

 Open access • Journal Article • DOI:10.1177/014662169001400305

Rasch models in latent classes: An integration of two approaches to item analysis

— [Source link](#) 

Jürgen Rost

Published on: 01 Sep 1990 - Applied Psychological Measurement (SAGE Publications)

Topics: Latent variable model, Latent class model, Polytomous Rasch model, Local independence and Probabilistic latent semantic analysis

Related papers:

- [Modeling Item Responses When Different Subjects Employ Different Solution Strategies.](#)
- [Estimating the Dimension of a Model](#)
- [A Rasch Model for Partial Credit Scoring.](#)
- [Probabilistic Models for Some Intelligence and Attainment Tests](#)
- [A logistic mixture distribution model for polychotomous item responses](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/rasch-models-in-latent-classes-an-integration-of-two-3vafzlssgk>

Rasch Models in Latent Classes: An Integration of Two Approaches to Item Analysis

Jürgen Rost
University of Kiel

A model is proposed that combines the theoretical strength of the Rasch model with the heuristic power of latent class analysis. It assumes that the Rasch model holds for all persons within a latent class, but it allows for different sets of item parameters between the latent classes. An estimation algorithm is outlined that gives conditional

maximum likelihood estimates of item parameters for each class. No a priori assumption about the item order in the latent classes or the class sizes is required. Application of the model is illustrated, both for simulated data and for real data. *Index terms: conditional likelihood, EM algorithm, latent class analysis, Rasch model.*

An increasing amount of research has been done on the integration of Rasch models and latent class analysis during the past five years (e.g., Langeheine & Rost, 1988). Using the concept of located classes by Lazarsfeld and Henry (1968), Formann (1985, 1989a) described the Rasch model (RM) as a latent class model (LCM) with constrained logistic parameters. Clogg (1988), and Clogg, Lindsay, and Grego (1989), employed similar restrictions for the LCM, and obtained a model that not only provides conditional maximum likelihood estimates for the item parameters, but also satisfactory person parameter estimates.

However, latent class analysis can do more than simply imitate the Rasch model. It can be used to weaken the strong assumptions of the Rasch model—specifically, the assumption of parallel item characteristic curves or item response functions (IRFs). Rost (1988a) pointed out that the latent classes must be ordered whenever the IRFs are monotone and vice versa. This makes the LCM a tool for checking the unidimensionality of a test without specifying the shape of its IRFs. Croon (1988), Heinen, Hagenars, and Croon (1989), and Formann (1989b) describe ways of imposing these order restrictions on the LCM.

Likewise, in the case of polychotomous generalizations of the Rasch model, an integration with latent class analysis has occurred. Rost (1988b) transferred the threshold approach of modeling multicategorical item responses in generalized Rasch models into the framework of latent class models.

This paper presents a new, straightforward way of combining Rasch and latent class models that overcomes the deficiencies of both approaches and retains their positive features. The idea is that the RM describes the response behavior of all persons within a latent class, but that different sets of item parameters hold for the different latent classes. Because the classes are not known beforehand (i.e., they are “latent”), such a model is heuristic in the sense that it identifies subpopulations of individuals as being Rasch scalable. On the other hand, the prominent feature of the RM—namely, allowing for distribution-free item parameter estimates—is a great advantage for parameter estimation in the proposed extension, because ability distributions within classes are unknown and should not be made subject to a priori assumptions.

Item analysis using the Rasch model requires the item difficulties to be constant for all persons.

APPLIED PSYCHOLOGICAL MEASUREMENT

Vol. 14, No. 3, September 1990, pp. 271-282

© Copyright 1990 Applied Psychological Measurement Inc.

0146-6216/90/030271-12\$1.85

This is a strong requirement that may be violated in some cases by poorly-constructed items. In other cases, the diagnostic intention is to have items with varying difficulties for different subgroups of persons, because they indicate alternative strategies of solving items or distinguishing cognitive structures. Hence, a latent class approach to item analysis would be desirable.

LC models require response probabilities, however, that hold for all individuals in a latent class. This is also a very strong requirement, and experience has shown that for every cognitive structure or solution strategy, several latent classes are needed in order to account for individual differences in ability (see example below). Therefore, a generalized LC model allowing for ability differences within classes would be desirable.

The Mixed Rasch Model

Let p_{vig} denote the probability of person v giving a correct answer or “yes” response to item i under the condition that this person belongs to latent class g . It is assumed that these response probabilities can be described by the dichotomous Rasch model

$$p_{vig} = \exp(\tau_{vg} + \sigma_{ig}) / [1 + \exp(\tau_{vg} + \sigma_{ig})] \quad , \quad (1)$$

where τ_{vg} is the person’s ability and σ_{ig} is the item easiness parameter. The usual indeterminacy constraint $\sum_i \sigma_{ig} = 0$ must hold within each class g . If it is further assumed that the latent classes are mutually exclusive and exhaustive, the well-known latent class structure appears:

$$p_{vi} = \sum_g \pi_g p_{vig} \\ = \sum_g \pi_g \exp(\tau_{vg} + \sigma_{ig}) / [1 + \exp(\tau_{vg} + \sigma_{ig})] \quad , \quad (2)$$

where p_{vi} is the unconditional response probability and π_g is the class size parameter or “mixing proportion” with constraints $0 < \pi_g < 1$ and $\sum_g \pi_g = 1$.

These equations do not yet define the entire model because they do not specify what to do with the person parameters τ_{vg} . The solution used here is to condition out the person parameters, which can only be done with a Rasch-like model structure. The advantages of this choice are outlined below.

In order to derive the likelihood function, the pattern probabilities $p(\mathbf{x})$ —the probabilities of a response vector $\mathbf{x} = (x_1, x_2, \dots, x_i, \dots, x_k)$, where $x_i = 0$ or 1 —are needed first. Under the usual assumption that the partition into latent classes holds for all k items of a test, the pattern probability can be written as

$$p(\mathbf{x}) = \sum_g \pi_g p(\mathbf{x}|g) \quad , \quad (3)$$

where the conditional pattern probability $p(\mathbf{x}|g)$ is simply the product of response probabilities defined by Equation 1 over all items. It is well known that in the Rasch model the number of correct responses (i.e., the score $r = \sum_i x_i$) is a sufficient statistic for estimating τ ; hence, all persons with the same score r obtain the same τ estimate. Consequently the pattern probability $p(\mathbf{x}|g)$ can be conditioned on the score r associated with a given pattern so that the following equation holds:

$$p(\mathbf{x}|g) = p(\mathbf{x}|g,r) \cdot p(r|g) \quad . \quad (4)$$

This is a useful factorization because only the first factor depends on the item parameters σ_{ig} ,

$$p(\mathbf{x}|g,r) = \exp(\sum_i x_i \sigma_{ig}) / \Phi_r[\exp(\sigma)] \quad , \quad (5)$$

where Φ_r are the symmetric functions of order r of the delogarithmized item parameters, and only the second factor depends on the ability distribution in class g . There is no need to restrict the score probabilities $p(r|g)$ in any case (e.g., by fitting a normal distribution). In the model presented here, the class-specific score probabilities $p(r|g)$ themselves are treated as model parameters, so that no further assumptions about the ability distributions within the latent classes are required. The “mixed

Rasch model” is “distribution-free” in the same sense as the simple Rasch model.

Combining these elements, the likelihood function defining the model is

$$L = \prod_x \left\{ \sum_g \pi_g \pi_{rg} \exp(\sum_i x_i \sigma_{ig}) / \Phi_r[\exp(\sigma)] \right\}^{n(x)} \quad (6)$$

where $n(\mathbf{x})$ denotes the observed number of response patterns \mathbf{x} , and the score probabilities $\pi_{rg} = p(r|g)$ have been renamed, in keeping with the convention of using greek letters to denote model parameters.

Hence the number of independent model parameters is composed as follows:

1. $h - 1$ class size parameters π_g , where h is the number of classes,
2. $(k - 1)h$ class-specific item parameters, where k is the number of items and must be decreased by 1 because of the norming constraint, and
3. $2 + h(k - 2)$ class-specific score probabilities, because one parameter in each class depends on the sample size and the class size, and the two parameters for the 0 and 1 vectors are class independent (see below).

It can easily be verified that the Rasch model is the 1-class solution of the proposed model in Equation 6. Similarly, the simple latent class model is a special case of Equation 6. It is obtained from the restriction that there is no ability variation within the classes (i.e., the individual parameters of all persons in a latent class are constant). In such a case, the response probability in a latent class (Equation 1) can be reparameterized in the following way:

$$p_{vig} = \exp(\tau_g + \sigma_{ig}) / [1 + \exp(\tau_g + \sigma_{ig})] = \pi_{ig} \quad (7)$$

Because only $k - 1$ item parameters σ_{ig} in a particular class g are independent, the additive constant τ_g serves as the k th parameter. The probability parameters π_{ig} defined in this way are the “latent probabilities” of regular latent class analysis. Consequently, the proposed model in Equation 6 is the common generalization of both the Rasch and the latent class models.

Kelderman and Macready (in press) treated the Rasch model as a special case of the general log-linear model with a latent categorical variable, as introduced by Haberman (1979). Their point of view is more general; the mixed Rasch model proposed here is only one of several ways of combining a latent trait variable with a latent class variable. The mixed Rasch model corresponds, in particular, to Model 5, Table 1, in Kelderman and Macready (in press). Yet this embedding into the most general model structure makes the model somewhat difficult to apply with currently-available software.

The algorithm proposed in the following section is capable of dealing with more than two (e.g., four or five) latent classes and a relatively large number of items (15 to 20), and can be directly generalized to polychotomous item response models (see Rost, in press). Furthermore, it makes no requirements on the starting values for the parameter estimates, and no a priori hypotheses about item difficulties within the latent classes are needed.

Mislevy and Verhelst (in press) proposed a mixed Rasch model based on such a priori hypotheses about item difficulties under different solution strategies. They employ a marginal likelihood method that seems to be limited in its applicability with regard to the number of latent classes and the number of items.

An EM Algorithm for Parameter Estimation

The EM algorithm (Dempster, Laird, & Rubin, 1977) has been shown to work well for all types of latent class models (cf. Rost, 1988a); consequently it was used in this research.

E Step

In the *Expectation* step of this iterative procedure, the expected pattern frequencies within each latent class are estimated on the basis of the observed global pattern frequencies and preliminary estimates of the model parameters. These pattern frequencies within classes $n(\mathbf{x},g)$ are obtained by weighting the observed frequencies with a proportion indicating the probability of observing this pattern in the class under consideration:

$$\hat{n}(\mathbf{x},g) = n(\mathbf{x})\pi_g p(\mathbf{x}|g) / [\sum_g \pi_g p(\mathbf{x}|g)] \quad , \quad (8)$$

and

$$p(\mathbf{x}|g) = \pi_{r_g} \exp(\sum_i x_i \sigma_{ig}) / \Phi_r[\exp(\sigma)] \quad . \quad (9)$$

This proportion depends only on the model parameters π_g , σ_{ig} , and π_{r_g} , as specified by Equations 4 and 5.

There is an exception, however, for the two response patterns with scores 0 and k (i.e., the 0 and 1 vectors). The proportion of these vectors belonging to one of the latent classes cannot be determined by means of the model parameters. The ratio of Equation 9 is 1 for both of these patterns because the symmetric functions of order 0 and k equal 1, and the numerator is 1 because of the norming condition. Consequently, the expected pattern frequencies $\hat{n}(\mathbf{0}|g)$ and $\hat{n}(\mathbf{1}|g)$ depend not on the item parameters σ , but only on the score parameters π_{r_g} . The 0 and 1 vectors are the only patterns in their score groups, $r = 0$ and $r = k$; thus, no improvement of their expected frequencies is obtained from Equation 8. The expected frequencies of these two patterns do not diverge from the initial estimates of their score probabilities.

This makes sense, of course: A “distribution-free” item response model cannot predict the class membership of a response pattern for which its occurrence depends only on the ability distribution and not on the item parameters. The situation calls for a practical solution, however. The solution used in this study makes no arbitrary assumption about the proportion of these two pattern frequencies in the latent classes. Instead, they were excluded from the estimation procedure. This did not affect the item parameter estimates because a conditional maximum likelihood method was employed (see M step below); nor did it affect the score parameter estimates where $r > 0$ and $r < k$. The only consequence was that the class size parameters π_g no longer summed to 1, but the following side constraint held:

$$\sum_g \pi_g = [N - n(\mathbf{0}) - n(\mathbf{1})] / N \quad , \quad (10)$$

where N is the sample size, and $n(\mathbf{0})$ and $n(\mathbf{1})$ are the observed frequencies of the 0 and 1 vectors, respectively.

M Step

In the M step, maximum likelihood (ML) estimates of the model parameters are computed on the basis of the expected pattern frequencies within the classes provided by Equation 8. The item parameters and score parameters can be estimated separately for each latent class by maximizing the log-likelihood function of the data in a class g :

$$\log L_g = \sum_x \hat{n}(\mathbf{x},g) [\log(\pi_{r_g}) + \sum_i x_i \sigma_{ig} - \log(\Phi_r)] \quad , \quad (11)$$

where $\Phi_r = \Phi_r[\exp(\sigma)]$ is an abbreviated notation for the symmetric functions of all item parameters. When the first partial derivatives of this function are set to zero, the following system of estimation equations is obtained for the item parameters (Gustafsson, 1980):

$$\exp(\hat{\sigma}_{ig}) = n_{ig} / \sum_r (n_{r_g} \Phi_{r-1,i} / \Phi_r) \quad , \quad (12)$$

In this equation, $\Phi_{r-1,i}$ denotes the symmetric functions of order $r - 1$ of all item parameters except item i (i.e., the first partial derivative of Φ). The sufficient statistics n_{ig} (i.e., the number of persons giving a 1 response to item i in class g) are simple summations of the expected pattern frequencies $\hat{n}(x,g)$, as are the score frequencies n_{ig} . The only difference from typical data is that these “counts” are not integer-valued because they are expected counts.

The system of equations described by Equation 12 can be used to improve the item parameter estimates iteratively. The computations of the symmetric functions are not problematic, even for large numbers of items, when the so-called “summation algorithm” described by Gustafsson (1980) is employed. One iteration of item parameter estimation within each M step proved to be sufficient, as the estimates converged during the iterations of the EM algorithm.

The estimators of the score probabilities are explicit:

$$\hat{n}_{ig} = n_{ig} / n_g \quad , \quad (13)$$

and

$$n_g = \sum_x n(x,g) \quad ; \quad (14)$$

they cause no problems at all. Likewise the class size parameters are estimated by the total number of patterns in a class divided by the sample size:

$$\hat{\pi}_g = n_g / N \quad . \quad (15)$$

It should be noted again that the frequencies of the 0 and 1 vectors are excluded from all these computations because their proportional count in a latent class cannot be predicted by the model. Although the EM algorithm converges very slowly, none of the computations reported below took more than a few minutes on a 24 MHz PC-AT. The program was applied to a maximum number of 15 items, but it should also work with higher numbers if the sample of persons is sufficiently large.

Example Application

Data

As part of a larger study on physics education (Rost, Häußler, & Hoffmann, 1989), the physics knowledge of adults between the ages 20 and 40 years was assessed by means of a paper-and-pencil test ($N = 869$). A subtest containing 10 items was analyzed with the Rasch model and latent class analysis.

Results

The test did not prove to be Rasch scalable. The conditional likelihood ratio statistic by Andersen (1973) resulted in $\chi^2 = 63$ with 9 degrees of freedom when the sample was divided into high- and low-scoring examinees. Table 1 shows that item fit indices defined by Wright and Masters (1982) suggest that misfit was not a question of selecting single items. This fit index is a standardized weighted squared residual with mean 0 and variance 1 (Wright & Masters, 1982, p. 101). For the data in Table 1, 6 out of 10 items had a poor fit (i.e., values above or below two standard deviations from the mean).

Latent class analysis of these data showed an interesting result. For up to three latent classes, the response probabilities of all items were in the same order across the classes (see Figure 1a). According to the concept of ordered classes, this could be taken as an indication of monotone IRFs (see Rost, 1988a), but even this does not indicate why the Rasch model fit so badly. The answer was given by the four-class solution (see Figure 1b).

Table 1
 Item Fit Indices for
 the Physics Knowledge Test

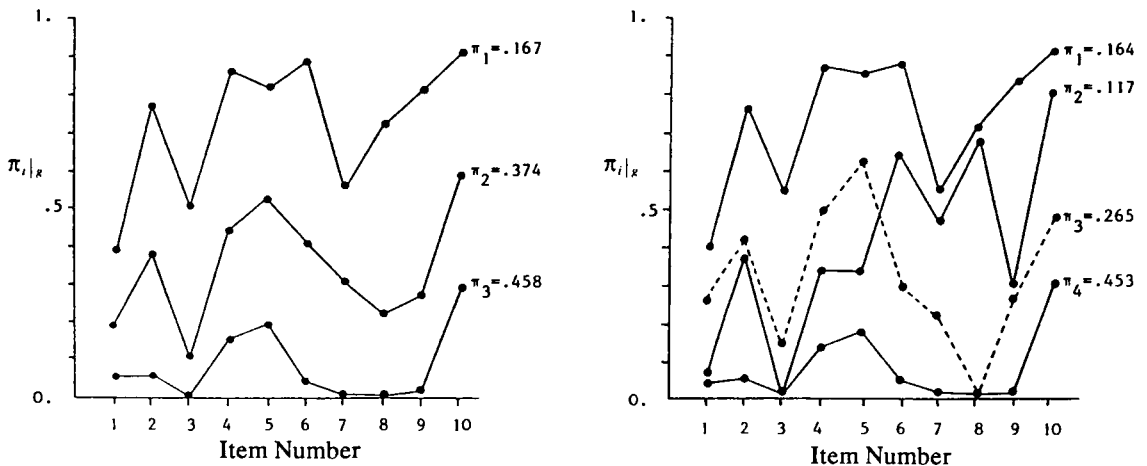
| Item No. | Item Fit |
|----------|----------|
| 1 | 3.0 |
| 2 | -1.6 |
| 3 | -.5 |
| 4 | .5 |
| 5 | 2.2 |
| 6 | -4.0 |
| 7 | 1.1 |
| 8 | -2.2 |
| 9 | -3.5 |
| 10 | 3.2 |

Although the upper and lower latent classes remained stable when moving from three to four classes, the middle class split up into two classes with intersecting profiles. Items 1 to 5 were easier for examinees in class 3, and items 6 to 10 were easier for class 2 (the items were reordered according to this result). This result has an obvious interpretation: The first five items asked for knowledge that is of a more theoretical nature (e.g., knowing that a “perpetuum mobile” would not work), and the last five items asked for more practical knowledge (e.g., drawing a circuit diagram for two lamps). This result suggests that the population seemed to be heterogeneous, in the sense that there are both practically-oriented and theoretically-oriented people with respect to the knowledge measured by this test.

Nevertheless, there are strong differences in people’s abilities, as shown by the two extreme classes. This example is typical of the analysis of achievement and attitude data by means of latent class models (e.g., Rost, 1985). The first three or four latent classes are usually used to represent quantitative differences between people’s abilities (i.e., they have nonintersecting profiles). Structural or “qualitative” differences (e.g., practical versus theoretical orientation) only occur in the 4- or 5-class solutions.

Figure 1

Profiles of Item Response Probabilities of Ordinary Latent Class Analysis for the Physics Knowledge Test
 (a) Three-Class Solution
 (b) Four-Class Solution



This effect should disappear when the mixed Rasch model in Equation 6 is applied. Figure 2 shows the item parameter profiles of the two-class solution. Taking the different scaling of response probabilities and logistic item parameters into account, the profiles look much the same as the two middle classes in Figure 1b. As expected, the structural differences in the examinees' abilities are already shown in the two-class solution.

Figure 3 gives the score distributions for both classes. It can be seen that the score distribution and the related ability distribution of the practically-oriented examinees (class 2) is flatter and less left-skewed than the abilities of the theoretically-oriented examinees (class 1). 46% of all examinees in the sample belonged to class 1 and 36% to class 2. 18% could not be classified because they solved none [$n(0) = 150$] or all [$n(1) = 8$] of the items.

The model fit can be compared by means of the AIC index (Bozdogan, 1987), which is a simple function of the log-likelihood (L) and the number of independent parameters (m) of a model:

$$AIC = -2 \ln L + 2m \quad (16)$$

According to this criterion, the four-class solution of the simple LC model fit the data better than the two-class solution of the mixed Rasch model (see Table 2). This is an indication that the test is not really Rasch scalable, even if examinee heterogeneity is allowed for in the sense of the proposed

Figure 2

Item Parameter Profiles of the Two-Class Solution of the Mixed Rasch Model for the Physics Knowledge Test

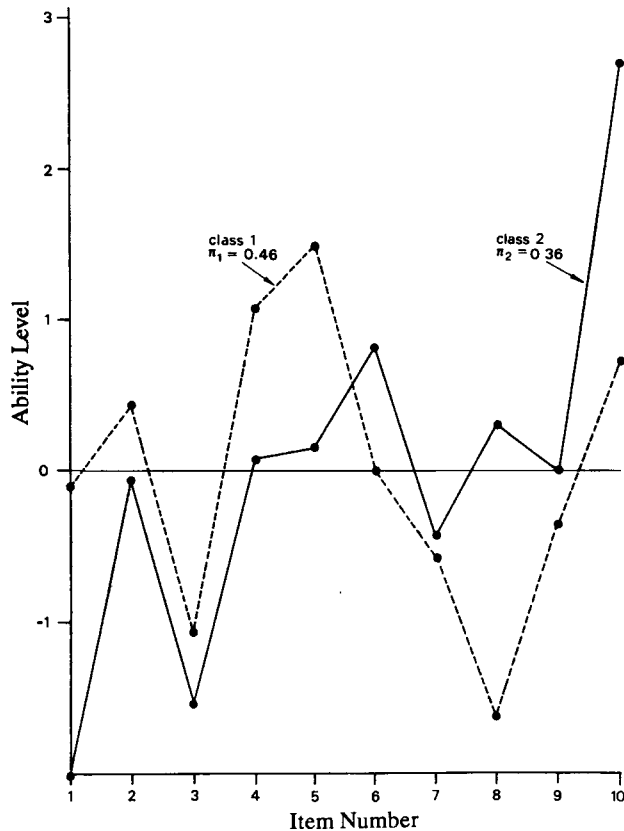
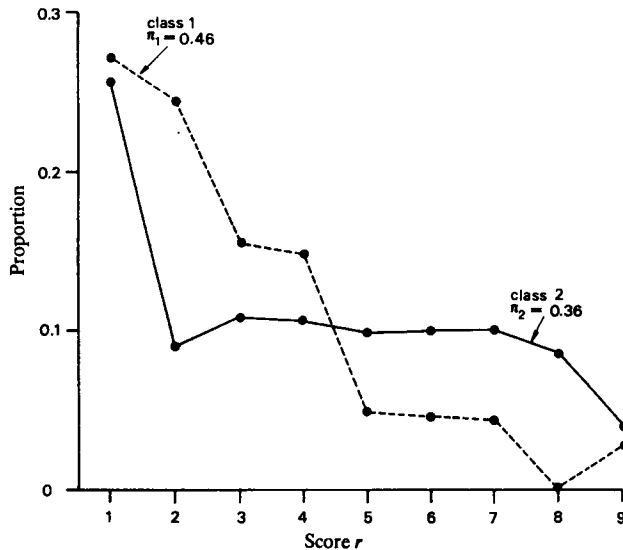


Figure 3
Score Distributions of the Two-Class Solution of the Mixed Rasch Model



model. It can be seen, however, that the improvement of model fit is most drastic if the two-class mixed Rasch model is compared with the simple Rasch model one-class solution) or the simple two-class model.

Simulation Analyses

Three sets of data were simulated in order to demonstrate the power of the algorithm in detecting the heterogeneity of the data or its ability to “recognize” that there is nothing to detect. One set was Rasch scalable, and one set consisted of two equally-large subgroups of Rasch scalable “examinees” (simulees) with reverse order of item difficulties. The third dataset was composed of three heterogeneous subgroups, each of a different size, and one with a zero variance of abilities and a near-to-zero variation of item parameters.

The sample size of all datasets was $N = 1,800$ with 10 items. Within each subgroup of Rasch homogeneous simulees, four levels of τ were specified, namely, 2.7, -.9, +.9, and -2.7 each for a quarter of the subpopulation. The item parameter sets used to simulate the data and the estimated parameter values under the mixed Rasch model with appropriate number of classes are shown in

Table 2
AIC Goodness-of-Fit Statistic, Log Likelihood (L), and
Number of Parameters (m) for the Latent Class Model
and the Mixed Rasch Model

| No. of Classes | Latent Class Model | | | Mixed Rasch Model | | |
|-------------------|--------------------|-----|------|-------------------|-----|------|
| | L | m | AIC | L | m | AIC |
| 1 | -4976 | 10 | 9972 | -4332 | 19 | 8702 |
| 2 | -4359 | 21 | 8760 | -4282 | 39 | 8642 |
| 3 | -4295 | 32 | 8654 | -4256 | 59 | 8630 |
| 4 | -4258 | 43 | 8602 | -4240 | 79 | 8638 |

Table 3; the parameter values were reproduced very well by the model, even in the most difficult third case.

Table 3
 Item Parameter Values Used for Data Simulation and Their Estimates of the Mixed Rasch Model With the Appropriate Number of Latent Classes: One Class for Dataset 1 (C_1), Two Classes for Dataset 2 (C_1, C_2), and Three Classes for Dataset 3 (C_1, C_2, C_3)

| Item No. | Sets of Item Parameters | | | Estimates of the Mixed Rasch Model | | | | | |
|----------|-------------------------|------|-----|------------------------------------|-------|-----------|-------|-----------|-------|
| | 1 | 2 | 3 | Dataset 1 | | Dataset 2 | | Dataset 3 | |
| | | | | C_1 | C_1 | C_2 | C_1 | C_2 | C_3 |
| 1 | -2.7 | 2.7 | .5 | -2.7 | -2.7 | 2.6 | -2.8 | 2.8 | .5 |
| 2 | -2.1 | 2.1 | -.5 | -2.1 | -2.2 | 2.1 | -2.2 | 2.2 | -.5 |
| 3 | -1.5 | 1.5 | .5 | -1.6 | -1.5 | 1.5 | -1.7 | 1.3 | .6 |
| 4 | -.9 | .9 | -.5 | -.9 | -.9 | .9 | -.7 | 1.0 | -.4 |
| 5 | -.3 | .3 | .5 | -.3 | -.3 | .3 | -.3 | .2 | .6 |
| 6 | .3 | -.3 | -.5 | .3 | .3 | .3 | .3 | -.6 | -.4 |
| 7 | .9 | -.9 | .5 | 1.0 | .8 | -1.0 | 1.0 | -.9 | .5 |
| 8 | 1.5 | -1.5 | -.5 | 1.5 | 1.5 | -1.4 | 1.6 | -1.5 | -.4 |
| 9 | 2.1 | -2.1 | .5 | 2.1 | 2.2 | -2.0 | 2.2 | -2.2 | .5 |
| 10 | 2.7 | -2.7 | -.5 | 2.7 | 2.6 | -2.8 | 2.7 | -2.6 | -.5 |

Dataset 1

The least difficult set of data for single datasets was Dataset 1, in which the same item parameters held for all examinees. Results for Dataset 1 in Table 3 show estimates that are nearly identical to the first set of item parameters. Table 3 does not provide information about the two-, three-, and four-class solution of this homogeneous dataset. However, when such a model with two or more classes was applied to this dataset, latent classes were constructed according to the random fluctuations in the data. This meant that the latent classes were substantial (i.e., $\pi_1 = .24$, $\pi_2 = .27$, and $\pi_3 = .41$ in the three-class solution), but the parameters had nearly the same (correct) order in all classes, except for some switches among adjacent item difficulties. In some cases, single parameter values became extremely high in order to fit perfectly an outlying pattern frequency.

These attempts at maximizing the likelihood of the data by assuming additional classes can be easily identified as an unsuitable fitting of random variation. As Table 4 shows, the log-likelihood increased much more slowly than the number of parameters, so that the one-class solution had the lowest AIC index in all cases.

Dataset 2

The second artificial example consisted of two subpopulations that were equal in size, but whose item difficulties were in reverse order (see Table 3). This caused the global item scores and, consequently, the item parameters of the one-class solution to fluctuate only randomly (i.e., all item scores were between 879 and 922, and all item parameters, including one latent class, were between $-.07$ and $.05$). Those unfamiliar with LC models may be surprised that the algorithm succeeded in “unmixing” these data and reproducing the correct distributions (see Table 3). Moreover, the solution was found after only 50 iterations, starting from random initial values and reaching the accuracy criterion $.0001$ for the log-likelihood. The estimates for the class sizes were $\pi_1 = .453$ and $\pi_2 = .457$,

Table 4
AIC Goodness-of-Fit Statistic, Log Likelihood (L), and Number of Parameters (m) for the Three Simulated Datasets

| No. of Classes | Dataset 1 | | | Dataset 2 | | | Dataset 3 | | |
|----------------|-----------|-----|-------|-----------|-----|-------|-----------|-----|-------|
| | L | m | AIC | L | m | AIC | L | m | AIC |
| 1 | -8338 | 19 | 16714 | -11080 | 19 | 22198 | -11565 | 19 | 23168 |
| 2 | -8325 | 39 | 16728 | -9407 | 39 | 18892 | -10972 | 39 | 22022 |
| 3 | -8312 | 59 | 16742 | -9397 | 59 | 18912 | -10830 | 59 | 21778 |
| 4 | -8300 | 79 | 16758 | -9377 | 79 | 18912 | -10820 | 79 | 21798 |

indicating that both classes had equal size; 9% of the individuals could not be assigned because they solved all items or none.

When three or four classes were assumed, the effects observed in the first data example recurred. Again, the AIC index clearly identified the correct number of latent classes (see Table 4).

Dataset 3

In this dataset, the correct number of latent classes was three, but there was one subpopulation with mean size $\pi_3 = .33$, which had practically no variation of item difficulties (the third set of item parameters in Table 3), and there was a constant level of ability; these examinees may be considered to be a "guessing" group. The remaining two classes had the same item profiles as in Dataset 2, but one of these groups was twice as large as the other. When the two-class solution was computed, the guessing group was mixed with the "profiled" groups where the "disturbance" was stronger for class 1, which was originally the smaller one of these two groups (see Table 5).

Table 5
Item Parameter Estimates of the Two-Class Solution for the Third Set of Simulated Data

| Item No. | Class 1 $\pi_1 = .48$ | Class 2 $\pi_1 = .46$ |
|----------|--------------------------|--------------------------|
| 1 | 1.2 | -2.5 |
| 2 | .3 | -1.9 |
| 3 | .9 | -1.2 |
| 4 | .1 | -.7 |
| 5 | .4 | 0.0 |
| 6 | -.4 | .2 |
| 7 | 0.0 | 1.0 |
| 8 | -.7 | 1.2 |
| 9 | -.3 | 1.9 |
| 10 | -1.2 | 2.0 |

The correct three-class solution was found without any problems (see Table 3), and the estimated class sizes were close to the correct parameter values: $\pi_1 = .39$, $\pi_2 = .20$, and $\pi_3 = .35$. From the AIC values (Table 4) it is clear that the two classes were not sufficient, and four classes were not necessary to explain the data.

Discussion

The EM algorithm for the proposed mixed Rasch model proved to be efficient in detecting ex-

aminee heterogeneity in the data, and in determining the item profiles, the latent score distributions, and the sizes of the latent classes. Thus one possible field of application is to test the fit of the Rasch model for a given set of data. The sample of individuals is usually split by means of manifest criteria (e.g., score, gender, or age) in order to check the invariance of item parameter estimates across subgroups of individuals, which is a central property of the Rasch model. The “power” of these model tests, however, depends on finding the appropriate splitting criterion, and that is not always the test score or any known criterion. The mixed Rasch model makes it possible to find a partition of the population for which the item parameters differ most, and between-group heterogeneity is maximized.

Testing the fit of the Rasch model, however, is only a side issue of the mixed Rasch model. The primary diagnostic potential of this model lies in its property to account for *qualitative* differences among examinees, and its simultaneous ability to *quantify* their abilities with respect to the same tasks. This might be a new idea for statistical models in psychological testing, where multidimensionality is often seen as the only way out of the inaccuracies of unidimensional models; yet it is an obvious idea for the psychological practitioner who knows that relevant individual differences are not only differences in *how well* somebody can do something, but also in *how* he/she does these things.

Mixture distribution models are a promising way of taking qualitative individual differences into account without losing the strong but necessary assumptions of the basic models—those models that hold for the unmixed data (i.e., the Rasch model in the present case). The Rasch model calls for this extension because its theoretical strength is better used for identifying groups of examinees who are really scalable, than for refuting the unidimensionality assumption for the entire population, and then moving on to a weaker model. Future applications of the model will show whether this promise is warranted.

Some theoretical elaborations would also be desirable. One of these extensions is the generalization of the model for polychotomous item responses (Rost, in press). This would make the model applicable to attitude questionnaires or other kinds of items with a graded response format, as well as for achievement items with partial credit scoring. Research (Rost, 1988b) has shown that the larger amount of information provided by polychotomous item responses can be well employed to find proper solutions in latent class analysis.

A further extension would be to introduce a distributional assumption of abilities within the latent classes. This is not consistent with the spirit of the “distribution-free” Rasch model, but it saves many parameters (nearly half), it may accelerate the algorithm, and it may also “smooth” the likelihood function.

References

- Andersen, E. B. (1973) A goodness of fit test for the Rasch model. *Psychometrika*, 38, 123–140.
- Bozdogan, H. (1987). Model selection and Aikake's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345–370.
- Clogg, C. C. (1988). Latent class models of measuring. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 173–206). New York: Plenum.
- Clogg, C. C., Lindsay, B., & Grego, J. (1989). *A simple latent class model for item analysis*. Unpublished manuscript, Pennsylvania State University, University Park.
- Croon, M. (1989). *Latent class analysis with ordered latent classes*. Tilburg: Social Science Department, Tilburg University (submitted for publication, 1989).
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1–22.
- Formann, A. K. (1985). Constrained latent class models: Theory and applications. *British Journal of Mathematical and Statistical Psychology*, 38, 87–111.
- Formann, A. K. (1989a). Constrained latent class models: Some further applications. *British Journal of Mathematical and Statistical Psychology*, 42, 37–54.

- Formann, A. K., (1989b). *Latent class models with order restrictions*. Research Bulletin 29. Wien [Vienna]: Institute of Psychology, University of Vienna.
- Gustafsson, J. E. (1980). A solution of the conditional estimation problem for long tests in the Rasch model for dichotomous items. *Educational and Psychological Measurement*, 40, 377-385.
- Haberman, S. (1979). *Analysis of qualitative data, Vol. 2: New developments*. New York: Academic Press.
- Heinen, T., Hagenars, J., & Croon, M. (1989). *Latent trait models in LCA perspective*. Tilburg: Department of Social Sciences, Tilburg University (submitted for publication, 1989).
- Kelderman, H., & Macready, G. B. (in press). The use of loglinear models for assessing item bias across manifest and latent examinee groups. *Journal of Educational Measurement*.
- Langeheine, R., & Rost, J. (1988). *Latent trait and latent class models*. New York: Plenum.
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Mislevy, R. J., & Verhelst, N. (in press). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*.
- Rost, J. (1985). A latent class model for rating data. *Psychometrika*, 50, 37-49.
- Rost, J. (1988a). Test theory with qualitative and quantitative latent variables. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 147-171). New York: Plenum.
- Rost, J. (1988b). Rating scale analysis with latent class models. *Psychometrika*, 53, 327-348.
- Rost, J. (in press). A logistic mixture distribution model for polychotomous item responses. *British Journal of Mathematical and Statistical Psychology*.
- Rost, J., Häußler, P., & Hoffmann, L. (1989). Long-term effects of physics education in the Federal Republic of Germany. *International Journal of Science Education*, 11, 213-226.
- Wright, D. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago IL: MESA Press.

Author's Address

Send requests for reprints or further information to
Jürgen Rost, IPN, Olshausenstraße 62, D-2300 Kiel
1, West Germany.