*Journal of Educational Statistics*
*Spring 1988, Vol. 13, No. 1, pp. 45–52*

# The Rasch Model and Multistage Testing

**C.A.W. Glas**
*National Institute for Educational Measurement (Cito)*

Key words: *Rasch model, latent trait theory, multistage testing*

*This paper concerns the problem of estimating the item parameters of latent trait models in a multistage testing design. It is shown that using the Rasch model and conditional maximum likelihood estimates does not lead to solvable estimation equations. It is also shown that marginal maximum likelihood estimation, which assumes a sample of subjects from a population with a specified distribution of ability, will lead to solvable estimation equations, both in the Rasch model and in the Birnbaum model.*

As has been pointed out by Lord (1980) two-stage and multistage testing procedures bear the possibility of improving the measurement of a subject's ability, if the number of items that can be administered is fixed, and we have a large pool of items at our disposal. The same holds for calibrating an item pool: if the difficulty of the test items is matched with the ability level of the different groups of subjects, estimation errors of the item parameters can be reduced.

In order to get an impression of the ability level of the subjects, a so-called routing test is administered first. In two-stage testing the performance on the routing test governs the choice of the next test to be administered: the better the performance, the more difficult the follow-up test. In multistage testing this procedure is generalized in the sense that more tests are subsequently administered and that the next test to be administered depends on the performance on the previous tests. Tailored testing can be viewed as a limiting case of this principle, where each subtest consists of one item only.

In a so-called incomplete design with common items, that is, a design where different samples of subjects take different, though overlapping, subtests, two possibilities are open to calibrate the complete set of item parameters on one scale. The first possibility starts with carrying out estimation in each sample separately, to arrive at disjunct scales. These scales are then combined to a common scale by transforming the estimates of the item parameters. The transformation, which must belong to the family of admissible transformations of the item response model under consideration, is such that the difference in the estimates of the parameters of common items is minimized. For the Rasch model this method is described in detail by Wright and Stone (1979). The second possibility, the one pursued here, is to maximize the likelihood of the total data set over all parameters simultaneously. The advantages of this approach are several. Given a maximum likelihood estimate of all item parameters on a common scale, asymptotic confidence intervals of estimates as well as the asymptotic distribution of tests of model fit can be derived. I will return to this point later.

The implications of collecting data in a multistage testing design on maximum likelihood estimation will be studied in the context of the Rasch model for dichotomous items, although it will be indicated that some findings apply to other models also.

Let the vector $x$ be a response pattern and let $\Psi$ be the set of all possible response patterns, given a certain test administration strategy. Let $n(x)$ be the number of persons obtaining response pattern $x$ and let $n$ be the vector of frequency counts for all possible patterns. Since every person is associated with one response pattern only, $n$ has a multinomial distribution with parameters N and $P(x|\lambda)$, where $\lambda$ is a vector of parameters. The object of this paper is to study maximum likelihood estimation, so $\Psi$ must be identified and a proper specification of $P(x|\lambda)$ must be found. Consider the example of a routing test and two follow-up tests. Response pattern $x$ will be partitioned $(x_1, x_2, x_3)$, where $x_1$ is the response pattern on the routing test; $x_2$ and $x_3$ will be defined later.

If the routing test consists of $m_1$ items, $x_1$ is a $m_1$-dimensional vector with elements

$$x_i = \begin{cases} 0 \text{ if item } i \text{ has been made incorrect,} \\ 1 \text{ if item } i \text{ has been made correct,} \end{cases} \tag{1}$$

for $i = 1, \ldots, m_1$.

Suppose a person gets the first follow-up test if the number of correct responses on the routing test $t_1(x_1)$, $t_1(x_1) = \sum_{i=1}^{m_1} x_i$, is less than a cut-off score $\alpha$. Otherwise the second follow-up test is administered. If the first follow-up test has $m_2$ items and the second one $m_3$ items, an administration function $d_i$, $i = 1, \ldots, m_1 + m_2 + m_3$, from $t_1(x_1)$ to $\{0, 1\}$ is defined by

$$d_i = \begin{cases} 1 \text{ for } i = 1, \ldots, m_1, \\ 1 \text{ for } i = m_1 + 1, \ldots, m_2 \text{ if } t_1(x_1) < \alpha, \\ 0 \text{ for } i = m_2 + 1, \ldots, m_3 \text{ if } t_1(x_1) < \alpha, \\ 1 \text{ for } i = m_2 + 1, \ldots, m_3 \text{ if } t_1(x_1) \geqq \alpha, \\ 0 \text{ for } i = m_1 + 1, \ldots, m_2 \text{ if } t_1(x_1) \geqq \alpha. \end{cases} \tag{2}$$

Obviously $d_i$ equals one if item i is responded to and zero otherwise. Let $x_2$ be a $m_2$-dimensional vector associated with the first, and $x_3$ a $m_3$-dimensional vector associated with the second follow-up test. The definition of the elements of $x$, given in (1) can now be generalized in the following manner:

$$x_i = \begin{cases} 0 \text{ if } d_i = 1 \text{ and item } i \text{ has been made incorrect,} \\ 1 \text{ if } d_i = 1 \text{ and item } i \text{ has been made correct,} \\ \beta \text{ if } d_i = 0, \end{cases} \tag{3}$$

for $i = 1, \ldots, m, m = m_1 + m_2 + m_3$, and $\beta$ is an arbitrary constant unequal to zero or one.

If an item is presented, it is assumed that the Rasch model holds, so the probability of observing $x_i, x_i \in \{0, 1\}$ is given by

$$P(x_i \mid d_i = 1, \xi, \delta_i) = \frac{\exp(x_i(\xi - \delta_i))}{1 + \exp(\xi - \delta_i)} \tag{4}$$

where $\delta_i$ is the item difficulty and $\xi$ the ability parameter. If the item is not presented,

$$P(x_i \mid d_i = 0) = \begin{cases} 1 \text{ if } x_i = \beta, \\ 0 \text{ if } x_i \neq \beta. \end{cases} \tag{5}$$

In this papr two approaches to estimating item parameters in the Rasch model, using data collected in a multistage testing design, will be studied. First an attempt will be made to develop a conditional maximum likelihood (CML) estimation procedure, which will prove to suffer from certain incurable flaws. Secondly it will be shown that marginal maximum likelihood (MML) estimation procedures do not suffer from these drawbacks.

## The CML Approach

The CML estimation equations for the item parameters in an incomplete design have been derived by Fischer (1981). Fischer also presented a necessary and sufficient condition for the existence of a solution to the CML estimation equations and showed that the estimates are unique.

The techniques presented by Fischer cover the situation where the test administration is a priori fixed. In a multistage testing design however, the test administration is governed by the subject's performance. In consequence of this, the set of possible response patterns in the last case is only a subset of the response patterns possible in the first situation.

Consider the two-stage testing situation described above with $m_1 = m_2 = m_3 = 4$ and a cut-off score $\alpha = 2$. Response vector $(1, 0, 0, 0, \beta, \beta, \beta, \beta, 1, 1, 0, 0)$ is not possible, since a person with one item correct on the routing test cannot get the second follow-up test. The way these restrictions on the sample space modify the CML estimation equations given by Fischer (1981) for a fixed incomplete design will be shown.

It is convenient to introduce the transformation $\theta = \exp(\xi)$ and $\epsilon_i = \exp(-\delta_i)$. The probability of response pattern $x$ as a function of $\theta$ and the item parameters is given by

$$P(x|\theta,\epsilon)=\frac{\theta^{t}\prod_{i=1}^{m}\epsilon_i^{d_i x_i}}{\prod_{i=1}^{m}(1+\theta\epsilon_i)^{d_i}} \tag{6}$$

with $t=\sum_{i=1}^{m} d_i x_i$ and $\epsilon$ has elements $\epsilon_i, i=1,\ldots,m$.

In an a priori fixed incomplete design, conditioning on t would lead to a probability $P(x|t,\epsilon)$ which is independent of $\theta$. This does not work in the present situation. Consider an example with $m_1=m_2=m_3=1$; the second item is given if the first one is incorrect and the third item is given if the first one is correct. So the possible response patterns are $(0,0,\beta)$, $(0,1,\beta)$, $(1,\beta,0)$, and $(1,\beta,1)$. The conditional probability of $(1,\beta,0)$ is equal to

$$\frac{P((1,\beta,0)|\theta,\epsilon)}{P(t=1|\theta,\epsilon)}=\frac{\dfrac{\theta\epsilon_1}{(1+\theta\epsilon_1)(1+\theta\epsilon_3)}}{\dfrac{\theta\epsilon_1}{(1+\theta\epsilon_1)(1+\theta\epsilon_3)}+\dfrac{\theta\epsilon_2}{(1+\theta\epsilon_1)(1+\theta\epsilon_2)}}.$$

It can be seen that the factor $(1+\theta\epsilon_1)(1+\theta\epsilon_3)$ does not cancel and the conditional probability of $(1,\beta,0)$ will not be independent of $\theta$. The problem is solved by conditioning on the sum score for every subtest separately. Let $t_h$ be the sum score on subtest h, $h=1,\ldots,3$, so if $I_h$ is the set of the item indices of subtest h, $t_h=\sum_{i\in I_h} d_i x_i$. Using (6) the joint probability of the sum scores on the three subtests can be shown to be

$$P(t_1,t_2,t_3|\theta,\epsilon)=\frac{\gamma_{t_1}(\epsilon_1)\gamma_{t_2}(\epsilon_2)\gamma_{t_3}(\epsilon_3)\theta^{t}}{\prod_{i=1}^{m}(1+\theta\epsilon_i)^{d_i}}. \tag{7}$$

In (7) $\gamma_{t_h}(\epsilon_h)$ stands for an elementary symmetric function of order $t_h$ with, as parameters, the elements of $\epsilon_h$, a vector of item parameters associated with subtest h. For a definition of an elementary symmetric function see Rasch (1960) or Fischer (1981). It is important to notice that $\gamma_{t_h}(\epsilon_h)=1$ if $t_h=0$ and $\gamma_{t_h}(\epsilon_h)=0$ if $t_h<0$.

From (6) and (7) it follows that the likelihood of the vector of frequency counts $n$ conditional on the vectors of sum scores $t_1, t_2, t_3$ is given by

$$L_c(n|t_1,t_2,t_3,\epsilon)\propto\prod_{x}\left[\frac{\prod_{i=1}^{m}\epsilon_i^{d_i x_i}}{\prod_{h=1}^{3}\gamma_{t_h}(\epsilon_h)}\right]^{n(x)} \tag{8}$$

("$\propto$" stands for "proportional to").

For items of the routing test, the equations maximizing (8) as a function of $(\epsilon_1)$ are given by

$$s_i=\sum_{v=1}^{N}\frac{\epsilon_i\gamma^{(i)}_{t_{1v}-1}(\epsilon_1)\gamma_{t_{2v}}(\epsilon_2)\gamma_{t_{3v}}(\epsilon_3)}{\prod_{h=1}^{3}\gamma_{t_{hv}}(\epsilon_h)}, \tag{9}$$

where $i = 1, \ldots, m_1$. In (9) $s_i$ stands for the number of correct responses given to item i, the summation on the right hand side is over respondents, and $t_{hv}$ stands for the sum score of person v on subtest h. The factor $\gamma^{(i)}_{t_{1v}-1}(\epsilon_1)$ stands for an elementary symmetric function of order $t_{1v} - 1$, where $\epsilon_i$ has been set equal to zero. Canceling $\gamma_{t_{2v}}(\epsilon_2)$ and $\gamma_{t_{3v}}(\epsilon_3)$ in (9) gives

$$s_i = \sum_{v=1}^{N} \frac{\epsilon_i \gamma^{(i)}_{t_{1v}-1}(\epsilon_1)}{\gamma_{t_{1v}}(\epsilon_1)}, \tag{10}$$

where $i = 1, \ldots, m_1$. In the same manner the CML estimation equations associated with $\epsilon_2$ and $\epsilon_3$ are given by

$$s_{i'} = \sum_{v=1}^{N} \frac{\epsilon_{i'} \gamma^{(i')}_{t_{2v}-1}(\epsilon_2)}{\gamma_{t_{2v}}(\epsilon_2)}, \tag{11}$$

where $i' = m_1 + 1, \ldots, m_2$ and

$$s_{i''} = \sum_{v=1}^{N} \frac{\epsilon_{i''} \gamma^{(i'')}_{t_{3v}-1}(\epsilon_3)}{\gamma_{t_{3v}}(\epsilon_3)}, \tag{12}$$

where $i'' = m_2 + 1, \ldots, m_3$. Inspection of (10), (11), and (12) shows that the estimation equations for the three subtests are independent, that is, the parameters $\epsilon_2$ and $\epsilon_3$ do not figure in equations (10), etc. So each set of equations can only be solved by choosing a separate normalization for the item parameters for every single subtest and the result does not calibrate all item parameters on the same latent continuum.

Shifting the three scales in the manner described in the opening of this paper will not work either because no common items exist. Also carrying out CML estimation in both subgroups separately and equating by shifting both scales using the estimates of the items of the routing test will not work, for the same reasons as those given above: each sample restriction on the sample space must be taken into account, and CML estimates for the item parameters of the routing test and the item parameters of a follow-up test simultaneously, do not exist. Obviously more assumptions are needed to achieve this. In the next section it will be shown that adopting the "random-effects" Rasch model is sufficient.

## The MML Approach

In the marginal latent trait models, it is assumed that subjects are a random sample from a population in which ability $\xi$ is distributed with density $g(\xi|\tau)$, where $\tau$ is a vector of so-called population parameters (Bock & Lieberman, 1970). A method for estimating both item and population parameters has been given by Bock and Aitkin (1981). Mislevy (1984) pointed out that this method can also be applied for simultaneously estimating all parameters in an incomplete design.

The objective here is to show that restrictions on the set of possible response patterns due to the multistage testing design will not lead to estimation equations and asymptotic confidence intervals which differ from those derived for an a priori fixed design.

The probability of a response pattern can be derived in the following manner. Consider the same example as given above: a routing test and two follow-up tests. Again response pattern $x$ is partitioned $(x_1, x_2, x_3)$, and the vector of item parameters $\delta$ is partitioned $(\delta_1, \delta_2, \delta_3)$ in the same way. Because of the dependence between the test administration and the responses on the routing test,

$$\begin{aligned} P(x|\delta, \tau) &= P(x_1, x_2, x_3|\delta, \tau) \\ &= P(x_1|\delta_1, \tau)P(x_2, x_3|x_1, \delta, \tau). \end{aligned} \tag{13}$$

Since it is assumed that the respondents are a random sample from a distribution $g(\xi|\tau)$, the first factor is given by

$$P(x_1|\delta_1, \tau) = \int_\Omega P(x_1|\delta_1, \xi)g(\xi|\tau)\partial\xi, \tag{14}$$

where $\Omega$ is the range of $\xi$.

The second factor is given by

$$P(x_2, x_3|x_1, \delta, \tau) = \int_\Omega P(x_2, x_3|x_1, \delta, \xi)h(\xi|x_1, \delta_1, \tau)\partial\xi$$

or due to local independence

$$P(x_2, x_3|x_1, \delta, \tau) = \int_\Omega P(x_2, x_3|\delta, \xi)h(\xi|x_1, \delta_1, \tau)\partial\xi \tag{15}$$

where $h(\xi|x_1, \delta_1, \tau)$ stands for the distribution of ability given the response pattern on the routing test, so

$$h(\xi|x_1, \delta_1, \tau) = \frac{P(x_1|\delta_1, \xi)g(\xi|\tau)}{\int_\Omega P(x_1|\delta_1, \xi)g(\xi|\tau)\partial\xi}. \tag{16}$$

Inserting (14), (15), and (16) in (13), it follows that

$$P(x|\delta, \tau) = \int_\Omega P(x|\delta, \xi)g(\xi|\tau)\partial\xi \tag{17}$$

and this probability has the same form as the marginal probability obtained in an experiment where the design is fixed a priori.

As a consequence of the validity of (17) in a multistage testing design, the computational devices for estimation, such as the EM algorithm (see Rigdon & Tsutakawa, 1983), can be used directly. Because of the multinomial form of the model, it is possible to make use of the well-established framework of multinomial estimation and testing (Bishop, Fienberg, & Holland, 1975) and expressions for asymptotic confidence intervals as given by Bock and Aitkin (1981) and Mislevy (1984) remain valid. Also, the test of fit proposed by these authors remains valid, although the number of degrees of freedom drops, due to the smaller number of response patterns possible.

The way that dependencies between subsets of a response pattern are handled in equations (14) through (17) is not limited to the dichotomous Rasch model. If for instance $\delta_h$ is redefined as a vector of item discrimination parameters and item difficulties associated with the two-parameter model (Birnbaum, 1968), the reasoning developed in equations (14) through (17) still holds. Also the fact that the response to an item is dichotomously scored is not essential; the reasoning applies to polychotomous items also. Finally it must be noticed that the exact test administration function is irrelevant. In the examples a test administration on the basis of sumscore $t_1$ was used, but it may for instance also be possible to administer subtests on the basis of other characteristics of the response pattern on the routing test. In all these instances, computational methods based on the assumption of an ability distribution can be used directly, whereas when using methods based on conditioning on sufficient statistics for the ability paremeters, the restrictions on the sample space have to be taken into account.

## Concluding Remarks

In this paper two methods for estimating the item parameters in a multistage testing design are discussed, the CML method and the MML method. In the usual analysis of an incomplete design, the CML estimation method is generally more powerful than the MML estimation method, because no assumptions are made with respect to the distribution of ability. So when using the CML procedure, lack of model fit due to improper modelling of the ability distribution cannot occur. However, in a multistage testing design, the CML estimation method breaks down. Not only does the estimation of all item parameters simultaneously not work, estimating the parameters in the different subgroups and calibrating the estimates on one scale via the common items is impossible. As for the MML estimation procedure, it was shown that the likelihood of the data can be written in a multinomial form, with response patterns as categories. Furthermore the specification of the probability of a response pattern is identical to the specification in latent trait models with a fixed design. Therefore the theory of estimation and testing can be directly applied to the designs discussed here.

## References

Birnbaum, A. (1968). Test scores, sufficient statistics and the information structures of tests. In F.M. Lord & M.R. Novick (Eds.), *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

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

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item

parameters: An application of an EM algorithm. *Psychometrika, 46,* 443–459.
Bock, R.D., & Lieberman, M. (1970). Fitting a response model for dichotomously scored items. *Psychometrika, 35,* 179–197.
Fischer, G.H. (1981). On the existence and uniqueness of maximum likelihood estimates in the Rasch model. *Psychometrika, 46,* 59–77.
Lord, F.M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum.
Mislevy, R.J. (1984). Estimating latent distributions. *Psychometrika, 49,* 375–381.
Rasch, G. (1960). *Probablistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research.
Wright, B.D., & Stone, M.H. (1979). *Best test design.* Chicago: MESA Press.

## Author

C.A.W. GLAS, National Institute for Educational Measurement (Cito), PO Box 1034, Arnhem, The Netherlands. *Specialization:* psychometrics.