A Minimum χ^2 Method for Equating Tests Under the Graded Response Model

Seock-Ho Kim and Allan S. Cohen

University of Wisconsin, Madison

The minimum χ^2 method for computing equating coefficients for tests with dichotomously scored items was extended to the case of Samejima's graded response items. The minimum χ^2 method was compared with the test response function method (also referred to as the test characteristic curve method) in which the equating coefficients were obtained by matching the

test response functions of the two tests. The minimum χ^2 method was much less demanding computationally and yielded equating coefficients that differed little from those obtained using the test response function approach. *Index terms: equating, graded response model, item response theory, minimum χ^2 method, test response function method.*

Under item response theory (IRT), the metrics obtained from item parameter estimation procedures are unique only up to a linear transformation. Therefore, to equate two tests that have been calibrated separately it is necessary to determine the slope and intercept coefficients of the linear equation required for the transformation.

Two general classes of equating methods have been developed for this purpose: mean and sigma methods and characteristic curve methods (referred to here as response function methods). Mean and sigma methods use the first two moments of the distribution of item difficulty estimates to determine the appropriate linear equation (Bejar & Wingersky, 1981; Cook, Eignor, & Hutten, 1979; Linn, Levine, Hastings, & Wardrop, 1981; Loyd & Hoover, 1980; Marco, 1977; Vale, 1986). One problem with these methods is that they do not use information available from the estimated item discrimination parameters in obtaining the equating coefficients. In addition, these methods use the summary statistics of the anchor item parameters and, consequently, are sensitive to the distributional characteristics of the item parameter estimates. Because of this, deviant estimates can distort the values of the equating coefficients obtained using the mean and sigma methods.

Response function methods (Divgi, 1980; Haebara, 1980; Stocking & Lord, 1983) use the information available from both the item discrimination and item difficulty parameters. These methods determine the equating coefficients by minimizing some measure of the difference between the test response functions (TRFs) estimated in each sample. In the minimum χ^2 method (Divgi, 1985) (a variation of the TRF method), the standard errors of the estimates of the item parameters are included in the linear equation.

In the Stocking & Lord (1983) TRF procedure, the equating coefficients are obtained by minimizing a quadratic loss function based on differences in "true scores" (the value of the TRF at a specified trait level) yielded by the two test calibrations. The TRF procedure, however, requires a complicated multivariate search technique to find the equating coefficients that minimize the quadratic loss function. Baker (1992a) extended the TRF procedure to the graded response model and implemented this extension in the updated version of the computer program EQUATE (Baker, Al-Karni, & Al-Dosary, 1991).

The minimum χ^2 method proposed by Divgi (1985) for use with dichotomously scored items uses esti-

APPLIED PSYCHOLOGICAL MEASUREMENT

Vol. 19, No. 2, June 1995, pp. 167-176

© Copyright 1995 Applied Psychological Measurement Inc.

0146-6216/95/020167-10\$1.75

167

mates of both the item discrimination and item difficulty parameters and their standard errors. This method is implemented more easily in a computer program than the TRF method. The small amount of research that has compared the equating results from the minimum χ^2 method with results from the TRF method indicates that there are few differences between the two methods (Kim & Cohen, 1992). The present paper extends the minimum χ^2 approach to the graded response model (Samejima, 1969, 1972) and compares its performance with that of the TRF method.

The Minimum χ^2 Method

Dichotomous Item Responses

Lord (1980) demonstrated that, under IRT, the relationship between the metric of any two item calibrations is linear. Thus, when estimates from a second calibration are to be transformed to the metric of the first, the transformed estimates of the item discrimination (a) and item difficulty (b) parameters of item i are given by

$$a_{i2}^* = a_{i2}/A \quad (1)$$

and

$$b_{i2}^* = Ab_{i2} + B, \quad (2)$$

respectively, where

- * indicates a transformed value,
- the subscript 2 refers to the second calibration,
- A is the slope, and
- B is the intercept.

The value of the transformed trait level (θ) estimate of person j can be expressed as

$$\theta_{j2}^* = A\theta_{j2} + B. \quad (3)$$

The task of equating the two metrics is to find the appropriate equating coefficients A and B .

The minimum χ^2 method for obtaining two equating coefficients is based on the quadratic form

$$\chi^2 = \sum_{i=1}^n \chi_i^2 = \sum_{i=1}^n \xi_i' \Sigma_i^{-1} \xi_i, \quad (4)$$

where

- n is the total number of items,
- ξ_i is the vector of the differences between the first and the transformed second item parameter estimates, and
- Σ_i is the variance-covariance matrix of the ξ_i .

In the two-parameter item response function model, for example,

$$\xi_i = \xi_{i1} - \xi_{i2}^* \quad (5)$$

and

$$\Sigma_i = \Sigma_{i1} + \Sigma_{i2}^*, \quad (6)$$

where

$$\xi_{i1} = (a_{i1}, b_{i1})', \quad (7)$$

and

$$\xi_{i2}^* = (a_{i2}^*, b_{i2}^*)'. \quad (8)$$

Σ_{i1} is the estimated 2×2 variance-covariance matrix of sampling errors for item i from the first calibration, and Σ_{i2}^* is the transformed variance-covariance matrix from the second calibration.

The χ^2 can be written as

$$\chi^2 = \sum_{i=1}^n (a_{i1} - a_{i2}^*, b_{i1} - b_{i2}^*) (\Sigma_{i1} + \Sigma_{i2}^*)^{-1} (a_{i1} - a_{i2}^*, b_{i1} - b_{i2}^*)' . \quad (9)$$

This χ^2 is a function of the two unknown equating coefficients A and B . To obtain B , take the partial derivative of Equation 9 with respect to B : $\partial\chi^2/\partial B = 0$. Note that this partial derivative is linear with respect to B and easily solved as a function of A . Denote S_{iab} and S_{ibb} as individual elements from the matrix

$$S_i = \Sigma_i^{-1} = (\Sigma_{i1} + \Sigma_{i2}^*)^{-1} . \quad (10)$$

Then

$$B = \frac{\sum_{i=1}^n [(a_{i1} - a_{i2}/A)S_{iab} + (b_{i1} - Ab_{i2})S_{ibb}]}{\sum_{i=1}^n S_{ibb}} . \quad (11)$$

When this value of B is substituted into the expression for χ^2 , this becomes a minimization problem with only a single unknown A . A procedure such as Newton's method can be used to solve for A .

The Graded Response Case

Under Samejima's (1969, 1972) graded response model, an item is comprised of m_i ordered response categories and the examinee selects only one of the categories. Item parameters are estimated under the graded response model using the $m_i - 1$ boundary response functions (BRFs). Each of these BRFs represents the cumulative probability of selecting response categories greater than the category of interest (Samejima, 1969). The BRFs for item i are characterized by an item discrimination parameter a_i and the $m_i - 1$ location parameters b_{ik} ($k = 1, \dots, m_i - 1$). The b_{ik} for the item are ordered, typically from low ($k = 1$) to high ($k = m_i - 1$); in the homogeneous case of the model, a_i is the same across all BRFs. The probability of selecting a given response category, the operating response function $P_{ik}^o(\theta)$, is given by

$$P_{ik}^o(\theta) = \begin{cases} 1 - P_{i1}(\theta) & \text{if } k = 1 \\ P_{i(m_i-1)}(\theta) & \text{if } k = m_i \\ P_{i(k-1)}(\theta) - P_{ik}(\theta) & \text{otherwise,} \end{cases} \quad (12)$$

where $P_{ik}(\theta)$ is the cumulative probability obtained from the BRF. It is defined as

$$P_{ik}(\theta) = \left\{ 1 + \exp[a_i(\theta - b_{ik})] \right\}^{-1} . \quad (13)$$

For equating two tests under the graded response model, there are two sets of item parameter estimates. The first set consists of estimates from the Test 1 calibration,

$$\xi_{im,1} = [a_{i1}, b_{i11}, \dots, b_{ik1}, \dots, b_{i(m_i-1)1}]' . \quad (14)$$

The second consists of transformed estimates from the Test 2 calibration,

$$\xi_{im,2}^* = [a_{i2}^*, b_{i12}^*, \dots, b_{ik2}^*, \dots, b_{i(m_i-1)2}^*]' , \quad (15)$$

where $a_{i2}^* = a_{i2}^*/A$ and $b_{ik2}^* = Ab_{ik2}^* + B$. The symmetric estimated variance-covariance matrices are defined as

$$\Sigma_{im_1} = \begin{bmatrix} \text{Var}(a_{i1}) & \text{Cov}(a_{i1}, b_{i11}) & \dots & \text{Cov}[a_{i1}, b_{i(m_1-1)1}] \\ & \text{Var}(b_{i11}) & \dots & \text{Cov}[b_{i11}, b_{i(m_1-1)1}] \\ & & \ddots & \vdots \\ & & & \text{Var}[b_{i(m_1-1)1}] \end{bmatrix} \quad (16)$$

and

$$\Sigma_{im_2}^* = \begin{bmatrix} \text{Var}(a_{i2}^*) & \text{Cov}(a_{i2}^*, b_{i12}^*) & \dots & \text{Cov}[a_{i2}^*, b_{i(m_1-1)2}^*] \\ & \text{Var}(b_{i12}^*) & \dots & \text{Cov}[b_{i12}^*, b_{i(m_1-1)2}^*] \\ & & \ddots & \vdots \\ & & & \text{Var}[b_{i(m_1-1)2}^*] \end{bmatrix} \quad (17)$$

for the Test 1 calibration and the transformed Test 2 calibration, respectively. The variance-covariance matrix for the transformed Test 2 calibration can be written as

$$\Sigma_{im_2}^* = \begin{bmatrix} \text{Var}(a_{i2})/A^2 & \text{Cov}(a_{i2}, b_{i12}) & \dots & \text{Cov}[a_{i2}, b_{i(m_1-1)2}] \\ & A^2 \text{Var}(b_{i12}) & \dots & \text{Cov}[b_{i12}, b_{i(m_1-1)2}] \\ & & \ddots & \vdots \\ & & & A^2 \text{Var}[b_{i(m_1-1)2}] \end{bmatrix}. \quad (18)$$

Let

$$\xi_{im_i} = \xi_{im_i1} - \xi_{im_i2}^* \quad (19)$$

and

$$\Sigma_{im_i} = \Sigma_{im_i1} - \Sigma_{im_i2}^*, \quad (20)$$

then χ^2 is defined as

$$\chi^2 = \sum_{i=1}^n \chi_{im_i}^2 = \sum_{i=1}^n \xi_{im_i}' \Sigma_{im_i}^{-1} \xi_{im_i}. \quad (21)$$

The linear equating coefficients are found by minimizing the χ^2 in Equation 21 with respect to A and B .

For simplicity, it was assumed that all common items in the equating problem had the same number of categories, m . Given m categories, there are m item parameter estimates for item i ; that is, ξ_i is an $m \times 1$ vector and Σ_i is an $m \times m$ symmetric matrix (note that the subscript m_i has been deleted because all items have m categories).

Let S_{ikl} be an individual element from the r th row and k th column of $S_i = \Sigma_i^{-1}$. Then the equating coefficient B can be obtained from

$$B = \frac{\sum_{i=1}^n \left\{ (a_{i1} - a_{i2}/A) \sum_{l=2}^m S_{il1} + \sum_{k=2}^m \sum_{l=2}^m [b_{i(k-1)1} - Ab_{i(k-1)2}] S_{ikl} \right\}}{\sum_{i=1}^n \left(\sum_{k=2}^m \sum_{l=2}^m S_{ikl} \right)} \quad (22)$$

As noted for equating dichotomous items, an important advantage of the minimum χ^2 method is that the partial derivative, $\partial\chi^2/\partial B = 0$, is linear in B , and hence, easily solved as a function of A . When this initial value of B is substituted into Equation 21, the resulting minimization problem has only a single unknown, A , which can be solved iteratively. Newton's method, which can be used to obtain a value for A , can be summarized as

$$A^{(s)} = A^{(s-1)} - [H^{(s-1)}]^{-1} f^{(s-1)}, \quad (23)$$

where s indexes the iteration,

$$f^{(s-1)} = \left. \frac{\partial\chi^2}{\partial A} \right|_{A^{(s-1)}}, \quad (24)$$

and

$$H^{(s-1)} = \left. \frac{\partial^2\chi^2}{\partial A^2} \right|_{A^{(s-1)}}. \quad (25)$$

Examples of Computing Equating Coefficients

Estimating Equating Coefficients

The equating coefficients for graded response tests using the minimum χ^2 method were obtained using two different approaches. First, the equating coefficients were obtained using only the main diagonal terms of the estimated variance-covariance matrix. This approach was used because most computer programs for the graded response model, such as MULTLOG (Thissen, 1991) and PARSCALE (Muraki & Bock, 1993), provide only the estimated standard errors of the item parameters. Note that most computer programs for the dichotomous item response model, such as BILOG (Mislevy & Bock, 1990) and LOGIST (Wingersky, Barton, & Lord, 1982), provide the complete variance-covariance matrices for item parameter estimates. For the second approach, the equating coefficients were obtained using all terms of the estimated variance-covariance matrix computed from the modified equations given by Baker (1992b).

The diagonal variance-covariance matrix approach. The minimum χ^2 method of estimating equating coefficients was implemented here using a computer program written in Professional FORTRAN for the IBM PC/AT. In the program implementation, two simplifying assumptions were made. First, all anchor items had the same number of response categories. Second, the off-diagonal terms of the variance-covariance matrix were not used. The mean value of a_{i2} was divided by the mean value of a_{i1} to obtain the initial estimate of A . The initial estimate of B was obtained using Equation 22. The program then evaluated $\partial\chi^2/\partial A$, and the obtained estimate of A was used to obtain a new estimate of B . The iterations were repeated until a prespecified convergence criterion was met for the differences of both values of A and B between two successive iterations. The convergence criterion used was .001.

Full estimated variance-covariance matrix approach. The second assumption was made in the former approach because computer programs for estimating item parameters under the graded response model, such as MULTLOG (Thissen, 1991) and PARSCALE (Muraki & Bock, 1993), provide only the estimated standard errors of item parameters. An anonymous reviewer, however, suggested that the estimated vari-

ance-covariance matrix could be computed using equations given by Baker (1992b, p. 236; see the Appendix below for a more detailed description). For comparison purposes, therefore, Baker's equations were implemented and a second set of equating coefficients for the minimum χ^2 method were calculated.

Horizontal Equating

Method. Equating using the minimum χ^2 method used simulated 30-item tests taken by 300 examinees. All items had four categories. The a_i s of the tests were generated from a uniform distribution ranging from 1.34 to 2.65 in the logistic metric. The three difficulty parameters for the BRFS of each item were generated from a normal distribution with mean 0 and unit variance. Each set of three ordered difficulty parameters for the BRFS for an item were randomly paired with a single a_i . The θ levels of the 300 simulated examinees were sampled from a unit normal distribution over the range -2.8 to 2.8 . GENIRV (Baker, 1986) was used to generate the vectors of examinee response category selections for Test 1. Then, using the same item parameter and θ specifications and a new random number generator seed, a new set of examinee item response vectors were generated for Test 2. Item parameter and θ estimates were obtained using MULTILOG (Thissen, 1991).

Item parameter and θ estimates from the Test 2 dataset then were transformed to the metric of the Test 1 dataset. Because horizontal equating was used here, the expected values of the equating coefficients were $A = 1.0$ and $B = 0.0$. The horizontal equating here involved test forms that were designed to measure the same trait at the equivalent difficulty level for the same population.

The computer program EQUATE (Baker et al., 1991) also was used on the same datasets to estimate equating coefficients using the TRF method. The TRF matching approach used 21 points equally spaced from -4 to 4 on the Test 1 θ metric and two sets of item parameter estimates.

Results. The equating coefficient values obtained after two iterations using the minimum χ^2 method were $A = 1.009$ and $B = .003$. The equating coefficients obtained for the estimated full variance-covariance matrix were $A = 1.014$ and $B = .005$. Both sets of equating coefficients were very close to the expected theoretical values.

Next, using these coefficients, the item parameter and θ estimates of Test 2 were transformed to the Test 1 metric. The means of the MULTILOG item parameter and θ estimates of Test 1 and those of Test 2 following the transformation using the diagonal or full variance-covariance matrix methods are reported in Table 1. Because the two datasets differed only with respect to the random number seed used to generate the examinees' item response category choice vectors, there should have been good agreement between Test 1 and the transformed Test 2 using the diagonal method or using the full variance-covariance method. In fact, this was the case: The two sets of results differed only in the second decimal place.

The obtained values of the equating coefficients for the TRF method after three iterations were $A = 1.010$

Table 1
Means of the Item and θ Parameter Estimates for
Test 1 and the Transformed Test 2 Parameter Estimates
From the Horizontal Equating for the Diagonal and
Full χ^2 Methods and the TRF Method

Parameter	Test 1	Transformed Test 2		
		Minimum χ^2		TRF
		Diagonal	Full	
$\bar{a}$	2.287	2.267	2.256	2.265
$\bar{b}_1$	-1.067	-1.074	-1.077	-1.065
$\bar{b}_2$	.042	.045	.048	.055
$\bar{b}_3$	1.351	1.366	1.375	1.377
$\bar{\theta}$	-.184	-.181	-.166	-.158

and $B = .012$. The values also agreed very closely with the theoretical values as well as with those values yielded by the two versions of the minimum χ^2 method. In terms of the values of the obtained equating coefficients, there were essentially negligible differences in the outcomes among equating methods.

Table 1 also shows the means of Test 2 after transformation using the TRF method. The differences between the mean values of the transformed parameter estimates obtained using the two minimum χ^2 methods and those using the TRF method were very small.

Vertical Equating

Method. The item parameters of the test used in the preceding example were transformed to a different metric using the values $A = .9$ and $B = .5$. These values then were used in the GENIRV computer program (Baker, 1986) to generate a Test 3 dataset with a sample size of 300. θ s were normally distributed with mean 0 and unit variance. The generated examinee item response vectors then were analyzed using MULTILOG to obtain the item parameter estimates. Test 2 then was equated to the metric of Test 3 using both implementations of the minimum χ^2 method. The vertical equating in this study used the anchor item approach (Marco, 1977) to place two tests of differing difficulty and two groups of differing trait level on a common metric.

Results. The obtained values of the equating coefficients were $A = .948$ and $B = .656$ using the diagonal variance-covariance matrix. The equating coefficients using the full variance-covariance matrix were $A = .969$ and $B = .617$. Both sets of equating coefficients agreed well with the underlying values of $A = .9$ and $B = .5$.

The means for item parameter and θ estimates are reported in Table 2. The two sets of underlying item parameters differed primarily in terms of the locations of the items. Thus, when the two sets of common items were equated, the change in location should have been reflected in the mean θ levels of the two groups of examinees. The results in Table 2 reflect this difference. The mean θ level of the examinees for the transformed Test 2 from the diagonal-only minimum χ^2 method was .496, which was approximately .739 above the mean of the Test 3 group (-.243). The difference between the θ means was slightly larger than the expected difference of .5. The mean θ of the examinees of the transformed Test 2 based on equating coefficients using the full variance-covariance matrix was .453. This value was approximately .696 above the mean of the Test 3 group. However, there was very good agreement between the mean values of the item parameter estimates for Test 3 and the transformed Test 2.

The equating coefficients obtained from the TRF method were $A = .959$ and $B = .671$. These results also were in reasonably good agreement with the underlying values. In addition, there was very close agreement between the results from the three equating methods (see Table 2).

Table 2
Means of the Item and θ Parameter Estimates for
Test 3 and the Transformed Test 2 Parameter Estimates
From the Vertical Equating for the Diagonal and
Full χ^2 Methods and the TRF Method

Parameter	Test 3	Transformed Test 2		
		Minimum χ^2		TRF
		Diagonal	Full	
$\bar{a}$	2.402	2.413	2.360	2.385
$\bar{b}_1$	-.374	-.355	-.417	-.353
$\bar{b}_2$	.676	.696	.658	.711
$\bar{b}_3$	1.993	1.936	1.926	1.966
$\bar{\theta}$	-.243	.496	.453	.509

Discussion

Results from the three equating methods studied were quite similar. In the horizontal equating example, results were nearly the same and the equating coefficients differed little from the theoretically expected values of $A = 1.0$ and $B = 0.0$. Mean item parameter and θ estimates were also nearly the same. Given data such as these, the results from any of these equating methods can be used with confidence with data from the graded response model that are similar to the data used here.

Results for the vertical equating example were nearly as good as for horizontal equating. Equating coefficients from the three methods were nearly the same. Likewise, transformed mean item parameter and θ estimates were also almost the same. The minor differences in the means of the item parameter and θ estimates were mainly due to sampling variation in the generated item response vectors.

The equating coefficients based on the two versions of the variance-covariance matrix were very similar. Note, however, that the obtained variance-covariance matrices based on the equations in the Appendix for the graded response model may not be the same as the actual variance-covariance matrices that might be output from MULTILOG or PARSCALE.

Finally, a comparison of the two equating methods clearly should be based on a more comprehensive analysis than provided in the two examples presented here. The primary purpose of this paper was to describe a minimum χ^2 method for equating graded response items. The data presented show that the results do not support selecting one equating method over another. Furthermore, there is no theoretical rationale for selecting one of these methods over another. The minimum χ^2 method using only the diagonal variance-covariance matrix has the advantage of ease of implementation over either the full matrix or the TRF method. A more comprehensive comparison under a variety of conditions may reveal specific strengths or weaknesses of each of these equating methods.

Appendix

Baker (1992b) presented the second derivatives of the log likelihood for the graded response model. These terms were used to obtain the approximate variance-covariance matrix. Logits in Baker's equations, however, were expressed in linear terms as

$$\lambda_i \theta_g + \zeta_{ik} = a_i (\theta_g - b_{ik}). \quad (26)$$

Because complicated transformations were required to obtain the variance-covariance matrix in terms of a_i and b_{ik} here, the second derivatives were expressed in terms of a_i and b_{ik} . The expected values of the negative of the second derivatives are:

$$E\left(-\frac{\partial^2 L}{\partial a_i^2}\right) = \sum_{g=1}^G f_g \left\{ w_{gi1}^2 (\theta_g - b_{i1})^2 (P_{gi1}^o)^{-1} + \sum_{k=2}^{m-1} \left\{ w_{gi(k-1)} \left[(\theta_g - b_{i(k-1)}) \right] - w_{gik} (\theta_g - b_{ik}) \right\}^2 (P_{gik}^o)^{-1} \right. \\ \left. + w_{gi(m-1)}^2 \left[\theta_g - b_{i(m-1)} \right]^2 (P_{gim}^o)^{-1} \right\}, \quad (27)$$

$$E\left(-\frac{\partial^2 L}{\partial a_i \partial b_{ik}}\right) = -\sum_{g=1}^G f_g a_i w_{gik} \times \left\{ \left\{ w_{gi(k-1)} \left[\theta_g - b_{i(k-1)} \right] - w_{gik} (\theta_g - b_{ik}) \right\} (P_{gik}^o)^{-1} \right. \\ \left. - \left\{ w_{gik} (\theta_g - b_{ik}) - w_{gi(k+1)} \left[\theta_g - b_{i(k+1)} \right] \right\} (P_{gi(k+1)}^o)^{-1} \right\}, \quad (28)$$

$$E\left(-\frac{\partial^2 L}{\partial b_{ik}^2}\right) = \sum_{g=1}^G f_g a_i^2 w_{gik}^2 \left\{ (P_{gik}^o)^{-1} + [P_{gi(k+1)}^o]^{-1} \right\}, \quad (29)$$

$$E\left(-\frac{\partial^2 L}{\partial b_{ik} \partial b_{i(k-1)}}\right) = \sum_{g=1}^G a_i^2 w_{gik} w_{gi(k-1)} (P_{gik}^o)^{-1}, \quad (30)$$

and

$$E\left(-\frac{\partial^2 L}{\partial b_{ik} \partial b_{i(k+1)}}\right) = \sum_{g=1}^G a_i^2 w_{gik} w_{gi(k+1)} [P_{gi(k+1)}^o]^{-1}, \quad (31)$$

where

θ_g is the quadrature point on the θ scale,

f_g is the number of examinees associated with the quadrature point based on the quadrature weight and the total number of examinees,

$w_{gik} = P_{ik}(\theta_g) Q_{ik}(\theta_g)$, and

$P_{gik}^o = P_{ik}^o(\theta_g)$.

MULTILOG's 20 quadrature points and associated quadrature weights were used here. The resulting 4×4 variance-covariance matrix can be obtained as:

$$\Sigma_i = \begin{bmatrix} E\left(-\frac{\partial^2 L}{\partial a_i^2}\right) & E\left(-\frac{\partial^2 L}{\partial a_i \partial b_{i1}}\right) & E\left(-\frac{\partial^2 L}{\partial a_i \partial b_{i2}}\right) & E\left(-\frac{\partial^2 L}{\partial a_i \partial b_{i3}}\right) \\ E\left(-\frac{\partial^2 L}{\partial a_i \partial b_{i1}}\right) & E\left(-\frac{\partial^2 L}{\partial b_{i1}^2}\right) & E\left(-\frac{\partial^2 L}{\partial b_{i1} \partial b_{i2}}\right) & E\left(-\frac{\partial^2 L}{\partial b_{i1} \partial b_{i3}}\right) \\ E\left(-\frac{\partial^2 L}{\partial a_i \partial b_{i2}}\right) & E\left(-\frac{\partial^2 L}{\partial b_{i1} \partial b_{i2}}\right) & E\left(-\frac{\partial^2 L}{\partial b_{i2}^2}\right) & E\left(-\frac{\partial^2 L}{\partial b_{i2} \partial b_{i3}}\right) \\ E\left(-\frac{\partial^2 L}{\partial a_i \partial b_{i3}}\right) & E\left(-\frac{\partial^2 L}{\partial b_{i1} \partial b_{i3}}\right) & E\left(-\frac{\partial^2 L}{\partial b_{i2} \partial b_{i3}}\right) & E\left(-\frac{\partial^2 L}{\partial b_{i3}^2}\right) \end{bmatrix}^{-1}. \quad (32)$$

References

- Baker, F. B. (1986). *GENIRV: A program to generate item response vectors* [Computer program]. Madison: University of Wisconsin, Laboratory of Experimental Design.
- Baker, F. B. (1992a). Equating tests under the graded response model. *Applied Psychological Measurement*, 16, 87-96.
- Baker, F. B. (1992b). *Item response theory: Parameter estimation techniques*. New York: Marcel Dekker.
- Baker, F. B., Al-Karni, A., & Al-Dosary, I. M. (1991). EQUATE: A computer program for the test characteristic curve method of IRT equating. *Applied Psychological Measurement*, 15, 78.
- Bejar, I. I., & Wingersky, M. S. (1981). *An application of item response theory to equating the Test of Standard Written English* (College Board Report No. 81-8). Princeton NJ: Educational Testing Service.
- Cook, L. L., Eignor, D. R., & Hutten, L. R. (1979, April). *Considerations in the application of latent trait theory to objectives-based criterion-referenced tests*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Divgi, D. R. (1980, April). *Evaluation of scales for multilevel test batteries*. Paper presented at the annual

- meeting of the American Educational Research Association, Boston.
- Divgi, D. R. (1985). A minimum chi-square method for developing a common metric in item response theory. *Applied Psychological Measurement*, 9, 413-415.
- Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, 22, 144-149.
- Kim, S.-H., & Cohen, A. S. (1992). Effects of linking methods on detection of DIF. *Journal of Educational Measurement*, 29, 51-66.
- Linn, R. L., Levine, M. V., Hastings, C. N., & Wardrop, J. L. (1981). Item bias in a test of reading comprehension. *Applied Psychological Measurement*, 5, 159-173.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale NJ: Erlbaum.
- Loyd, B. H., & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, 17, 179-193.
- Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, 14, 139-160.
- Mislevy, R. J., & Bock, R. D. (1990). *BILOG 3: Item analysis and test scoring with binary logistic models* [Computer program]. Mooresville IN: Scientific Software.
- Muraki, E., & Bock, R. D. (1993). *PARSCALE: IRT based test scoring and item analysis for graded open-ended exercises and performance tasks* [Computer program]. Chicago: Scientific Software.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, No. 17.
- Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph Supplement*, No. 18.
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201-210.
- Thissen, D. (1991). *MULTILOG user's guide: Multiple, categorical item analysis and test scoring using item response theory* [Computer program]. Chicago: Scientific Software.
- Vale, C. D. (1986). Linking item parameters onto a common scale. *Applied Psychological Measurement*, 10, 333-344.
- Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). *LOGIST user's guide*. Princeton NJ: Educational Testing Service.

Acknowledgments

The authors express their gratitude to Frank B. Baker for making his EQUATE and GENIRV programs available for this study. The authors also thank the Editor and anonymous reviewers for their insightful comments and suggestions.

Authors' Addresses

Send requests for reprints or further information to Seock-Ho Kim, The University of Georgia, 325 Aderhold Hall, Athens GA 30602, U.S.A., or Allan S. Cohen, Testing and Evaluation Services, University of Wisconsin, 1025 West Johnson Street, Madison WI 53706, U.S.A. Internet: skim@bob.coe.uga.edu or cohen@tne.edsci.wisc.edu.

47:4809