The Rasch Poisson Counts Model for Incomplete Data: An Application of the EM Algorithm

Margo G. H. Jansen
University of Groningen

Rasch's Poisson counts model is a latent trait model for the situation in which K tests are administered to N examinees and the test score is a count [e.g., the repeated occurrence of some event, such as the number of items completed or the number of items answered (in)correctly]. The Rasch Poisson counts model assumes that the test scores are Poisson distributed random variables. In the approach presented here, the Poisson parameter is assumed to be a product of a fixed test difficulty and a gamma-distributed random examinee latent trait parameter. From these assumptions, marginal maximum likelihood estimators can be derived for the test difficulties and the parameters of the prior gamma distribution. For the examinee parameters, there are a number of options. The model can be applied in a situation in which observations result from an incomplete design. When examinees are assigned to different subsets of tests using background information, this information must be taken into account when using marginal maximum likelihood estimation. If the focus is on test calibration and there is no interest in the characteristics of the latent traits in relation to the background information, conditional maximum likelihood methods may be preferred because they are easier to implement and are justified for incomplete data for test parameter estimation. *Index terms: EM algorithm, incomplete designs, latent trait models, marginal maximum likelihood estimation, Rasch Poisson counts model.*

Most of the currently used latent trait models are designed for a testing situation in which examinees are presented with multiple-choice, dichotomously scored items. A particularly well-known model for this kind of test data is the one-parameter logistic model [1PLM; also known as the Rasch model (Birnbaum, 1968; Spray, 1990)]. In item response theory models, the probability of an examinee answering an item correctly is a function of an examinee latent trait parameter and one or more item parameters (e.g., the difficulty parameter in the 1PLM or the difficulty and discrimination parameters in the two-parameter logistic model). The Rasch Poisson counts model (RPCM) is a unidimensional latent trait model for tests rather than items. The RPCM can be used in testing situations in which tests consist of multiple attempts of a single item within a given time limit, as is often the case in psychomotor skills assessment (Spray, 1990), and the test score is the number of successful attempts. Another situation in which the RPCM might be appropriate is as a limiting case for a binomial trials model when the test score is the number of incorrect responses out of n items with low error probabilities (Jansen & Van Duijn, 1992; Lord & Novick, 1968; Rasch, 1960). If the number of items is large there is no need for the items to have the same error probability within examinees. Meredith (1971) discussed the Poisson model in relation to classical test theory, in particular true-score theory. A number of applications to language and arithmetic tests are described in Rasch's book, and another application to a time-limit test was given by Meredith.

The situation considered here is the case in which the same test is given repeatedly, under different conditions, or alternatively, several test versions are administered to the same sample of examinees. It is useful to think of the observed test scores y_{jk} as organized in an $N \times K$ matrix, with N rows for examinees ($j = 1, ..., N$) and K columns for tests or occasions ($k = 1, ..., K$). Let y_{j+} and y_{+k}, the row and column sums,

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 19, No. 3, September 1995, pp. 291–302
© Copyright 1995 Applied Psychological Measurement Inc.
0146-6216/95/030291-12$1.85

denote the examinee and test scores. The model assumes that the test scores y_{jk} correspond to independent random variables distributed as Poisson variates with means μ_{jk}, which are assumed to be products of two parameters—the latent trait of the examinee and the difficulty of the test. In Rasch's original formulation, both sets of parameters were considered fixed; thus, standard methods for analyzing two-way contingency tables without interaction could be used. Jansen & Van Duijn (1992) expanded Rasch's model by assuming a latent distribution with fixed but unknown parameters for the trait parameters. Jansen and Van Duijn demonstrated how marginal maximum likelihood (MML) methods could be used for estimating the parameters of the trait distribution and the tests. The idea of a random examinee parameter or "true score," once one of the cornerstones of classical test theory, also can be found in Meredith (1971). Originally, the model was formulated for complete data and, implicitly, tests of equal length. These restrictions lead to a very attractive model but are impractical in view of empirical applications in which there is often missing data or systematically incomplete data. The latter is a data collection design in which not all tests (or items) are administered to all examinees. Typical examples are test equating studies and studies in which a targeted testing procedure is used (discussed below).

Rasch's Poisson Counts Model

The probability of observing a response y_{jk} of examinee j on test k is given by:

$$P\left(y_{jk} = y_{jk}\right) = \frac{\exp\left(-n_k\mu_{jk}\right)\left(n_k\mu_{jk}\right)^{y_{jk}}}{y_{jk}!}, \tag{1}$$

where n_k is the length of the test, which may be the number of items or the duration of the observation period. The error rate μ_{jk} is a product of the test difficulty parameter (β) and the examinee trait level parameter (θ):

$$\mu_{jk} = \theta_j\beta_k. \tag{2}$$

The βs are considered fixed. A latent distribution with fixed but unknown parameters is assumed for the θs. A matrix $\mathbf{V} = \{v_{jk}\}$ can be defined with elements $v_{jk} = 1$ if test k is administered to examinee j and $v_{jk} = 0$ otherwise. Examinees for which $v_{jk} = 0$ on all K tests are excluded.

From Equations 1 and 2, the likelihood of $\mathbf{Y}$ (the matrix of responses of N examinees on K tests considered as a function of θ and β) can be written as

$$L\left(\mathbf{Y}|\theta,\beta\right) = \frac{\exp\left(-\sum_{jk} v_{jk}\theta_j n_k\beta_k\right)\left\{\prod_j \theta_j^{\Sigma_k v_{jk}y_{jk}}\right\}\left\{\prod_k n_k\beta_k^{\Sigma_j v_{jk}y_{jk}}\right\}}{\prod_{jk} y_{jk}!^{v_{jk}}}. \tag{3}$$

Let $p(\theta|\Phi)$ denote the multivariate prior density. Then the posterior density function is

$$p(\theta|\beta,\Phi,\mathbf{Y}) = \mathrm{H}^{-1}L(\mathbf{Y}|\theta,\beta)\,p(\theta|\Phi), \tag{4}$$

where the constant of proportionality H is the marginal likelihood

$$L_n(\mathbf{Y}|\beta,\Phi) = \int L(\mathbf{Y}|\theta,\beta)\,p(\theta|\Phi)\,d(\theta). \tag{5}$$

Because the θs are nonnegative quantities, a two-parameter gamma distribution can be assumed. The gamma distribution, which is conjugate to the Poisson distribution, is also very convenient. However, although the gamma distribution is fairly flexible, it is not applicable to all the empirical data for which the RPCM could be used. Although the RPCM generalizes to other mixing distributions, the results will in general be considerably less tractable (Anscombe, 1950; Jansen, 1994; Johnson & Kotz, 1969). The discussion here is restricted to the gamma distribution.

Letting the θs be sampled from a gamma distribution with shape parameter c and scale parameter $1/s$, it can be shown that the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{\beta}, c, s, \mathbf{Y})$ of $\boldsymbol{\theta} = (\theta_1, ..., \theta_N)$ is the product of independent gamma distributions, each with parameters $(s + \Sigma v_{jk} m_k \beta_k)$ and $(c + \Sigma v_{jk} y_{jk})$. For identifiability, the constraint $\Sigma\beta_k = 1$ is introduced. The joint posterior density of $(\mathbf{Y}, \boldsymbol{\theta})$ for a given (β, c, s) can be written as

$$\prod_j \frac{\exp\left[-\left(\sum_k v_{jk} n_k \beta_k + s\right)\theta_j\right] \theta_j^{\Sigma_k v_{jk} y_{jk} + c - 1} \prod_k \left(n_k \beta_k\right)^{v_{jk} y_{jk}}}{\Gamma(c) \prod_k \left(y_{jk}!\right)^{v_{jk}}}. \tag{6}$$

The marginal likelihood can be derived from Equation 6 by integrating with respect to θ:

$$L_m(\mathbf{Y}|\boldsymbol{\beta}, c, s) = \frac{s^{Nc}\left[\prod_j \Gamma\left(\sum_k v_{jk} y_{jk} + c\right)\right] \prod_k \left(n_k \beta_k\right)^{\Sigma_j v_{jk} y_{jk}}}{\Gamma(c)^N \prod_j \left(\sum_k v_{jk} n_k \beta_k + s\right)^{\Sigma_k v_{jk} y_{jk} + c} \prod_k \left(y_{jk}!\right)^{v_{jk}}}. \tag{7}$$

Estimators for the model parameters are derived by taking the log of Equation 7, obtaining the partial derivatives with respect to the parameters, and setting the resulting expressions equal to 0.

$$\psi(c) - \log(s) = \frac{1}{N} \sum_j \psi\left(\sum_k v_{jk} y_{jk} + c\right) - \log\left(\sum_k v_{jk} n_k \beta_k + s\right), \tag{8}$$

$$\frac{c}{s} = \frac{1}{N} \sum_j \frac{\left(\sum_k v_{jk} y_{jk} + c\right)}{\left(\sum_k v_{jk} n_k \beta_k + s\right)}, \tag{9}$$

and

$$\frac{\left(\sum_j v_{jk} y_{jk} / N_k\right)}{n_k \beta_k} = \frac{1}{N_k} \sum_j \frac{\left(\sum_k v_{jk} y_{jk} + c\right)}{\left(\sum_k v_{jk} n_k \beta_k + s\right)} v_{jk}, \tag{10}$$

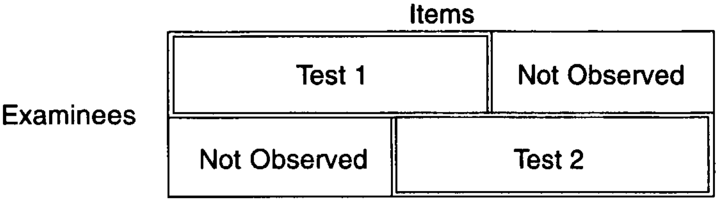
where $N_k = \Sigma_j v_{jk}$ is the number of examinees for which there are observed scores on test k. Note that by allowing for missing data, it is no longer possible to estimate the hyperparameters c and s independently of the βs and vice-versa (Jansen & Van Duijn, 1992).

Marginal Maximum Likelihood Estimation in Rasch's Model for Incomplete Designs

Figure 1 is an example of an incomplete data collection design. Two tests with a common set of items are administered to two examinee groups, but each group takes only one of the two tests.

Figure 2 shows a related procedure known as *targeted testing*. If the interest is focused on item or test calibration, targeted examinee sampling might be a better term. In this design, there may be a number of tests designed to measure the same ability, and the tests are ordered in ascending difficulty. The tests are grouped into subsets so that two adjacent subsets have tests in common. The examinees are assigned to the

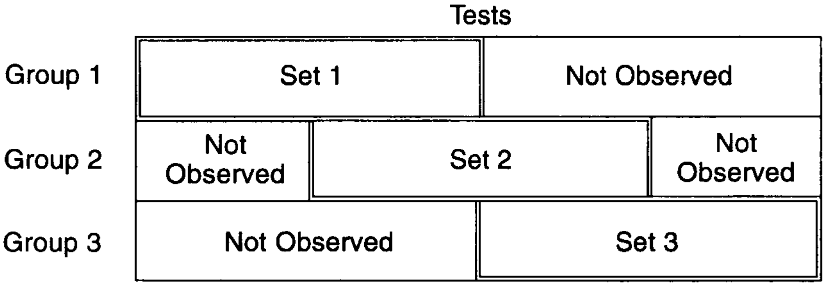
Figure 1
An Incomplete Data Design Linked by Common Items



subsets of tests that are assumed to match their θ level. Examinees, therefore, are not assigned to subgroups at random but according to background information, which is presumed to be (positively) related to their **θ**.

Mislevy & Sheehan (1989) demonstrated, in an item response theory context, that collateral information can be ignored if this information is not used in item sampling and/or in examinee sampling. However, ignoring background information in the analysis that has been used to assign examinees to test items will lead to biased estimates of the item parameters (Eggen, 1993; Mislevy & Sheehan, 1989). In this case, the groups of examinees taking different subsets of tests will have different θ levels; thus, the mechanism that caused the missing data will have to be taken into account in the analysis.

Figure 2
A Design With Nonequivalent Groups Linked by Common Tests



The RPCM can be extended easily to the situation in which collateral information is used to assign examinees to specific sets of tests. To accomplish this, additional notation is needed. Assume that the total sample has been subdivided into G subsamples depending on their scores on a background variable X, and let y_{gjk} denote the score of examinee j ($j = 1, ..., N_g$) in group g ($g = 1, ..., G$) on test k. If MML estimation is implemented using the background variable X, the MML function conditional on X is a product over subpopulations.

The marginal likelihood is derived by integrating the joint posterior density of $(\mathbf{Y},\boldsymbol{\theta})$ given (β, c, s, X) with respect to $\boldsymbol{\theta}$ ($\theta = \{\theta_{gj}\}$, $g = 1, ..., G; j = 1, ..., N_g$),

$$L_n(\mathbf{Y}|\beta,c,s,X) = \prod_g \frac{s_g^{N_g c_g}\left[\prod_j \Gamma\left(\sum_k v_{gk} y_{gjk} + c_g\right)\right]\prod_k (n_k\beta_k)^{\sum_j v_{gk} y_{gjk}}}{\Gamma(c_g)^{N_g} \prod_j \left(\sum_k v_{gk} n_k \beta_k + s_g\right)^{\sum_k v_{gk} y_{gjk} + c_g} \prod_k y_{gjk}!^{v_{gk}}} . \tag{11}$$

Estimators for β and (c, s) are obtained by setting the derivatives of log L_n, with respect to the parameters, equal to 0. This results in the following equations, which have to be solved simultaneously:

$$\psi(c_g) - \log(s_g) = \frac{1}{N_g}\sum_j \psi\left(\sum_k v_{gk} y_{gjk} + c_g\right) - \log\left(\sum_k v_{gk} n_k \beta_k + s_g\right), \tag{12}$$

$$\frac{c_g}{s_g} = \frac{1}{N_g}\sum_j \frac{\left(\sum_k v_{gk} y_{gjk} + c_g\right)}{\left(\sum_k v_{gk} n_k \beta_k + s_g\right)}, \tag{13}$$

and

$$\sum_{gj} v_{gk} y_{gjk} / n_k \beta_k = \sum_{gj} \frac{\left(\sum_k v_{gk} y_{gjk} + c_g\right)}{\left(\sum_k v_{gk} n_k \beta_k + s_g\right)} v_{gk}\,, \tag{14}$$

for $g = 1, \ldots, G$; $j = 1, \ldots, N_g$; and $k = 1, \ldots, K - 1$.

Computational Procedures

Equations 12–14 can be solved using the Newton-Raphson algorithm. The asymptotic covariance matrix can be obtained by inverting the observed information matrix. An alternative is to apply the EM algorithm, which in this case is easy to implement. The expectation step consists of the computation of the sufficient statistics for the parameters of the prior distribution c_g, s_g, given provisional values of c_g, s_g, and β:

$$\sum_j \theta_{gj}^{(p+1)} | \beta^{(p)},\ c_g^{(p)}, s_g^{(p)} \tag{15}$$

and

$$\sum_j (\log \theta_j)^{(p+1)} | \beta^{(p)},\ c_g^{(p)}, s_g^{(p)}\,. \tag{16}$$

From this, the estimates of c_g and s_g are updated in the M (maximization) step using Equations 12 and 13, and β is re-estimated from Equation 14 until a specified criterion of convergence is met.

Conditional Maximum Likelihood Estimation in Targeted Testing

In targeted testing, the value on the background variable X, which can take G distinct values, determines which subset of tests is administered to an examinee. If the only objective is test calibration and there is no further interest in the θ distribution in the subpopulations from which the examinees were sampled, conditional maximum likelihood (CML) estimation methods might be preferred.

Using the same notation as before, the joint probability of selecting examinee j having observed score x_g on background variable X and observing the response vector $\mathbf{y}_j$ can be factored as

$$P(\mathbf{y}_j, x_g) = P(\mathbf{y}_j | y_{j+}, x_g) \cdot P(y_{j+} | x_g) \cdot P(x_g), \tag{17}$$

where $\mathbf{y}_j$ is the vector of scores y_{jk}, and y_{j+} is the sum of these scores ($y_{j+} = \Sigma_k v_{gk} y_{gjk}$). All the information relevant with respect to estimating the βs resides in the first term of the right-hand side of Equation 17 (it can be shown that the second part can be neglected). Furthermore, $P(X_j = x_g)$ does not depend on β. Thus, CML estimation of the βs by maximizing the conditional likelihood is justified. For $g = 1, \ldots, G$ and $j = 1, \ldots, N_g$ the conditional likelihood can be written as

$$L_c = \prod_g \prod_j \left(\sum_k v_{gk} y_{gjk} \right)! \prod_k \frac{\left(n_k \beta_k \Big/ \sum_k v_{gk} n_k \beta_k \right)^{v_{gk} y_{gjk}}}{\left(y_{gjk}! \right)^{v_{gk}}} , \qquad (18)$$

and the likelihood equations are given by

$$\sum_g \sum_j v_{gk} y_{gjk} / n_k \beta_k - \sum_g \sum_j \frac{\left(\sum_k v_{gk} y_{gjk} \right)}{\left(\sum_k v_{gk} n_k \beta_k \right)} v_{gk} = 0 . \qquad (19)$$

It is expected that, in general, the CML test parameter estimates ($\hat{\beta}$s) will be virtually identical to the MML estimates.

Both MML estimation and CML estimation, in the context of incomplete data collection designs, have been applied to parameter estimation in other latent trait models, such as the 1PLM (Glas, 1989), with basically comparable results. A more comprehensive discussion of the application of Rubin's ignorability principle (Little & Rubin, 1987) to estimation problems in the context of item response models and structurally incomplete data can be found in Eggen (1993).

Estimating the Examinee Parameters

To estimate an examinee's θ parameter from the scores on the tests that he or she has taken, the same procedure can be followed (i.e., calculating the posterior means). The expected a posteriori (EAP) estimator is given by

$$\hat{\theta}_{gj} = \frac{\sum_k v_{gk} y_{gjk} + c_g}{\sum_k v_{gk} n_k \beta_k + s_g} . \qquad (20)$$

Note that the examinee's θ estimate ($\hat{\theta}$) is now a function not only of the individual observed score but also of the parameters of the population to which the examinee belongs. This implies that if two examinees with the same observed scores (on the same tests) belong to different populations their $\hat{\theta}$s will be different, unless the population parameters are identical. This presents a problem if θs are being estimated in order to equate test scores using the θ scale (Glas, 1989). Another possibility, which avoids this complication, is to estimate the θs using maximum likelihood (ML) and treating the $\hat{\beta}$s as known constants. The ML estimator of θ of examinee j in group g is

$$\hat{\theta}_{gj} = \frac{\sum_k v_{gk} y_{gjk}}{\sum_k v_{gk} n_k \beta_k} , \qquad (21)$$

and the expected score on test q for an examinee with $\hat{\theta}_{gj}$, for all g and j, is given by

$$\mathrm{E}\left(y_{gjq} \middle| \hat{\theta}_{gj}, \beta_q \right) = n_q \theta_{gj} \beta_q . \qquad (22)$$

In the context of CML estimation, the same expression as in Equation 21 for the ML estimator of θ will be derived.

Modeling the Test Parameters

It may be desirable to model the βs in terms of explanatory variables. For example, the following loglinear model could be modeled:

$$\log\beta_k = \eta_0 + \mathbf{x}_k'\,\boldsymbol{\eta}, \tag{23}$$

where $\mathbf{x}_k$ is the *k*th column of a design matrix, and $\boldsymbol{\eta}$ is a parameter vector of length *r* (McCullagh & Nelder, 1989; Van Duijn, 1993). The parameter η_0 ensures that the $\boldsymbol{\beta}_k$s will sum to 1. η_0 can be expressed in terms of $\mathbf{x}_k'\,\boldsymbol{\eta}$:

$$\eta_0 = -\ln\left[\sum_k \exp\left(\mathbf{x}_k'\,\boldsymbol{\eta}\right)\right]. \tag{24}$$

The RPCM also can be applied to the analysis of longitudinal count data obtained by administering the same test at *K* occasions (time points). A suitable structure that models changes over time then can be imposed on the $\boldsymbol{\beta}_k$s. Incorporating missing data and unequal test lengths yields slightly more complicated, but basically similar, likelihood equations (Van Duijn, 1993). Fischer (1991) described a similar model for counted data. An important difference is that Fischer used CML estimation methods.

Examples

Example 1: Longitudinal Count Data With Randomly Missing Values

The methods presented here were applied to the analysis of an empirical dataset derived from a larger study on the development of arithmetic and language skills of Dutch primary school students (van der Velde, 1982). The data were the scores on a reading test [the Brustest (Brus & Voeten, 1987)] that was administered about every three months during the school year when the students were in the second, third, fourth, and fifth grade. Only the first six scores (from the first two years) were used from these 12 administrations. The Brustest is a speed test; the score is the number of words that an examinee can read in one minute. The sample consisted of 154 examinees. There was only a small percentage of missing data; thus, it was assumed that the missing data were missing at random.

The means, standard deviations (SDs), and *N* of the reading scores for each occasion are given in Table 1. Tables 2 and 3 show the test parameter estimates and the parameter estimates of the θ distribution from the RPCM. The results in Table 3 indicate that an "average" examinee with a $\theta \approx 4.6$ will have an expected reading score of approximately .5 words per second on the first and 1.0 on the last occasion.

Example 2: The Effect of Ignoring *X* in a Simulated Targeted Design

When using a targeted design, the background variables must be taken into account in the MML procedure by simultaneously estimating the βs and the θ distribution parameters of all groups involved. In the

Table 1
Descriptive Statistics (Total *N* = 154)

Occasion	Mean	SD	*N*
1	27.1	14.52	147
2	35.7	15.82	148
3	45.6	16.07	148
4	49.4	16.79	153
5	55.3	17.01	153
6	61.1	16.46	153

Table 2
Test Parameter Estimates

Occasion	β	SE(β)
1	.099	.0014
2	.130	.0017
3	.165	.0018
4	.181	.0019
5	.202	.0020
6	.223	.0020

following example, "error" scores were generated for two groups of 500 examinees [data were generated so that higher values of θ corresponded to lower ability and were associated with higher (error) scores]. Examinees in the first group had values of x_1 on X and were randomly drawn from a gamma distribution with parameters ($c_1 = 2.5$, $s_1 = 10$). Examinees in the second group had values of x_2 on X and were randomly drawn from a gamma distribution with parameters ($c_2 = 2.5$, $s_2 = 15$).

The group with the highest θ levels (Group 1) had higher expected scores than Group 2 (and the highest expected variance). Because the scores were error scores, Group 1 was less "able" than Group 2. The test with the highest β was the most difficult test. The first two (and least difficult) tests were given only to examinees in Group 1, and the last two and most difficult tests were given to Group 2 only. The middle tests were given to both groups. All six tests had a length of 100 items.

Table 3
Examinee θ Distribution

Parameter	Estimate	SE
c	7.47	.867
s	1.64	.197

The following proportional test difficulties were used: $\beta_k = 1.25\beta_{k-1}(k = 2, ..., K)$. These data were analyzed twice using MML. In the first analysis, information on the background variable X was ignored by treating all examinees as if they were sampled from the same distribution. In the second analysis, the information on background variable X was included. In a third analysis, CML estimates were obtained. The results of these analyses are presented in Table 4. The estimates of the parameters of the θ distribution ignoring X were $c = 2.33$ [standard error (SE) = .122] and $s = 11.74$ (SE = .669). Estimates of the parameters from the analysis including X were $c_1 = 2.70$ (SE = .164), $s_1 = 9.16$ (SE = .755), $c_2 = 2.84$ (SE = .221), and $s_2 = 17.26$ (SE = 1.439).

Table 4 shows that ignoring the background variable X resulted in biased MML $\hat{\beta}$s. The $\hat{\beta}$s of the easier tests were overestimated and the $\hat{\beta}$s of the more difficult tests were underestimated. The CML estimates proved to be identical to the MML estimates that were obtained when X was included in the analysis. Compare, for example, the $\hat{\beta}$ of the first and easiest test, which had true $\beta = .089$ and MML estimates of .095 (ignoring X) and .090 (using X). The CML estimate was also .090. For the most difficult test (Test 6), $\beta = .271$, and the MML estimates were .257 ignoring X (which was too low) and .268 using X. Again, the CML estimate was identical to the MML estimate using X.

Simulation Studies

Recovery of Parameters With Incomplete Data

Method. A simulation study was conducted to examine how well βs and θs are recovered when there is incomplete data because of targeted testing. The sample sizes and number of tests were varied. Error

Table 4
True β and MML Estimated β ($\hat{\beta}$) and the Standard Errors (SEs) of the Estimates Ignoring and Including X, and CML Estimates and Their SEs

		Ignoring X		Including X		CML	
Test	β	$\hat{\beta}$	SE($\hat{\beta}$)	$\hat{\beta}$	SE($\hat{\beta}$)	$\hat{\beta}$	SE($\hat{\beta}$)
1	.089	.095	.0030	.090	.0030	.090	.0030
2	.111	.121	.0035	.114	.0034	.114	.0034
3	.139	.141	.0027	.138	.0027	.138	.0027
4	.173	.177	.0030	.174	.0030	.174	.0030
5	.217	.208	.0046	.216	.0049	.216	.0049
6	.271	.257	.0050	.268	.0054	.268	.0054

scores were simulated so that the more difficult tests had higher scores. Data were simulated for two examinee groups (Group 1 and Group 2) of equal size on two sets of tests. The first set consisted of five tests and the second set consisted of seven tests. Each test contained 25 items (n = 25). Two different sample sizes were used for each group: $N = 50$ and $N = 200$. The following proportional βs were used: $\beta_k = 1.25\beta_{k-1} (k = 1, ..., K)$. For c and s, values compatible with $E(y_{j+}) = 10$ for Group 1 and $E(y_{j+}) = 5$ for Group 2 ($c_1 = c_2 = 2$, $s_1 = 5$, and $s_2 = 10$) were selected. The observed scores were generated as follows:

Step 1: For each group, θ_j was drawn from a gamma distribution with mean c/s and variance c/s^2.

Step 2: For each examinee and test, $\beta\theta$ was computed.

Step 3: Independent realizations of a Poisson distributed random variate were generated using $\beta\theta$ (Step 2) as the error rate.

Step 4: To obtain incomplete data, the scores on selected test-group combinations were coded as missing. The complete dataset for the set of five tests was made incomplete by coding the generated scores on the first test for Group 2 and on the last test for Group 1 as missing data; for the second set of seven tests, the first two for Group 2 and the last two tests for Group 1 were coded as missing for the incomplete case.

The complete and incomplete datasets were analyzed, and the resulting parameter estimates ($\hat{\theta}$ and $\hat{\beta}$) were compared. Results are based on 100 replications. Within each replication, the β_ks and θ_js were estimated. ML estimation was used for the θs. The $\hat{\beta}_k$ were averaged over replications. For the θs, the means and SDs of the $\hat{\theta}_j$s were calculated for each group as well as the correlations between θ and $\hat{\theta}$, which were also averaged over replications.

Results. The results presented in Table 5 suggest that the recovery of β_k was on average very good, both for the complete and incomplete datasets, even for $N = 50$. As expected, for the incomplete data the SDs of the $\hat{\beta}$ were generally somewhat larger; they were substantially larger for the first and last tests for the set of 5 tests and for the first two and last two tests for the set of 7 tests. For example, the SD of the $\hat{\beta}$s of the last test of the set of five tests was .016 for the complete and .035 for the incomplete case.

For the θs, the means and variances of the $\hat{\theta}_j$s were calculated for each group, as well as the correlations between θ and $\hat{\theta}$. The results in Table 6 show that there was fairly low correlation between $\hat{\theta}$ and θ. The highest correlations were found in the examinee groups with the larger θ variances (here also the high θ groups). The correlations averaged slightly lower for the incomplete than for the complete datasets. On average, the means of $\hat{\theta}$ and θ values were almost identical; however, the SDs of the $\hat{\theta}$s were systematically too large. Thus, the $\hat{\theta}$s were biased to some extent. The average means were almost equal to the expected means (.4 for Group 1 and .2 for Group 2). The average variance for $\hat{\theta}$ in Group 1 for five tests with $N = 50$ was .095 in the complete case; the expected variance was .08. This same trend also was observed for larger sample sizes and for seven tests.

Table 5
True β and Average (Ave) and SD of Estimated βs for
Complete and Incomplete Data for Five Tests and Seven Tests

		$N_1 = N_2 = 50$				$N_1 = N_2 = 200$			
		Complete		Incomplete		Complete		Incomplete	
Test	β	Ave	SD	Ave	SD	Ave	SD	Ave	SD
Five Tests									
1	.122	.119	.012	.121	.018	.123	.006	.123	.008
2	.152	.155	.014	.155	.014	.152	.006	.152	.006
3	.190	.189	.014	.190	.016	.189	.008	.190	.009
4	.238	.237	.016	.238	.018	.238	.008	.238	.008
5	.298	.299	.016	.295	.035	.298	.008	.297	.015
Seven Tests									
1	.066	.067	.007	.068	.012	.067	.004	.067	.006
2	.083	.084	.010	.085	.015	.083	.005	.084	.008
3	.104	.103	.011	.104	.013	.104	.006	.104	.006
4	.130	.129	.013	.131	.014	.129	.006	.129	.007
5	.162	.160	.011	.161	.013	.162	.007	.162	.007
6	.202	.202	.014	.199	.028	.203	.007	.202	.013
7	.253	.255	.016	.253	.029	.252	.008	.252	.013

Comparison of EAP and ML Estimates

Method. A second simulation study compared EAP and ML $\hat{\theta}$s. The same sample sizes ($N = 50$ and $N = 200$) and the same proportional βs ($\beta_k = 1.25\beta_{k-1}$) were used as in the first study with five tests. The same procedure was used as described above for obtaining complete and incomplete datasets. Three different specifications for c and s compatible with $E(y_{j+}) = 10$ for Group 1 and $E(y_{j+}) = 5$ for Group 2 were selected ($c_1 = 1$ and $s_1 = 2.5$, $c_2 = 1$ and $s_2 = 5$; $c_1 = 2$ and $s_1 = 5$, $c_2 = 2$ and $s_2 = 10$; $c_1 = 5$ and $s_1 = 12.5$, $c_2 = 5$ and $s_2 = 25$). All tests were 25 items long.

For both EAP and ML $\hat{\theta}$, the mean squared differences (MSDs) and the correlations between θ and $\hat{\theta}$ were calculated. These outcomes were also averaged over 100 replications.

Results. Table 7 shows that EAP $\hat{\theta}$s showed a consistent, although very modest, superiority over ML

Table 6
Average (Ave) and SD of the Means and Variances (Var) of the $\hat{\theta}$s and of the Correlations Between the $\hat{\theta}$s and θs (r) for Five Tests and Seven Tests (For Group 1 the Expectation of the Mean was $c/s = .4$ and the Expectation of the Variance was $c/s^2 = .08$; for Group 2 These Values Were .2 and .02, Respectively)

	Complete Data						Incomplete Data					
Tests, N,	Mean		Var		r		Mean		Var		r	
and Group	Ave	SD	Ave	SD	Ave	SD	Ave	SD	Ave	SD	Ave	SD
Five Tests, $N = 50$												
Group 1	.400	.043	.095	.029	.909	.032	.398	.049	.101	.032	.878	.042
Group 2	.200	.026	.029	.010	.847	.051	.201	.027	.030	.011	.832	.053
Five Tests, $N = 200$												
Group 1	.399	.021	.093	.013	.913	.014	.399	.024	.099	.015	.882	.019
Group 2	.199	.012	.028	.005	.840	.028	.199	.012	.029	.005	.823	.031
Seven Tests, $N = 50$												
Group 1	.405	.046	.096	.027	.911	.030	.401	.056	.109	.034	.855	.050
Group 2	.197	.025	.028	.009	.843	.049	.198	.026	.029	.009	.820	.053
Seven Tests, $N = 200$												
Group 1	.403	.022	.096	.015	.913	.014	.402	.025	.109	.018	.857	.022
Group 2	.200	.011	.028	.004	.842	.023	.201	.012	.030	.004	.822	.026

$\hat{\theta}$s. The average correlations were marginally larger for the EAP estimates, although the differences were generally confined to the third decimal place. The correlations between the $\hat{\theta}$s and θs were at best moderately high, depending on the variance of the true values. The smaller the variance, the lower the correlations. The highest average correlation was approximately .95 for the $c_1 = 1$, $s_1 = 2.5$ and $c_2 = 1$, $s_2 = 5$ conditions, and ranged from .81 to .87 (bottom rows) for the set of specifications associated with the smallest expected θ variances ($c_1 = 5, s_1 = 12.5$ and $c_2 = 5, s_2 = 25$). ML $\hat{\theta}$s overestimated the true θ SDs; EAP $\hat{\theta}$s underestimated the true θ SDs. The MSDs between $\hat{\theta}$ and θ values were generally smaller for EAP than for ML $\hat{\theta}$s. The smaller the θ variance was, the more pronounced this effect became. The largest differences were observed for the set of specifications associated with the smallest expected θ variances.

Table 7
Average Mean Squared Error [(MSE = $\Sigma_j(\theta_j - \hat{\theta}_j)^2/N$, Averaged Over Replications] and Average Correlations Between the $\hat{\theta}$s and θs (*r*) for Five Tests

	Complete Data				Incomplete Data			
N and	MSE		*r*		MSE		*r*	
Estimator	Ave	SD	Ave	SD	Ave	SD	Ave	SD
$N_1 = N_2 = 50$; $c_1 = 1$, $s_1 = 2.5$; $c_2 = 1$, $s_2 = 5$								
ML	1.20	.287	.946	.017	1.63	.434	.929	.029
EAP	1.07	.270	.946	.017	1.42	.367	.930	.028
$N_1 = N_2 = 200$; $c_1 = 1$, $s_1 = 2.5$; $c_2 = 1$, $s_2 = 5$								
ML	1.21	.120	.949	.009	1.62	.186	.933	.013
EAP	1.08	.112	.950	.009	1.41	.170	.935	.013
$N_1 = N_2 = 50$; $c_1 = 2$, $s_1 = 5$; $c_2 = 2$, $s_2 = 10$								
ML	1.14	.211	.916	.029	1.58	.363	.890	.029
EAP	.89	.171	.920	.022	1.19	.263	.895	.028
$N_1 = N_2 = 200$; $c_1 = 2$, $s_1 = 5$; $c_2 = 2$, $s_2 = 10$								
ML	1.20	.123	.914	.010	1.62	.169	.890	.013
EAP	.96	.100	.918	.009	1.22	.130	.894	.012
$N_1 = N_2 = 50$; $c_1 = 5$, $s_1 = 12.5$; $c_2 = 5$, $s_2 = 25$								
ML	1.23	.210	.846	.034	1.68	.305	.809	.041
EAP	.76	.147	.868	.030	.96	.201	.835	.038
$N_1 = N_2 = 200$; $c_1 = 5$, $s_1 = 12.5$; $c_2 = 5$, $s_2 = 25$								
ML	1.22	.100	.844	.020	1.61	.151	.808	.025
EAP	.75	.067	.867	.016	.90	.088	.840	.019

Discussion

The mixed model of the RPCM considered here, in which the examinee parameters θ_j were assumed random and MML estimation methods were used to estimate the parameters, can allow for missing data. When data are systematically incomplete (i.e., examinees are assigned to subsets of tests on the basis of information presumably related to their trait level), a numerical example using simulated data illustrated that in order to obtain correct MML estimates for the test parameters, the collateral information must be taken into account when performing the analysis. CML estimation methods also can be used to provide the desired results, independently of the distribution of the examinee parameters. CML estimation is computationally less cumbersome and has good asymptotic properties, but the extra information on the subpopulation distributions is not available.

A simulation study examined how well test and examinee parameters were recovered in relation to sample size, differences in the subpopulation distributions, and the number of tests, for systematically incomplete data. The results indicated that test parameters were fairly accurately estimated, even with a relatively small sample size ($N = 50$), in both the complete and incomplete data case. Some bias appeared in the ML estimates of the examinee parameters. The parameter estimates had a larger SD than the true values and the correlations between estimated and true θ values were only moderately high, as shown in Table 6.

A second simulation study compared the differences in precision between MML and ML θ estimates. The EAP estimates, which showed a smaller variance than the ML estimates, performed somewhat better, but the differences were not large. To more precisely establish the extent of these differences in relation to such variables as the variability in the test parameters and the number of tests, the degree of incompleteness and the characteristics of the examinee distributions, additional simulation studies based on empirically realistic specifications are needed. In simulation studies not presented here, test length was found to be an important factor—higher correlations were observed with tests of 50 or more items. At present, it is somewhat doubtful if in practical applications the possible gain in precision justifies the more cumbersome MML estimation procedure, especially if other distributions than the convenient gamma distribution have to be considered. Note, however, that in this study ML estimation for the individual examinee parameters did not perform very well either. Thus, it might also be useful to consider methods for correcting the bias in the ML estimator.

References

Anscombe, F. J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika, 37,* 358–382.

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. 397–479). Reading MA: Addison-Wesley.

Brus, B. Th., & Voeten, M. J. M. (1987). *Een-Minuut-Test, vorm A en B. Verantwoording en Handleiding* [One-Minute-Test. Reference manual]. Nijmegen: Berkhout Testmateriaal.

Eggen, T. J. H. M. (1993). Itemresponsetheorie en onvolledige gegevens (Item response theory and incomplete data). In T. J. H. M. Eggen & P. F. Sanders (Eds.), *Psychometrie in de Praktijk* [Psychometrics in practice] (pp. 239–284). Arnhem: CITO (National Institute for Educational Measurement).

Fischer, G. H. (1991). On power series models and the specifically objective assessment of change in event frequencies. In J. P. Doignon & J. C. Falmagne (Eds.), *Mathematical psychology: Current developments* (pp. 293–310). New York: Springer-Verlag.

Glas, C. A. W. (1989). *Estimating and testing Rasch models*. Doctoral dissertation, University of Twente, Enschede, The Netherlands.

Jansen, M. G. H. (1994). Parameters of the latent distribution in Rasch's Poisson counts model. In G. H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics and methodology* (pp. 319–326). New York: Springer-Verlag.

Jansen, M. G. H., & Van Duijn, M. A. J. (1992). Extensions of Rasch's multiplicative Poisson model. *Psychometrika, 57,* 405–414.

Johnson, N. L., & Kotz, S. (1969). *Distributions in statistics. Discrete distributions.* Boston: Houghton-Mifflin.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Wesley.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models.* London: Chapman & Hall.

Meredith, W. (1971). Poisson distribution of error in mental test theory. *British Journal of Mathematical and Statistical Psychology, 24,* 49–82.

Mislevy, R. J., & Sheehan, K. M. (1989). The role of collateral information about examinees in item parameter estimation. *Psychometrika, 54,* 661–680.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research. Expanded edition, University of Chicago Press, 1980.

Spray, J. A. (1990). One-parameter item response theory models for psychomotor tests involving repeated independent attempts. *Research Quarterly for Exercise and Sport, 61,* 162–168.

Van der Velde, A. (1982). *Lezen en spellen in klas twee tot zes* (Reading and spelling in grade two to six) (Research Report). Groningen: University of Groningen, The Netherlands.

Van Duijn, M. A. J. (1993). *Mixed models for repeated count data*. Doctoral dissertation, University of Groningen, Groningen, The Netherlands.

Author's Address

Send requests for reprints or further information to Margo G. H. Jansen, Department of Education, University of Groningen, Grote Rozenstraat 38, 9712 TJ Groningen, The Netherlands. Internet: g.g.h.jansen@ppsw.rug.nl.