Chapter 37

Applications of Rasch's Poisson Counts Model to Longitudinal Count Data

Margo G.H. Jansen

University of Groningen

1. The Rasch Poisson Counts Model

Most of the currently used latent trait models are designed for testing situations, where subjects are presented with multiple written, dichotomously scored (multiple choice) items. A particularly well known model for these kind of test data is the one-parameter logistic (1-PML) or Rasch model. In such a model, the chance of a subject answering an item correctly is a function of a subject ability parameter and one or more item parameters, a difficulty parameter in Rasch's model and a difficulty and a discrimination parameter in the two-parameter logistic model (2-PML). The Rasch Poisson Counts Model (RPCM) is an unidimensional latent trait model for tests where the scores can be regarded as realizations of a Poisson process. It can be used, for example, in testing situations where the same item or task is given repeatedly, and the test score is the number of (un)successful attempts. With only a single item there is just one (item) difficulty parameter, characterizing the whole test, and no need for an item discrimination parameter since it is absorbed in the difficulty. The model can also be viewed as a limiting case for a binomial trials model (Rasch, 1960; Masters & Wright, 1984), when the test score is the number of incorrect responses out of n items with small error probabilities. These error probabilities need not be equal, as was pointed out by Lord & Novick (1968).

Summarizing the observed scores of N subjects on k tests designed to measure the same trait, or equivalently on the same test on k occasions, as an N by k table of counts, and let the entries of the table x_{vi}, denote the score of subject v on test i, and the row and column sums x_{v+} and x_{+i}, the subject- and test scores. The RPCM assumes the x_{vi}'s to be Poisson istributed random variables with means λ_{vi}. The λ_{vi}'s are supposed to be products of two other parameters, a row parameter which refers to the ability of the vth subject and a column parameter referring to the „difficulty" of the test given on the ith occasion. To allow for unequal test lengths, the rates can be modelled instead of the means. So, according to the RPCM the probability that subject v has a score of x_{vi} on test i, of length n_i is given by:

$$P\left(X_{vi} = x_{vi}\right) = \frac{\exp\left(-n_i \lambda_{vi}\right) n_i \lambda_{vi}{}^{x_{vi}}}{x_{vi}!} \quad , \quad \lambda_{vi} > 0 \, , \tag{1}$$

and

$$\lambda_{vi} = \varepsilon_v \, \tau_i \, . \tag{2}$$

Since the Poisson parameters are positive, taking positive values for the test parameters implies that the subject parameters have to be positive also. To avoid indeterminacy, we constrain the τ_i to sum to 1. In Rasch's original model both the ability and the occasion or test parameter were assumed to be fixed. In our model we consider the abilities as random variables, or in other words as randomly sampled from some population. Introducing the assumption of random ability parameters has as a consequence that the unconditional variances of the scores are now larger than their means, while in the Poisson model means and variances are equal. Thus, the extended model allows for a phenomenon, often observed in count data, which is known as overdispersion (Thall & Vail, 1990; Lindsey, 1993). For the random ability parameters we will assume a gamma distribution with shape parameter c and mean m. Since the subject parameters are non-negative quantities, the gamma distribution which is conjugate to the Poisson is a convenient choice. Using other, and therefore non-conjugate distributions, such as the lognormal distribution leads to less tractable results (Jansen, 1994). The parameters c and m and the τ's can be estimated using (marginal) maximum likelihood procedures. The estimation of the τ's however, is not affected by the choice of the latent distribution. Conditional and marginal maximum likelihood procedures lead to identical estimation equations for the τ's.

The marginal likelihood can be written in the following convenient way:

$$L_m\left(\mathbf{X} \middle| \boldsymbol{\tau}, c, m\right) = \prod_v \frac{\Gamma\left(\Sigma_i x_{vi} + c\right) \quad \left(c/m\right)^c \quad \left(\Sigma_i n_i \tau_i\right)^{\Sigma_i x_{vi}}}{\left(\Sigma_i x_{vi}\right)! \quad \Gamma\left(c\right) \quad \left(\Sigma_i n_i \tau_i + c/m\right)^{\Sigma_i x_{vi} + c}} \quad \text{X}$$

$$\prod_v \quad \left(\Sigma_v x_{vi}\right)! \quad \prod_i \frac{\left(n_i \tau_i / \Sigma_i n_i \tau_i\right)^{x_{vi}}}{x_{vi}!} . \tag{3}$$

The first part in the above expression is a negative binomial likelihood for the row sums x_{v+}, which depends on m and c and on the test parameters through their (weighted) sum only. The second term is a multinomial likelihood for the vectors of scores $\mathbf{x}_v = (x_{v1},\ldots,x_{vk})$, conditional on the row sums x_{v+}, depending on the test parameters $(\tau_1,\ldots,\tau_k)$ only.

2. Longitudinal Data

The data which we have considered in the previous section can also be seen as resulting from observing a number of relatively short time-series's, where the observations are counts. In an educational context, longitudinal data will arise when we observe a learning or growth process and the occasion parameters $\boldsymbol{\tau} = (\tau_1,\ldots,\tau_k)$, can then be interpreted as learning parameters. When modelling longitudinal data, we may wish to incorporate a further structure on the occasion parameters in the model, reflecting some plausible pattern of changes over time. This can be done by assuming a log-linear model for the τ_i's, for instance:

$$\log \tau_i = \mathbf{z}_i^T \boldsymbol{\eta} , \tag{4}$$

where $\boldsymbol{\eta}$ is a parameter vector of length p, and $\mathbf{z}_i$ a known design vector. In particular, the simplest longitudinal model, assuming a linear decrease (increase) of $\log(\tau)$ over time, can be written as

$$\tau_i = \exp\left(\eta_0 + \eta_1 t_i\right) , \tag{5}$$

where η_0 is a „nuisance parameter" ensuring that the τ's sum to one, and t is the „time" at occasion i. Note that the model implies that individual subjects share the same „growth" parameters, and that therefore the individual estimated growth curves will be parallel, which might be un-realistic for some types of (longitudinal) count data. A dependency of the fixed occasion parameters on categorical between-subjects factors can of course be modelled easily as was shown by Jansen & van Duijn (1992). An index g can be added to denote the gth level of the between-subjects factor(s), but this does not result in a truly subject-specific model, since within-groups the in dividual growth curves are still parallel.

Obviously, the applicability of the RPCM is not limited to count data observed in (psychological) testing situations. In an educational context, we might also consider data involving school absenteeism, and possibly drop-out rates and other discrete school achievement measures.

3. Incorporating Explanatory Variables for the Ability Parameters

When the subjects are cross-classified according to one or more between-subjects factors we can model group differences by allowing the parameters c and m of the ability distribution to vary over groups (Jansen & van Duijn, 1992). Here we will take a slightly different approach which enables us to incorporate the relation between manifest, categorical as well as continuous or measured, explanatory variables and the random subject effects into our model. This approach which ties in nicely with general linear modelling is also advantageous from an interpretational point of view. We will now assume that the index parameter c is the same for all subjects, while m_v varies over subjects and can be further modelled as a function of fixed between-subjects factors or continuous covariates. Let us write $\mathbf{y}_v$ for a q vector of fixed covariates for subject v, and assume the following log-linear model for the mean m_v of the random effects:

$$\log m_v = \mathbf{y}_v^T \boldsymbol{\beta} \ . \qquad (6)$$

The inclusion of explanatory variables for the random subject parameters can be seen from two points of view. Namely, modelling the subject parameters can be seen as formulating a GLIM type model for a gamma distributed dependent variable (Aitkin et al., 1990), which is now latent instead of manifest. Alternately, we can consider our approach as an application of a negative binomial regression model to the observed row sums x_{v+}. For the implementation of the principles of negative binomial regression in our situation, we can also (partly) rely on known results, see for instance Lawless (1987). Lawless also describes a number of statistical tests for extra Poisson variation.

4. Subject Specific Learning Parameters

In many situations the assumption that the occasion parameters are fixed is too restrictive. We have formulated a second model in which we allow the condition or learning parameters to vary over subjects. A tractable and reasonably flexible way to describe this type of variation is to assume that the subject's vectors of occasion parameters $(\tau_{v1},\ldots,\tau_{vk})$ are sampled from a Dirichlet distribution. For the sake of notational simplicity we will assume the n_i's to be equal to 1. As before, x_{vi} denotes the score of subject v on occasion i and the scores are again independently Poisson distributed for given intensity parameter λ_{vi}. But λ_{vi} is now assumed to be a product of two parameters, which are both considered random, an ability parameter ε_v, and a subject-specific occasion parameter τ_{vi}. The vectors of parameters $\boldsymbol{\tau}_{i1} = (\tau_{i1},\ldots,\tau_{ik})$ with $\tau_{v+} = 1$, and v = 1,…,N, are assumed to be sampled from a common Dirichlet distribution, with parameters $\mathbf{b} = (b_1,\ldots,b_k)$.

The τ_{vi}' s and b_i' s are related in the following way:

$$E(\tau_{vi}) = b_i/b_+ , \quad \text{and}$$
$$\text{Var}(\tau_{vi}) = b_i(b_+ - b_i)\Big/\left\{b_+^2 (b_+ + 1)\right\}, \quad \text{where} \quad b_+ = \Sigma(b_i). \tag{7}$$

Again the marginal likelihood consists of a product of two terms, a negative binomial likelihood for the row sums depending on the parameters of the gamma distribution as before and a so-called compound multinomial likelihood for the vectors of scores $\mathbf{x}_v = (x_{v1},\ldots,x_{vk})$, given the row sums x_{v+}, depending on the Dirichlet parameters $b_1,\ldots,b_k$ only. The fixed (hyper)parameters of the Dirichlet or their reparametrizations $\pi_1,\ldots,\pi_k$, where $\pi_i = b_i /b_+$, which can be interpreted as mean „difficulty“ parameters, are now the parameters of interest. Further restrictions, for instance in the form of a log-linear model, can be imposed on them. Additionally, we also have the parameter b_+, or $\rho = 1/b_+$. When $\rho \rightarrow 0$ (b_+ goes to infinity) the Poisson-Gamma model returns. The subject specific occasion parameters $\tau_{v1},\ldots,\tau_{vk}$ are a posteriori also Dirichlet distributed, with parameters $(x_{vi} + b_1,\ldots,x_{vk} + b_k)$. The posterior means can be used to obtain point estimators

$$\tau_{vi} = \frac{x_{vi} + b_i}{x_{v+} + b_+}, \; v = 1,\ldots, N; \; i = 1,\ldots, k \; . \tag{8}$$

The estimated growth curves of the individual subjects will no longer be parallel, as was the case in the Gamma-Poisson model. Including explanatory variables for the Dirichlet involves somewhat tedious but otherwise straightforward derivations. Details are given in Van Duijn & Jansen (1995). Random missing data can be handled in the Gamma-Poisson model with fixed occasion parameters. In the case of random subject dependent occasion parameters however, missing data are more problematic. Models for event frequencies based on compounding a Poisson distribution with Gamma and or Dirichlet distributions have also been applied in studies devoted to topics such as crime victimization, consumer behaviour, and many others (Nelson, 1984). A Poisson model for repeated count data, with fixed effects only, is given by Fischer (1992). Models very similar to our own can be found in the work of Böckenholt (Böckenholt, 1993) on recurrent choice data.

5. Applications to Reading Development Data

In our application we analyze data which are derived from a larger study on the development of arithmetic and language skills of (Dutch) primary school pupils. The data we use concern scores on a reading test that was administered at approximately three monthly intervals during the school year when the pupils were in the second, third, fourth and fifth grades.

From these twelve administrations, we will only use the first six, since some pupils were not further tested after they had reached an arbitrary cutting score of 80. The test that was used in the study is the Brus test, which is a speed test where the score consist of the number of words that a subject can read within a time-limit of one minute.

Our sample consisted of 153 subjects. Among the background variables were sex, a socio-economic background indicator and a home-language variable (since the study was conducted in the province of Frisia the home-language for some of the pupils was Frisian). In our analysis we will look at the influence of the socio-economic status (SES) variable on the development of reading skills only, and restrict our analysis to the subjects with complete data (N = 144). Table 1 presents the means and standard deviations of the reading scores summed over occasions for the high, medium and low SES groups.

Group	*Mean*	*Std Dev*	*N*
1. High	294.3	92.94	43
2. Medium	276.6	93.41	56
3. Low	257.3	76.90	45

Table 1: Means and standard deviations of the total subject scores per socio-economic background (complete data only, N = 144)

For the analysis SES was coded using dummy variables. The three SES groups differed in overall ability level but the differences in the expected direction were not significant. The results are given in Table 2.

Parameter	*Estimate*	*Std Error*
C	8.70	1.044
β_0	5.59	0.053
β_1	-0.06	0.070
β_2	-0.13	0.073

Table 2: Parameter estimates for the regression of ability on SES

Means and standard deviations of the scores per occasion for the total group of subjects are given in columns 2 and 3 in Table 3. The last two columns contain estimates for the occasion parameters and their approximate standard errors, based on the un-restricted Gamma-Poisson model. Occasion parameters can be seen as the proportional distribution of the total score over the occasions.

Occasion	*Mean*	*Std Dev*	τ	*SE*(τ)
1	27.2	14.38	0.099	0.0016
2	35.7	15.67	0.130	0.0018
3	45.5	15.81	0.165	0.0020
4	49.9	15.99	0.181	0.0021
5	56.0	16.32	0.203	0.0023
6	61.5	15.85	0.223	0.0027

Table 3: Means and standard deviations of the subject scores per occasion and occasion parameter estimates

The results of fitting different 'growth' models to the (fixed) log occasion parameters can be found in Table 4. The occasions were assumed to be equally spaced. Significant differences related to the SES variable were not encountered.

We have considered four different log-linear models, a model with only a linear time-effect, and models with a linear time-effect combined with a quadratic time-effect and/or a year effect. The „year“ variable was introduced to account for the possible effects of the somewhat larger gap between occasion 3 and 4, due to summer holidays. As a goodness-of-fit criterion we used the deviance. The deviance is a linear function of the loglikelihood of the current model and the maximum loglikelihood, achievable for an exact fit.

Model	Deviance	df
Null-Model: Saturated	0	0
$\eta_0 + \eta_1 t_j$	151.9	4
$\eta_0 + \eta_1 t_j$ + year	145.0	3
$\eta_0 + \eta_1 t_j + \eta_2 t_j^2$	20.5	3
$\eta_0 + \eta_1 t_j + \eta_2 t_j^2$ + year	1.6	2

Table 4: The results of fitting linear and quadratic functions to the (log) occa sion parameters. Year has two levels, and time has been rescaled to have a zero mean.

From Table 4 we infer that the deviance increases significantly as we exchange the saturated model, which fits exactly, for the most simple longitudinal model, assuming a linear increase of log(τ) over time. The inclusion of a year effect leads to a significant but small increase of the fit. Adding a quadratic time-component produces a relatively large reduction in deviance. It seems therefore reasonable to choose either the third or the fourth model as the final model. The fouth model however, is already rather comlex given the limited number of occasions, and we are probably over-fitting the data. Comparing the unrestricted estimates of the occasion parameters with the estimates based respectively, on the model assuming a linear time-component only and the model with a quadratic component (see also Figures 1 and 2), we might conclude that, form a drescriptive point of view, the latter (model 3), is acceptable.

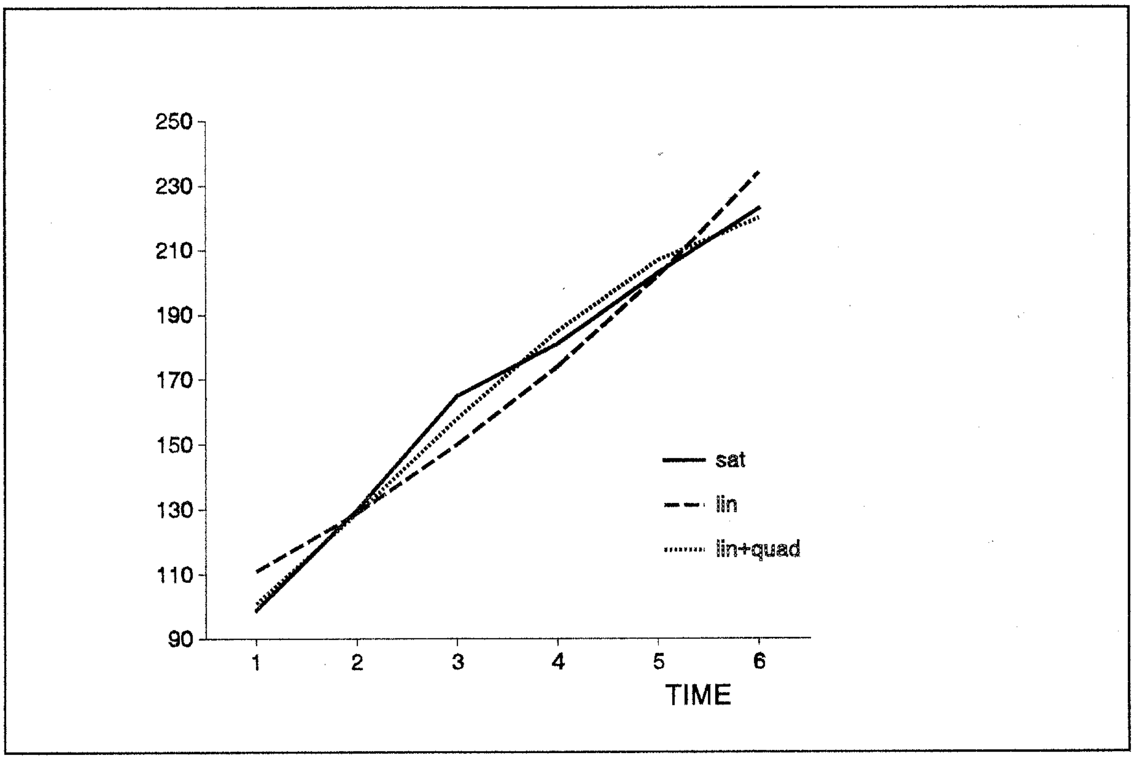

Figure 1: Estimated growth curves

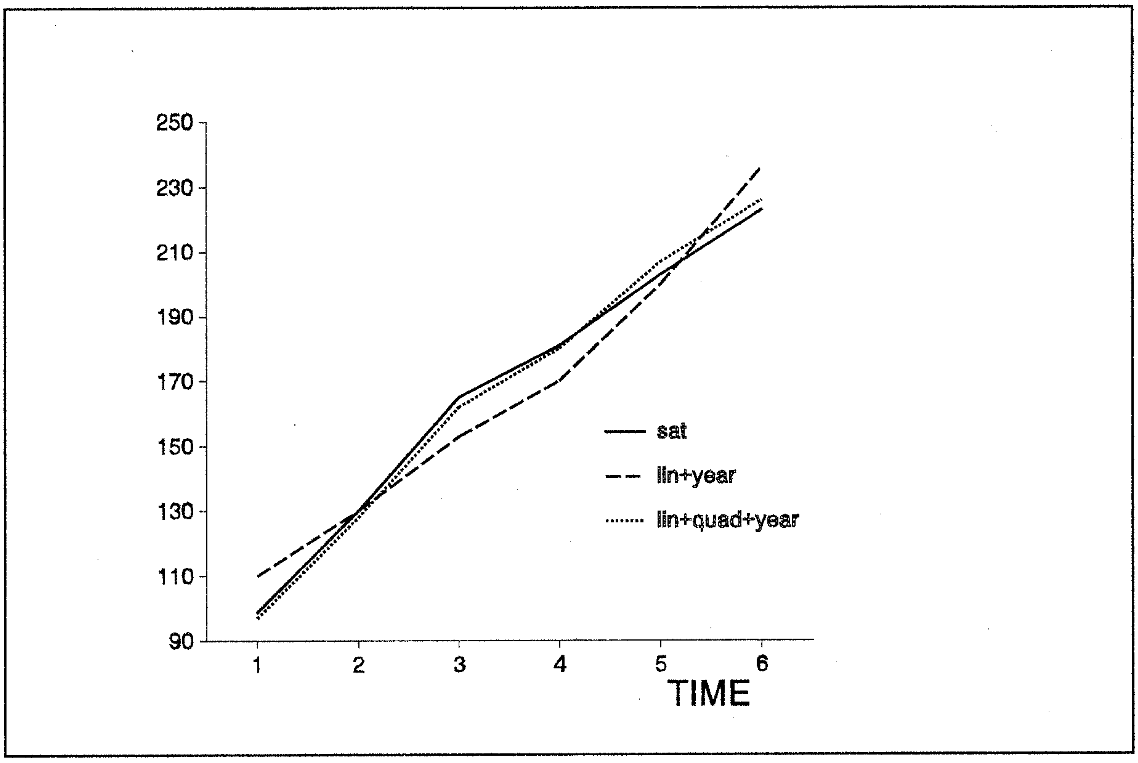

Figure 2: Estimated growth curves

The Dirichlet-Gamma-Poisson model was considered next. As it turned out, the model with subject-dependent occasion parameters did not fit the data better than the model assuming fixed occasion parameters. That, in this case, the Dirichlet-Gamma-Poisson model will not fit the data better than the Gamma-poisson model is also clear from the estimated value of ρ, the interaction strength, which is very small indeed $(\rho = 0.0007)$.

6. Summary and Discussion

Longitudinal count data obtained by giving the same reading test repeatedly were organized in a two-way table with subjects as rows and occasions as columns. The scores were analyzed using a Poisson model for repeated count data with a multiplicative parameter structure. In our first approach we used the RPCM with fixed occasion (test) parameters. The random subject parameters account for the differences in overall performance between subjects, while the occasion parameters refer to the relative performance (relative to the total performance over occasions), which is assumed to differ over occasions, in the same way for each subject.

In our analysis we studied the relation between socio-economic status (SES) and the subject (ability) and occasion parameters. Somewhat surprisingly, the effect of SES on reading ability, although in the expected direction turned out to be very small. It should be noted however, that although the sample has been divided into three SES-classes, the actual differences in SES between the classes might be small. Secondly, a number of different log-linear models were fitted to the occasion parameters representing possible patterns of change over time, among these a simple model assuming a (log) linear increase in (relative) reading performance over time. The model combining a linear and a quadratic time effect and an effect for year was judged to be adequate from a descriptive point of view. Inter-individual differences in growth pattern proved to be negligible.

The models we have applied here are obviously related to multi-level models for discrete outcome variables (Goldstein, 1991), although more restrictive. We allow here for a fixed number of time-points and moreover, the number of time-points has to be relatively small. But, although in general the multi-level approach seems to work well and is more flexible than ours, the appropriateness for the analysis of discrete data is not yet fully established.

Appendix: Estimation equations for the Gamma Poisson model with manifest between-subjects covariates

The log-likelihood and the partial derivatives with respect to the regression parameters are given by equations (A1) and (A2).

$$\begin{aligned}\log L_m &= Nc \log(c) - N \log \Gamma(c) \\ &- \sum_v c \log(m_v(\mathbf{y})) + \sum_v \log \Gamma\left(\Sigma_i a_{vi} x_{vi} + c\right) \\ &- \sum_v \left(\sum_i a_{vi} x_{vi} + c\right) \log\left(\Sigma_i a_{vi} n_i \tau_i + c/m_v(\mathbf{y})\right) \\ &+ \sum_i \left(\sum_v a_{vi} x_{vi}\right) \log(n_i \tau_i) - \sum_{vi} a_{vi} \log(x_{vi}!),\end{aligned} \tag{A1}$$

and

$$\frac{d\, l_m}{d\, \beta_r} = c \sum_v y_r \left\{ \frac{\left(c + \Sigma_i a_{vi} x_{vi}\right)}{\left(c/m_v(\mathbf{y})\right) + \Sigma_i a_{vi} n_i \tau_i} \Big/ m_v(\mathbf{y}) - 1 \right\}, \tag{A2}$$

for $r = 1, \dots, q$.

The MLE $\hat{c}$ of c can be obtained by solving Equation (A3) and the MLE's for the τ_i's by solving the Equations (A4). Since the τs sum to 1 there are only K-1 independent equations.

$$1 - \psi(c) + \frac{1}{N}\sum_v \log\left(c / m_v(\mathbf{y})\right)$$

$$+ \frac{1}{N}\sum_v \left\{\psi\left(c + \Sigma\, a_{vi} x_{vi}\right) - \log\left(c/m_v(\mathbf{y}) + \Sigma\, a_{vi} n_i \tau_i\right)\right\}$$

$$- \frac{1}{N}\sum_v \frac{\left(c + \Sigma\, a_{vi} x_{vi}\right)}{\left(c/m_v(\mathbf{y}) + \Sigma\, a_{vi} n_i \tau_i\right)} \Big/ m_v(\mathbf{y}) = 0 \;. \qquad \text{(A3)}$$

$$\frac{\Sigma_v\, a_{vi} x_{vi}}{\tau_i} - \sum_v \frac{\left(c + \Sigma_i a_{vi} x_{vi}\right)}{\left(c/m_v(\mathbf{y}) + \Sigma_i a_{vi} n_i \tau_i\right)} a_{vi} n_i = 0 \;. \qquad \text{(A4)}$$

With ψ we denote the digamma function. The indicator variables a_{vi}, have values of 1 if a score has been observed of subject v on occasion (or test) i and 0 otherwise.

References

Aitkin, M., Anderson, D., Francis, B. & Hinde, J. (1990). *Statistical Modelling in GLIM*. Oxford: Oxford University Press.

Böckenholt, U. (1993). A latent-class regression approach for the analysis of recurrent choice data. *British Journal of Mathematical and Statistical Psychology, 46*, 95-118.

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

Goldstein, H. (1991). Nonlinear multilevel models, with an application to discrete response data. *Biometrika, 78*, 45-51.

Jansen, M.G.H. (1994). Parameters of the latent distribution in Rasch's Poissons Counts Model. In: Fischer, G.H. & Laming, D. (Eds.*) Contributions to Mathematical Psychology, Psychometrics, and Methodology*, 315-322. New York: Springer-Verlag.

Jansen, M.G.H. & van Duijn, M.A.J. (1992). Extensions of Rasch's multiplicative Poisson model. *Psychometrika, 54*, 405-414.

Lawless, J.F. (1989). Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics, 15*, 209-225.

Lindsey, J.K. (1993). *Models for repeated measurements*. Oxford: Clarendon Press.

Masters, G.N. & Wright, B.D. (1984). The essential process in a family of measurement models. *Psychometrika, 49*, 529-544.

Nelson, J.F. (1984). The Dirichlet-Gamma-Poisson model of repeated events. *Sociological Methods & Research, 12*, 347-373.

Rasch, G., (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press. (Original work published in 1960).

Thall, P.F., & Vail, S.C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics, 46*, 657-671.

van Duijn, M.A.J. & Jansen, M.G.H. (1995). Modelling repeated count data: Some extensions of the Rasch Poisson counts model. *Journal of Educational and Behavioral Statistics, 20*, 241-258.

Van der Velde, A. (1982). *Lezen en spellen in klas twee tot zes. (Reading and spelling in class two to six)*. Groningen: University of Groningen, internal report.