Correcting Unconditional Parameter Estimates in the Rasch Model for Inconsistency

Paul G. W. Jansen
The Netherlands Telecommunications Services

Arnold L. van den Wollenberg
University of Nijmegen

Folkert W. Wierda
Postbank

Results of simulation studies indicate that the unconditional maximum likelihood method is commonly regarded as an appropriate substitute for the theoretically superior conditional method for estimating the parameters of the Rasch model. To this end, the unconditional estimates are "corrected" by a factor $(K-1)/K$, where K is the number of items. In this paper, the simulation study of Wright and Douglas (1977b), which seemed to corroborate this correction term, is critically discussed. It appears to contain a puzzling assumption, and to rest on inadequate logic. Accordingly, there is a need for new simulation studies on the validity of the correction term $(K-1)/K$ for unconditional maximum likelihood estimation in the Rasch model. *Index terms: Item response theory, item parameter estimation; Item response theory, one-parameter logistic model; Maximum likelihood estimation, unconditional; One-parameter logistic model; Rasch model.*

The Rasch model (Rasch, 1960/1980) has a number of attractive properties, including consistency of estimators (Andersen, 1973b), model tests with known and desirable properties (Andersen, 1973a; Gustafsson, 1980b; Molenaar, 1983; van den Wollenberg, 1982), and formalization of a specific type of "objective" measurement (Rasch, 1966, 1977; Roskam & Jansen, 1984), which are, however, only achievable when its parameters are estimated by conditional maximum likelihood (CML). This type of estimation may encounter some technical difficulties, however, and therefore procedures of "ordinary" maximum likelihood have been proposed and applied.

Item estimators obtained under the latter procedure have been proven to be inconsistent (Andersen, 1973b); it seems, however, that the bias can be corrected for by the term $(K-1)/K$, where K is the number of items. Andersen (1971) proved this term to be correct for $K=2$; to date, however, the evidence for the validity of the correction for $K>2$ has been empirical rather than analytical. The simulation investigations of Wright and Douglas (1977a, 1977b) have corroborated the "correctness" of the correction term. The present paper presents an analytical study of the simulation of Wright and Douglas, which is replicated in purely algebraic terms.

Maximum Likelihood Estimation in the Rasch Model

Suppose N persons respond to K dichotomous items. The probability of person v responding correctly (+) to item i (denoted as the binary indicator variable a_{vi} taking the value 1; otherwise 0) may be

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 12, No. 3, September 1988, pp. 297–306
© Copyright 1988 Applied Psychological Measurement Inc.
0146-6216/88/030297-10\$1.75

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

represented as the parametric probability function

$$P(+|v,i) = \frac{\theta_v \varepsilon_i}{1 + \theta_v \varepsilon_i} = \pi_{vi} \quad , \tag{1}$$

which is the Rasch model. Because π_{vi} is monotonically increasing in θ, θ_v represents the "ability" or trait level of person v; correspondingly, ε_i represents the "easiness" of item i. According to common practice, the restraint $\prod_{i=1}^{K} \varepsilon_i = 1$ will be assumed throughout. Estimated values of ε_i are denoted as e_i. Also, the reparameterizations $\xi_v = \ln(\theta_v)$ and $\sigma_i = -\ln(\varepsilon_i)$ will be used.

It is well-known for the Rasch model that $\sum_i a_{vi} = r_v$ is a sufficient statistic for ξ_v, and $\sum_v a_{vi} = s_i$ is a sufficient statistic for σ_i (Fischer, 1974). Estimates of ξ and σ are denoted as b and d respectively (the latter notation was adopted by Wright & Douglas, 1977b).

Conditional Maximum Likelihood Estimation

The property of sufficiency makes it possible to use the procedure of CML estimation in the Rasch model. In this method, the number of parameters is reduced by conditioning the likelihood of the data on the sufficient statistics for the parameters to be eliminated. For instance, when the aim is to estimate the ε_i, the unknown θ_v can be eliminated by conditioning the likelihood on r_v. Because the statistic r_v, the number-correct raw score, is sufficient for the parameter θ_v, no relevant information is lost in the conditional approach to estimation.

In this case the conditional estimation equation for the ε_i is obtained by equating the sufficient statistics for ε_i to their conditional expectations:

$$s_i \triangleq E(S_i|r_v,\varepsilon) \qquad (i = 1, \ldots, K) \quad , \tag{2}$$

which is a general property of exponential distributions (Andersen, 1971, 1980; Fischer, 1974, p. 236). Equation 2 can be rewritten as

$$s_i \triangleq \sum_{v=1}^{N} \mathrm{E}(A_{vi}|r_v,\varepsilon) = \sum_{v=1}^{N} \pi_{r_v i} \qquad (i = 1, \ldots, K) \quad , \tag{3}$$

where A_{vi} denotes the random variable a_{vi}. Because of sufficiency, the conditional probability $\pi_{r_v i}$ only depends on v through r_v; it is the same for all persons having this raw score. Therefore, the index v may be dropped and the conditional probability can be written as π_{ri}. The unconditional probability was written as π_{vi} above. Thus

$$\pi_{r_v i} = \pi_{ri} = P(a_{vi} = 1|r_v = r,\varepsilon) \tag{4}$$

and

$$\pi_{vi} = P(a_{vi} = 1|\theta_v,\varepsilon) \quad . \tag{5}$$

Estimated values are written as p_{ri} and p_{vi} respectively.

The Symmetric Functions

The conditional probability π_{ri} is defined as the ratio of $P(a_{vi} = 1, R_v|\theta_v,\varepsilon)$ to $P(R_v|\theta_v,\varepsilon)$, where R_v denotes the random variable r_v. Eventually the following expression is found:

$$\pi_{ri} = \varepsilon_i \frac{\gamma_{r-1}^{(i)}}{\gamma_r} \tag{6}$$

(Fischer, 1974, pp. 221–232), so that the estimation Equation 3 becomes

$$s_i = \sum_{r=1}^{K-1} n_r e_i \frac{g_{r-1}^{(i)}}{g_r} \quad , \tag{7}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

where g denotes the estimate of γ, and n_r represents the number of persons with score r. Persons with scores of 0 or K are omitted because they do not contribute to estimation; it will be assumed below that these persons' scores are not contained in the item total s_i. The factor γ_r in Equation 6 is the elementary symmetric function of order r in the parameters $\varepsilon_1, \ldots, \varepsilon_K$, and $\gamma_{r-1}^{(i)}$ is the elementary symmetric function of order $r - 1$ (the latter does not contain ε_i).

Estimation of the ε_i occurs by solving Equation 7 iteratively. The greatest problem in solving this equation lies in the computation of the symmetric functions in every iteration. Although Rasch (1960/1980, p. 180) presented some recursive formulas that reduce the burden of computing all the symmetric functions in every iteration, such as

$$\gamma_r = \varepsilon_i \gamma_{r-1}^{(i)} + \gamma_r^{(i)} \quad , \tag{8}$$

and Gustafsson (1980a) devised an algorithm for rather large K, CML estimation in the Rasch model still requires respectable computational work for large K. This holds even more in the case of multiparameter generalizations of the Rasch model (see, e.g., Andrich, 1978; Jansen & Roskam, 1986; Masters, 1982; Stegelmann, 1983). Thus, it seems worthwhile to investigate unconditional estimation as an alternative to conditional estimation.

Person parameters are usually estimated by first computing d_i by means of Equation 7 and inserting this for d_i^* in the unconditional Equation 10 below (Fischer, 1974).

Joint Maximum Likelihood Estimation

The unconditional counterpart of Equation 7 is the estimation equation

$$s_i = \sum_{v=1}^{N} \frac{\exp(b_v^* - d_i^*)}{1 + \exp(b_v^* - d_i^*)} \qquad (i = 1, \ldots, K) \tag{9}$$

(Fischer, 1974; Wright & Panchapakesan, 1969), where an asterisk denotes an unconditional estimate. Equation 9 shows that the person and item parameters must be estimated simultaneously in the unconditional approach; that is, both Equation 9 and

$$r_v = \sum_{i=1}^{K} \frac{\exp(b_v^* - d_i^*)}{1 + \exp(b_v^* - d_i^*)} \qquad (v = 1, \ldots, N) \tag{10}$$

must be solved at the same time. Therefore, the unconditional method typically can be denoted as joint maximum likelihood estimation.

Consistency

Andersen (1971; quoted in Fischer, 1974, p. 260) has proven that joint maximum likelihood estimation produces, for moderate and fixed K, inconsistent estimators of the item parameters. Haberman (1977, pp. 817, 835) has proven that unconditional maximum likelihood (UML) estimators in the Rasch model are consistent if, for the series N_s, K_s ($N_s > K_s$, for all s; N_s, $K_s \to \infty$ for $s \to \infty$),

$$\frac{\log N_s}{K_s} \to 0 \quad \text{for} \quad s \to \infty \quad . \tag{11}$$

Expression 11 can be reduced to the requirement that $N^{1/K}$ approaches 1, which is the case if both N and K become very large.

The inconsistency in the estimators of the item parameters is caused by the presence of the person parameters in Equation 9. By enlarging the sample (i.e., increasing N), the number of latter parameters is also increased, resulting in no effective increase in information for estimating the item parameters. This situation would, of course, be avoided if in some way the unknown b^* could be eliminated from Equation 9. This is done by conditional estimation.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Other Approaches to Estimation

Another recent approach would be to integrate out the person parameter from the joint likelihood to be computed from Equation 1, using either an assumed $g(\theta)$ or some sort of empirical characterization of $g(\theta)$. The resulting estimation method is called marginal maximum likelihood estimation (MMLE; Bock & Aitkin, 1981; Thissen, 1982). When $g(\theta)$ is assumed normal, it reduces to the procedure proposed by Andersen and Madsen (1977).

In CML estimation, response probabilities are independent of the person parameter at the level of an individual person (by conditioning on that person's raw score), whereas in MMLE such independence is obtained at the level of an "average person" (by integrating over ξ), that is, at the level of a person sampled randomly from the reference population.

Thus, sample independence is obtained for an "average" person only in MMLE. Also, the validity of MMLE depends on the validity of the assumed distribution of person parameters. Finally, it seems that at present, computer programs for this method are not widely available (and certainly are not incorporated into packages for the Rasch model). Consequently, at present marginal estimation is not widely used in applications of the Rasch model.

Still another approach to estimation is the minimum chi-square method (Fischer, 1974). Simulation studies of van den Wollenberg, Wierda, and Jansen (1988) show it to be a good alternative to CML estimation.

Correcting for Inconsistency

Andersen (1971) proved that the bias in the unconditional estimates d_i^* can be corrected for $K = 2$ by

$$d_i = \frac{K-1}{K} d_i^* \quad , \tag{12}$$

where d_i is the conditional estimate and d_i^* is the unconditional estimate. The validity of this correction for $K > 2$ has been investigated in a number of subsequent studies (e.g., Fischer & Scheiblechner, 1970; Wright & Douglas, 1977a, 1977b), but the correction term appears not to have been *proven* to be equal to $(K - 1)/K$. Both for practical (as argued above) and for theoretical reasons, it would be important if such a proof could be found (see Wainer, Morgan, & Gustafsson, 1980, p. 60). In this study, this was attempted by repeating in analytical terms the empirical study of Wright and Douglas (1977b).

The bias of d_i^* with respect to σ_i is defined as $\mathrm{E}(d_i^*) - \sigma_i$. Because $\mathrm{E}(d_i) = \sigma_i$, bias can be studied as the asymptotic difference $\mathrm{E}(d_i^* - d_i)$. Wright and Douglas adopted the latter criterion for studying empirically the inconsistency of d_i^*.

The Wright and Douglas Simulations

The Bias Equation

Because of sufficiency, persons with raw score r obtain the same b_r in the Rasch model. Therefore, Equation 9 can be rewritten as

$$s_i = \sum_{r=1}^{K-1} n_r \frac{\exp(b_r - d_i^*)}{1 + \exp(b_r - d_i^*)} \quad . \tag{13}$$

Note that as a result of Equation 13, the bias is located in the item estimate d_i^*.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Applying the identity Equation 8 and using $e_i = \exp(-d_i)$, Equation 7 can be rewritten as

$$s_i = \sum_{r=1}^{K-1} n_r \frac{\dfrac{g_{r-1}^{(i)}}{g_r^{(i)}} \exp(-d_i)}{1 + \dfrac{g_{r-1}^{(i)}}{g_r^{(i)}} \exp(-d_i)} \quad , \tag{14}$$

where g denotes the estimate of γ.

Next, Wright and Douglas equated the right sides of Equations 13 and 14, eliminating s_i, and assumed that the resulting equality $\sum_r n_r p_{ri}^* = \sum_r n_r p_{ri}$ $(i = 1, \ldots, K)$ holds for all n_r (1977b, p. 582). This is a crucial assumption, as it leads directly to $p_{ri}^* = p_{ri}$ $(i = 1, \ldots, K; r = 1, \ldots, K - 1)$, that is, to the equality

$$\exp(b_r - d_i^*) = \frac{g_{r-1}^{(i)}}{g_r^{(i)}} \exp(-d_i) \qquad (i = 1, \ldots, K; r = 1, \ldots, K-1) \tag{15}$$

where the term on the left represents unconditional estimation and the term on the right represents conditional estimation. Taking logarithms results in the *bias equation*

$$d_i^* - d_i = b_r - \ln\left(\frac{g_{r-1}^{(i)}}{g_r^{(i)}}\right) \qquad (i = 1, \ldots, K; r = 1, \ldots, K-1) \quad . \tag{16}$$

Note that Equation 16 is based on the assumption that the estimate d_i^* in Equation 13 and the estimate d_i in Equation 14 have the same values irrespective of n_r. This point will be further elaborated below.

The bias $d_i^* - d_i$ is denoted as a_{ri} by Wright and Douglas; a_{ri} measures the bias in the unconditional estimate of σ at score r, relative to the (consistent) conditional estimate. The symmetric functions $g_r^{(i)}$ and $g_{r-1}^{(i)}$ are functions of e (i.e., d) exclusively. The only problem in evaluating Equation 16 is, therefore, the presence of the unknown b_r.

Elimination of b_r

Wright and Douglas (1977b) stated: "Were we to extend our analysis of [the conditional procedure] to derive functions of d_i which yield optimal ability estimates b_r we would find $b_r = \log(g_{r-1}/g_r)$ to be a maximum likelihood solution" (p. 582). Using this result (which is, however, puzzling to the present authors; see below), Equation 16 can be rewritten as

$$a_{ri} = \ln\left(\frac{g_{r-1}}{g_r}\right) - \ln\left(\frac{g_{r-1}^{(i)}}{g_r^{(i)}}\right) \quad . \tag{17}$$

Defining the relative bias c_{ri} as a_{ri}/d_i, a number of simulations showed that invariably the mean relative bias was approximately equal to

$$\bar{c} = \frac{\sum_{r=1}^{K-1} \sum_{i=1}^{K} c_{ri}}{K(K-1)} \simeq \frac{1}{K-1} \quad , \tag{18}$$

which leads to Andersen's exact correction term Equation 12, as is easily verified. In the simulations, test length and the standard deviation of item difficulties were varied; however, in all cases the item difficulties σ_i were sampled from a normal density with mean 0.

Analytical Study

In this section the study of Wright and Douglas is repeated, but in a completely analytical way. First, an "algebraic" simulation is presented in which the mean relative bias $\bar{c}$ is expressed as a function

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

of the conditional parameter estimates d_i (and of the estimates g_i, which are a function of d_i). The resulting equation can be evaluated with respect to the empirical finding of Equation 18, specifically for the conditions in which $\bar{c}$ is expected to be equal to $1/(K-1)$. This reveals that the choice of a symmetric distribution of σ_i values in the simulation studies was essential to observing Equation 18; that is, Equation 18 would not have been found if Wright and Douglas had applied a distribution of σ_i values that was asymmetrical around 0 (e.g., $\sigma_i = -3, 1, 2$ for $i = 1, 2, 3$). This implies that following the Wright and Douglas study, the correction $1/(K-1)$ cannot be obtained as a *general* result.

In the next section, it is demonstrated that the equality $b_r = \ln(g_{r-1}/g_r)$ is a hindrance to obtaining the correction factor $1/(K-1)$. If this correction is valid, then $b_r \neq \ln(g_{r-1}/g_r)$; but if the latter equality holds, the correction factor cannot *generally* be $1/(K-1)$. Another possibility would be, of course, that *both* are incorrect (as will be seen below). Later, it is shown that by dropping $b_r = \ln(g_{r-1}/g_r)$, an analytical proof of the correction factor $1/(K-1)$ is easily obtained, again following the reasoning of Wright and Douglas (that is, by applying their definition of bias as expressed in Equation 15); this is followed by a critical discussion of the logic of Wright and Douglas' study, in particular their definition of bias. It will be argued that the derivation of Equation 15 from Equations 13 and 14 is not valid, and poses conceptual problems as well.

Algebraic Simulation

Instead of assessing the mean relative bias (Equation 18) empirically, it was reduced algebraically to an interpretable expression:

$$\bar{c} = \frac{1}{K(K-1)} \sum_{i=1}^{K} \frac{1}{d_i} \left(\sum_{r=1}^{K-1} a_{ri} \right)$$

$$= \frac{1}{K(K-1)} \sum_{i=1}^{K} \frac{1}{d_i} \left[\sum_{r=1}^{K-1} \ln \left(\frac{g_{r-1}}{g_r} \right) - \sum_{r=1}^{K-1} \ln \left(\frac{g_{r-1}^{(i)}}{g_r^{(i)}} \right) \right] \quad . \tag{19}$$

It is easily verified that

$$\sum_{r=1}^{K-1} \ln \left(\frac{g_{r-1}^{(i)}}{g_r^{(i)}} \right) = \sum_{r=1}^{K-1} [\ln (g_{r-1}^{(i)}) - \ln (g_r^{(i)})] = -\ln g_{K-1}^{(i)} \quad . \tag{20}$$

Analogously, $\sum \ln (g_{r-1}/g_r) = -\ln (g_{K-1})$. Inserting this and Equation 20 in Equation 19, it follows that

$$\bar{c} = \frac{1}{K(K-1)} \sum_{i=1}^{K} \frac{1}{d_i} [\ln (g_{K-1}^{(i)}) - \ln (g_{K-1})]$$

$$= \frac{1}{K(K-1)} \left[\sum_{i=1}^{K} \frac{1}{d_i}(d_i) - \ln (g_{K-1}) \sum_{i=1}^{K} \frac{1}{d_i} \right] \quad , \tag{21}$$

because

$$\ln g_{K-1}^{(i)} = \ln \prod_{\substack{j=1 \\ j \neq i}}^{K} e_j = -\sum_{\substack{j=1 \\ j \neq i}}^{K} d_j = d_i \quad . \tag{22}$$

It also follows that

$$\bar{c} = \frac{1}{K(K-1)} \left[K - \ln (g_{K-1}) \sum_{i=1}^{K} \frac{1}{d_i} \right]$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

$$= \frac{1}{K-1} - \ln (g_{K-1}) \sum_{i=1}^{K} \frac{1}{d_i} \quad . \tag{23}$$

For $K = 2$, the mean relative bias (Equation 23) is exactly equal to $1/(K-1)$, because then $(1/d_1) + (1/d_2) = 0$ due to the restraint $d_1 + d_2 = 0$.

For $K > 2$, however, even and odd values of K must be distinguished. When K is even, the expected value of $\sum(1/d_i)$ would be 0 when the density of σ_i is symmetric around 0 (again taking $\sum d_i = 0$). When K is odd, however, there are no sets of values of d_i that simultaneously satisfy $\sum(1/d_i) = 0$, $\sum d_i = 0$, and $d_i \neq 0$ for all i. Note that in the Wright and Douglas simulations, the density of σ_i was invariably symmetric around 0. Consequently, it seems that Wright and Douglas obtained the empirical result $(K-1)/K$ because of their choice of a symmetric σ_i distribution. In case of an asymmetric set of σ_i values, their procedure would not have yielded this correction term, as is demonstrated by the simulation studies of van den Wollenberg et al. (in press).

The Choice of b_r

The choice of $b_r = \ln (g_{r-1}/g_r)$ is puzzling. It can be shown that this assumption is not consistent with the correction factor $(K-1)/K$. That is, assuming the relative bias to be $1/(K-1)$, it can be proven that in that case b_r cannot, in general, be equal to $\ln (g_{r-1}/g_r)$. It will be shown below that it is indeed possible, by dropping this assumption, to derive the correction factor following the logic of Wright and Douglas.

Suppose the relative bias is $1/(K-1)$. Then (see Equation 16):

$$\begin{aligned} b_r - \ln\left(\frac{g_{r-1}}{g_r}\right) &= \ln\left(\frac{g_{r-1}^{(i)}}{g_r^{(i)}}\right) + \frac{d_i}{K-1} - \ln\left(\frac{g_{r-1}}{g_r}\right) \\ &= \ln\left[\frac{p_{ri}}{e_i(1-p_{ri})}\right] + \frac{d_i}{K-1} - \ln\left[\frac{p_{ri}}{e_i(1-p_{r-1,i})}\right] \\ &= \ln (1-p_{r-1,i}) - \ln (1-p_{ri}) + \frac{d_i}{K-1} \qquad (i = 1, \ldots, K;\ r = 1, \ldots, K-1) \quad . \end{aligned} \tag{24}$$

Under the assumption $b_r = \ln (g_{r-1}/g_r)$, the right side of Equation 24 should be 0. Because p_{ri} is monotonically increasing in r (for a formal proof see Jansen, 1983, pp. 243–245), the term $\ln (1 - p_{r-1,i}) - \ln (1 - p_{ri})$ is positive for all i, r. This implies that a necessary condition for the right side of Equation 24 to be 0 is that $d_i < 0$, and this should be true for all i. However, the sign of d_i is arbitrary, and is in fact determined by the choice of the norming restraint.

In view of this, the requirement $d_i < 0$ for all i, and consequently the initial assumption $b_r = \ln (g_{r-1}/g_r)$, is absurd. Thus, it appears that b_r cannot be, generally, equal to $\ln (g_{r-1}/g_r)$ when the relative bias is $(K-1)/K$. If this correction is valid, then $b_r \neq \ln (g_{r-1}/g_r)$; but if the latter equality holds, then the correction $(K-1)/K$ is invalid.

An Analytical Proof, Following Wright and Douglas

It is possible to derive analytically the correction $(K-1)/K$ *following the logic of Wright and Douglas* by dropping $b_r = \ln (g_{r-1}/g_r)$. The assumption $b_r = b_r^*$ is unnecessary and is also dropped. Applying Equations 20 through 22 to Equation 16 allows the following expressions:

$$b_r^* = d_i^* - d_i + \ln\left(\frac{g_{r-1}^{(i)}}{g_r^{(i)}}\right) \quad , \tag{25}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

$$\sum_{r=1}^{K-1} b_r^* = (K-1)(d_i^* - d_i) - \ln(g_{K-1}^{(i)}) = (K-1)(d_i^* - d_i) - d_i \quad , \tag{26}$$

and

$$d_i = \frac{K-1}{K} d_i^* - \frac{\sum_{r=1}^{K-1} b_r^*}{K} \quad . \tag{27}$$

This correction differs from the expected result (Equation 12), but this can be shown to be caused by norming restraints. In deriving Equations 25 through 27, $\sum d_i$ was assumed equal to 0. This implies for d_i^* and b_r^* that

$$\sum_{i=1}^{K} d_i = 0 = \frac{K-1}{K} \sum_{i=1}^{K} d_i^* - \sum_{r=1}^{K-1} b_r^* \tag{28}$$

and

$$\sum_{i=1}^{K} d_i^* = \frac{K}{K-1} \sum_{r=1}^{K-1} b_r^* \quad . \tag{29}$$

Thus d_i^* and b_r^* are supposed to be "renormed" according to Equations 28 and 29, instead of the usual choice $\sum d_i^* = 0$. When the d_i^* are transformed according to

$$\hat{d}_i^* = d_i^* - \bar{b}^* \tag{30}$$

where $\bar{b}^* = \sum b_r^*/(K-1)$, the $\hat{d}_i^*$ have the property $\sum \hat{d}_i^* = 0$. Rewriting Equation 30:

$$d_i^* = \hat{d}_i^* + \bar{b}^* \quad . \tag{31}$$

Substituting this for d_i^* in Equation 27 yields

$$d_i = \frac{K-1}{K} \hat{d}_i^* \quad , \tag{32}$$

which is Andersen's result (Equation 12) generalized for $K > 2$. Note that $\bar{b}^*$ is simply a location value; it is not, for example, the mean of the distribution of ability in the sample. Thus, under the restraint of Equations 28 and 29, the correction of Equation 27 should be applied, and under the restraint $\sum d_i^* = 0$ the correction is given by Equation 32. Both corrections are compatible with the restraint $\sum d_i = 0$. When in simulation studies d_i are computed in the CML approach under $\sum d_i = 0$, and d_i^* are computed in the UML approach with $\sum d_i^* = 0$, the relationship between these estimates should comply with Equation 32.

The Logic of Wright and Douglas' Simulations

The validity of Equation 32 depends, of course, on the validity of Wright and Douglas' primary logic, that is, on the validity of deriving Equation 15 from Equations 13 and 14. Equation 15 is based on the assumption that in Equations 13 and 14 the estimates d_i^* and d_i have the same value irrespective of n_r.

With respect to the d_i, there is some intuitive plausibility in this assumption. In the conditional approach to estimation, σ_i is estimated independently of the person distribution. Taking the distribution of n_r to represent the latter, it would only seem reasonable that the same σ_i is estimated (i.e., about the same d_i should be obtained for varying n_r). Naturally, the size of n_r may affect the precision of the estimators, but not the latent parametric structure to be estimated conditionally on a specific raw score

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

r. With respect to the d_i^*, however, this reasoning is not correct because the d_i^* are estimated in conjunction with the b_r^*.

Taken together, this implies that the derivation of Equation 15 from Equations 13 and 14 is not valid. Moreover, there is also a conceptual problem with Equation 15. When UML is to be used as a substitute for CML, it may seem reasonable to require that UML be as good as CML in reproducing the basic probabilities π_{ri}, that is, to require $p_{ri}^* = p_{ri}$ (which is Equation 15). But p_{ri} and p_{ri}^* represent different probabilistic events, as becomes clear from Equations 4 and 5. The CML probability p_{ri} is dependent on r (i.e., on the parameter values of all items), whereas p_{ri}^* is not. That this may lead to dramatic differences in probability values is illustrated by van den Wollenberg (1980), who for instance obtained $p_{ri}^* = .50$ but $p_{ri} = .09$ in one example. Thus it does not seem justifiable to formulate Equation 15 as a requirement of the UML and CML approaches to estimation in the Rasch model.

Conclusions

The simulation study of Wright and Douglas is based on two assumptions. The first is that in the derivation of the bias Equation 16 from Equations 13 and 14, $\sum_r n_r p_{ri}^* = \sum_r n_r p_{ri}$ ($i = 1, \ldots, K$) is assumed valid for all n_r; the second is the assumption that $b_r = \ln(g_{r-1}/g_r)$ ($r = 1, \ldots, K - 1$). In this study it appeared that both assumptions lead to the derivation of the correction term $(K - 1)/K$ only in special circumstances (e.g., a density of σ that is symmetric around 0). Generally, however, even within the framework of these two assumptions, the correction factor is of a much more complicated nature.

Moreover, it has been demonstrated that the second assumption is an impediment to deriving the correction term $(K - 1)/K$ following the logic of Wright and Douglas. That is, dropping this second assumption, the correction term $(K - 1)/K$ is easily derived as a result of the first assumption. Regrettably, however, the first assumption is also untenable.

To verify these conclusions, a number of simulation studies were undertaken; these are reported in van den Wollenberg et al. (1988). Van den Wollenberg et al. demonstrated that the bias of the unconditional maximum likelihood method is affected by the range of the item parameters and the asymmetry of the item parameter distribution, as well as the number of items (K); this implies that substitution of the correction factor $(K - 1)/K$ by another, more appropriate factor would be virtually impossible.

Also, de Gruijter (1986) developed an approximation to the unconditional estimation procedure, in which the unconditional estimates are corrected in a rather complicated way: The approximate correction factor does not depend only on K, but also on the distributions of ε and θ.

References

Andersen, E. B. (1971). *Conditional inference for multiple-choice questionnaires* (Report No. 8). Copenhagen: Copenhagen School of Economics and Business Administration.

Andersen, E. B. (1973a). A goodness-of-fit test for the Rasch model. *Psychometrika, 38*, 123–140.

Andersen, E. B. (1973b). Conditional inference for multiple-choice questionnaires. *British Journal of Mathematical and Statistical Psychology, 26*, 31–44.

Andersen, E. B. (1980). *Discrete statistical models with social science applications*. Amsterdam: North-Holland.

Andersen, E. B., & Madsen, B. (1977). Estimating the parameters of the latent population distribution. *Psychometrika, 42*, 357–374.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561–573.

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

de Gruijter, D. N. M. (1986). The bias of an approximation to the unconditional maximum likelihood procedure for the Rasch model. *Kwantitatieve Methoden, 20*, 49–55.

Fischer, G. H. (1974). *Einführung in die Theorie psy-*

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

chologischer Tests. Bern: Huber.

Fischer, G. H., & Scheiblechner, H. (1970). Algorithmen und Programme für das probabilistische Testmodell von Rasch. *Psychologische Beiträge, 12*, 23–51.

Gustafsson, J.-E. (1980a). A solution of the conditional estimation problem for long tests in the Rasch model for dichotomous items. *Educational and Psychological Measurement, 40*, 377–385.

Gustafsson, J.-E. (1980b). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology, 33*, 205–233.

Haberman, S. J. (1977). Maximum likelihood estimates in exponential response models. *Annals of Statistics, 5*, 815–841.

Jansen, P. G. W. (1983). *Rasch analysis of attitudinal data*. Unpublished doctoral dissertation, Rijks Psychologische Dienst, The Hague, The Netherlands.

Jansen, P. G. W., & Roskam, E. E. (1986). Latent trait models and dichotomization of graded responses. *Psychometrika, 51*, 69–91.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174.

Molenaar, I. W. (1983). Some improved diagnostics for failure of the Rasch model. *Psychometrika, 48*, 49–72.

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

Rasch, G. (1966, July). An informal report on a theory of objectivity in comparisons. In L. J. van der Kamp & C. A. J. Vlek (Eds.), *Proceedings of the NUFFIC (Netherlands Universities Foundation Fund for International Cooperation) international summer session in science at "Het Oude Hof," The Hague* (unpublished).

Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy, 14*, 58–94.

Roskam, E. E., & Jansen, P. G. W. (1984). A new derivation of the Rasch model. In E. Degreef & J. van Buggenhaut (Eds.), *Trends in mathematical psychology* (pp. 293–307). Amsterdam: Elsevier (North Holland).

Stegelmann, W. (1983). Expanding the Rasch model to a general model having more than one dimension. *Psychometrika, 48*, 259–267.

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

van den Wollenberg, A. L. (1980). *On the Wright-Panchapakesan goodness-of-fit test for the Rasch model* (Internal Report No. 80 MA 02). Nijmegen, The Netherlands: Katholieke Universiteit.

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

van den Wollenberg, A. L., Wierda, F. W., & Jansen, P. G. W. (1988). Consistency of Rasch model parameter estimation: A simulation study. *Applied Psychological Measurement, 12*, 307–313.

Wainer, H., Morgan, A., & Gustafsson, J.-E. (1980). A review of estimation procedures for the Rasch model with an eye toward longish tests. *Journal of Educational Statistics, 5*, 35–64.

Wright, B. D., & Douglas, G. A. (1977a). Best procedures for sample-free item analysis. *Applied Psychological Measurement, 1*, 281–295.

Wright, B. D., & Douglas, G. A. (1977b). Conditional versus unconditional procedures for sample-free item analysis. *Educational and Psychological Measurement, 37*, 573–586.

Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement, 29*, 23–48.

Acknowledgments

The authors thank Erling Andersen and Ivo Molenaar for suggestions for improving the manuscript.

Author's Address

Send requests for reprints or further information to Paul G. W. Jansen, Afdeling Bedrijfspsychologische Zaken, PTT Telecommunicatie, P.O. Box 30 000, 2500 GA The Hague, The Netherlands.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/