The Measurement of Latent Traits by Proximity Items

Herbert Hoijtink
University of Groningen

A probabilistic parallelogram model for the measurement of latent traits by proximity items (the PARELLA model) is introduced. This model assumes that the responses of persons to items result from proximity relations: the smaller the distance between person and item, the larger the probability that the person will agree with the content of the item. The model is unidimensional and assigns locations to items and persons on the latent trait. The parameters of the PARELLA model are estimated by marginal maximum likelihood and expectation maximization. The efficiency of the estimation procedure is illustrated, a diagnostic for the fit of items to the model is presented, and the PARELLA model is used for the analysis of three empirical datasets. *Index terms: expectation maximization, latent trait theory, marginal maximum likelihood, nonmonotone trace lines, single-peaked preference functions, unfolding.*

Coombs, Dawes, and Tversky (1970) define measurement as "the process through which numbers are assigned to objects such that the relation between the objects is represented by the relation between the numbers" (p. 12). A latent trait generally is specified as a unidimensional characteristic of a person that is not directly observable, such as an ability or an attitude.

If the measurement of latent traits is considered, the "objects" as mentioned in the definition of Coombs et al. (1970) are items and persons. The items represent different levels of the latent trait. Their locations constitute the grid of the instrument with which the latent trait can be measured. For example, if the trait to be measured is an ability, an item can be constructed that a person with a certain ability can solve. Persons who answer the item correctly are expected to have a location on the latent trait above the item, which indicates that the person is more able than the item is difficult. Persons who answer the item incorrectly are expected to have a location below the item.

If the trait to be measured is an attitude, an item can be constructed for which a person would have to indicate agreement or disagreement. In this situation, the relation between person response and location relative to the item will be different. If a person endorses an item, person location will be close to the location of the item, and if a person does not endorse an item, his or her location will differ substantially from the item's location.

If only one item is available, the locations of persons can be compared to only one value of the latent trait. The measurement instrument created also will not be very discriminating, because only two groups of persons can be distinguished. However, more items are usually available, which permits the construction of a measurement instrument with a larger discriminatory power.

A second aspect of the definition of Coombs et al. (1970) that needs to be explained within the context of the measurement of latent traits is "the process through which numbers are assigned to objects" (p. 12). This passage can be restated as "the process through which the responses of persons to items (correct or incorrect in the case of ability items, and agree or disagree in the case of attitude

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 15, No. 2, June 1991, pp. 153-169

0146-6216/91/020153-17$2.10

items) are related to locations of persons and items on the latent trait." The locations of the items and persons on the latent trait are unknown. What is needed is a measurement model that does exactly what the stated passage demands: relates the answers of persons to items, to locations of persons, and to items on the latent trait.

If a restriction is made to models designed for dichotomous responses that are stochastically independent for each single person between items, and that characterize persons and items by only one (location) parameter, then there are two measurement models that can be seen as the roots of the models currently used. With the introduction of his scalogram model, Guttman (1950) made an important contribution to the theory of the measurement of latent traits. His measurement model has a deterministic nature and is formalized as

$$X_{ai} = 1 \text{ if } (\beta_a - \delta_i) \geq 0 \quad ,$$
$$X_{ai} = 0 \text{ if } (\beta_a - \delta_i) < 0 \quad , \tag{1}$$

where X_{ai} is the response of person a, $a = 1, \ldots, N$ to item i, $i = 1, \ldots, n$,
β_a is the location of person a, and
δ_i is the location of item i.

Coombs' (1964) parallelogram model also has a deterministic nature, but it differs structurally from Guttman's model:

$$X_{ai} = 1 \text{ if} |\beta_a - \delta_i| \leq \tau \quad ,$$
$$X_{ai} = 0 \text{ if} |\beta_a - \delta_i| > \tau \quad , \tag{2}$$

where τ is a threshold (of equal size for each item) governing the maximum distance between β_a and δ_i for which a person still renders a positive response.

Guttman's (1950) model is based on dominance relations between persons and items: a person will answer correctly ($X_{ai} = 1$) if β_a is greater than δ_i, and will answer incorrectly ($X_{ai} = 0$) if δ_i is greater than β_a. Coombs' (1964) model is based on proximity relations: a person will endorse the item ($X_{ai} = 1$) when within the threshold τ, and will not endorse the item ($X_{ai} = 0$) when not within the threshold τ. Although applications partly overlap, the model formalized in Equation 1 is most suited for the measurement of abilities, whereas the model formalized in Equation 2 is most suited for the measurement of attitudes and preferences.

The last part of the Coombs et al. (1970) definition of measurement, "The relation between the objects is represented by the relation between the numbers," can now be easily interpreted. If an item is more difficult than a person is able in Guttman's (1950) model, the number assigned to the item will be greater than the number assigned to the person. If a person selects an item in the model of Coombs (1964), the difference between the numbers assigned to the person and the item will be less than τ.

A drawback of the Guttman (1950) and Coombs (1964) models is their deterministic nature. For sensibly distanced items, empirical data will never be so perfect that Equations 1 or 2 will be accurate descriptions of the response process. What is needed are probabilistic models. In the case of Guttman's model,

$$P(X_{ai} = x_{ai}|\beta_a, \delta_i) = \mathrm{f}[(\beta_a - \delta_i), x_{ai}] \quad , \tag{3}$$

and in the case of Coombs' model,

$$P(X_{ai} = x_{ai}|\beta_a, \delta_i) = \mathrm{f}(|\beta_a - \delta_i|, x_{ai}) \quad . \tag{4}$$

The Rasch (1960; Fischer, 1974) model can be seen as the probabilistic counterpart of Guttman's model. The two- and three-parameter logistic models (Birnbaum, 1968; Hambleton & Swaminathan, 1985) and their multicategory extensions (Thissen & Steinberg, 1986) are extensions of the Rasch model, because more item parameters are added or a different kind of data can be analyzed. Models based on Equation 4 have only recently been developed: the PIRT model (Andrich, 1988), its multidimensional, multiparameter extensions (DeSarbo & Hoffman, 1986; Takane, 1983), and the model that is introduced below.

The PARELLA model

The PARELLA model is a probabilistic analogue of Equation 2 based on the Coombs (1964) parallelogram model. The choice process is to a large extent determined by the distance between person and item, but random characteristics of either person or item are included in the model.

The PARELLA model is formulated as:

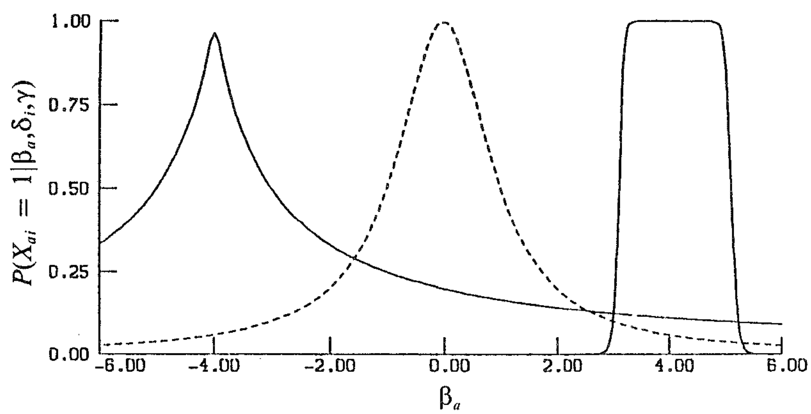
$$P(X_{ai} = x_{ai} \mid \beta_a, \delta_i, \gamma) = \frac{[(\beta_a - \delta_i)^2]^{\gamma(1 - x_{ai})}}{1 + [(\beta_a - \delta_i)^2]^{\gamma}} \quad . \tag{5}$$

In this equation, the squared distances are raised to the power γ instead of 2γ in order to assure that the resulting number will be non-negative.

Main features of the PARELLA model are: the single peakedness of the item response function (IRF), which is a feature of any proximity relation-based measurement model; and a choice probability equal to 1.0 for $\beta_a = \delta_i$. The parameter γ—also referred to as the power parameter—determines the relative importance of the distance between person and item, and the random characteristics of either.

PARELLA IRFs are plotted for different values of γ in Figure 1. The larger the γ, the more important is the distance to an item for evaluation of the item, and the smaller the γ, the more important are random characteristics. In the model for $\gamma = 0$, the distance between person and item is of no importance, and each item has a choice probability of .5, except for the situation in which $\beta_a = \delta_i$. In the models for γ positive and increasing, the distance between person and item becomes increasingly important for a person's choice, until the model can barely be distinguished from the deter-

Figure 1
Item Response Functions Defined by the PARELLA Model for Items with Location and Power Parameters (From Left to Right) of (-4,.5), (0,1) and (4,10).



ministic Coombsian model for large values of γ (e.g., 10 or more). In this situation, the random characteristics of an item (or person) do not have any importance; choice is determined completely by distance between person and item.

It is assumed here that the power parameter is greater than or equal to 0. Negative values of the power parameter will lead to single-dipped instead of single-peaked IRFs. These could be used for the analysis of "non-preference" data.

The PARELLA model contains three parameters: the locations of persons and items, and the power parameter. This implies that each IRF has the same threshold (i.e., the same width). Independent of the value of the power parameter, a person located at a distance of 1.0 from an item always has a probability of .50 of giving a positive response to the item. For large γ, persons will give positive responses when within a distance of 1.0 of the item, and negative responses when outside of this distance. For smaller values of γ (e.g., .5 or 1.0), persons will give positive responses with a probability greater than .50 when within a distance of 1.0 from an item, and negative responses with a probability less than .50 when outside of this distance.

Note that the choice of the threshold value of 1.0 is rather arbitrary. Other values would only change the scale unit of the estimates without affecting the interpretation or the likelihood of the parameter configuration to be estimated.

Estimating the Parameters of the Marginal PARELLA Model

Since its introduction, marginal maximum likelihood (MML; Bock & Aitkin, 1981) is becoming increasingly popular as an estimation procedure for item response models (Bock, 1989; Thissen, 1982). The estimation procedure for the PARELLA model uses MML. One main feature of MML is the estimation of the probability density of the person parameters [$g(\beta)$] instead of estimating the location of each person. This does not imply that location estimates for individual persons cannot be obtained; several methods exist to estimate β_a using the estimates for δ, γ, and $g(\beta)$ (Bock & Aitkin, 1981; Mislevy & Bock, 1982; Wainer & Wright, 1980).

The likelihood equation of the marginal PARELLA model is

$$\mathrm{L}(\boldsymbol{\delta},\gamma,\boldsymbol{\pi},\mathbf{B}\,|\,\mathbf{X}) = \prod_{a=1}^{N} \sum_{q=1}^{Q} P(X_a\,|\,B_q)\,\pi_q \quad , \tag{6}$$

where δ denotes the vector with item locations and γ is the power parameter. The probability density of the person parameters is estimated through a step function with nodes B_q, $q = 1, \ldots, Q$, and weights π_q ($\Sigma\pi_q = 1$). Due to the assumption of local stochastic independence,

$$P(\mathbf{X}_a\,|\,B_q) = \prod_{i=1}^{n} P(X_{ai} = x_{ai}\,|\,B_q) = \prod_{i=1}^{n} \frac{[(B_q - \delta_i)^2]^{\gamma(1-x_{ai})}}{1 + [(B_q - \delta_i)^2]^{\gamma}} \quad . \tag{7}$$

The likelihood function in Equation 6 is not easy to work with (for an equivalent situation, see Bock & Lieberman, 1970). Bock and Aitkin (1981) demonstrate how the expectation maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977) can be used to replace equations such as Equation 6 by more tractable equations.

The EM algorithm is especially suited to estimating parameters in situations with missing information. Missing information can be constituted either by missing data elements or by parameters with unknown values, and is constituted in this study by $g(\beta)$.

The expected value of the "missing information" is estimated in the expectation step, given the data and the current values of δ and γ (i.e., the initial values the first time the expectation step is executed). π is updated here by iteration between the following two equations:

$$N_q = \sum_{a=1}^{N} \left[\frac{P(\mathbf{X}_a \mid B_q)\ \pi_q}{\sum_{q=1}^{Q} P(\mathbf{X}_a \mid B_q)\ \pi_q} \right] \quad , \tag{8}$$

and

$$\pi_q = N_q / N \quad , \tag{9}$$

where N_q denotes the number of persons expected at node q.

During this iterative procedure, the nodes are located at equal distances such that B_1 and B_Q have a weight of approximately .01. The number of nodes needed to obtain an accurate step function estimate of $g(\beta)$ depends on the shape of $g(\beta)$. Experience thus far indicates that 10 nodes are sufficient if $g(\beta)$ resembles a uniform, skewed, or normal distribution, and that 16 or more nodes are needed if $g(\beta)$ is bimodal (Hoijtink, 1991).

After updating $\boldsymbol{\pi}$ and $\mathbf{B}$, the number of persons located at each node that are expected to select each item (N_{qi}) is computed, because these statistics are needed in the maximization step:

$$N_{qi} = \sum_{a=1}^{N} \left[\frac{X_{ai} P(\mathbf{X}_a \mid B_q)\ \pi}{\sum_{q=1}^{Q} P(\mathbf{X}_a \mid B_q)\ \pi_q} \right] \quad , \tag{10}$$

Given the updates for $\boldsymbol{\pi}$ and $\mathbf{B}$, $\boldsymbol{\delta}$ and γ are updated in the maximization step using the likelihood function

$$L = \prod_{i=1}^{n} \prod_{q=1}^{Q} \left[\binom{N_q}{N_{qi}} P(X_{qi} = 1 \mid B_q)^{N_{qi}} P(X_{qi} = 0 \mid B_q)^{(N_q - N_{qi})} \right] \quad . \tag{11}$$

The final estimates for $\boldsymbol{\delta}$, γ, $\boldsymbol{\pi}$, and $\mathbf{B}$ are obtained by iterating between the expectation and the maximization step until convergence occurs. See Hoijtink (1991) for further details on this estimation procedure.

When the EM algorithm is used to estimate model parameters, it is not always easy to compute the standard errors of the estimates. See Louis (1982) for general solutions to this problem, and Glas (1989) for a solution specific to item response models.

Estimating Locations for Individual Persons

Given the estimates for $\boldsymbol{\delta}$ and γ, the maximum likelihood estimate of β_a can be obtained through maximization of $L(\beta_a \mid \mathbf{X}_a)$ for β_a (Kiewiet, 1988):

$$\mathrm{L}(\beta_a \mid \mathbf{X}_a) = \prod_{i=1}^{n} P(X_{ai} = 1 \mid \beta_a)^{X_{ai}} P(X_{ai} = 0 \mid \beta_a)^{(1 - X_{ai})} \quad . \tag{12}$$

Another way of estimating the person parameter is the expected a posteriori estimate, which takes $g(\beta)$ into consideration (Bock & Aitkin, 1981):

$$E(\beta_a | \mathbf{X}_a) = \frac{\sum_{q=1}^{Q} [B_q[\mathrm{P}(\mathbf{X}_a | B_q)\ \pi_q]}{\sum_{q=1}^{Q} [P(\mathbf{X}_a | B_q)\ \pi_q]} \quad . \tag{13}$$

The accuracy of both estimators has not been fully evaluated. Other estimates for β_a are possible, such as robust estimates of the kind considered by Wainer and Wright (1980) or biweight estimates (Mislevy & Bock, 1982).

Investigation of the Accuracy of the Estimates

To illustrate the accuracy of the estimates of δ_i, γ, and the step function estimate of $g(\beta)$, the parameter vector termed "generated value" in Table 1 was used to generate 100 samples of 300 response patterns each. A value for β_a was randomly sampled from $g(\beta)$ [an N(0,1) density for this simulation study] to simulate a response pattern. Given the value of β_a, $\boldsymbol{\delta}$, and γ, the probability that each item would be selected was computed from Equation 5. Subsequently, each probability was compared to a pseudo-random number drawn from a U(0,1) density. If the probability exceeded the random number, the response was set to 1; otherwise, the response was set at 0.

For each sample, $\boldsymbol{\delta}$, γ, and $g(\beta)$ were estimated using a step function. The means and variances of the estimates over 100 replications for $\boldsymbol{\delta}$ and γ are presented in Table 1. As can be seen, the differences between the generated and mean estimated values are negligible, and the estimates are apparently unbiased. Note that the variances vary between .05 and .10, which implies that there were substantial differences between the generated and estimated values of the parameters in some of the 100 replications. These differences are easily explained as sample fluctuations, however, and it may be possible that sample size should not be much less than 300 to obtain reliable estimates.

The first four moments of the step function estimate of $g(\beta)$ were compared to the first four moments of the generated N(0,1) density in order to assess the bias of the step function estimate.

Table 1
Generated Values, Mean, and Variance of the Estimates for 10 Items Over 100 Replications, and Power and Descriptive Statistics of the N(0,1) Density (N = 300)

Item Number	Generated Value	Mean of Estimates	Variance of Estimates
1	-1.60	-1.60	.09
2	-1.40	-1.42	.08
3	-1.00	-1.01	.07
4	-.80	-.80	.07
5	-.60	-.60	.06
6	-.30	-.30	.07
7	.80	.82	.06
8	1.20	1.21	.07
9	1.70	1.70	.09
10	2.00	2.01	.10
Power	1.00	.98	.05
N(0,1) Density			
Mean	.00	.01	.07
Variance	1.00	1.00	.12
Skewness	.00	.02	.21
Kurtosis	.00	-.04	.42

As can be seen in the lower portion of Table 1, the difference between generated and estimated moments was negligible, and the step function estimate appears to be unbiased. The results were equally good when $g(\beta)$ was not N(0,1). In a more elaborate simulation study, Hoijtink (1991) showed that skewed, uniform, and bimodal probability densities for $g(\beta)$ can be estimated with equal accuracy.

Comparing Expected and Empirical IRFs

An aspect of items that is worthwhile investigating is the equality of expected and empirical IRFs: the expected curve given the model with the parameter estimates inserted $[P(X_{ai} = 1|\beta_a)]$, and the empirical curve reconstructed from the observations. The following quantities provide the opportunity to perform such an investigation (a standardized version of the idea to be discussed in this section was developed by Mislevy & Bock, 1990).

The expected number of persons located at node q that select item i, computed from data and estimated parameters, is given by:

$$O_{qi} = \sum_{a=1}^{N} \left[\frac{x_{ai} P(\mathbf{x}_a | B_q)\ \pi_q}{\sum_q P(\mathbf{x}_a | B_q)\ \pi_q} \right] \quad , \tag{14}$$

This quantity is designated O_{qi} to denote that its computation involves observed data. Note that the same quantity appeared in the estimation procedure as N_{qi}.

The expected number of persons located at node q that select item i can also be computed using only estimated parameters:

$$M_{qi} = N\,\pi_q\, P(X_{qi} = 1\,|\,B_q) \quad . \tag{15}$$

This quantity is called M_{qi} to denote that it is the expected number according to the model, and observed data are not involved in its computation.

The quantity O_{qi} is closely related to the empirical IRF (reconstructed from the observations). At each node, O_{qi}/N_q gives the estimate of $P(X_{ai} = 1|B_q)$ according to the data. Thus, the value of the empirical IRF at each node can be observed. The quantity M_{qi} is closely related to the expected IRF. The quantity M_{qi}/N_q at each node gives the estimate of $P(X_{ai} = 1|B_q)$ according to the model.

M_{qi} and O_{qi} are both divided by N_q to arrive at the empirical and expected IRFs, respectively. Consequently, to investigate the equality of empirical and expected IRFs, a direct comparison of the M_{qi} and O_{qi} is equivalent to a comparison of the M_{qi}/N_q with the O_{qi}/N_q.

If the empirical IRF of item i resembles the expected IRF, the difference $|O_{qi} - M_{qi}|$ for all q will be small. Consequently, the sum of these differences (denoted by SUM) will be small. If item i does not fit, the sum of the differences will be substantially greater than 0. A test of fit with known distribution that combines the differences is not available. Two aspects that would have been incorporated in a test of fit—the number of nodes and the sample size—must be evaluated independently of the test.

If the number of nodes increases, the differences per node will probably decrease and have to be judged more strictly. This problem can be avoided, however, by examining the sum of differences: Under the hypothesis of equality of the expected and empirical IRFs, this sum will probably not change very much if the number of nodes increases. If the sample size increases, the sum of differences will also increase. In the simulation study reported below, a sample of 300 was used to derive rules of thumb for the interpretation of the sum of differences. For sample sizes larger than 300 these rules should be relaxed; for sample sizes smaller than 300 they should be used more strictly.

Simulation Study

Method. A small simulation study was executed to determine the power of the proposed diagnostic and to derive rules of thumb for the interpretation of the sum of differences. The design contained nine cells, the response patterns in each cell were generated according to the mechanism explained above, and each cell consisted of 300 response patterns. The generating parameter vector contained a N(0,1) density for $g(\beta)$, $\gamma = 1$, and the $\delta_i = (-1.6, -1.4, -1.0, -.8, -.6, -.3, .8, 1.2, 1.7, 2.0)$. Furthermore, an item with a deviating trace line was added in eight of the nine cells. Rasch items with locations of -1.5, 0, and 1.5 were added to the first three cells. An item (called a Broad item) with an IRF with a larger width than the IRF of the PARELLA model (location at 0.0 and $\delta = 1.0$) was added to the fourth cell:

$$P(X_{ai} = 1 \mid \beta_a, \delta_i) = 1/[1 + .5(\beta_a - \delta_i)^2] \quad . \tag{16}$$

An item (Small) with a smaller IRF than that defined by the PARELLA model (located at 0.0 and with $\delta = 1.0$) was added to the fifth cell:

$$P(X_{ai} = 1 \mid \beta_a, \delta_i) = 1/[1 + 2(\beta_a - \delta_i)^2] \quad . \tag{17}$$

Finally, an item was added in three cells with a constant choice probability of .3, .5, and .7, respectively. Parameters were estimated and O_{qi}, M_{qi}, and SUM were calculated from the data matrix containing no deviant items (i.e., 10 items) and the eight data matrices containing one deviant item (i.e., the 10 PARELLA homogeneous items plus one deviant item).

Results. Table 2 summarizes the fit values for the 10 PARELLA homogeneous items and the eight deviant items. Results for the items in accordance with the PARELLA model can be used as a reference when evaluating the results for the aberrant items. The differences between the O_{qi} and M_{qi} were always less than 3, except for Item 5, which had differences of 4.4 and 5.1, and Item 6, which had a difference of 4.0. The largest sum of differences observed was 17.1 for Item 5. For the aberrant items, there were two or more differences of approximately 6 for the items with constant choice probabilities of .3, .5, and .7 for the Small item and for the Rasch item located at -1.5. The sum of differences observed for these items (24.8 or larger) was clearly worse than those observed for the PARELLA homogeneous items. This is a strong indication that these items were not in accordance with the PARELLA model.

Nothing peculiar was observed for the Rasch items with locations at 0.0 and 1.5. The individual differences and the sum of differences resembled those of the PARELLA homogenous items. Because only a few persons were located in the region where the IRF was supposed to be decreasing (above 1.0—the estimated location according to the PARELLA model—for the Rasch item located at 0.0, and above 2.4 for the Rasch item located at 1.5), the diagnostic proposed did not detect that the IRF was monotone increasing instead of single peaked.

Two observations can be made about the sum of differences for the PARELLA homogeneous items in the analysis that contained a deviant item. First, the largest sum of differences observed was always (with the exception of Rasch item 0, Rasch item 1.5, and Broad) substantially less than the sum of differences for the deviant item. Second, the size of this largest sum of differences observed (between 8.0 and 15.9) was well within the range of what was expected for PARELLA homogeneous items (sum of differences less than 17.1). It appears that the fit of the PARELLA homogeneous items was almost unaffected by the presence of a deviant item, at least for an item set containing one deviant item.

These results suggest that the power of the proposed diagnostic appears to be satisfactory. That Rasch items located at the upper end of the latent trait could not be distinguished from PARELLA

Table 2

Fit Values, [O_{qi} (O), M_{qi} (M), Absolute Difference (D) Between O_{qi} and M_{qi}], and N_q (N), for the 10 PARELLA Homogeneous Items From an Analysis Without Deviant Items, and Fit Values for the Deviant Items From an Analysis in Which Items Were, in Turn, Added to the Original 10 Items (d_i is the Generating Location for the PARELLA Homogeneous Items, and Estimated Location for the Deviant Items; SUM is the Sum of Differences; and MAX is the Largest Sum of Differences Observed Among the PARELLA Homogeneous Items)

	Node									
	1	2	3	4	5	6	7	8	9	10
PARELLA Homogeneous Items										
Item 1: d_i = -1.6, SUM = 12.9										
O	1.6	3.4	16.5	38.1	23.9	17.9	11.9	5.0	3.6	.2
M	1.0	3.1	16.5	39.8	24.8	15.2	9.1	3.2	1.6	.1
D	.6	.3	.0	1.7	.9	2.7	2.8	1.8	2.0	.1
N	2.5	4.6	16.6	52.0	57.1	60.0	56.6	29.4	20.1	1.1
Item 2: d_i = –1.4, SUM = 5.7										
O	1.1	2.5	15.4	47.6	28.5	20.3	10.8	4.0	1.8	.1
M	.8	2.5	15.2	47.0	30.9	18.5	10.7	3.7	1.8	.1
D	.3	.0	.2	.6	2.4	1.8	.1	.3	.0	.0
N	2.5	4.6	16.6	52.0	57.1	60.0	56.6	29.4	20.1	1.1
Item 3: d_i = –1.0, SUM = 10.2										
O	.1	.8	9.7	50.9	47.0	24.8	12.6	3.9	2.1	.1
M	.6	1.7	10.7	51.0	46.3	27.9	15.2	5.0	2.3	.1
D	.5	.9	1.0	.1	.7	2.9	2.6	1.1	.2	.0
N	2.5	4.6	16.6	52.0	57.1	60.0	56.6	29.4	20.1	1.1
Item 4: d_i = –.8, SUM = 6.3										
O	.1	.3	1.3	5.9	7.2	16.6	22.8	20.7	20.1	1.0
M	.1	.3	1.3	5.5	8.7	14.3	22.9	21.1	20.1	.8
D	.0	.0	.0	.4	1.5	2.3	.1	.4	.0	.2
N	2.5	4.6	16.6	52.0	57.1	60.0	56.6	29.4	20.1	1.1
Item 5: d_i = –.6, SUM = 17.1										
O	.3	1.1	6.3	37.6	55.7	36.1	15.5	4.7	1.7	.1
M	.4	1.2	7.7	42.0	56.1	38.8	20.6	6.4	2.9	1.1
D	.1	.1	1.4	4.4	1.6	2.7	5.1	1.7	1.2	.0
N	2.5	4.6	16.6	52.0	57.1	60.0	56.6	29.4	20.1	1.1
Item 6: d_i = –.3, SUM = 14.2										
O	.2	1.0	8.6	37.6	56.1	51.1	26.3	10.1	5.8	.2
M	.4	1.0	6.0	33.6	56.0	48.5	26.3	7.9	3.4	.1
D	.2	.0	2.6	4.0	.1	2.6	.0	2.2	2.4	.1
N	2.5	4.6	16.6	52.0	57.1	60.0	56.6	29.4	20.1	1.1
Item 7: d_i = .8, SUM = 7.7										
O	.2	.4	2.0	11.4	18.8	32.3	52.4	25.1	8.3	.2
M	.2	.4	2.2	10.4	18.7	34.6	53.0	25.9	10.5	.3
D	.0	.1	.2	1.0	.1	2.3	.6	.8	2.2	.1
N	2.5	4.6	16.6	52.0	57.1	60.0	56.6	29.4	20.1	1.1
Item 8: d_i = 1.2, SUM = 6.3										
O	.1	.4	1.8	9.7	14.8	26.7	49.5	28.9	14.6	.5
M	.2	.4	1.9	9.0	15.7	28.6	46.8	28.6	12.7	.4
D	.1	.0	.1	.7	.9	1.9	2.7	.3	1.9	.1
N	2.5	4.6	6.6	52.0	57.1	60.0	56.6	29.4	20.1	1.1
Item 9: d_i = 1.7, SUM = 7.1										
O	.0	.2	1.4	5.5	11.9	21.7	30.9	26.4	17.6	.4
M	.1	.3	1.5	6.7	11.1	19.0	31.6	27.1	18.1	.6
D	.1	.1	.1	1.2	.8	2.7	.7	.7	.5	.2
N	2.5	4.6	16.6	52.0	57.1	60.0	56.6	29.4	20.1	1.1

continued on the next page

Table 2, continued

Fit Values, [O_{qi} (O), M_{qi} (M), Absolute Difference (D) Between O_{qi} and M_{qi}], and N_q (N), for the 10 PARELLA Homogeneous Items From an Analysis Without Deviant Items, and Fit Values for the Deviant Items From an Analysis in Which Items Were, in Turn, Added to the Original 10 Items (d_i is the Generating Location for the PARELLA Homogeneous Items, and Estimated Location for the Deviant Items; SUM is the Sum of Differences; and MAX is the Largest Sum of Differences Observed Among the PARELLA Homogeneous Items)

	Node									
	1	2	3	4	5	6	7	8	9	10
Item 10: d_i = 2.0, SUM = 5.9										
O	.1	.3	1.3	5.9	7.2	16.6	22.8	20.7	20.1	1.0
M	.1	.3	1.3	5.5	8.7	14.3	22.9	21.1	20.1	.8
D	.0	.0	.0	.4	1.5	2.3	.1	.4	.0	.2
N	2.5	4.6	16.6	52.0	57.1	60.0	56.6	29.4	20.1	1.1
Deviant Items										
Item Rasch 1.5: d_i = 0.0, SUM = 41.0, MAX = 8.0										
O	.9	.7	9.4	23.3	55.5	96.3	54.1	5.4	8.1	.2
M	.2	.2	4.6	12.7	51.0	95.8	41.5	3.3	3.5	.1
D	.7	.5	4.8	10.6	4.5	.5	12.6	2.1	4.6	.1
N	3.5	2.0	20.7	26.2	56.5	96.5	60.5	10.2	23.0	1.0
Item Rasch 0: d_i = 1.0, SUM = 11.8, MAX = 9.7										
O	.2	.9	5.4	7.8	25.7	46.8	25.7	28.0	3.2	.4
M	.1	.7	2.8	6.1	23.7	48.4	27.0	30.1	3.1	.2
D	.1	.2	2.6	1.7	2.0	1.6	1.3	2.1	.1	.2
N	1.6	8.1	22.7	33.0	76.0	87.0	29.0	34.8	6.3	.8
Item Rasch 1.5: d_i = 2.4, SUM = 5.3, MAX = 12.9										
O	.3	1.5	.6	5.0	4.9	8.5	12.6	7.5	13.7	1.5
M	.1	.4	.2	3.2	4.2	8.2	12.7	8.0	13.8	1.6
D	.2	1.1	.4	1.8	.7	.3	.1	.5	.1	.1
N	2.3	10.4	4.3	38.5	47.2	65.7	67.9	26.2	25.7	1.7
Item Broad: d_i = 0.0, SUM = 18.7, MAX = 15.9										
O	.4	6.0	11.8	26.5	74.0	84.0	15.4	5.0	2.6	.2
M	.1	2.3	7.1	22.8	73.6	80.7	15.1	3.3	2.0	.2
D	.3	3.7	4.7	3.7	.4	3.3	.3	1.7	.6	.0
N	1.1	14.5	23.2	37.0	75.6	89.0	29.6	13.3	14.7	2.1
Item Small: d_i = 0.0, SUM = 24.8, MAX = 9.0										
O	.1	.7	7.1	22.6	64.4	28.1	18.3	5.0	.6	.1
M	.3	.9	8.7	30.3	66.3	29.8	26.6	6.7	1.9	.3
D	.2	.2	1.6	7.7	1.9	1.7	8.3	1.7	1.3	.2
N	2.0	5.0	30.4	61.5	76.6	34.5	54.0	23.4	10.2	2.3
Item Constant .3: d_i = 2.1, SUM = 45.0, MAX = 11.8										
O	.3	.3	1.9	13.9	21.2	19.7	23.3	16.5	2.3	1.6
M	.1	.1	.6	3.1	8.7	12.5	21.0	26.6	2.3	2.0
D	.2	.2	1.3	10.8	12.5	7.2	2.3	10.1	.0	.4
N	2.5	1.8	8.6	35.0	69.2	66.1	66.5	44.9	2.3	3.1
Item Constant .5: d_i = .2, SUM = 31.1, MAX = 10.3										
O	.6	7.2	6.4	23.1	54.9	5.9	28.5	11.1	7.2	.2
M	.1	2.6	4.2	25.3	64.7	5.9	37.0	14.2	7.0	.2
D	.5	4.6	2.2	2.2	9.8	.0	8.5	3.1	.2	.0
N	.9	14.1	15.8	60.6	93.3	5.9	50.7	32.4	25.3	1.0
Item Constant .7: d_i = .1, SUM = 26.4, MAX = 11.4										
O	1.9	.3	12.2	20.1	52.8	31.6	77.2	8.8	9.7	.4
M	.7	.1	5.7	14.4	53.9	31.8	75.3	5.3	3.9	.1
D	1.2	.2	6.5	5.7	1.1	.2	1.9	3.5	5.8	.3
N	6.0	.6	20.4	30.6	66.9	32.3	112.1	13.6	16.7	.9

items is easily explained, which leaves only the marginal (non)fit of the broad item. Thus, two or more differences between O_{qi} and M_{qi} greater than 4.0, or a sum of differences greater than 20, is indicative of an inequality between the empirical and expected IRFs for a sample size of 300 with 10 nodes used for the step function estimate of $g(\beta)$. As noted above, for sample sizes larger than 300 these rules should be relaxed. For sample sizes smaller than 300 they should be used more strictly.

Applications

Measurement of Attitudes Toward Capital Punishment

Andrich (1988) used a small dataset consisting of the responses of 54 graduate students who took an introductory course in educational measurement and statistics. Eight items were used to measure their attitude toward capital punishment. The responses to the items were dichotomous, with 1 indicating agreement and 0 indicating disagreement. The phrasing of the items (Table 3) shows clearly that a person's response will be determined by proximity relations (i.e., the closer a person's location to the location of the item, the larger the probability that the person will endorse the item). Consequently, the PARELLA model might be appropriate to describe the data. (A sample size of 54 persons is not large enough to allow for reliable estimates. Because Andrich used the same data to illustrate his model, however, this was ignored for illustrative purposes.)

The results of the PARELLA analysis of Andrich's (1988) data are displayed in Table 3. The order of the statements as found by Andrich was the initial order. The estimated order of the statements according to the PARELLA model agrees with the order found by Andrich (the results in Table 3 are a linear transformation of his actual results, but the transformation parameters are unfortunately unknown). The fit of the statements to the PARELLA model appears to be good: the largest difference between the O_{qi} and E_{qi} observed was 1.6, and the largest sum of differences was 3.4.

Table 3
Choice Percentage, Item Location, SUM, Two Largest Observed Values of Fit Diagnostic (D_1 and D_2), and Andrich's Scale Values for the PARELLA Homogeneous Statements Concerning Capital Punishment (N = 54, Q = 20)

Statement	Percent	Location	SUM	D_1	D_2	Andrich
1. Capital punishment is one of the most hideous practices of our time.	44	−1.99	2.8	.5	.4	1.7
2. The state cannot teach the sacredness of human life by destroying it.	65	−1.17	3.1	.8	.8	1.9
3. Capital punishment is not an effective deterrent to crime.	67	−1.15	3.4	1.3	.6	1.9
4. I don't believe in capital punishment but I am not sure it isn't necessary.	46	−.30	3.3	.8	.7	4.1
5. I think capital punishment is necessary but I wish it were not.	48	.73	1.2	.5	.2	7.4
6. Until we find a more civilized way to prevent crime we must have capital punishment.	44	.86	2.3	1.4	.3	7.9
7. Capital punishment gives the criminal what he deserves.	35	1.45	2.9	1.6	.4	8.4
8. Capital punishment is justified because it does act as a deterrent to crime.	35	1.56	.6	.2	.1	8.7
Power		1.25				

Table 4
Step Function Estimates of $g(\beta)$ for Statements Concerning Capital Punishment

Node	Weight
−1.58	.00
−1.41	.08
−1.24	.38
−1.07	.09
−.90	.00
−.73	.00
−.55	.00
−.38	.00
−.21	.00
−.04	.00
.13	.00
.31	.01
.48	.04
.65	.12
.82	.13
.99	.03
1.17	.01
1.34	.04
1.51	.06
1.68	.01

The order of the statements constitutes a dimension ranging from "against" to "for" capital punishment. This can be easily verified from the content of the statements. The estimated power was 1.25. This suggests the presence of a moderately strong parallelogram structure (i.e., not too many 0s between 1s). The step function estimate of $g(\beta)$ is shown in Table 4. Two groups of persons can be distinguished: a group located around −1.20 that could be labeled "moderately against capital punishment," and a group located around .80 that could be labeled "moderately for capital punishment."

Measurement of Attitudes Toward Nuclear Power

Formann (1988) reported a dataset consisting of the responses of 600 persons to five statements concerning nuclear power. The responses were dichotomous, with 1 indicating agreement and 0 indicating disagreement. The phrasing of the statements makes it probable that proximity relations will largely determine a person's response. Consequently the PARELLA model might provide a good description of the structure underlying the data.

The initial order of the statements (see Table 5) on the latent trait was the order Formann (1988) expected the statements to have because of their content (ranging from "for" to "against" nuclear power). This order was in agreement with the order of the locations estimated by the PARELLA model. Table 5 also shows that the fit of the data to the model was not satisfactory. The sum of differences for Items 2 and 5 were 18.5 and 21.6, respectively. Two additional analyses were performed—one with Item 2 removed, and one with Item 5 removed.

Table 5 shows that the results after removal of Item 2 were encouraging: the largest sum of differences observed was 12.2, which is an acceptable value, considering the sample size of 600; the power

Table 5
Choice Percentage, Item Location, and Sum of Differences, for Statements Concerning Nuclear Power With All Items in the Analysis and With Item 2 Removed (–2) and Item 5 Removed (–5) (N = 600, Q = 20)

Statement	Percent		Analysis All	–2	–5
1. In the near future, alternate sources of energy will not be able to substitute nuclear energy.	32	Loc	–1.11	–1.35	–.85
		Sum	11.2	9.3	9.6
2. It is difficult to decide between the different types of power stations if one carefully considers all their pros and cons.	47	Loc	–.84		–.46
		Sum	18.5		8.0
3. Nuclear power stations should not be put into operation before the problems of radioactive waste have been solved.	83	Loc	–.02	–.19	.44
		Sum	5.6	3.5	6.1
4. Nuclear power stations should not be put into operation before it is proven that the radiation caused by them is harmless.	81	Loc	.65	.42	.87
		Sum	5.1	5.8	4.1
5. The foreign power stations now in operation should be closed down.	50	Loc	1.32	1.12	
		Sum	21.6	12.2	
Power			1.42	1.75	1.03

parameter increased from 1.42 to 1.75, which indicates that a strong parallelogram structure was present in the data. The results after removal of Item 5 were less encouraging: although the sum of differences indicated a good fit (i.e., the largest value observed was 9.6), the value of the power parameter decreased from 1.42 to 1.03.

Experience with PARELLA indicates that removal of a badly fitting item always leads to an increase in the value of the power parameter. Consequently, it may be concluded that Item 2 is the item deviating from the PARELLA model, and that the results without that item are the optimal solution for this

Table 6
Step Function Estimates of $g(\beta)$ for Statements Concerning Nuclear Power

Node	Weight
–1.28	.01
–1.05	.02
–.83	.04
–.60	.09
–.37	.13
–.14	.14
.09	.14
.31	.15
.54	.13
.77	.07
1.00	.03
1.23	.02
1.45	.01
1.68	.00
1.91	.00
2.14	.00
2.37	.00
2.59	.00
2.82	.00
3.05	.01

dataset. The ambivalent phrasing of Item 2 makes it likely that both those who are and those who are not in favor of nuclear power might agree with this statement. Thus, the threshold of this item will probably be greater than the threshold predicted by the PARELLA model.

The step function estimates of $g(\beta)$ for the optimal solution is presented in Table 6. The step function was unimodal and symmetric around 0.0. This indicates that most people had a slightly negative attitude toward nuclear power, but some people were in favor of or definitely against nuclear power.

Evaluation of Political Variables

Szirmai (1986) discussed the evaluation by 952 persons of 11 political items. The evaluations were made on a five-point scale (1 = "I agree strongly" to 5 = "I disagree strongly"; 0 was reserved for the response "I don't know"). These data were recoded (1,2 = 1) (0,3,4,5 = 0) in order to process them with the PARELLA model. This dichotomization was selected because it resulted in sensible choice percentages (i.e., not too small, not too large). Missing values were recoded 0, because it seemed unlikely that a person would not answer to a variable he/she agreed with. Szirmai (1986) expected the order of the statements on the latent trait to be interpretable as a left-wing to right-wing dimension of political attitude. It is very likely that proximity relations should determine a person's choice process, because some measures are expressions of a left-wing attitude and others are expressions of a right-wing attitude, and each person will be oriented either toward the right or the leftwing.

An initial order for the items (Table 7) was obtained with a nonparametric parallelogram model: the MUDFOLD model (Van Schuur, 1984). All 11 items were analyzed in the first run. Items 3, 7, 8, and 10 appear to have the worst fit (i.e., with the sum of differences substantially greater than 20). Due to the presence of other fit diagnostics at the time this dataset was analyzed, Item 10 was the first item to be removed from the dataset instead of Item 8. The remaining 10 items were analyzed in the second run. Item 8 had a bad fit and was removed from the item set. After removal of Item 2 in the third run, eight items remained.

There might be two reasons why 3 of the 11 items did not fit the PARELLA model. First, they may not have been appropriate for the measurement of left- to right-wing political attitudes as Szirmai intended. Second, they may have been appropriate, but their threshold or power parameter may have differed from those of the other items. The software of the PARELLA model does not permit different threshold or power parameters, however, so this assumption could not be tested.

The order of the remaining items did reflect a left-wing to right-wing dimension of political attitude (Table 8), with items that expressed a positive attitude toward social security and income leveling to the left side of the dimension and opposite measures to the right side of the dimension.

Table 9 shows that most persons were located between -.75 and .25, which indicates that most persons in the sample supported social security and income leveling. The value of the power parameter for the final item set was .69, which indicates that a weak parallelogram structure was present in the data, and that more "0s between 1s" would be observed.

Discussion

These results illustrated that the PARELLA model can be used to analyze data expected to result from proximity relations. Hoijtink (1991) presents additional analysis of empirical datasets. A number of issues remain, however, concerning the PARELLA model. The estimation procedure is the aspect of the PARELLA model that has been best developed. Further research is still needed in determining correct standard errors of the estimates, and a test of fit is needed for the adequacy of the IRFs defined by the PARELLA model with known distribution. Future research should also concern PARELLA-like models with a different power parameter and threshold parameter for each item (see Hoijtink, 1991).

Table 7
Item Location, Two Largest Observed Values of Fit Diagnostic (D_1 and D_2), and SUM, for a PARELLA Homogeneous Set of Szirmai's Political Items (N = 952, Q = 10)

	Run 1				Run 2				Run 3				Run 4			
Item	Loc.	D_1	D_2	SUM	Loc.	D_1	D_2	SUM	Loc.	D_1	D_2	SUM	Loc.	D_1	D_2	SUM
1	-1.19	1.9	1.8	7.5	-1.08	4.0	3.0	12.9	-.99	3.8	2.3	11.3	-1.12	2.7	2.3	7.8
2	-1.10	7.9	3.1	15.7	-.99	18.4	3.9	27.8	-.89	15.2	10.2	37.1				
3	-.96	16.7	7.2	26.5	-.87	13.5	2.0	19.9	-.76	7.5	6.6	21.1	-.85	6.4	2.8	11.5
4	-.80	6.8	3.2	7.7	-.67	6.1	2.6	13.1	-.60	4.0	1.7	8.1	-.80	3.2	1.7	6.3
5	-.76	11.1	5.4	24.1	-.66	12.5	4.2	26.1	-.54	5.3	5.0	15.0	-.67	2.2	1.0	6.8
6	-.57	9.6	7.2	23.8	-.32	2.5	1.9	8.5	-.19	5.5	2.0	10.4	-.39	2.0	.5	3.7
7	.57	14.1	8.3	29.3	.61	4.8	3.5	13.1	.77	11.9	4.3	22.0	.59	7.7	3.5	15.2
8	.75	20.2	19.0	55.2	.84	26.0	16.2	50.6								
9	1.18	2.1	.9	4.8	1.31	2.8	1.7	5.7	1.38	.6	.5	1.7	1.33	4.9	1.5	8.2
10	1.26	14.4	12.1	38.6												
11	1.61	5.2	5.1	14.0	1.83	5.2	4.3	14.3	1.81	2.9	2.6	8.7	1.91	5.5	4.3	11.8
Power	.81				.74				.82				.69			

Table 8
Choice Percentage, Item Location, and Sum of Differences for Each of Szirmai's Political Items

Statement	Percent	Loc.	SUM
1. Price compensation in flat rates, instead of percentages.	64	-1.12	7.8
3. Introduction of a minimum income for self-employed people.	75	-.85	11.5
4. Extra cuts in the incomes of higher public employees.	80	-.80	6.3
5. Cutting the incomes of the free professions.	84	-.67	6.8
6. Extra remuneration for dirty work and shift work.	87	-.39	3.7
7. Sacrificing purchasing power for an improvement in the employment situation.	50	.59	15.2
9. Decoupling state old age pensions from the minimum wage.	31	1.33	8.2
11. The abolition of rent subsidies.	23	1.91	11.8
2. Introduction of a ceiling for all incomes.	66		
8. Cutting the wages of public employees relative to non-public sector incomes.	44		
10. Cutting social security payments.	29		
Power		.69	

Table 9
Step Function Estimates of $g(\beta)$ for Szirmai's Political Measures

Node	Weight
-1.07	.01
-.90	.03
-.73	.11
-.56	.21
-.38	.17
-.21	.17
-.04	.16
.14	.10
.31	.04
.48	.01

References

Andrich, D. (1988). The application of an unfolding model of the PIRT type to the measurement of attitude. *Applied Psychological Measurement, 12,* 33–51.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores* (pp. 195–479). Reading MA: Addison-Wesley.

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., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35,* 179–197.

Coombs, C. H. (1964). *A theory of data.* Ann Arbor MI: Mathesis Press.

Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: An elementary introduction.* Englewood Cliffs NJ: Prentice-Hall.

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 B, 39,* 1–38.

DeSarbo, W. S., & Hoffman, D. L. (1986). Simple and weighted unfolding threshold models for the spatial representation of binary choice data. *Applied Psychological Measurement, 10,* 247–264.

Fischer, G. H. (1974). *Einfuhrung in die theorie psy-*

chologischer tests: Grundlagen und Anwendungen [Introduction to the theory of psychological tests: Principles and application]. Bern, Switzerland: Huber.

Formann, A. K. (1988). Latent class models for nonmonotone dichotomous items. *Psychometrika, 53,* 45–62.

Glas, C. A. W. (1989). *Contributions to estimating and testing Rasch models.* Unpublished doctoral dissertation, University of Twente, The Netherlands.

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Measurement and prediction: Studies in social psychology in World War II, Vol IV.* Princeton NJ: Princeton University Press.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston MA: Kluwer-Nijhoff.

Hoijtink, H. (1991). *PARELLA: Measurement of latent traits by proximity items.* Leiden, The Netherlands: DSWO Press.

Kiewiet, D. J. (1988). *Een programma voor de schatting van persoons-parameters in het PPPA-model* [A program for the estimation of person parameters in the PARELLA model] (Report No. HB-88-937-SW). Groningen, The Netherlands: Heymans Bulletins Psychologische Instituten.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society B,* 226–223.

Mislevy, R. J., & Bock, R. D. (1982). Biweight estimates of latent ability. *Educational and Psychological Measurement, 42,* 725–737.

Mislevy, R. J., & Bock, R. D. (1990). *BR063.* Mooresville IN: Scientific Software, Inc.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: Paedagogiske Institut.

Szirmai, A. (1986). *Inequality observed: A study of attitudes toward income inequality.* Unpublished doctoral dissertation, University of Groningen, The Netherlands.

Takane, Y. (1983). *Choice model analysis of the pick-any/n type of binary data.* Paper presented at the European Psychometric and Classification meeting, Jouy-en-Josas, France.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika, 47,* 175–186.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika, 4,* 567–577.

Van Schuur, H. (1984). *Structure in political beliefs: A new model for stochastic unfolding with application to European party activists.* Amsterdam, The Netherlands: CT Press.

Wainer, H., & Wright, B. D. (1980). Robust estimation of ability in the Rasch model. *Psychometrika, 45,* 373–391.

Acknowledgments

The author gratefully acknowledges Ivo Molenaar and Wijbrandt van Schuur for their advice and encouragement during the course of the investigation, Derk Jan Kiewiet for writing the program for the ML estimator for the person parameter, and Anne Boomsma, Wendy Post, Tom Snijders, and David Thissen for their comments on aspects of the research.

Author's Address

Send requests for reprints or further information to Herbert Hoijtink, Department of Statistics and Measurement Theory, University of Groningen, Grote Kruisstraat 2/1, 9712 TS, Groningen, The Netherlands. Requests for information on the software for the PARELLA model should be sent to: iec ProGamma, P.O. Box 841, 9700 AV Groningen, The Netherlands.