PSYCHOMETRIKA—VOL. 53, NO. 4, 525–546
DECEMBER 1988

THE DERIVATION OF SOME TESTS FOR THE RASCH MODEL FROM THE MULTINOMIAL DISTRIBUTION

CEES A. W. GLAS

NATIONAL INSTITUTE FOR EDUCATIONAL MEASUREMENT (CITO)
ARNHEM, THE NETHERLANDS

The present paper is concerned with testing the fit of the Rasch model. It is shown that this can be achieved by constructing functions of the data, on which model tests can be based that have power against specific model violations. It is shown that the asymptotic distribution of these tests can be derived by using the theoretical framework of testing model fit in general multinomial and product-multinomial models. The model tests are presented in two versions: one that can be used in the context of marginal maximum likelihood estimation and one that can be applied in the context of conditional maximum likelihood estimation.

Key words: Item response theory, Rasch model, model test.

1. Introduction

The problem of evaluating model fit in latent trait models is often solved within the well established framework of testing model fit in the general multinomial model (see for instance Bock & Aitkin, 1981). This can be done by recognizing that if $\{\mathbf{x}\}$ stands for the set of all possible response patterns, the vector of frequency counts $\mathbf{n}$ with elements $n_{\mathbf{x}}$ has a multinomial distribution with parameters N and $\boldsymbol{\pi}$, where N is the number of respondents, $\boldsymbol{\pi}$ a vector with as elements the theoretical probabilities $p(\mathbf{x} \mid \boldsymbol{\lambda})$ of the response patterns and $\boldsymbol{\lambda}$ a vector of model parameters.

The likelihood ratio statistic for testing the assumed model against a general multinomial alternative is given by

$$G^2 = 2 \sum_{\{\mathbf{x}\}} n_{\mathbf{x}} \ln \left(\frac{n_{\mathbf{x}}}{N p(\mathbf{x} \mid \boldsymbol{\lambda})} \right).$$

It can be shown (see for instance Bishop, Fienberg & Holland, 1975) that G^2 has an asymptotic χ^2 distribution.

If however the number of possible response patterns is large, the vector of frequency counts $\mathbf{n}$ will have a number of very small or even zero elements. In such cases it is often suggested to pool patterns to obtain observed and expected frequencies which are sufficiently large. However this pooling is a function of the data itself, so the asymptotic distribution of a test of model fit based on the pooled data can hardly be derived. Another drawback in this approach lies in the fact that interpreting the causes of a possible misfit is hampered by the aggregation level of the test: the influence of particular items on the outcome of the test and other specific causes of misfit cannot be identified. The present paper is concerned with presenting a remedy for this situation for the dichotomous Rasch model. The essence of the technique is finding some function of $\mathbf{n}$ leading to a test which has power against specific model violations. The first test discussed focusses on a table with as entries counts of the number of subjects who attain a certain sum score r and

I am indebted to Norman Verhelst and Niels Veldhuijzen for their helpful comments. Requests for reprints should be sent to Cees A. W. Glas, Cito, PO Box 1034, 6801 MG Arnhem, THE NETHERLANDS.

0033-3123/88/1200-9044$00.75/0

make item i correct. These counts, denoted by m_{ri}, will be referred to as "first order realizations". It will be shown that a test based on these observations is sensitive to variation in the slopes of the item characteristic curves.

The second test focusses on a table with counts of the number of subjects who make both item i and item j correct. These counts will be denoted by m_{ij}^* ($i \neq j$). The idea of evaluating these so-called "second order realizations" is due to van den Wollenberg (1982), who has shown that tests based on these observations have power against violations of the axiom of unidimensionality.

The tests proposed in the present paper are based on the comparison of expected and observed frequencies. For the estimation of the expected value of m_{ri} and m_{ij}^* one must deal with the person parameters in the Rasch model, which cause inconsistencies in the estimation of the item parameters (Andersen, 1973). The problem is solved in two ways. The first method maximizes the likelihood of the data conditional upon the sufficient statistics for the person parameters (Rasch, 1961). This is generally known as conditional maximum likelihood estimation. The second approach, generally known as marginal maximum likelihood estimation, considers the person parameters as independent, identically distributed random variables (Rigdon & Tsutakawa, 1983).

The paper will consider both estimation procedures. So in the next sections a test based on the first order realizations will be presented for respectively the marginal and the conditional approach and an example of the testing procedure will be given. In the following sections the same will be done for second order realizations.

2. First Order Realizations, the Marginal Approach

Consider the set of all possible response patterns on a test of k items. Suppose the number of possible response patterns is t and N persons made the test. Let $\mathbf{n}$ be a t-dimensional vector of frequency counts which will be partitioned $\mathbf{n}' = (n_0, \mathbf{n}_1', \ldots, \mathbf{n}_r', \ldots, \mathbf{n}_{k-1}', n_k)$, such that $\mathbf{n}_r$ is a vector of the frequency counts of response patterns leading to a sum score r, n_0 stands for the number of persons attaining a zero score and n_k stands for the number of persons attaining a perfect score.

In the Rasch model it is assumed that the probability of a correct response as a function of the ability parameter ϑ is given by

$$p(x_i = 1 \mid \vartheta, \delta_i) = \frac{\exp(\vartheta - \delta_i)}{1 + \exp(\vartheta - \delta_i)}, \tag{1}$$

where δ_i stands for the difficulty of item i. If it is also assumed that the ability parameters are randomly sampled from a normal distribution with parameters μ, $\mu = 0$, and σ, the expectation of the number of persons producing response pattern $\mathbf{x}$ is given by

$$\begin{aligned} E(n_{\mathbf{x}} \mid \boldsymbol{\delta}, \sigma) \triangleq N\pi_{\mathbf{x}} &= N \int_{-\infty}^{\infty} \frac{\exp(r\vartheta - \mathbf{x}'\boldsymbol{\delta})}{\prod_{i=1}^{k}(1 + \exp(\vartheta - \delta_i))} g(\vartheta \mid \sigma)\, \partial\vartheta \\ &= N \int_{-\infty}^{\infty} \exp(r\vartheta - \mathbf{x}'\boldsymbol{\delta}) p_0(\vartheta) g(\vartheta \mid \sigma)\, \partial\vartheta, \end{aligned} \tag{2}$$

with $\boldsymbol{\delta}' = (\delta_1, \ldots, \delta_k)$, $g(\vartheta \mid \sigma)$ the normal probability density function and r the sum score associated with $\mathbf{x}$.

The frequency distribution of the sum scores is a sufficient statistic for the ability distribution (Rigdon & Tsutakawa, 1983) and it can be assumed that persons with the same sum score form a rather homogeneous group on the latent continuum. For every item, checking m_{ri} against its expected value across various score levels may reveal deviations of the empirical item characteristic curve from the one predicted by the model. If for instance m_{ri} is too small at low score levels and too large at high score levels, it can be concluded that the item contributes to a possible lack of model fit because its item characteristic curve is too steep. Therefore the objective of this section is to develop a statistical test based on the observations m_{ri}.

Let X_r be a matrix with as columns all response patterns leading to a sum score r for $r = 1, \ldots, k-1$ and let $\mathbf{m}_r$ be a vector of the observations m_{ri}, $i = 1, \ldots, k$. It can be easily verified that $X_r \mathbf{n}_r = \mathbf{m}_r$. Using this it can also be shown that

$$E(m_{ri} \mid \boldsymbol{\delta}, \sigma) \triangleq N\Psi_{ri} = N \int_{-\infty}^{\infty} \varepsilon_i \gamma_{r-1}^{(i)}(\boldsymbol{\varepsilon}) \exp(r\vartheta) p_0(\vartheta) g(\vartheta \mid \sigma)\, \partial\vartheta, \tag{3}$$

with $\varepsilon_i = \exp(-\delta_i)$ and $\gamma_{r-1}^{(i)}(\boldsymbol{\varepsilon})$ an elementary symmetric function of order $r-1$ with as parameters the elements of $\boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}' = (\varepsilon_1, \ldots, \varepsilon_k)$, where ε_i has been set equal to zero. For the definition of an elementary symmetric function one is referred to Fischer (1981).

Let the k-vector of deviances in score group r, $\mathbf{d}_r$, be defined as $\mathbf{d}_r = \mathbf{m}_r - E(\mathbf{m}_r \mid \hat{\boldsymbol{\delta}}, \hat{\sigma})$ for $r = 1, \ldots, k-1$, $d_0 = n_0 - E(n_0 \mid \hat{\boldsymbol{\delta}}, \hat{\sigma})$ and $d_k = n_k - E(n_k \mid \hat{\boldsymbol{\delta}}, \hat{\sigma})$, where $\hat{\boldsymbol{\delta}}$ and $\hat{\sigma}$ stand for maximum likelihood estimates. These deviances will be combined into a quadratic form using matrices $\hat{W}_r = X_r \hat{D}_{\pi(r)} X_r'$, $r = 1, \ldots, k-1$ with $\hat{D}_{\pi(r)}$ a diagonal matrix with as diagonal elements estimates of the probabilities $\pi_{\mathbf{x}}$ as defined in (2), for all response patterns leading to a sum score r. It can be verified that W_r has diagonal elements Ψ_{ri} as defined in (3) and off-diagonal elements

$$\Psi_{rij} = \int_{-\infty}^{\infty} \varepsilon_i \varepsilon_j \gamma_{r-2}^{(i,j)}(\boldsymbol{\varepsilon}) \exp(r\vartheta) p_0(\vartheta) g(\vartheta \mid \sigma)\, \partial\vartheta, \tag{4}$$

with $\gamma_{r-2}^{(i,j)}(\boldsymbol{\varepsilon})$ an elementary symmetric function of order $r-2$, where the parameters ε_i and ε_j have been set equal to zero.

Using these definitions the following theorem can be stated.

Theorem 1. If $N \to \infty$

$$R_m = \frac{d_0^2}{E(n_0 \mid \hat{\boldsymbol{\delta}}, \hat{\sigma})} + N^{-1} \sum_{i=1}^{k-1} \mathbf{d}_r' \hat{W}_r^{-1} \mathbf{d}_r + \frac{d_k^2}{E(n_k \mid \hat{\boldsymbol{\delta}}, \hat{\sigma})} \tag{5}$$

has an asymptotic χ^2 distribution with $k(k-2)$ degrees of freedom.

A detailed proof of the theorem is given in Appendix A. In this appendix it is shown that the number of degrees of freedom is equal to the number of deviances on which the test is based, which is $k(k-1)+2$, minus the number of parameters that have to be estimated, which is $k+1$, minus one, so the degrees of freedom are given by $k(k-2)$. Inspection of (5) reveals that the number of items is the main restriction on the computability of the test. This restriction works in two ways. First, elementary symmetric functions have to be computed. For this recursion formulas are available and the numerical problems are solved (see Verhelst, Glas & van der Sluis, 1984). Secondly the $k \times k$ matrices $\hat{W}_r$ have to be inverted. It is beyond the scope of the present paper to derive a general rule for the number of items possible, but experience shows that the computation does not run into problems up to 90 items. The theorem given above refers to a situation where a sample from one population is confronted with one test. The result can be generalized to a situation where samples from different populations are confronted with different, though possibly overlapping tests. With respect to the generalization to these situations, the following considerations must be made.

Let $A = D_\pi^{-1/2}(\partial\pi/\partial\lambda')$, with π a vector of the probabilities $\pi_{\mathbf{x}}$ of all possible response patterns, D_π a diagonal matrix of these probabilities, and $\lambda = (\delta_1, \ldots, \delta_k, \sigma)$. In the derivation given in Appendix A, it is essential that the $t \times (k + 1)$ matrix A is of full column rank. If this is not the case the model is not identified. For a situation of G populations and H tests the definition of A must be extended. Let $\mathbf{x}_{gh}$ be a response pattern on test h given to the sample of population g. The exact design will not be specified, but the restrictions on the design imposed by the necessity of constructing an identified model will be discussed below.

Let π be the vector of theoretical probabilities of all possible response patterns for all relevant combinations of a population and a test. Suppose that π has elements

$$\pi_{\mathbf{x}_{gh}} = \int_{-\infty}^{\infty} \prod_{i \in B_h} \frac{\exp(x_i \alpha(\vartheta - \delta_i))}{(1 + \exp(\alpha(\vartheta - \delta_i)))} \frac{\exp\left(\dfrac{-1/2(\vartheta - \mu_g)^2}{\sigma_g^2}\right)}{\sigma_g(2\pi)^{1/2}}, \tag{6}$$

where B_h is the set of indices of all items in test h.

The model given in (6) is not identified because there exists a transformation of the parameters which leaves the theoretical probabilities unchanged: $\tilde{\alpha} = \alpha/d$ $(d > 0)$, $\tilde{\delta}_i = d\delta_i + c$, $\tilde{\sigma}_g = d\sigma_g$ and $\tilde{\mu}_g = d\mu_g + c$.

It can be easily shown that the functional dependence among the parameters causes the matrix $\partial\pi/\partial\lambda'$ with λ a vector of all parameters, to be of incomplete column rank. For one test and one population the indeterminacy can be resolved by imposing the restrictions $\alpha = 1$ and $\mu = 0$. It would go beyond the scope of the present paper to treat all possible test administration designs, but some situations often arising in practice will be given as examples of how the identification problem must be solved. It is always assumed that $\alpha = 1$.

1. In the case of one test and several populations $\mu_1 = 0$ and μ_g free as $g > 1$ is a sufficient restriction to identify the model.

2. In the case of several tests and one population the restriction $\mu = 0$ may be chosen. As an example one may think of the instance of test equating where different samples make different and possibly nonoverlapping tests. In this case the item parameters can be calibrated on a common scale by assuming that all samples are drawn from a common ability distribution.

3. Consider a situation where one test is administered to a sample of one population and another test is given to a sample from another population. If $B_1 \cap B_2 \neq \varnothing$, that is, if the two tests have common items, the restriction $\mu_1 = 0$ and μ_2 free is sufficient. So this covers the well known situation of common item equating. If however $B_1 \cap B_2 = \varnothing$, the restriction $\mu_2 = 0$ must also be imposed. In this case the likelihood of the data for the first sample is independent of the likelihood of the data in the second sample and the item parameters are not calibrated on one common scale.

Given a matrix $A = D_\pi^{-1/2}(\partial\pi/\partial\lambda')$ of full column rank, that is, given a full rank parametrization of the model, generalization of the statistical test R_m to a situation of more tests and more populations is straightforward. Let N_{gh} be the size of the sample of population g taking test h. Test h has k_h items. For every relevant combination of a test and a population, a table with entries m_{righ} will be analyzed, m_{righ} stands for the number of persons in the sample of population g, who take test h, attain a score r and make item i, $i \in B_h$, correct. If $\boldsymbol{\delta}_h$ is the vector of item parameters of test h and μ_g and σ_g are the parameters of population g, $E(m_{righ} \mid \boldsymbol{\delta}_h, \mu_g, \sigma_g)$ can be evaluated using expression (3) with the proper substitution of arguments. If $\mathbf{m}_{rgh}$ is a vector with elements m_{righ}, $i = 1, \ldots, k_h$, a vector of deviances $\mathbf{d}_{rgh} = \mathbf{m}_{rgh} - E(\mathbf{m}_{rgh} \mid \hat{\boldsymbol{\delta}}_h, \hat{\mu}_g, \hat{\sigma}_g)$ can be defined for all relevant combinations of r, g and h. The deviances for a zero score and a perfect score, d_{0gh} and $d_{k_h gh}$ are defined analogous to the one test and one population case.

It must be stressed that the deviances are evaluated using maximum likelihood estimates of the parameters obtained by maximizing the likelihood of the complete data set, that is, by maximizing $\sum_{g,h} \sum_{\{\mathbf{x}_{gh}\}} n_{\mathbf{x}_{gh}} \ln \pi_{\mathbf{x}_{gh}}$ as a function of all parameters. The symbol $\sum_{g,h}$ stands for a summation that runs over all combinations of g and h that are relevant to the test administration design under consideration. As in the case of one test and one population all deviances are combined into a quadratic form, this time by applying $k_h \times k_h$ matrices $\hat{W}_{rgh}$. $\hat{W}_{rgh}$ has diagonal elements $\hat{\Psi}_{righ}$ $(i = 1, \ldots, k_h)$ and off-diagonal elements $\hat{\Psi}_{rijgh}$ $(i \neq j)$ which are defined by (3) and (4) with the proper substitution of the item parameters $\hat{\boldsymbol{\delta}}_h$ and the population parameters $\hat{\mu}_g$ and $\hat{\sigma}_g$. Now the following generalization of Theorem 1 can be given.

Theorem 2. Consider G populations and H tests, from every population at least one sample is drawn and every test is administered at least once. Let the parametrization be such that $\partial\boldsymbol{\pi}/\partial\boldsymbol{\lambda}'$ is of full column rank. If for all relevant combinations of g and h, $N_{gh} \to \infty$,

$$R_m = \sum_{g,h} \left(\frac{d_{0gh}^2}{E(n_{0gh} \mid \hat{\boldsymbol{\lambda}}_{gh})} + N_{gh}^{-1} \sum_{r=1}^{k_h - 1} \mathbf{d}'_{rgh} \hat{W}_{rgh}^{-1} \mathbf{d}_{gh} + \frac{d_{k_h gh}^2}{E(n_{k_h gh} \mid \hat{\boldsymbol{\lambda}}_{gh})} \right)$$

has an asymptotic χ^2 distribution with $\sum_{g,h} (k_h(k_h - 1) + 1) - \dim(\boldsymbol{\lambda})$ degrees of freedom.

The proof of this theorem is a straightforward generalization of the proof given in Appendix A. The degrees of freedom can be determined by counting the number of deviances, subtracting the number of parameters that have to be estimated and subtracting one degree of freedom for every multinomial distribution under consideration, that is, every combination of a test and a sample that is relevant to the design. In the application section of this paper an example of the use of this theorem will be given.

3. First Order Realizations, The Conditional Approach

Neyman and Scott (1948) have shown that the presence of so-called incidental parameters, that is, parameters who's number tends to infinity as the sample size tends to infinity, causes maximum likelihood estimates of the structural parameters to lack consistency and efficiency. In the Rasch model the person parameters act as incidental parameters. One way of dealing with these parameters is used in the previous section. Rasch (1960, 1961) suggested dealing with the problem by a conditional maximum likelihood estimation method: maximizing the likelihood of the data given minimal sufficient statistics for the incidental parameters. Andersen (1973) has shown that the well-known theorems concerning the asymptotic normality of maximum likelihood estimates are also valid for these conditional estimates, the estimates are consistent and uniformly converge to a normal distribution. Furthermore the estimates are efficient in the sense that the lower bound for the asymptotic variance of the estimator is attained.

Fischer (1981) has given a necessary and sufficient condition for the existence and uniqueness of the conditional maximum likelihood estimates of the item parameters. As for testing the model in the conditional case, the idea of comparing the expected and observed values of m_{ri} $(r = 1, \ldots, k - 1$ and $i = 1, \ldots k)$ is not new. Van den Wollenberg (1982) presented a test statistic, Q_1, based on the evaluation of these realizations. However the asymptotic distribution of the test could, as yet, not be derived, though simulation studies support the conjecture that it is χ^2 distributed (van den Wollenberg). Another test based on the first order realizations was suggested by Martin Löf (1973) and in fact it is the same test as the one proposed in the present paper. Re-introducing the test is motivated by the fact that Martin Löf assumes that the number of persons attaining a score r is a Poisson distributed random variable. The test proposed here does not need this assumption and generalizations to the situations considered in the previous section are easily made. But first the case of one test administered to one sample will be treated.

The conditional probability of response pattern $\mathbf{x}$ is given by

$$p(\mathbf{x} \mid r, \boldsymbol{\varepsilon}) \triangleq \pi_{\mathbf{x}\cdot r} = \frac{\prod_{i=1}^{k} \varepsilon_i^{x_i}}{\gamma_r(\boldsymbol{\varepsilon})}, \tag{8}$$

where $\boldsymbol{\varepsilon}' = (\varepsilon_1, \ldots, \varepsilon_k)$ and $\gamma_r(\boldsymbol{\varepsilon})$ an elementary symmetric function of order r ($r = \sum_{i=1}^{k} x_i$). with as arguments the elements of $\boldsymbol{\varepsilon}$.

As in the previous section, the vector of frequency counts of patterns leading to a sum score r, $\mathbf{n}_r$, is transformed $X_r \mathbf{n}_r = \mathbf{m}_r$, with X_r a matrix with as columns all response patterns with sum score r ($r = 1, \ldots, k - 1$). The resulting vector $\mathbf{m}_r$ is given by $\mathbf{m}_r' = (m_{r1}, \ldots, m_{ri}, \ldots, m_{rk})$. It can be verified that

$$E(m_{ri} \mid N_r, \boldsymbol{\varepsilon}) \triangleq N_r \Gamma_{ri} = N_r \frac{\varepsilon_i \gamma_{r-1}^{(i)}(\boldsymbol{\varepsilon})}{\gamma_r(\boldsymbol{\varepsilon})}, \tag{9}$$

with $\gamma_{r-1}^{(i)}(\boldsymbol{\varepsilon})$ as in the previous section and N_r the number of persons scoring r. If

$$\Gamma_{rij} \triangleq \frac{\varepsilon_i \varepsilon_j \gamma_{r-2}^{(i,j)}(\boldsymbol{\varepsilon})}{\gamma_r(\boldsymbol{\varepsilon})}, \tag{10}$$

and $D_{\pi\cdot r}$ a diagonal matrix with on the diagonal all probabilities $\pi_{\mathbf{x}\cdot r}$ leading to a sum score r ($r = 1, \ldots, k - 1$), it can also be verified that the $k \times k$ matrix $W_{\cdot r}$, $W_{\cdot r} \triangleq X_r D_{\pi\cdot r} X_r'$, has diagonal elements Γ_{ri} and off-diagonal elements Γ_{rij}.

Let $\hat{\boldsymbol{\varepsilon}}$ stand for the conditional maximum likelihood estimate of $\boldsymbol{\varepsilon}$. The vector of deviances $\mathbf{d}_{\cdot r}$ is defined by $\mathbf{d}_{\cdot r} = \mathbf{m}_r - E(\mathbf{m}_r \mid r, \hat{\boldsymbol{\varepsilon}})$ and $\hat{W}_{\cdot r}$ stands for $W_{\cdot r}$ evaluated at $\hat{\boldsymbol{\varepsilon}}$. The following theorem is equivalent with theorem 1 for the marginal case.

Theorem 3. If $N_r \rightarrow \infty$ for $r = 1, \ldots, k - 1$

$$R_c = \sum_{r=1}^{k-1} N_r^{-1} \mathbf{d}_{\cdot r}' \hat{W}_{\cdot r}^{-1} \mathbf{d}_{\cdot r} \tag{11}$$

has an asymptotic χ^2 distribution with $(k - 1)(k - 2)$ degrees of freedom.

Appendix B gives a detailed proof of this theorem. As in the case of Theorem 1, it is convenient to have some rule of thumb to determine the degrees of freedom of the test. To give this rule, it must first be noticed that there is an important difference between the conditional model and the marginal model. In the first case the model is multinomial with parameters N and $\boldsymbol{\pi}$. In the conditional case the model is a product of $k - 1$ multinomial models with parameters N_r and $\boldsymbol{\pi}_{\cdot r}$, where $\boldsymbol{\pi}_{\cdot r}$ has elements as defined in (8). This gives rise to $k - 1$ additional restrictions which must be accounted for. So the degrees of freedom are now given by the number of deviances minus the number of parameters that have to be estimated, which is $k - 1$, minus the additional restrictions, so the degrees of freedom of R_c is $(k-1)(k-2)$. For the logic behind this rule one is referred to the derivation in Appendix B.

The statistical test proposed here can be generalized in much the same way as R_m, only a generalization to more than one population has no meaning, since the conditional likelihood is only function of the item parameters. As mentioned above Fischer (1981) has given a necessary and sufficient condition for the existence of a solution to the estimation equations and this condition also applies to a situation where different tests are administered to different groups. Among some more technical requirements, the tests must be linked by common items. If every selection of items that is administered to one particular group is called a test, common person equating is not different from common item equating, so the conditions given by Fischer and the generalization of R_c that will be given below also apply to this kind of test equating.

Let B_h be the index set of test h and let $B_{h'}$ be the index set of test h'. The tests indicated by h and h' are linked if there exists a sequence of index sets $B_{h_1}, B_{h_2}, \ldots, B_{h_z}$ such that $B_h \cap B_{h_1} \neq \emptyset$, $B_{h_1} \cap B_{h_2} \neq \emptyset, \ldots, B_{h_z} \cap B_{h'} \neq \emptyset$. This condition is necessary but not sufficient. However the more technical requirements specified by Fischer (1981) are in almost all instances fulfilled and if the sample size goes to infinity they are almost surely fulfilled. So let the design be such that a solution to the estimation equations exists. For every test h, a table with entries m_{rih} ($i = 1, \ldots, k_h$ and $r = 1, \ldots, k_h - 1$) is analyzed. Let N_{rh} be the number of persons attaining a score r on test h and let m_{rih} stand for the number of persons who attain a score r on test h and make item i, $i \in B_h$, correct. A vector of deviances is defined by $\mathbf{d}_{\cdot rh} = \mathbf{m}_{rh} - E(\mathbf{m}_{rh} \mid N_r, \hat{\boldsymbol{\varepsilon}})$, where $\mathbf{m}'_{rh} = (m_{r1h}, \ldots, m_{rih}, \ldots, m_{rk_hh})$ and $E(\mathbf{m}_{rh} \mid N_r, \hat{\boldsymbol{\varepsilon}})$ must be evaluated by (9) with the obvious substitution of arguments.

In the same manner, the definition of the $k_h \times k_h$ matrix $\hat{W}_{\cdot rh}$ is derived from the definition of $\hat{W}_{\cdot r}$ given above.

Theorem 3 can now be generalized in the following manner.

Theorem 4. Consider H tests which are linked by common items. If $N_{rh} \to \infty$ for $h = 1, \ldots, H$ and $r = 1, \ldots, k_h - 1$,

$$R_c = \sum_{h=1}^{H} \sum_{r=1}^{k_h - 1} N_{rh}^{-1} \mathbf{d}'_{\cdot rh} \hat{W}_{\cdot rh}^{-1} \mathbf{d}_{\cdot rh} \tag{12}$$

has an asymptotic χ^2 distribution with $\sum_{h=1}^{H} (k_h - 1)^2 - \dim(\boldsymbol{\varepsilon}) - 1$ degrees of freedom.

The proof of Theorem 4 is a straightforward generalization of the proof of Theorem 3. The degrees of freedom can be recovered by noticing that every test contributes $k_h(k_h - 1)$ deviances, $k_h - 1$ restrictions are imposed for every test by the product-multinomial form of the model and $\dim(\boldsymbol{\varepsilon}) - 1$ parameters have to be estimated.

As for the computation of the test, it must be noticed that the maximum number of items in one test, $\max_h (k_h)$, is the main restriction, because the dimension of $\hat{W}_{\cdot rh}$ and the maximal order of the elementary symmetric functions which must be computed are both bounded by $\max_h (k_h)$.

4. An Example

In order to clarify the technique suggested in this paper an example will be presented. The example concerns six arbitrarily chosen multiple choice items from an examination in

TABLE 1

Parameter Estimation For The MAVO-C Level

r	n_r	i	MML s_i	MML $\hat{\delta}_i$	MML $SE(\hat{\delta}_i)$	CML s_i^*	CML $\hat{\delta}_i$	CML $SE(\hat{\delta}_i)$
0	2							
1	20	1	289	-.801	.131	223	-.794	.141
2	47	2	289	-.801	.131	223	-.794	.141
3	61	3	176	.952	.110	110	.937	.121
4	77	4	189	.774	.111	123	.774	.122
5	87	5	230	.196	.113	164	.187	.123
6	66	6	263	-.320	.119	197	-.329	.129

N=360 $\hat{\mu}$= .900 $SE(\hat{\mu})$=.078

$\hat{\sigma}$=1.061 $SE(\hat{\sigma})$=.085

reading comprehension in English for two levels of Dutch secondary education, called MAVO-C and MAVO-D. For the analysis presented here, only a sample of the complete examination data was available, so $N = 360$ for MAVO-C and $N = 367$ for MAVO-D. Since the data were collected in the actual examination situation, it was assumed that guessing would not occur too frequently and that the Rasch model would be appropriate. The analysis presented here gives support to this assumption.

Table 1 and Table 2 give the results for the conditional and the marginal estimation procedure for both levels separately. The column marked N_r gives the frequency distribution of the sum scores r, the column marked s_i gives the number of correct responses to the items. Since the number of persons achieving a perfect score does not influence the conditional estimation procedure, a column marked s_i^* is added, which contains the number of correct responses to an item in the subgroup of examinees who did not achieve a perfect score. For the marginal case the estimation of the parameters was carried out by the method described by Thissen (1982) and Rigdon and Tsutakawa (1983), for the conditional case the techniques described by Fischer (1974) were applied.

The columns marked MML and CML give respectively the marginal and the conditional estimates. A normalization $\sum_i \delta_i = 0$ is chosen. At the bottom of both tables the estimated mean and variance of the ability distribution are displayed. Table 3 gives the observed and expected values of m_{ri} ($r = 1, \ldots, k - 1$ and $i = 1, \ldots, k$) for the marginal model. The last five columns of Table 3 are marked "scaled deviance". The entries are given by $(m_{ri} - E(m_{ri} \mid \hat{\lambda}))/\text{var}\,(m_{ri} \mid \hat{\lambda})^{1/2}$. The interpretation of the size of the "scaled deviance" may be aided by the fact that if only one item and one score level are considered and no parameters have to be estimated, the "scaled deviance" would be a standardized binomial variable.

TABLE 2

Parameter Estimation For The MAVO-D Level

			MML			CML		
r	n_r	i	s_i	$\hat{\delta}_i$	$SE(\hat{\delta}_i)$	s_i^*	$\hat{\delta}_i$	$SE(\hat{\delta}_i)$
0	1							
1	4	1	332	-.983	.170	219	-.980	.176
2	16	2	328	-.851	.168	215	-.849	.182
3	46	3	225	1.075	.113	112	1.077	.128
4	74	4	241	.850	.113	128	.850	.120
5	113	5	275	.323	.118	162	.318	.128
6	113	6	312	-.413	.115	199	-.415	.131

N=367 $\hat{\mu}$=1.629 SE($\hat{\mu}$)=.074

$\hat{\sigma}$= .986 SE($\hat{\sigma}$)=.086

Computation of the model test for the data in Table 3 resulted in $R_m = 35.6$ with $k(k-1) = 24$ degrees of freedom. Table 4 gives an evaluation of m_{ri} ($i = 1, \ldots, k$ and $r = 1, \ldots, k-1$) using conditional estimates. In this case $R_c = 18.3$ with $(k-1)(k-2) = 15$ degrees of freedom. The same computations were also carried out for the MAVO-D level, the results are summarized in Table 5, row one, two, five and six.

The model fit for the MAVO-D level is somewhat better that the model fit for the MAVO-C level. Inspection of Table 3 and Table 4 shows that the first item gives the largest scaled deviances. The estimation procedure was repeated using only the last five items, the results of the evaluation of the model fit are given in Table 5, rows three, four, seven and eight. For the MAVO-C level the model fit clearly improves: $R_m = 18.3$ with 15 degrees of freedom and $R_c = 12.2$ with 12 degrees of freedom. The next question which may be answered is whether the five remaining items form a Rasch homogeneous scale for both levels combined. Therefore marginal maximum likelihood estimates of the model parameters were made on the data of both levels simultaneously, under the assumption that both levels have normal ability distributions with different means and variances. Table 6 gives the estimation results in the column marked "MML-2", where μ_1 and σ_1 are the mean and variance for MAVO-C and μ_2 and σ_2 are the mean and the variance for MAVO-D. Table 5 gives the result of the R_m test in the row marked MML-2: $R_m = 33.0$ with 34 degrees of freedom.

Next it was assumed that both levels have the same ability distribution, so $\mu = \mu_1 = \mu_2$ and $\sigma = \sigma_1 = \sigma_2$. The row marked "MML-1" in Table 5 gives the result of the model test: $R_m = 76.0$ with 32 degrees of freedom, so the hypothesis that both levels have the same ability distribution must be rejected.

TABLE 3

Observed And Expected Frequencies m_{ri} For The MAVO-C Level, Evaluated Using Marginal M.L. Estimates

	observed					expected					scaled deviance				
r=	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
i															
1	7	31	46	65	74	5.50	20.87	45.22	71.07	83.50	.64	2.28	.14	-.80	-1.18
2	8	27	41	65	83	5.50	20.87	45.22	71.07	83.50	1.07	1.16	-.67	-.81	-.00
3	1	7	16	28	58	.95	4.88	14.92	34.06	58.49	.00	.96	.28	-1.09	.00
4	0	4	15	33	71	1.14	5.77	17.36	38.61	63.43	-1.07	-.74	-.58	-.95	1.05
5	0	12	26	54	72	2.03	9.76	27.18	53.81	74.54	-1.43	.73	-.24	.00	-.33
6	4	14	39	63	77	3.41	14.99	37.26	64.32	80.27	.32	-.26	.30	-.17	-.41

TABLE 4

Observed And Expected Frequencies m_{ri} For The MAVO-C Level, Evaluated Using Conditional M.L.Estimates

	observed					expected					scaled deviance				
r=	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
i															
1	7	31	46	65	74	5.91	25.33	44.11	65.67	81.83	.54	1.65	.53	.22	-3.55
2	8	27	41	65	83	5.91	25.33	44.11	65.67	81.83	1.02	.21	-.88	-.22	-.53
3	1	7	16	28	58	1.03	5.92	14.52	31.37	57.20	.00	.47	.45	-.78	.17
4	0	4	15	33	71	1.23	7.03	16.96	35.69	62.18	-1.14	-.24	-.57	-.62	2.09
5	0	12	26	54	72	2.22	11.98	26.72	46.96	73.20	-1.58	.00	-.17	.96	-.35
6	4	14	39	63	77	3.71	18.39	36.58	59.63	78.76	.17	-1.32	.63	.92	-.65

TABLE 5

Hypothesis Testing

method	population	items	R	df	crit.val.95%
MML	MAVO-C	1-6	35.62	24	36.4
	MAVO-D	1-6	14.88	24	36.4
	MAVO-C	2-6	18.27	15	25.0
	MAVO-D	2-6	14.21	15	25.0
CML	MAVO-C	1-6	27.51	20	31.4
	MAVO-D	1-6	12.95	20	31.4
	MAVO-C	2-6	12.18	12	21.0
	MAVO-D	2-6	11.72	12	21.0
MML-2	combined	2-6	33.05	34	45.9
MML-1	combined	2-6	76.05	32	48.3

5. Second Order Realizations

Van den Wollenberg (1982) has shown that statistical tests for the Rasch model which are based on analyzing first order realizations are in many instances insensitive to violation of the axiom of unidimensionality. The argument given by van den Wollenberg can be summarized in the following manner. Suppose unidimensionality is violated. If the subject's position on one latent trait is fixed, the assumption of local stochastic independence requires that the association between the items vanishes. In the case of more than one latent trait however, the subject's position in the latent space is not sufficiently described by one unidimensional ability parameter and as a consequence the association between the responses to the items given the ability parameter will not vanish. Therefore, van den Wollenberg (1982) proposed a test that focusses on the observed and the expected association between the items. The asymptotic distribution of the test can, as yet, not be derived, although simulation studies point in the direction of a χ^2 distribution. Practical application of the test also has its limitations, for its computation needs the conditional maximum likelihood estimation of the item parameters for every score level.

The present section will propose a model test which is based on the heuristic developed by van den Wollenberg and that can be derived by the same principle as used in the previous section: constructing a linear function of the frequencies of the response patterns to obtain deviances which may show the model violations under consideration. This is achieved in the following manner.

Let $\mathbf{m}^*$ be a $1/2k(k-1)$ dimensional vector with elements m^*_{ij}, $i = 1, \ldots, k-1$, $j = i+1, \ldots, k$, and let $\mathbf{n}^*$ be the vector of frequency counts of response patterns with a sum score $2 \leq r \leq k-1$, so $\mathbf{n}^{*\prime} = (\mathbf{n}'_2, \ldots, \mathbf{n}'_{k-1})$. The objective is to construct a matrix Y

TABLE 6

Parameter Estimation For The MAVO-C And MAVO-D levels Combined (Item 1 is removed)

	MML-2		MML-1		CML	
i	$\hat{\delta}_i$	$SE(\hat{\delta}_i)$	$\hat{\delta}_i$	$SE(\hat{\delta}_i)$	$\hat{\delta}_i$	$SE(\hat{\delta}_i)$
2	-1.030	.101	-1.028	.102	-1.016	.105
3	.858	.077	.858	.077	.860	.085
4	.650	.077	.649	.077	.647	.082
5	.077	.080	.075	.080	.065	.085
6	-.554	.085	-.555	.080	-.556	.088
	$\hat{\mu}_1$ = .723		$\hat{\mu}$ =1.126			
	$SE(\hat{\mu}_1)$= .082		$SE(\hat{\mu})$= .057			
	$\hat{\sigma}_1$ =1.230		$\hat{\sigma}$ =1.196			
	$SE(\hat{\sigma}_1)$= .095		$SE(\hat{\sigma})$= .068			
	$\hat{\mu}_2$ =1.439					
	$SE(\hat{\mu}_2)$= .076					
	$\hat{\sigma}_2$ =1.022					
	$SE(\hat{\sigma}_2)$= .096					

such that $\mathbf{m} = Y\mathbf{n}^*$. The reason for not including score level $r = 1$ and the exact definition of the elements of m^*_{ij} will become apparent in the sequel. Let Y be a matrix where every column is associated with a response pattern $\mathbf{x}$ leading to a sum score $2 \leq r \leq k - 1$. If $\mathbf{y}_\mathbf{x}$ is the column associated with $\mathbf{x}$, $\mathbf{y}_\mathbf{x}$ has elements

$$y_{\mathbf{x}(ij)} = \begin{cases} 1 & \text{if } x_i = 1 \text{ and } x_j = 1, \\ 0 & \text{for all other combinations of } x_i \text{ and } x_j. \end{cases}$$

Notice that if the same definition is also applied to a response pattern with a score $r = 1$, $\mathbf{y}_\mathbf{x} = 0$. Using $\mathbf{m}^* = Y\mathbf{n}^*$ it can be verified that m^*_{ij} is equal to the number of persons who obtained a sum score $2 \leq r \leq k - 1$ and made both item i and item j correct. The testing procedure will be developed for the marginal approach first.

Using $E(\mathbf{m}^* \mid \boldsymbol{\delta}, \sigma) = E(Y\mathbf{n}^* \mid \boldsymbol{\delta}, \sigma)$, it can be shown that $E(\mathbf{m}^* \mid \boldsymbol{\delta}, \sigma)$ has elements

$$E(m^*_{ij} \mid \boldsymbol{\delta}, \sigma) = N \sum_{r=2}^{k-1} \int_{-\infty}^{\infty} \varepsilon_i \varepsilon_j \gamma^{(i,j)}_{r-2}(\boldsymbol{\varepsilon}) \exp(r\vartheta) p_0(\vartheta) g(\vartheta \mid \sigma)\, \partial\vartheta, \tag{13}$$

so $E(m^*_{ij} \mid \boldsymbol{\delta}, \sigma) = N \sum_{r=2}^{k-1} \Psi_{rij}$ with Ψ_{rij} as defined in (3). Let $\hat{D}^*_\pi$ be a block-diagonal matrix diag $(\hat{D}_{\pi(2)}, \ldots, \hat{D}_{\pi(r)}, \ldots, \hat{D}_{\pi(k-1)})$ with $\hat{D}_{\pi(r)}$ as defined in section 2. If $\hat{U} = Y\hat{D}^*_\pi Y'$

it can be verified that U is a $1/2k(k-1) \times 1/2k(k-1)$ matrix with elements

$$u_{(ij)(i'j')} = \begin{cases} \sum_{r=2}^{k-1} \int_{-\infty}^{\infty} \varepsilon_i \varepsilon_j \varepsilon_{i'} \varepsilon_{j'} \gamma_{r-4}^{(iji'j')}(\boldsymbol{\varepsilon}) \exp(r\vartheta) p_0(\vartheta) g(\vartheta \mid \sigma)\, \partial\vartheta, & \text{if } i \neq i' \text{ and } j \neq j', \\ \sum_{r=2}^{k-1} \int_{-\infty}^{\infty} \varepsilon_i \varepsilon_j \varepsilon_{i'} \gamma_{r-3}^{(iji')}(\boldsymbol{\varepsilon}) \exp(r\vartheta) p_0(\vartheta) g(\vartheta \mid \sigma)\, \partial\vartheta, & \text{if } j' = i \text{ or } j' = j. \end{cases}$$

For the derivation of the asymptotic distribution of the model test all score levels have to be taken into account. So let d_0, $\mathbf{d}_1$ and $\mathbf{d}_k$ be the deviance for the score levels $r = 0$, $r = 1$ and $r = k$ as defined in section 2 for the marginal case. Suppose $\mathbf{f} = \mathbf{m}^* - E(\mathbf{m}^* \mid \hat{\boldsymbol{\delta}}, \hat{\sigma})$. Using these definitions the following theorem can be given.

Theorem 5. If $N \to \infty$,

$$R_{2m} = \frac{d_0^2}{E(n_0 \mid \hat{\boldsymbol{\delta}}, \hat{\sigma})} + N^{-1}\mathbf{d}_1'\hat{W}_1^{-1}\mathbf{d}_1 + N^{-1}\mathbf{f}'\hat{U}^{-1}\mathbf{f} + \frac{d_k^2}{E(n_k \mid \hat{\boldsymbol{\delta}}, \hat{\sigma})} \tag{14}$$

has an asymptotic χ^2 distribution with $1/2k(k-1)$ degrees of freedom.

The proof of this theorem generally follows the same lines as the proof of Theorem 1 as it is given in Appendix A, only some modifications need to be made. Appendix C indicates these modifications.

For the derivation of an analogous theorem for the conditional case, it is convenient to define a vector of frequencies of sum scores $\mathbf{N}' = (N_1, \ldots, N_r, \ldots, N_{k-1})$. If $\mathbf{m}^* = Y\mathbf{n}^*$ with Y as defined above, the conditional expectation $E(\mathbf{m}^* \mid \mathbf{N}, \boldsymbol{\varepsilon})$ is a vector with elements $\sum_{r=2}^{k-1} N_r \Gamma_{rij}$ where Γ_{rij} is defined by (10) in section 3.

Let Y be partitioned as $(Y_2, \ldots, Y_r, \ldots, Y_{k-1})$ and $\hat{U}_{..} = \sum_{r=2}^{k-1} N_r Y_r \hat{D}_{\pi \cdot r} Y_r'$ with $\hat{D}_{\pi \cdot r}$ as defined in section 3 for the conditional case. The elements of $\hat{U}_{..}$ can be recovered in the same manner as in the marginal case. Finally $\mathbf{f} = \mathbf{m}^* - E(\mathbf{m}^* \mid \mathbf{N}, \hat{\boldsymbol{\varepsilon}})$ and the definition of $\mathbf{d}_{.1}$ and $\hat{W}_{.1}$ remains unchanged. The following theorem is the conditional equivalent of Theorem 5.

Theorem 6. If for $r = 1, \ldots, k-1$, $N_r \to \infty$

$$R_{2c} = \mathbf{d}_1'\hat{W}_1^{-1}\mathbf{d}_1 + \mathbf{f}'\hat{U}_{..}^{-1}\mathbf{f}$$

has an asymptotic χ^2 distribution with $1/2k(k+1) - 2(k-1)$ degrees of freedom.

The proof of Theorem 6 runs along the same lines as the proof of theorem 5 given in Appendix C.

The computability of R_{2m} and R_{2c} is limited by the dimension of U and $U_{..}$. If $k = 15$ U and $U_{..}$ are 95×95 matrices, so 15 items must be considered as an upper bound for the number of items that can be analyzed in one run. A larger number of items would not only cause computational problems, the vector of "second order deviances" $\mathbf{f}$ would become too large to identify the results of individual items. Therefore a larger test must be split up in a number of subtests and these subtests must be analyzed separately. In the next section a simulated example of the use of the techniques presented here will be given.

6. A Simulated Example

In order to validate the claims concerning the power of the R_{2m} and R_{2c} test in cases of violation of the axiom of unidimensionality, a small simulation study was carried out.

TABLE 7

Hypothesis Testing For Data Generated Using One and Two Dimensions

	CML				MML			
	R_c	df	R_{2c}	df	R_m	df	R_{2m}	df
one dimension	19.34	20	11.42	11	18.95	24	13.95	15
two dimensions	28.44	20	112.18	11	31.94	24	138.56	15

The studies were replicated for both the marginal and the conditional approach and for several values of k and ε. Only the results for a test of six items will be presented here, because the results of the other simulation studies follow exactly the same pattern.

The studies were carried out under two conditions, in the first condition the items appealed to one latent trait, in the second condition the first three items appealed to one latent trait and the last three items appealed to another latent trait. The item parameters are given by $\varepsilon' = (2.0, 1.0, 0.5, 2.0, 1.0, 0.5)$. The data were generated in the following manner.

For the first condition a person parameter ϑ was drawn from the standardized normal distribution after which the probability $p(x_i = 1 \mid \vartheta, \delta_i)$ could be calculated using (1). A value ξ_i was drawn from the uniform distribution on (0, 1) and the response to item i was calculated according to the following rule:

$$\textbf{if } p(x_i = 1 \mid \vartheta, \delta_i) > \xi_i \textbf{ then } x_i = 1 \textbf{ else } x_i = 0.$$

This was repeated for all items.

For the second condition the procedure was the same, only in this case for every simulated response pattern two person parameters were drawn. The first three items appealed to the first person parameter and the last three items appealed to the second person parameter.

Table 7 gives the results for two typical simulation runs with $N = 4000$. The rows marked "one dimension" give the results of the hypothesis testing for a one dimensional data set, the rows marked "two dimensions" give the results for a two dimensional data set. It can be seen that in the last case the R_c and the R_m tests do not reject the model, while the R_{2c} and R_{2m} tests are indeed sensitive to the specific model violation that was initiated.

Table 8 gives the observed values of $\mathbf{m}^*$ and the expected values of $\mathbf{m}^*$ under the marginal model for the one dimensional data set. The column marked "scaled deviance" gives the values of $(m_{ij}^* - E(m_{ij}^* \mid \hat{\boldsymbol{\delta}}, \hat{\sigma}))/\text{var}\,(m_{ij}^* \mid \hat{\boldsymbol{\delta}}, \hat{\sigma})^{1/2}$. These values may aid the identification of items which violate the model. For every item i, this can be done by inspecting the scaled deviances indexed $j = 1, \ldots, i-1, i+1, \ldots, k$.

The simulation study described here far from exhausts all possible patterns of violation of the dimensionality axiom. In fact the simulation study was rather extreme, for in practical testing situations the data are almost never generated by two completely independent latent dimensions. Still, the results presented show that test statistics based on first order realizations are in some instances insufficient to test the fit of the Rasch model

TABLE 8

Evaluation of m_{ij}^* For The Marginal Model

i	j	observed	expected	scaled deviance
1	2	294	292.054	-.052
1	3	205	204.564	-.018
1	4	409	403.051	-.055
1	5	311	295.912	3.372
1	6	199	203.882	.786
2	3	158	156.959	.029
2	4	319	309.404	1.842
2	5	224	228.211	.600
2	6	149	156.420	1.111
3	4	211	216.727	1.149
3	5	168	159.101	1.052
3	6	112	106.688	.763
4	5	312	313.504	.241
4	6	223	216.007	1.373
5	6	158	158.556	.018

and it may be wise to supplement the testing procedure with a test based on the second order realizations.

7. Discussion

In the present paper a statistical testing procedure was proposed, which can serve both as an overall test of model fit and as a diagnostic tool for identifying violations of the Rasch model. The asymptotic distribution of the tests can be derived by making use of the well established theory of multinomial testing. Once the relation to this theoretical framework is recognized, a generalization of the testing procedure to incomplete designs is easily derived.

The first two tests, R_c and R_m, are sensitive to differences in the form of the item characteristic curves, so these tests have power against the two and the three parameter model. The other two tests, R_{2c} and R_{2m}, are based on the simultaneous realization of answers of a respondent to pairs of items and these tests have power against multidimensional models.

As a final remark it must be mentioned that the present paper is only one more contribution to the topic of testing the fit of the Rasch model, other well founded testing strategies exist (see for instance Andersen, 1971; and Kelderman 1984). This forms a striking contrast with the literature on the testing of model fit for the two and three parameter model. Workers in the field of the Rasch model are at advantage because the model is a member of the exponential family and sufficient statistics for the person parameters exist. Still item response theory would be helped a lot if also for the more complex models well founded testing procedures would be derived.

Appendix A: The proof of Theorem 1

The proof of Theorem 1 is closely related to the theory of the asymptotic distribution of Pearson's χ^2 as it is given by Bishop, Fienberg and Holland (1975) and Rao (1973). An important theorem which will be used in the sequel has been proved by Rao (1973, page 186 and following) and is summarized by Bishop, Fienberg and Holland (1975) in the Theorem 14.3-7 and 14.3-8 on page 473. For convenience it is reproduced below.

Theorem A. Let $\mathbf{a}$ be a s-dimensional stochastic variable which has a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \Sigma)$ and $Q = \mathbf{a}'V\mathbf{a}$ for some symmetric matrix V. Then the distribution function of Q is equal to the distribution function of $\sum_{i=1}^{s} \beta_i Z_i^2$, where $Z_1^2, \ldots, Z_s^2$ are independent χ^2 variables with one degree of freedom each and $\beta_1, \ldots, \beta_s$ are the eigenvalues of B, $B = V^{1/2}\Sigma V^{1/2}$. If $B^2 = B$, all the eigenvalues are either 0 or 1 and the number of degrees of freedom is equal to trace (B).

Before using this theorem, it is convenient to write R_m in a compact form. To achieve this the following notation is introduced. Let X be a $v \times t$ matrix ($v = 1/2k(k-1) + 2$ and t the number of possible response patterns) which is given by

$$X = \begin{bmatrix} x_0 & & & & & & \\ & X_1 & & & & 0 & \\ & & \ddots & & & & \\ & & & X_r & & & \\ & & & & \ddots & & \\ & 0 & & & & X_{k-1} & \\ & & & & & & x_k \end{bmatrix},$$

where X_r ($r = 1, \ldots, k-1$) is a matrix with as columns all response patterns leading to a score r, $x_0 = 1$ and $x_k = 1$. In section 2 the t-dimensional vector of frequency counts $\mathbf{n}$ was partitioned (n_0, $\mathbf{n}_1, \ldots, \mathbf{n}_{k-1}$, n_k). Let $\hat{\mathbf{p}} = \mathbf{n}/N$ and let $\boldsymbol{\pi}$ be the vector of theoretical probabilities. D_π is defined as a diagonal matrix of the elements of $\boldsymbol{\pi}$, $W = XD_\pi X'$ and $\mathbf{z} = N^{1/2}(\hat{\mathbf{p}} - \hat{\boldsymbol{\pi}})$. Then the model test R_m is given by $R_m = (X\mathbf{z})'\hat{W}^{-1}(X\mathbf{z})$.

Under certain regularity conditions (see Birch, 1964), which are easily shown to be fulfilled for the marginal Rasch model, for multinomial models in general the following expression holds:

$$N^{1/2}\begin{pmatrix} \hat{\mathbf{p}} - \boldsymbol{\pi} \\ \hat{\boldsymbol{\pi}} - \boldsymbol{\pi} \end{pmatrix} \xrightarrow[a]{} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} I_t \\ L \end{pmatrix}(D_\pi - \boldsymbol{\pi}\boldsymbol{\pi}')(I_t, L)\right), \tag{A1}$$

where $\rightarrow_a$ stands for convergence in distribution, I_t is a $t \times t$ identity matrix, $L = D_\pi^{1/2}A(A'A)^{-1}A'D_\pi^{-1/2}$ and A is a $t \times (k+1)$ matrix defined by $A = D_\pi^{-1/2}(\partial\boldsymbol{\pi}/\partial\boldsymbol{\lambda}')$, $\boldsymbol{\lambda}' =$

$(\delta_1, \ldots, \delta_k, \sigma)$. Using (A1) the asymptotic covariance matrix of $\mathbf{z}$ is given by

$$\underline{\Sigma} = D_\pi - \pi\pi' - (D_\pi - \pi\pi')L' - L(D_\pi - \pi\pi') + L(D_\pi - \pi\pi')L'. \tag{A2}$$

However $A'\pi^{1/2} = \mathbf{0}$, since $A'\pi^{1/2}$ is a vector with elements

$$\sum_{\{\mathbf{x}\}} \frac{\partial \pi_{\mathbf{x}}}{\partial \lambda_i} = \partial \sum_{\{\mathbf{x}\}} \frac{\pi_{\mathbf{x}}}{\partial \lambda_i} = 0$$

and so (A2) simplifies to

$$\Sigma = D_\pi - \pi\pi' - D_\pi^{1/2}A(A'A)^{-1}A'D_\pi^{1/2}. \tag{A3}$$

The asymptotic covariance matrix of $X\mathbf{z}$ is equal to $X\Sigma X'$, so from the theorem given at the beginning of this appendix it follows that R_m has an asymptotic χ^2 distribution if B, $B = W^{-1/2}X\Sigma X'W^{-1/2}$, is idempotent and the degrees of freedom can be derived by identifying trace (B). To achieve this, a number of propositions must be given first. Let $\mathbf{u}$ be defined by $\mathbf{u}' = (1, 1, 1, \ldots, 1)$ and let $\pi' = (\pi_0, \pi_1', \ldots, \pi_r', \ldots, \pi_{k-1}', \pi_k)$.

Proposition 1. $W_r\mathbf{u} = rX_r\pi_r$, for $r = 1, \ldots, k-1$.

Proof. In the main text, W_r was defined as $X_r D_{\pi(r)} X_r'$. Since $X_r'\mathbf{u} = r\mathbf{u}$ by the definition of X_r, $D_{\pi(r)}X_r'\mathbf{u} = rD_{\pi(r)}\mathbf{u} = r\pi_r$ and the result follows. □

Proposition 2. $W_r^{-1}X_r\pi_r = (1/r)\mathbf{u}$, for $r = 1, \ldots, k-1$.

Proof. Follows directly from Proposition 1. □

Proposition 3. Trace $(\pi'X'W^{-1}X\pi) = 1$.

Proof.

$$\sum_{r=1}^{k-1} \text{trace}\,(\pi'X_r'W_r^{-1}X_r\pi_r) + \pi_0 + \pi_k = \sum_{r=1}^{k-1} \frac{1}{r}\,\text{trace}\,(\pi_r'X_r'\mathbf{u}) + \pi_0 + \pi_k$$

by proposition 2, but the last expression is equal to $\sum_{\{\mathbf{x}\}} \pi_{\mathbf{x}}$ and the sum of all probabilities in the model in one. □

Proposition 4. $D_\pi^{1/2}X'W^{-1}XD_\pi^{1/2}A(A'A)^{-1}A' = A(A'A)^{-1}A'$.

Proof. Since $D_\pi^{1/2}X'W^{-1}XD_\pi^{1/2}$ has a block-diagonal form, which can be given by diag $(1, T_1, \ldots, T_r, \ldots, T_{k-1}, 1)$, with $T_r = D_{\pi(r)}^{1/2}X_r'W_r^{-1}X_rD_{\pi(r)}^{1/2}$, it is sufficient to show that $T_rD_{\pi(r)}^{1/2}(\partial\pi_r/\partial\lambda') = D_{\pi(r)}^{1/2}(\partial\pi_r/\partial\lambda')$. Since T_r is idempotent, it is a projection matrix and its manifold $M(T_r)$ is given by $M(D_{\pi(r)}^{1/2}X')$ [see for instance Rao, 1973, sec. 1c].

Using $D_{\pi(r)}^{1/2}\pi_r^{1/2} = \pi_r$ and proposition 2 it can be verified that $T_r\pi_r^{1/2} = \pi_r^{1/2}$. But if $\alpha_r = \exp(-X_r'\delta)$ and $g_r = \int \exp(r\vartheta)p_0(\vartheta)g(\vartheta \mid \sigma)\,\partial\vartheta$ it follows that $\pi_r = g_r\alpha_r$, so $\alpha_r^{1/2}$ is an element of $M(T_r)$.

This can be used in the following manner. Let $\beta(\vartheta)$ be a k-dimensional vector with elements $\exp(\vartheta - \delta_i)/(1 + \exp(\vartheta - \delta_i))$, $i = 1, \ldots, k$. Then it can be shown that

$$D_{\pi(r)}^{-1/2}\left(\frac{\partial \pi_r}{\partial \delta'}\right) = D_{\pi(r)}^{-1/2}\left(-D_{\pi(r)}X_r' + \int \alpha_r\beta(\vartheta)'\exp(r\vartheta)p_0(\vartheta)g(\vartheta \mid \sigma)\,\partial\vartheta\right)$$

$$= -D_{\pi(r)}^{1/2}X' + \alpha_r^{1/2}g_r^{-1/2}\int \beta(\vartheta)'\exp(r\vartheta)p_0(\vartheta)g(\vartheta \mid \sigma)\,\partial\vartheta.$$

So both the terms in the above summation are matrices composed of column-vectors belonging to $M(T_r)$.

In the same manner it can be shown that

$$D_{\pi(r)}^{-1/2}\left(\frac{\partial \pi_r}{\partial \sigma}\right) = D_{\pi(r)}^{-1/2}(\sigma^{-3} \int \vartheta^2 \alpha_r \exp (r\vartheta) p_0(\vartheta) g(\vartheta \mid \sigma)\, \partial\vartheta - \sigma^{-1}\pi_r)$$

$$= \alpha_r^{1/2} g_r^{-1/2} \sigma^{-3} \int \vartheta^2 \exp (r\vartheta) p_0(\vartheta) g(\vartheta \mid \sigma)\, \partial\vartheta - \alpha_r^{1/2} g_r^{1/2} \sigma^{-1},$$

which also belongs to $M(T_r)$. □

If $B_1 = W^{-1/2} X \pi\pi' X' W^{-1/2}$, $B_2 = W^{-1/2} X D_\pi^{1/2} A(A'A)^{-1} A' D_\pi^{1/2} X' W^{-1/2}$ and I_v a $v \times v$ identity matrix, $B = I_v - B_1 - B_2$.

To use Theorem A, it must be proved that B is idempotent.

Proposition 5. $B^2 = B$.

Proof. $(I_v - B_1 - B_2)^2 = (I_v - B_1 - B_2)$ because

(i). $B_1^2 = B_1$ as a consequence of $\pi' X' W^{-1} X \pi = 1$ (see Proposition 1),

(ii). $B_2^2 = B_2$ as a consequence of proposition 4,

(iii). $B_1 B_2 = 0$, which can be shown by noticing that $\pi' = \pi'^{1/2} D_\pi^{1/2}$ and using proposition 4 and $A'\pi^{1/2} = \mathbf{0}$ in that order. □

To determine the degrees of freedom of the χ^2 distribution of R_m, trace (B) must be identified.

Proposition 6. Trace $(B) = k(k - 2)$.

Proof.

(i). trace $(I_v) = v = k(k - 2) + 2$.

(ii). trace (B_1) = trace $(\pi' X' W^{-1} X \pi) = 1$ (see Proposition 2),

(iii). trace (B_2) = trace $(A(A'A)^{-1} A' D_\pi^{1/2} X' W^{-1} X D_\pi^{1/2})$ = (by Proposition 4) trace $(A(A'A)^{-1}A')$ = trace $((A'A)^{-1}(A'A))$ = k + 1. □

Theorem 1 can now be proved by using Theorem A.

From the well known fact that

$$\mathbf{z}'\hat{D}_\pi^{-1}\mathbf{z} \xrightarrow[a]{} \mathbf{z}' D_\pi^{-1} \mathbf{z}$$

(see for instance Bishop, Fienberg & Holland, 1975, p. 515) it follows that

$$R_m = (X\mathbf{z})' \hat{W}^{-1} (X\mathbf{z}) \xrightarrow[a]{} (X\mathbf{z})' W^{-1} (X\mathbf{z})$$

and R_m has an asymptotic χ^2 distribution because, by proposition 5, B is idempotent and the degrees of freedom are equal to $k(k - 2)$ by proposition 6.

This concludes the proof of Theorem 1.

Appendix B: The proof of Theorem 3

The proof of Theorem 3 generally follows the same lines as the proof of Theorem 1. Like the proof of Theorem 1, the proof will be based on Theorem A, given in Appendix A.

However, most of the elements used in appendix A need a slightly different definition. Let the vector of frequency counts n be re-defined as $(\mathbf{n}'_1, \ldots, \mathbf{n}'_r, \ldots, \mathbf{n}'_{k-1})$. The frequency of the pattern leading to a zero score and the frequency of the pattern leading to a perfect score are omitted. This is motivated by the fact that the conditional probability of these patterns equals one. Suppose $\hat{\mathbf{p}}' = (\mathbf{n}'_1/N_1, \ldots, \mathbf{n}'_r/N_r, \ldots, \mathbf{n}'_{k-1}/N_{k-1})$ and $\boldsymbol{\pi}'_{\cdot} = (\boldsymbol{\pi}'_{\cdot 1}, \ldots, \boldsymbol{\pi}'_{\cdot k-1})$, $\boldsymbol{\pi}_{\cdot r}$ has elements $\pi_{x \cdot r}$ as defined in (8). Using the results of Birch (1964), it can be shown that the distribution of $\mathbf{z}'_{\cdot} = (N_1^{1/2}(\hat{\mathbf{p}}_1 - \hat{\boldsymbol{\pi}}_{\cdot 1})', \ldots, N_{k-1}^{1/2}(\hat{\mathbf{p}}_{k-1} - \hat{\boldsymbol{\pi}}_{\cdot k-1})')$ converges to a multivariate normal distribution with expectation $\mathbf{0}$ and covariance matrix

$$\Sigma = V - LV - L'V + LVL', \tag{B1}$$

with V a block diagonal matrix diag $(V_1, \ldots, V_r, \ldots, V_{k-1})$, $V_r = D_{\pi \cdot r} - \boldsymbol{\pi}_{\cdot r}\boldsymbol{\pi}'_{\cdot r}$, $D_{\pi \cdot r}$ and $\boldsymbol{\pi}_{\cdot r}$ as defined in section 3. If $D_{\pi \cdot \cdot} = \text{diag}\ (D_{\pi \cdot 1}, \ldots, D_{\pi \cdot k-1})$, L is given by $L = D_{\pi \cdot \cdot}^{1/2} A(A'A)^{-1}A'D_{\pi \cdot \cdot}^{-1/2}$. Let A be partitioned $(A_1, \ldots A_r, \ldots A_{k-1})$. If the normalization $\delta_1 = 0$ is chosen and $\boldsymbol{\delta}' = (\delta_2, \ldots, \delta_k)$ it follows that $A_r = D_{\pi \cdot r}^{-1/2}(\partial \boldsymbol{\pi}_{\cdot r}/\partial \boldsymbol{\delta}')$. However $A'_r \boldsymbol{\pi}_r^{1/2} = \mathbf{0}$, so (B1) simplifies to

$$\Sigma = D_{\pi \cdot \cdot} - H_{\pi \cdot \cdot} - D_{\pi \cdot \cdot}^{1/2} A(A'A)^{-1}AD_{\pi \cdot \cdot}^{1/2}, \tag{B2}$$

with $H_{\pi \cdot \cdot}$ a block diagonal matrix diag $(\boldsymbol{\pi}_1\boldsymbol{\pi}'_1, \ldots, \boldsymbol{\pi}_r\boldsymbol{\pi}'_r, \ldots, \boldsymbol{\pi}_{k-1}\boldsymbol{\pi}'_{k-1})$. Let X be defined as in appendix A, only this time x_0 and x_k are omitted. Then R_c can be written as $R_c = (X\mathbf{z}_{\cdot})'\hat{W}^{-1}(X\mathbf{z})$ with $W = XD_{\pi \cdot \cdot}X'$. To use Theorem A it must be shown that B, $B = W^{-1/2}X\Sigma X'W^{-1/2}$, is idempotent. The propositions needed will be indexed by the same number as the analogous propositions for the marginal case and proofs will only be given if they cannot be easily recovered from their marginal counterpart.

Proposition 1'. $W_{\cdot r}\mathbf{u} = rX_r\boldsymbol{\pi}_{\cdot r}$ for $r = 1, \ldots, k-1$.

Proposition 2'. $W_{\cdot r}^{-1}X_r\boldsymbol{\pi}_{\cdot r} = (1/r)\mathbf{u}$ for $r = 1, \ldots, k-1$.

Proposition 3'. Trace $(\boldsymbol{\pi}'_{\cdot r}X'_r W_{\cdot r}^{-1}X_r\boldsymbol{\pi}_{\cdot r}) = 1$ for $r = 1, \ldots, k-1$.

Proposition 4'. $D_{\pi \cdot \cdot}^{1/2}X'W^{-1}XD_{\pi \cdot \cdot}^{1/2}A(A'A)^{-1}A' = A(A'A)^{-1}A'$.

Proof. $D_{\pi \cdot \cdot}^{1/2}X'W^{-1}XD_{\pi \cdot \cdot}^{1/2}$ has a block-diagonal form, diag $(T_1, \ldots, T_r, \ldots, T_{k-1})$ with $T_r = D_{\pi \cdot r}^{1/2}X'_r W_{\cdot r}^{-1}X_r D_{\pi \cdot r}^{1/2}$. Using Proposition 2' it can be shown that $T_r\boldsymbol{\pi}_{\cdot r} = \boldsymbol{\pi}_{\cdot r}$. If $\boldsymbol{\delta}' = (\delta_1, \ldots, \delta_k)$, some computational labour will show that $\partial \boldsymbol{\pi}_{\cdot r}/\partial \boldsymbol{\delta}' = -D_{\pi \cdot r}X'_r + \boldsymbol{\pi}_{\cdot r}\boldsymbol{\pi}'_{\cdot r}X'$. So $D_{\pi \cdot r}^{-1/2}(\partial \boldsymbol{\pi}_{\cdot r}/\partial \boldsymbol{\delta}')$ has columns belonging to $M(T_r)$ and the same goes for $D_{\pi \cdot r}^{-1/2}(\partial \boldsymbol{\pi}_{\cdot r}/\partial \boldsymbol{\lambda}')$ with $\boldsymbol{\lambda}' = (\delta_1, \ldots, \delta_{k-1})$. So the projection $D_{\pi \cdot \cdot}^{1/2}X'W^{-1}XD_{\pi \cdot \cdot}^{1/2}$ will map A onto A. □

If $B_1 = W^{-1/2}XH_\pi X'W^{-1/2}$, $B_2 = W^{-1/2}XD_{\pi \cdot \cdot}^{1/2}A(A'A)^{-1}A'D_{\pi \cdot \cdot}^{1/2}X'W^{-1/2}$ and I_v is a $v \times v$ identity matrix, $v = k(k-1)$, B is given by $I_v - B_1 - B_2$.

Proposition 5'. $B^2 = B$.

Proof.

(i). $B_1^2 = B_1$ because B_1 is a block diagonal matrix diag $(W_{\cdot r}^{-1/2}X_r\boldsymbol{\pi}_{\cdot r}\boldsymbol{\pi}'_{\cdot r}W_{\cdot r}^{-1/2})$ and all matrices on the diagonal are idempotent,

(ii). $B_2^2 = B_2$ as a consequence of Proposition 4',

(iii). $B_1B_2 = 0$ as a consequence of Proposition 4'. □

Proposition 6'. Trace $(B) = (k-1)(k-2)$.

Proof.

(i). trace $(I_v) = k(k-1)$,

(ii). trace $(B_1) = \text{trace}\,(X'W^{-1}XH_\pi) = \sum_{r=1}^{k-1} \text{trace}\,(X'_r W_{\cdot r}^{-1} X_r \pi_{\cdot r} \pi'_{\cdot r})$

$$= \sum_{r=1}^{k-1} \mathbf{u}'\pi_{\cdot r} = k-1,$$

(iii). trace $(B_2) = k-1$. □

Using Theorem A of Appendix A, R_c has an asymptotic χ^2 distribution as a consequence of Proposition 5. From Proposition 6 it follows that the test has $(k-1)(k-2)$ degrees of freedom.

Appendix C. Proof of Theorem 5

Let Y be partitioned $(Y_2, \ldots, Y_r, \ldots, Y_{k-1})$ Suppose $c_r = \sum_{h=1}^{r-1} h$. Then every column in the matrix Y_r $(r = 2, \ldots, k-1)$ has c_r elements equal to one and all other elements are zero. This will be used in the sketch of the proof of Theorem 5 given below. The numbers of the propositions correspond with the analogous propositions given in Appendix A. It will prove convenient to introduce $Y_1 = X_1$.

Proposition 1″. $W_r \mathbf{u} = c_r Y_r \pi_r$ for $r = 1, \ldots, k-1$.

Proposition 2″. $W_r^{-1} Y_r \pi_r = (1/c_r)\mathbf{u}$ for $r = 1, \ldots, k-1$.

Let Y^* be given by

$$\begin{bmatrix} y_0 & & & 0 \\ & Y_1 & & \\ & & Y_2, \ldots, Y_{k-1} & \\ 0 & & & y_k \end{bmatrix},$$

with $y_1 = y_k = 1$. If X is replaced by Y^* in the Propositions 3, 4, and 5 in Appendix A, these propositions are still valid. Only Proposition 6 needs some modification. If $B = W^{-1/2}Y^*\Sigma Y^{*\prime}W^{-1/2}$, this proposition must read: trace $(B) = 1/2k(k-1)$. This concludes the sketch of the proof.

References

Andersen, E. B. (1971). The asymptotic distribution of conditional likelihood ratio tests. *Journal of the American Statistical Association, 66*, 630–633.

Andersen, E. B. (1973). *Conditional inference and models for measuring.* Copenhagen: Mentalhygiejnisk Forskningsinstitut.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* Cambridge, MA: MIT Press.

Birch, M. W. (1964). A new proof of the Pearson-Fisher theorem. *Annals of Mathematical Statistics, 35*, 817–824.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika, 46*, 443–459.

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests* [*Introduction to the theory of psychological tests*]. Bern: Verlag Hans Huber.

Fischer, G. H. (1981). On the existence and uniqueness of maximum likelihood estimates in the Rasch model. *Psychometrika, 46*, 59–77.

Kelderman, H. (1984). Loglinear Rasch model tests. *Psychometrika, 49*, 223–245.

Martin Löf, P. (1973). *Statistika Modeller. Anteckningar från seminarier Lasåret 1969–1970 utarbetade av Rolf Sunberg obetydligt ändrat nytryk, oktober 1973 [Statistical models. Proceedings of the Lasåret seminar 1969–1970, edited by Rolf Sunberg]*. Stockholm: Institutet för Försäkringsmatematik och Matematisk Statistik vid Stockholms Universitet.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika, 16*, 1–32.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications* (2nd ed.). New York: Wiley.

Rasch, G. (1960). *Probablistic models for some intelligence and attainment tests.* Kopenhagen: Danish Institute for Educational Research.

Rasch, G. (1961). On the general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), *Proceedings of the Fourth Symposium on Mathematical Statistics and Probability, 4*, 321–333.

Rigdon, S. E., & Tsutakawa, R. K. (1983). Parameter estimation in latent trait models. *Psychometrika, 48*, 567–574.

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

van den Wollenberg, A. L. (1982). Two new test statistics for the Rasch model. *Psychometrika, 47*, 123–140.

Verhelst, N. D., Glas, C. A. W., & van der Sluis, A. (1984). Estimation problems in the Rasch model: The basic symmetric functions. *Computational Statistics Quarterly, 1*, 245–262.

Manuscript received 11/12/86
Final version received 8/13/87