going to name, tell me the number that shows how much satisfaction you get from that area." The respondent's numerical answers from 1 to 7 correspond to seven ratings labeled as "A very great deal," "A great deal," "Quite a bit," "A fair amount," "Some," "A little," and "None." The items evaluated by this Likert scale are:

- A. The city or place you live in. (City)
- B. Your non-working activities—hobbies and so on. (Hobby)
- C. Your family life. (Family)
- D. Your friendships. (Friend)
- E. Your health and physical condition. (Health)

The directions for the Country items are: "You will notice that the boxes on this card go from the highest position of 'plus 5' for a country that you like very much, to the lowest position of 'minus 5' for a country you dislike very much. How far up the scale or how far down the scale would you rate the following countries?" The original 10-point scale was collapsed into a 5-point Likert scale. The items evaluated by this scale are Russia (U.S.S.R.), Japan, England, Canada, Brazil, China, Israel, and Egypt.

There were 1,484 completed interviews in 1974, 1,530 in 1977, and 1,599 in 1983. The respondents were randomly sampled each year. After excluding the cases with missing responses, 500 male and 500 female respondents were randomly sampled from each year-cohort. The precision of convergence was set at .001 in all analyses. The distribution of  $\theta$  for each year-cohort was assumed to be standard normal. Ten quadrature points were used for the integration; lognormal prior distributions for  $a_j$  and normal distributions for  $b_j$  parameters were used for the constraints of parameter estimation (Mislevy, 1986).

The combined data of 3,000 respondents to the Satisfaction items were analyzed with the highest and lowest  $c_k$  parameters fixed to impose an origin and scaling constant on the category scale. Therefore, the standard errors of these estimates are simply 0, as shown in Table 3. Satisfaction with the quality of life was assumed to be a unidimensional latent trait. Table 3 gives category frequency distributions, the category means, and the parameter estimates for the Satisfaction items. The signs of the  $b_j$  estimates are reversed, so that a higher  $b_j$  value indicates more satisfaction from that area of life.

The data indicate that people are more likely to be satisfied in the areas of family life and friendships than in the areas of living place and hobbies. Separate analyses of three year-cohorts showed that the relative order of the items was quite consistent over the years, although the magnitude of the  $b_j$  estimates increased for some of the items and decreased for others.

Although a set of items shares a common Likert scale, respondents perceive the scale differently. A single Likert scale may elicit more discriminating responses to some items than to others. An  $a_j$  parameter indicates this sensitivity of category changes for each item, and analysis shows that the Friend item has a slightly higher discrimination than the others. When the discriminations of items vary to a great extent, collapsing the category points of a Likert scale without considering their differential effect on the items is not generally recommended.

The modified scale for Country items has 5 points. The lowest category boundary was fixed and the remaining three  $c_k$  estimates were obtained from the data for 3,000 respondents. Because  $a_j$  parameters were not estimated, the scaling factor of the category metric was automatically fixed. It was assumed that the underlying trait dimension was a positive attitude toward foreign countries. Table 4 gives the  $b_j$  estimates as ordered by their corresponding category means. Each year-cohort was then analyzed with the common  $c_k$  estimates. This allowed for direct comparison of the  $b_j$  estimates over the years. These  $b_j$  estimates are plotted in Figure 1.

As Figure 1 shows, Americans' positive attitudes toward Canada and England are quite high in comparison to the other countries, and remain constant over the years. On the other hand, it is quite interesting to observe the trend of Americans' attitude toward Egypt, China, and Russia over the years. Many historical events may account for these dramatic shifts in popularity trends (Weaver, Alais, Franz, & Carpenter, 1986).

Table 3  
Response Frequency, Category Mean, and Parameter Estimates (Est.) and Their Standard Errors (SE) for Items and Categories of the Datasets From 1974, 1977, and 1983

Item	Category Weight							Category Mean	Location		Slope	
	7	6	5	4	3	2	1		Est.	SE	Est.	SE
Family	1251	1029	345	199	80	51	45	5.95	.405	.008	1.270	.012
Friend	867	1234	471	267	90	56	15	5.76	.133	.004	1.689	.012
Health	861	993	438	433	130	96	49	5.51	.021	.009	1.144	.011
Hobby	705	945	550	417	174	109	100	5.29	-.158	.008	1.228	.011
City	537	809	525	701	207	157	64	5.01	-.378	.008	1.182	.011
Total N	4221	5010	2329	2017	681	469	273					
Category	-.578	.376	.848	1.430	1.750	2.093						
SE	0.000	.004	.004	.005	.006	.000						

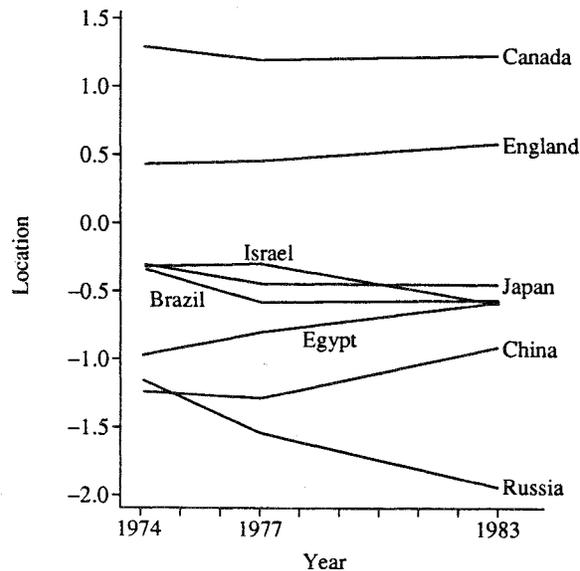
Table 4  
Response Frequency, Category Mean, and Parameter Estimates (Est.)  
and Their Standard Errors (SE) for Items and Categories of the  
Datasets From 1974, 1977, and 1983

Item	Category Weight					Category Mean	Location	
	5	4	3	2	1		Est.	S.E.
Canada	1941	762	241	33	23	4.52	1.237	.014
England	1213	1123	476	101	87	4.09	.489	.012
Israel	660	819	948	310	263	3.43	-.408	.012
Japan	484	1121	833	295	267	3.42	-.401	.012
Brazil	394	944	1243	275	144	3.39	-.491	.012
Egypt	322	752	1213	406	307	3.13	-.771	.012
China	217	598	1083	522	580	2.78	-1.136	.012
Russia	125	524	853	496	1002	2.42	-1.542	.012
Total <i>N</i>	5356	6643	6890	2438	2673			
Category	-.761	.423	1.595	2.210				
S.E.	0.000	.005	.006	.008				

### Discussion and Conclusions

Although arbitrariness exists in the item difficulty estimates in a dichotomous item response model, and although origin and scaling indeterminacies exist in the trait dimension as well as in the item parameter estimates, the origin and scaling of an item parameter dimension can be fixed by applying the MML method with a prior distribution on  $\theta$ . However, the polytomous item response model developed here introduces another indeterminacy, as shown in Equation 28. If multiple Likert scales with different numbers

Figure 1  
Trend of Item Location Estimates for the Country Items  
From the 1974, 1977, and 1983 Datasets



of points are used in a single questionnaire, item estimates for the whole set of items cannot be compared because of the arbitrariness of both the origin and the scaling of the  $c_k$  estimates of each Likert scale.

If an assumption of unidimensionality of the latent trait distribution is met, each set of items with a common Likert scale can be treated as a subtest. Test equating techniques developed for achievement testing can be applied to the method of category scale adjustment and all item estimates can be equated. This internal equating method needs to be investigated to take advantage of the specific features of the present model. Another extension of the categorical item response model incorporates a research design matrix (Muraki, 1983). A multidimensional polytomous item response model based on this model has also been formulated, and its MML-EM algorithm is available (Muraki, 1985).

Likert-type data are often analyzed by assigning numeric scores to the response categories based on the assumption that the observed categorical responses are quantitative and continuous. However, the actual intervals between adjacent categories are generally unknown in advance. When the ordered categorical responses are treated as a discrete variable, a log-linear model has been frequently used for the analysis of cross-classification tables (Clogg, 1979). According to this method, however, the number of cells to be analyzed increases exponentially along with the increment in the number of items. For example, in the Country data,  $5^8$  (= 390,625) cells in the cross-classification table must be analyzed. In addition to the computational difficulty of data with a large number of items or categories, the prediction of categorical responses for a given trait value with sets of  $b_j$  and  $c_k$  parameter estimates is not possible.

Another major advantage of polytomous item response models, as compared to the log-linear approach, is the ease of scoring persons or groups. For example, the rating-scale model is used for the analysis of the California Direct Writing Assessment data (Cooper & Breneman, 1988). In the assessment program, writing tasks (prompts) are grouped into subsets by their different types of writing. From the set of tasks as a whole, each student is assigned a prompt to respond by writing a short essay. These essays are then scored by three Likert-type scales. Muraki (1989) computed school-level scaled writing scores for each writing type and for each Likert scale after the model parameters were estimated. Year-by-year comparisons of the assessment results were greatly facilitated by the use of this particular item response model.

The Rasch dichotomous item response model is often referred to as the one-parameter logistic item response model. However, each of these models was formulated independently according to its own distinctive assumptions. The one-parameter model is a logistic approximation to the normal ogive model, in which the  $a_j$  parameter is a known constant and does not have to be estimated. On the other hand, there is no concept of an  $a_j$  parameter in the Rasch dichotomous item response model. The  $a_j$  parameter or discrimination is a unique concept in a Thurstone family of item response models, because the parameterization originates from the discriminial process specific to the randomly selected individual (Bock & Jones, 1968).

The Rasch family of polytomous item response models is based on the assumption that each probability of selecting a category  $k$  over a category  $k - 1$  is governed by the Rasch dichotomous response model. Rost (1988) formulated a Rasch polytomous item response model, called the successive interval model, by introducing the item-specific "dispersion" parameter independent of the  $c_k$  parameters. Rost's dispersion parameters were estimated by the joint and conditional maximum likelihood methods and were shown to be functionally analogous to the  $a_j$  parameter of the Thurstone models. Because remarkable similarities exist between the proposed rating-scale model and the successive intervals model, features of each model must be carefully examined and compared.

Unlike the joint and conditional maximum likelihood estimation methods, the marginal maximum likelihood method treats the person component,  $\theta$ , in the model as a random variable. In the analysis of the General Social Surveys data, it was assumed that the variable was normally distributed, because respondents were randomly selected. However, for more rigorous analyses, the prior can be updated by

the use of the distribution computed after each estimation cycle. This simultaneous estimation of a latent trait distribution and model parameters slows the convergence rate of the estimation process but minimizes the effect of initial misspecification about a prior distribution on the parameter estimates. An empirical distribution can be substituted for a theoretical prior if it is known before parameter estimation. The effects of priors for the marginal maximum likelihood method, including a multiple-group situation, need to be studied further.

Investigation has only recently begun on polytomous item response models. The EM algorithm is rapidly gaining popularity for various estimation problems. Combining these recently developed techniques should allow more flexible applications of item response theory to a wider variety of psychological and sociological research problems. A major obstacle to wider acceptance of more general item response models for qualitative data analyses is that a considerable amount of computational time is required to obtain converged parameter estimates with reasonable precision. Although the rapid development of computer technology will eventually shorten computational time, an effective numerical technique to speed up the convergence process should also be developed.

### References

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561–573.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, *Statistical theories of mental test scores*. Reading MA: Addison-Wesley.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29–51.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443–459.
- Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, 12, 261–280.
- Bock, R. D., & Jones, L. V. (1968). *The measurement and prediction of judgment and choice*. San Francisco: Holden-Day.
- Bock, R. D., & Lieberman, M. (1970). Fitting a response model for  $n$  dichotomously scored items. *Psychometrika*, 35, 179–197.
- Bock, R. D., Muraki, E., & Pfeifferberger, W. (1988). Item pool maintenance in the presence of item parameter drift. *Journal of Educational Measurement*, 25, 275–285.
- Clogg, C. C. (1979). Some latent structure models for the analysis of Likert-type data. *Social Science Research*, 8, 287–301.
- Cooper, C., & Breneman, B. (1988). *Writing achievement of California eighth graders: A first look*. Sacramento CA: State Department of Education.
- Davis, J. A. (1988). *General Social Surveys, 1972–1988: Cumulative codebook*. Chicago: National Opinion Research Center.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1–38.
- Edwards, A. L., & Thurstone, L. L. (1952). An interval consistency check for scale values determined by the method of successive intervals and the method of graded dichotomies. *Psychometrika*, 17, 169–180.
- Kendall, M. G., & Stuart, A. (1973). *The advanced theory of statistics* (Vol. 2). New York: Hafner.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, No. 140.
- Lord, F. M. (1980). *Application of item response theory to practical testing problems*. Hillsdale NJ: Erlbaum.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika*, 51, 177–195.
- Muraki, E. (1983). *Marginal maximum likelihood estimation for three-parameter polychotomous item response models: Application of an EM algorithm*. Doctoral dissertation, University of Chicago.
- Muraki, E. (1985, March). *Full-information factor analysis for polychotomous item response*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Muraki, E. (1989). *Maintaining consistency of performance assessment scales over several years*. Paper presented at the 19th Annual Assessment Conference, Boulder CO.
- Muraki, E., & Bock, R. D. (1988). *PARSCALE: Parameter scaling of rating data*. Mooresville IN: Scientific Software, Inc.

- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research. [Expanded edition, University of Chicago Press, 1980.]
- Rost, J. (1988). Measuring attitudes with a threshold model drawing on a traditional scaling concept. *Applied Psychological Measurement*, 12, 397-409.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, No. 17.
- Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph*, No. 18.
- Stroud, A. H., & Secrest, D. (1966). *Gaussian quadrature formulas*. Englewood Cliffs NJ: Prentice-Hall.
- Thissen, D. (1982). Marginal maximum likelihood estimation for the one parameter logistic model. *Psychometrika*, 47, 175-186.
- Thissen, D. (1986). *MULTILOG Version 4.0 user's guide*. Mooresville IN: Scientific Software, Inc.
- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Weaver, C. N., Alais, R. M., Franz, R. S., & Carpenter, J. B. (1986, April). *U.S. public attitudes and trade relations with eight other countries: Further tests for relationship*. Paper presented at the 22nd Interventional Atlantic Economic Conference, St. Thomas, Virgin Islands.

**Author's Address**

Send requests for reprints or further information to Eiji Muraki, Educational Testing Service, Princeton NJ 08541, U.S.A.