Journal of Educational Measurement
Winter 2000, Vol. 37, No. 4, pp. 357–373

Multidimensional Linking: Four Practical Approaches

T. C. Oshima
Georgia State University
T. C. Davey
Educational Testing Service
K. Lee
S.I.T.A.

This paper introduces and evaluates four practical multidimensional linking procedures that are based on the theoretical framework recently proposed by Davey, Oshima, and Lee (1996): (a) the Direct method, (b) the Equated Function method, (c) the Test Characteristic Function (TCF) method, and (d) the Item Characteristic Function (ICF) method. The evaluation was conducted using simulated data. As anticipated, the competing procedures yielded different linking parameter estimates. The TCF and ICF methods were found to be more stable and recovered the true linking parameters better than the other two methods. Furthermore, all procedures were found to be acceptable under almost any of the minimization criteria and offered dramatic improvement over not linking at all. It is recommended that the choice of a linking procedure should depend on the purpose of linking.

Many applications of item response theory (IRT), such as test equating or differential item functioning (DIF) analysis, depend on successful linking procedures. Various linking procedures have been introduced and used in the context of unidimensional IRT (e.g., Divgi, 1985; Haebara, 1980; Stocking & Lord, 1983). In the case of multidimensional linking, however, only a few studies have been conducted and its applications have been minimal. Many researchers agree that educational and psychological test data are often multidimensional. In such cases, one viable option is to use multidimensional IRT models instead of unidimensional IRT models. However, the use of multidimensional IRT in practice has been limited in part because of the lack of appropriate multidimensional linking procedures.

Davey (1991) introduced a theoretical background of multidimensional linking. Davey, Oshima, and Lee (1996) offered theoretical frameworks for three multidimensional linking procedures. An additional multidimensional linking procedure is introduced in this paper. The purpose of the research described here was to investigate empirically these procedures using simulated data. The simulation studies are carried out using a new computer program for multidimensional linking called IPLINK,[1] developed by Lee and Oshima (1996).

The authors thank the editors and anonymous reviewers for their helpful comments. Special thanks are due to Dr. Nambury Raju, Distinguished Professor at the Illinois Institute of Technology, for his careful review of the manuscript.

Multidimensional Linking

According to a multidimensional extension of the two-parameter logistic (M2PL) model (McKinley & Reckase, 1983), the probability of success on item i for an examinee j can be written as

$$P_i(\boldsymbol{\theta}_j) = \frac{1}{1 + e^{-1.7(\mathbf{a}'_i \boldsymbol{\theta}_j + d_i)}}, \tag{1}$$

where $\mathbf{a}_i$ is an $m \times 1$ vector of item discrimination parameters, d_i is a scalar parameter related to the difficulty of the item, $\boldsymbol{\theta}_j$ is an $m \times 1$ vector of ability parameters for the examinee, and m is the number of ability dimensions. Equation 1 is commonly referred to as the item response function (IRF). In this compensatory model, there is m discrimination parameters but only one difficulty parameter as a composite parameter representing the difficulty of the item. Therefore, being low on one trait can be compensated for by being high on another trait.

In the framework of IRT, the scale is determined only up to a linear transformation. Recall that for the unidimensional IRT models with the exponent expressed as $a_i(\theta_j - b_i)$, where a_i is the discrimination parameter and b_i is the difficulty parameter, the probability of correct response is not altered by the following transformations (Hambleton, Swaminathan, & Rogers, 1991):

$$a_i^* = \frac{a_i}{\alpha} \tag{2}$$

$$b_i^* = \alpha b_i + \beta \tag{3}$$

$$\theta_j^* = \alpha \theta_j + \beta, \tag{4}$$

where α is the multiplicative linking coefficient and β is the additive linking coefficient. Mathematically, it can be shown that $a_i^*(\theta_j^* - b_i^*) = \frac{a_i}{\alpha}\left((\alpha\theta_j + \beta) - \left(\alpha b_i + \beta\right)\right) = \frac{a_i}{\alpha}(\alpha(\theta_j - b_i)) = a_i(\theta_j - b_i)$. This property of IRT, known as model indeterminancy, also extends to multidimensional models. For the multidimensional models with the exponent expressed as $\mathbf{a}'_i\boldsymbol{\theta}_j + d_i$, extensions of the above three equations are:

$$\mathbf{a}_i^* = (A^{-1})'\mathbf{a}_i \tag{5}$$

$$d_i^* = d_i - \mathbf{a}'_i A^{-1}\boldsymbol{\beta} \tag{6}$$

$$\boldsymbol{\theta}_j^* = A\boldsymbol{\theta}_j + \boldsymbol{\beta}, \tag{7}$$

where the $m \times m$ rotation matrix A adjusts the variances and covariances of the ability dimensions (scale), and the $m \times 1$ translation vector $\boldsymbol{\beta}$ alters the means (location). As was shown in the unidimensional case, model indeterminancy in the multidimensional case can be mathematically shown as $\mathbf{a}'^*_i\boldsymbol{\theta}^*_j + d^*_i = ((A^{-1})'\mathbf{a}_i)'(A\boldsymbol{\theta}_j + \boldsymbol{\beta}) + (d_i - \mathbf{a}'_i A^{-1}\boldsymbol{\beta}) = \mathbf{a}'_i\boldsymbol{\theta} + \mathbf{a}'_i A^{-1}\boldsymbol{\beta} + d_i - \mathbf{a}'_i A^{-1}\boldsymbol{\beta} = \mathbf{a}'_i\boldsymbol{\theta}_j + d_i$.

The role of linking is to place the item parameter estimates from separate calibrations on a common ability metric. Assuming that two nonequivalent groups of examinees took some common items, the goal is to make the two sets of item parameter estimates as "similar" as possible by transforming the item parameter estimates from one of the groups to the metric underlying the item parameter estimates from the other group using the above equations. The two sets of item parameter estimates can be "similar" but not identical because two sets differ not only due to model indeterminancy but also due to random error. Depending on how one defines "similarity" between two sets of item parameters, different values of $\boldsymbol{A}$ matrices and $\boldsymbol{\beta}$ vectors may be found. Each of the following linking methods is based on the minimization of a targeted criterion function f with respect to $\boldsymbol{A}$ and $\boldsymbol{\beta}$. The description of each linking method assumes that item parameter estimates from the first sample are taken as defining as the base metric. Then, the intent of linking is to transform the item parameter estimates from the second sample so that they are similar to those from the first. That is, with two sets of estimated parameters ($\mathbf{a}_{1i}$, d_{1i}) and ($\mathbf{a}_{2i}$, d_{2i}) where the first subscript indicates the group, the second set ($\mathbf{a}_{2i}$, d_{2i}) is transformed into ($\mathbf{a}^*_{2i}$, d^*_{2i},) using Equations 5 and 6. The goal is to make ($\mathbf{a}^*_{2i}$, d^*_{2i},) as similar as possible to ($\mathbf{a}_{1i}$, d_{1i}) by choosing the "right" ($\boldsymbol{A}$, $\boldsymbol{\beta}$).

The Direct Method

This procedure is the multivariate extension of a unidimensional method offered by Divgi (1985). Linking parameters are estimated by minimizing the sum of squared differences between corresponding elements of $\mathbf{a}_{1i}$ and $\mathbf{a}^*_{2i}$, d_{1i} and d^*_{2i}, respectively, for all i. The function to be minimized with respect to ($\boldsymbol{A}$, $\boldsymbol{\beta}$) is as follows:

$$f_1(\boldsymbol{A}, \boldsymbol{\beta}) = \frac{1}{n\,(m+1)}\left(\sum_{i=1}^{n}\sum_{k=1}^{m}(a_{1ik} - a^*_{2ik})^2 + \sum_{i=1}^{n}(d_{1i} - d^*_{2i})^2\right), \tag{8}$$

where n is the number of items.

The Equated Function Method

This is the multidimensional analog of the "b equating" commonly used in the unidimensional case. This method minimizes the sum of squared differences between functions defined on sets of selected elements of $\mathbf{a}_{1i}$ and $\mathbf{a}^*_{2i}$, d_{1i} and d^*_{2i}. The number of functions needed (p) is determined by the number of unknown values in $\boldsymbol{A}$ and $\boldsymbol{\beta}$. In the unidimensional case ($m = 1$), $p = 2$ since the unknown values are only two (α and β). Then the two functions are typically the combination of the means and/or the standard deviations of certain parameters. For example, as described in Kolen and Brennan (1995), the means and the standard deviations of the common item difficulty parameters are set equal across calibration to estimate the α and β values (referred to as the mean/sigma method), or the means of the discrimination parameter and the means of the difficulty parameter can be set equal to achieve the same purpose (referred to as the mean/mean method). In the two-dimensional case ($m = 2$), the number of functions is six ($p = 6$), that is, four values of the $2 \times 2\boldsymbol{A}$ matrix plus the two values of $2 \times 1\boldsymbol{\beta}$ vector. Suppose the mean is the chosen function, and let $\mu_1, \mu_2, \ldots, \mu_p$ be the means of p separate sets of elements of $\mathbf{a}_{1i}$ and d_{1i}. Let $\mu^*_1, \mu^*_2, \ldots, \mu^*_p$ be the same set of functions defined on the transformed parameters. Then

$$f_2(\boldsymbol{A}, \boldsymbol{\beta}) = \frac{1}{p} \sum_{i=1}^{p} (\mu_p - \mu_p^*)^2. \tag{9}$$

The choice of the type of function and the method of blocking the items into p sets of groups are arbitrary. An example of a two-dimensional case is given in the method section.

The Test Characteristic Function (TCF) Method

This procedure extends Stocking and Lord's (1983) procedure that minimized the sum of squared differences between unidimensional test characteristic curves:

$$f_3(\boldsymbol{A}, \boldsymbol{\beta}) = \frac{1}{q} \sum_{\boldsymbol{\theta}} W_{\boldsymbol{\theta}} \left(\sum_{i=1}^{n} P_{1i}(\boldsymbol{\theta}) - \sum_{i=1}^{n} P_{2i}^*(\boldsymbol{\theta}) \right)^2, \tag{10}$$

where $P(\bullet)$ is the IRF in Equation 1 with corresponding parameter sets ($\mathbf{a}_{1i}$, d_{1i}) and ($\mathbf{a}_{2i}^*$, d_{2i}^*), and the outside summation approximates the multiple integral over the ability space $\boldsymbol{\theta}$. The q is the number of matching $\boldsymbol{\theta}$ vectors. The $W_{\boldsymbol{\theta}}$ are allowed differentially to weight differences taken at different $\boldsymbol{\theta}$ values to recognize that some regions of the $\boldsymbol{\theta}$ scale are more important than others (i.e., the weighted version), or the weight may be considered equal (for example $W_{\boldsymbol{\theta}} = 1$) in the unweighted version, as is typically done in the unidimensional case.

The Item Characteristic Function (ICF) Method

This method is similar to the TCF method, except that it minimizes the squared differences betwen item rather than test characteristic functions. A unidimensional procedure was proposed by Haebara (1980); the multidimensional extension is as follows:

$$f_4(\boldsymbol{A}, \boldsymbol{\beta}) = \frac{1}{n \bullet q} \sum_{\boldsymbol{\theta}} W_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left(P_{1i}(\boldsymbol{\theta}) - P_{2i}^*(\boldsymbol{\theta}) \right)^2. \tag{11}$$

All but the last of the above methods are described in more detail in Davey, Oshima, and Lee (1996). None, however, has been implemented for the multidimensional case. The following section describes the implementation of these methods.

Minimization Methods

In the first two methods, after proper transformation of the input variables ($\boldsymbol{A}$, $\boldsymbol{\beta}$), the functions to be minimized can be changed to linear functions of the transformed variables. Therefore, in essence the minimization process becomes equivalent to that for finding the solution for a linear function. For example, the minimization problem in Equation 8 is equivalent to the following multivariate linear problem:

$$\mathbf{a}_{1i} = \mathbf{a}^*_{2i} + error \text{ } and \text{ } d_{1i} = d^*_{2i} + error. \tag{12}$$

After substituting $\mathbf{a}^*_{2i}$ and d^*_{2i} above from Equations 5 and 6, respectively, and making the parameter transformation,

$$\boldsymbol{M} = \boldsymbol{A}^{-1} \text{ and } \boldsymbol{e} = -\boldsymbol{A}^{-1}\boldsymbol{\beta}, \tag{13}$$

we turn Equation 12 into the following multivariate linear model with ($\boldsymbol{M}$, $\boldsymbol{e}$) being the parameters to be estimated,[2] which can be readily transformed back to ($\boldsymbol{A}$, $\boldsymbol{\beta}$):

$$(\mathbf{a}_{1i}', d_{1i} - d_{2i}) = \mathbf{a}_{2i}'(\boldsymbol{M}, \boldsymbol{e}) + error. \tag{14}$$

In the second method (Equation 9), since μ is a linear function of ($\boldsymbol{a}_i$, d_i), a similar approach can be taken to solve the minimization problem using multivariate linear model techniques.

In the final two methods, however, the functions to be minimized are nonlinear functions of ($\boldsymbol{A}$, $\boldsymbol{\beta}$). Therefore, a weighted nonlinear least squares method is used to obtain the estimate ($\boldsymbol{A}$, $\boldsymbol{\beta}$) in this case. More precisely, a second-derivative (modified Newton) method is used. In the cases of the last two methods, $X = (\boldsymbol{A}, \boldsymbol{\beta})$, for a given function $F(X)$, the following is a simplified iterative procedure for finding the minimum of $F(X)$:

1. Start with initial point X^0.
2. Calculate the next $X^{n+1} = X^n + \rho\boldsymbol{\delta}^n$, where $\boldsymbol{\delta}^n$ is the direction and ρ (normally between 0 and 1) is the step length. There are many ways of choosing $\boldsymbol{\delta}^n$ and ρ such that $F(X^{n+1})$ is smaller than $F(X^n)$. We adopted the Goldfeld, Quardt, and Trotter (1966) approach: $\boldsymbol{\delta}^n = -(H^n + vI)^{-1}g^n$, where H^n is the second derivative of $F(X)$ at X^n, I is an identity matrix, and g^n is the first derivative of $F(X)$ at $\mathbf{X}^n$. In addition, v is a scalar variable and is chosen so that $(H^n + vI)$ is a positive definite matrix.
3. If, by predetermined criteria, $F(X^{n+1})$ is close enough to $F(X^n)$, or g^{n+1} is close to zero, stop; otherwise repeat Step 2.

Method

Although the methods described above can be used with any number of dimensions, only the simplest multidimensional case ($m = 2$) will be considered in the following section. The four linking procedures were evaluated through application to simulated data. Each simulation used item parameters estimated from real data as generating parameters to create responses to a 40-item, two-dimensional test. These parameters are estimates from a 1992 form of the ACT Assessment Mathematics test. Table 1 shows the item parameters. For each simulation, data from two examinee groups were simulated, with the first group arbitrarily chosen as defining the base ability metric onto which parameters estimated from the second group were to be scaled. Item parameters were then calibrated from the two groups

TABLE 1
Item Parameters Used for Data Generation

Item	a_1	a_2	d
1	2.37	0.00	2.49
2	0.58	0.38	1.00
3	0.63	0.27	0.66
4	0.99	0.47	0.76
5	0.60	0.18	0.29
6	0.77	0.64	−0.01
7	0.75	0.41	0.57
8	1.64	0.15	1.24
9	1.04	0.52	0.61
10	1.26	0.51	0.23
11	1.17	0.20	1.11
12	1.26	0.39	0.92
13	1.71	0.48	0.24
14	0.69	0.91	−0.68
15	0.57	0.72	−0.43
16	0.33	0.43	−0.28
17	2.10	0.69	0.59
18	1.19	1.16	−0.99
19	0.63	0.40	−0.21
20	1.11	1.31	−0.69
21	1.02	1.18	−0.04
22	0.96	1.26	−0.49
23	0.60	0.87	−0.60
24	1.01	0.47	−0.15
25	0.83	0.79	−1.47
26	0.81	0.77	−1.08
27	0.87	0.89	−0.97
28	1.71	1.76	−0.07
29	1.12	1.16	−0.84
30	0.93	1.38	−1.20
31	0.79	1.36	−0.90
32	1.87	1.52	−1.35
33	0.60	0.48	−0.61
34	0.44	0.40	−0.96
35	1.15	2.15	−1.89
36	0.63	0.78	−0.87
37	0.53	0.97	−1.31
38	0.31	0.99	−1.53
39	0.57	2.27	−3.77
40	0.56	0.46	−0.82
Mean	0.96	0.80	−0.34
SD	0.48	0.53	1.07

independently using NOHARM (Fraser, 1987). Finally, the second (scaled) group estimates were linked to those of the base group by each of the four linking procedures. This was done using the Windows-based computer program, IPLINK (Lee & Oshima, 1996).

In the Equated Function method (Equation 9), the six equated functions used in this study are as follows:

$$\mu_1 = \sum_{i=1}^{20} \mathbf{a}_{1i},\ \mu_2 = \sum_{i=21}^{40} \mathbf{a}_{1i},\ \mu_3 = \sum_{i=1}^{20} \mathbf{a}_{2i},\ \mu_4 = \sum_{i=21}^{40} \mathbf{a}_{2i},\ \mu_5 = \sum_{i=1}^{20} d_i,\ \mu_6 = \sum_{i=21}^{40} d_i.$$

For the TCF method (Equation 10), seven equally spaced θ_1 points from -4 to 4 and seven equally spaced θ_2 points from -4 to 4, making 49 (7×7) grid points, were used to calculate the difference between the two TCFs. All weights were set equal to unity. For the ICF method (Equation 11), the same grid used in the TCF method was used for $\boldsymbol{\theta}$ vectors. Again, unit weights were used.

Data were generated under each of six conditions that varied the difference between the ability distributions of the base and scaled examinee groups. These conditions allowed the means, variances, and correlations between the ability dimensions to differ across groups as follows: (1) No difference in the base and scaled groups (the null condition), (2) differences in $\boldsymbol{\theta}$ variances, (3) differences in $\boldsymbol{\theta}$ correlations, (4) differences in $\boldsymbol{\theta}$ means, (5) differences in $\boldsymbol{\theta}$ means and variances, and (6) differences in $\boldsymbol{\theta}$ means and dimension correlations.

In all conditions, correlated θ_1 and θ_2 were simulated by first generating two independent, normally distributed pseudorandom variables z_1 and z_2 and then transforming them to θ_1 and θ_2 by weighted linear transformations. The weights were the elements of $\mathbf{T}'$, a matrix which satisfied $\mathbf{R} = \mathbf{T}'\mathbf{T}$, where $\mathbf{R}$ is the target correlation matrix. The base group was held constant across conditions, with a mean vector of (0, 0) and a variance covariance matrix of $\begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$. The mean vector used for the scaled group in Conditions 4, 5, and 6 was $(-.5, -.5)$, while the variances in Conditions 2 and 5 were (.8, .8). The correlation was set equal to .7 in conditions 3 and 6. Because of the difference in correlations, Conditions 3 and 6 theoretically required an oblique rotation matrix, while Conditions 2 and 5 needed only an orthogonal axes transformation. Table 2 summarizes the six conditions.

Each simulation condition was replicated twenty times to produce a distribution of linking parameter estimates. The means and standard deviations of these distributions were computed to determine the stability of each procedure. Then the relationships among the item parameter estimates from different methods were explored. Next, the recovery of item parameter estimates was investigated by comparing the estimated parameters and the true parameters. Finally, the similarity of the two sets of item parameter estimates was evaluated by different criteria.

Results

Table 3 shows means and standard deviations of the estimates of linking coefficients obtained from four different minimizations over 20 replications. The 2×2 rotation matrix $\boldsymbol{A}$ is expressed in its elements $\begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix}$. The 2×1 translation vector $\boldsymbol{\beta}$ is expressed in its elements $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$.-

TABLE 2
Mean Vector $\boldsymbol{\theta}$ and Variance-Covariance Matrix Σ of the Ability Distribution for the Scaled Group

Condition											
C1		C2		C3		C4		C5		C6	
$\boldsymbol{\theta}$	Σ	$\boldsymbol{\theta}$	Σ	$\boldsymbol{\theta}$	Σ	$\boldsymbol{\theta}$	Σ	$\boldsymbol{\theta}$	Σ	$\boldsymbol{\theta}$	Σ
$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}1 & .5\\.5 & 1\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}.8 & .4\\.4 & .8\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}1 & .7\\.7 & 1\end{bmatrix}$	$\begin{bmatrix}-.5\\-.5\end{bmatrix}$	$\begin{bmatrix}1 & .5\\.5 & 1\end{bmatrix}$	$\begin{bmatrix}-.5\\-.5\end{bmatrix}$	$\begin{bmatrix}.8 & .4\\.4 & .8\end{bmatrix}$	$\begin{bmatrix}-.5\\-.5\end{bmatrix}$	$\begin{bmatrix}1 & .7\\.7 & 1\end{bmatrix}$

Note: The mean vector and variance-covariance matrix for the base group is always

$$\begin{bmatrix}0\\0\end{bmatrix}\begin{bmatrix}1 & .5\\.5 & 1\end{bmatrix}.$$

As expected, different minimization procedures yielded different linking coefficients within each condition. However, also as expected, the general trend over Conditions 1–6 was the same for all the procedures. Several observations can be made in Table 3. Across all the conditions, the standard deviations of the TCF and ICF methods were smaller than those of the Direct or Equated Function method, suggesting the TCF and ICF methods produced more stable estimates than the other two methods. Furthermore, the ICF method tended to produce slightly smaller standard deviations than the TCF method did, again indicating the superior stability for the ICF method.

Table 4 shows the correlation and mean absolute difference (MAD) of linking coefficients estimates across different methods. The correlation and MAD were calculated for 120 (20 replications × 6 conditions) linking situations. As anticipated, the Direct method and Equated Function method tended to produce similar estimates, and the TCF and the ICF methods tended to produce similar estimates as indicated by both the correlation (.72−.96) and MAD (.02−.13) indices. For two distinct methods (i.e., the Direct method or the Equated Function method as opposed to the TCF method or the ICF method), the correlation ranged from .45 to .92 and MAD ranged from .05 to .23. It is noteworthy here that differences in estimates expressed by MAD are not trivial, indicating that estimates obtained from each method can be quite different in a given situation. This is not unexpected, because each method minimizes a different operational definition of "similarity."

Table 5 shows bias and root mean square error (RMSE) of the estimated versus true linking coefficients. It should be noted that, because of scale indeterminancy, the direct comparison of those coefficients is not possible in the context of multidimensional linking. Therefore, several steps were taken to make the indirect comparison possible. (A detailed explanation is given in the Appendix.) First, the true rotation matrix was calculated using the original distributions of the two groups (Σ_1 and Σ_2). By defining $\Sigma_1 = \boldsymbol{B}_1\boldsymbol{B}_1'$ and $\Sigma_2 = \boldsymbol{B}_2\boldsymbol{B}_2'$, and further by limiting $\boldsymbol{B}_1$ or $\boldsymbol{B}_2$ to be a lower triangular matrix with the upper right corner element of zero, the unique rotation matrix was calculated to be $\boldsymbol{B}_1^{-1}\boldsymbol{B}_2$. The estimated rotation matrix was also transformed to have the same restriction (upper right corner element of zero). For

TABLE 3
Estimates of Linking Coefficients Obtained From Four Different Minimization Procedures

Condition	*A* β	Direct M	Direct SD	Equated Function M	Equated Function SD	TCF M	TCF SD	ICF M	ICF SD
C1	α_1	.98	(.13)	.98	(.14)	.99	(.07)	1.08	(.06)
	α_2	−.04	(.07)	−.02	(.10)	−.00	(.07)	−.01	(.04)
	α_3	.10	(.26)	.06	(.23)	.02	(.13)	−.05	(.10)
	α_4	1.13	(.22)	1.07	(.20)	1.02	(.10)	1.07	(.08)
	β_1	.00	(.09)	−.00	(.08)	−.00	(.09)	.07	(.07)
	β_2	−.08	(.17)	−.04	(.14)	−.02	(.13)	−.11	(.08)
C2	α_1	.92	(.11)	.91	(.09)	.89	(.10)	.98	(.09)
	α_2	−.00	(.08)	.04	(.09)	.02	(.06)	.00	(.06)
	α_3	−.12	(.19)	−.10	(.15)	−.03	(.11)	−.10	(.10)
	α_4	.87	(.16)	.81	(.14)	.88	(.05)	.94	(.05)
	β_1	.02	(.10)	−.01	(.06)	−.02	(.09)	.08	(.06)
	β_2	.01	(.15)	.05	(.11)	−.04	(.13)	−.11	(.08)
C3	α_1	1.01	(.28)	.96	(.19)	1.05	(.10)	1.09	(.11)
	α_2	−.09	(.13)	−.08	(.13)	−.01	(.08)	−.03	(.08)
	α_3	.29	(.42)	.27	(.30)	.12	(.17)	.10	(.15)
	α_4	1.25	(.26)	1.17	(.20)	1.05	(.11)	1.09	(.10)
	β_1	.12	(.11)	.08	(.08)	.11	(.08)	.11	(.07)
	β_2	−.24	(.20)	−.16	(.15)	−.17	(.12)	−.16	(.10)
C4	α_1	1.01	(.12)	.96	(.11)	.96	(.07)	1.08	(.06)
	α_2	.03	(.08)	.00	(.09)	.00	(.06)	.01	(.06)
	α_3	.11	(.22)	.10	(.21)	.01	(.10)	−.10	(.10)
	α_4	1.11	(.16)	1.05	(.15)	.99	(.07)	1.01	(.07)
	β_1	−.53	(.09)	−.50	(.12)	−.50	(.09)	−.45	(.07)
	β_2	−.55	(.25)	−.48	(.23)	−.43	(.16)	−.52	(.11)
C5	α_1	.86	(.10)	.88	(.10)	.83	(.09)	.98	(.09)
	α_2	.00	(.06)	.06	(.07)	−.01	(.04)	.00	(.03)
	α_3	.00	(.20)	−.04	(.17)	.03	(.11)	−.11	(.12)
	α_4	.86	(.17)	.75	(.14)	.84	(.10)	.88	(.09)
	β_1	−.52	(.10)	−.58	(.11)	−.55	(.11)	−.44	(.07)
	β_2	−.35	(.26)	−.25	(.24)	−.33	(.21)	−.53	(.12)
C6	α_1	1.08	(.24)	.94	(.14)	.93	(.08)	1.12	(.07)
	α_2	.02	(.10)	−.01	(.09)	.01	(.06)	.02	(.04)
	α_3	.29	(.41)	.32	(.32)	.17	(.13)	−.01	(.10)
	α_4	1.29	(.28)	1.16	(.24)	1.00	(.10)	1.04	(.09)
	β_1	−.43	(.16)	−.41	(.14)	−.50	(.10)	−.44	(.07)
	β_2	−.90	(.45)	−.74	(.37)	−.50	(.21)	−.61	(.13)

TABLE 4

Correlation (r) and Mean Absolute Difference (MAD) of Linking Coefficient Estimates Across Different Methods

		Equated Function		TCF		ICF	
		r	MAD	r	MAD	r	MAD
α_1	Direct	.76	.08	.45	.12	.52	.14
	Equated Function			.54	.09	.49	.13
	TCF					.74	.11
α_2	Direct	.72	.06	.71	.05	.74	.05
	Equated Function			.63	.07	.67	.06
	TCF					.90	.02
α_3	Direct	.91	.10	.70	.18	.61	.23
	Equated Function			.75	.15	.66	.19
	TCF					.78	.11
α_4	Direct	.92	.10	.77	.17	.71	.16
	Equated Function			.79	.13	.77	.13
	TCF					.91	.05
β_1	Direct	.96	.06	.90	.10	.92	.10
	Equated Function			.91	.10	.92	.11
	TCF					.96	.08
β_2	Direct	.96	.10	.70	.22	.72	.21
	Equated Function			.74	.18	.74	.19
	TCF					.84	.13

the transformation vector, the estimated value of $\boldsymbol{\beta}'\boldsymbol{\beta}$ (a scaler value) was compared to the corresponding true value.

Bias was then calculated by taking the mean of the differences between estimated and true values over 20 replications. This resulted in a 2×2 matrix of bias for the rotation matrix. To facilitate the understanding of results in Table 5, the 2×2 matrix was reduced to a scaler value by taking the mean of the absolute values of bias for the four elements. RMSE was calculated by taking the square root of the mean of the squared differences over 20 replications. Again, the 2×2 matrix of RMSE for the rotation matrix was reduced to a scaler value by taking the mean of the four elements.

In terms of the rotation matrix, it is quite evident that the TCF method had the least bias (except one condition) and least RMSE over other methods. For the translation vector, the ICF method ranked the best in half of the conditions for both bias and RMSE. The TCF method, however, ranked very close to the ICF method in most conditions.

As expected, bias and RMSE became larger as more differences were introduced into the rotation matrix and translation vector. However, it is encouraging to observe that bias and RMSE stayed in a reasonable range for the TCF and ICF methods across all the conditions including the most difficult condition, Condition 6.

TABLE 5

Bias and Root Mean Square Error (RMSE) of the Estimated versus True Rotation Matrix and Translation Vector

		Direct		Equated Function		TCF		ICF	
		Bias[1]	RMSE[2]	Bias	RMSE	Bias	RMSE	Bias	RMSE
C1	Rotation	.050	.144	.026	.122	.009	.060	.038	.076
	Translation	.042	.068	.028	.035	.024	.033	.029	.036
C2	Rotation	.045	.112	.042	.093	.009	.065	.057	.084
	Translation	.030	.042	.017	.027	.024	.034	.028	.038
C3	Rotation	.164	.270	.129	.203	.105	.135	.136	.160
	Translation	.127	.182	.058	.087	.058	.082	.051	.075
C4	Rotation	.066	.134	.046	.106	.015	.055	.045	.076
	Translation	.069	.237	−.030	.165	−.112	.144	−.092	.124
C5	Rotation	.015	.108	.040	.111	.036	.082	.054	.092
	Translation	−.120	.180	−.123	.176	−.127	.147	−.095	.154
C6	Rotation	.152	.286	.115	.205	.073	.108	.137	.152
	Translation	.630	1.051	.284	.606	−.046	.177	−.002	.161

[1] The bias value reported for the 2 × 2 rotation matrix is the mean of the absolute values of bias for the four elements.

[2] The RMSE reported for the 2 × 2 rotation matrix is the mean of RMSE for the four elements.

Although it is of theoretical interest to examine the estimated linking coefficients against the true linking coefficients, a more pressing issue is to examine how much each procedure minimized what it was supposed to minimize. Listed in Table 6 is the comparison of minimized function values evaluated by four different criteria, (f_1, f_2, f_3, and f_4), corresponding to Equations 8 through 11, respectively. The linking method used to calculate the linking coefficient estimates is listed in each column with an addition of "No Link". As in Table 3, reported values in Table 6 are the mean and the standard deviation of estimates over 20 replications.

For example, the first panel (Criterion $= f_1$) indicates how similar two sets of item parameters are in terms of the average squared difference of item parameter values after linking is performed using one of the four methods. See Equation 8 for the exact formula for f_1. The best (i.e., smallest) value in each row is indicated by the bold figures. As expected, the Direct method produced the smallest f_1 in all conditions when the criterion was f_1. For example, in Condition 1, the average squared difference of item parameters is .114 (see under No Linking) before linking. After linking by the Direct method, the difference was reduced to .069; with other linking methods, the difference was reduced to .076, .093, and .094 for the Equated Function, TCF, and ICF methods, respectively. Recall that Condition 1 was the null condition where linking was not necessary beyond adjusting for random errors. The more dramatic effect of linking can be observed in the remaining conditions. For example, in Condition 6, the average squared difference of item

TABLE 6
Comparison of Minimized Function Values Evaluated by Four Different Criteria

Evaluation Criterion		Direct	Equated Function	TCF	ICF	No Link
Criterion = f_1						
C1	Mean	**.069**	.076	.093	.094	.114
	S.D.	.047	.056	.067	.068	.084
C2	Mean	**.068**	.073	.089	.092	.121
	S.D.	.045	.046	.056	.060	.076
C3	Mean	**.065**	.075	.103	.101	.175
	S.D.	.036	.046	.081	.082	.156
C4	Mean	**.083**	.095	.126	.123	.747
	S.D.	.031	.037	.078	.073	.255
C5	Mean	**.106**	.114	.129	.133	.595
	S.D.	.106	.108	.124	.135	.119
C6	Mean	**.102**	.125	.207	.182	1.120
	S.D.	.062	.082	.147	.130	.574
Criterion = f_2						
C1	Mean	1.59	**9.29e-11**	2.47	3.12	6.53
	S.D.	2.57	4.31e-11	2.62	3.32	4.74
C2	Mean	.83	**1.20e-10**	2.48	4.17	13.40
	S.D.	.58	9.17e-11	3.10	4.04	8.23
C3	Mean	2.05	**1.16e-10**	3.05	3.13	15.20
	S.D.	2.64	1.04e-10	5.04	4.96	14.00
C4	Mean	2.63	**7.95e-11**	3.98	3.89	180.00
	S.D.	2.57	4.85e-11	5.23	4.53	39.00
C5	Mean	1.35	**1.29e-10**	3.31	5.03	155.00
	S.D.	1.04	9.52e-11	4.37	6.91	23.00
C6	Mean	4.93	**1.01e-10**	8.27	6.29	241.00
	S.D.	5.00	8.94e-11	7.36	6.17	93.40
Criterion = f_3						
C1	Mean	.38	.17	**.03**	.08	1.77
	S.D.	.35	.09	.02	.06	2.24
C2	Mean	.37	.21	**.03**	.09	3.08
	S.D.	.39	.26	.02	.04	4.07
C3	Mean	.40	.19	**.03**	.07	3.19
	S.D.	.44	.15	.01	.03	3.35
C4	Mean	.78	.30	**.04**	.12	24.00
	S.D.	.57	.23	.03	.06	4.87
C5	Mean	.49	.33	**.05**	.18	30.30
	S.D.	.27	.18	.02	.13	4.03
C6	Mean	1.35	.47	**.06**	.19	23.10
	S.D.	1.02	.28	.03	.10	2.91

Note: Direct through No Link columns are under the spanning header "Linking Method".

TABLE 6—continued
Comparison of Minimized Function Values Evaluated by Four Different Criteria

Evaluation Criterion		Linking Method: Direct	Equated Function	TCF	ICF	No Link
Criterion $= f_4$						
C1	Mean	.0040	.0039	.0035	**.0033**	.0053
	S.D.	.0010	.0009	.0008	.0007	.0026
C2	Mean	.0044	.0044	.0040	**.0037**	.0073
	S.D.	.0011	.0014	.0009	.0008	.0041
C3	Mean	.0044	.0042	.0038	**.0037**	.0075
	S.D.	.0013	.0012	.0009	.0009	.0042
C4	Mean	.0055	.0050	.0043	**.0041**	.0280
	S.D.	.0015	.0015	.0012	.0010	.0030
C5	Mean	.0062	.0068	.0060	**.0054**	.0338
	S.D.	.0015	.0021	.0020	.0014	.0045
C6	Mean	.0076	.0070	.0061	**.0055**	.0272
	S.D.	.0022	.0021	.0017	.0015	.0033

parameters before linking was 1.120. This difference was reduced to .102, .125, .207, and .182 for the Direct, Equated Function, TCF, and ICF methods respectively. It seems that the minimized value increases as two groups differ more; the increase, however, does not appear to be substantial given the function values under No Link condition. A graphic illustration of this is shown in Figure 1 for the criterion function f_1. Also evident in Figure 1 is the similarity between the first two methods and also the similarity between the last methods.

Similar observations can be made for the remaining criteria. As expected, when the evaluation criterion for similarity matched the procedure used for calculating the linking coefficients the smallest function values were obtained. In summary, each linking method minimized the function value that it was supposed to minimize and the other linking methods also minimized the function value to an acceptable degree.

Discussion

In the present study, four different minimization procedures were implemented in the computer program IPLINK to accomplish item parameter linking when two sets of two-dimensional data needed to be put on a common metric. Estimates of linking coefficients were compared and the recovery of these parameters was evaluated against the true parameters. Then the minimized function values were evaluated under several criteria of "similarity" of two data sets of item parameters.

As anticipated, the competing procedures yielded different linking parameter estimates. The TCF and ICF methods were found to be somewhat more stable than the Direct and Equated Function methods. Furthermore, the recovery of true linking parameters was better for the TCF and ICF methods. As expected, each procedure minimized what it was supposed to minimize. For example, by using the TCF

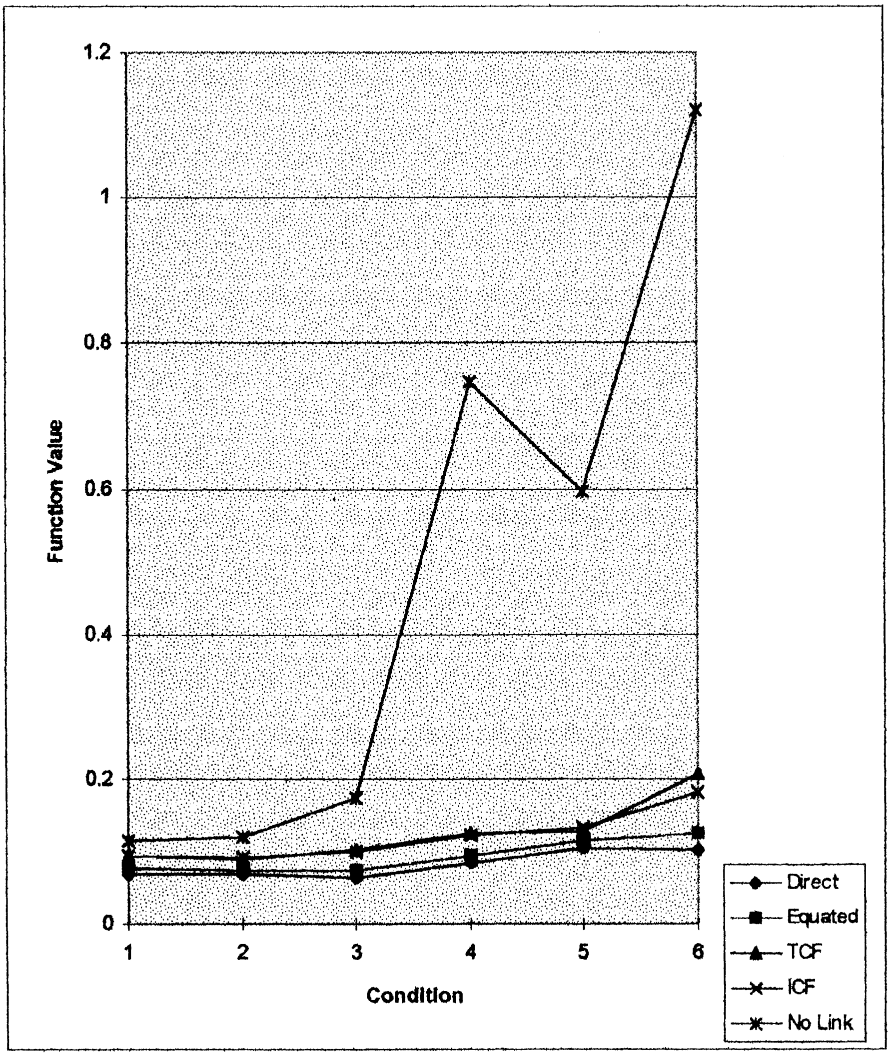

FIGURE 1. *Comparison of Function Values Under Criterion* f_1

method, the difference between two test characteristic functions was minimized the most as compared to the other four methods. However, the difference between the minimized function values across the four methods were fairly small, indicating that any of the four methods may be considered acceptable under any of the minimization criteria. All procedures offered dramatic improvements over the No Link condition.

Which, then, is the best procedure? The answer depends on the purpose of linking. In a test equating situation, where the main concern is with the equivalence of the examinees' true scores regardless of the test forms, the TCF method may be the best choice because the true score difference is minimized over the examinees. On the other hand, when one is interested in investigating differential item functioning (DIF), procedures based on item parameters or IRF may be more suitable. Consider the following examples:[3] Lord's chi-square method (Lord, 1980), in which item parameters were directly compared may benefit from the Direct method; Raju's area method (Raju, 1988), in which two item characteristic curves are compared may benefit from the ICF method; and the DFIT framework (Raju, van der Linden, & Fleer, 1995), in which DIF is derived from the difference in true scores may behave better with the TCF method. Furthermore, if one is interested in certain groups of items more than in others, the Equated Function method may be the most reasonable choice.

There are other choices to be made even after selecting a linking method. Suppose the TCF method was selected for use. Should one use weighted or unweighted TCF? How many grid points should one use? What is the range of thetas to be included? Again, the answers to these questions depend on the nature of linking. As a first study to investigate various multidimensional linking methods, the current simulation was confined to only the two-dimensional case and the number of replication was limited to 20. Now that a computer program (IPLINK) that performs various types of linking with various options is available for any m-dimensional data, further studies are needed to investigate the effect of various linking techniques on various applications of IRT in a wide range of conditions.

Notes

[1] IPLINK is a Windows-based program written in Turbo C++. The user creates an initial file on an Edit screen by simply selecting options such as the number of ability dimensions and the type of linking method as well as specifying the input and output files. The program is excuted by selecting Run from a pull-down menu. The program is available for downloading on the web page www.gsu.edu/~epstco.

[2] The $(\boldsymbol{M}, \boldsymbol{e})$ parameters were estimated using the general linear method: Suppose the general linear model is Y = X(BETA) + error, and the error has zero mean and identity covariance matrix. Then the least-squared estimate for BETA is $(X'X)^{-1}X'Y$.

[3] The examples are given in the unidimensional context, since multidimensional DIF techniques are still not fully developed. It should be noted that each of the procedures implemented here is a general procedure to handle m dimensional data, which certainly includes $m = 1$.

Appendix

The following section describes how the estimated and true linking coefficients were calculated and compared.

For a given set of item parameters (**a**, d), the probability functions or IRFs are

$$p_1 = \frac{1}{1 + e^{-1.7(\mathbf{a}'\boldsymbol{\theta}_1 + d)}},\ p_2 = \frac{1}{1 + e^{-1.7(\mathbf{a}'\boldsymbol{\theta}_2 + d)}}, \tag{A1}$$

where $\boldsymbol{\theta}_1$ has a normal distribution N $(\boldsymbol{\mu}_1, \Sigma_1)$ and $\boldsymbol{\theta}_2$ has a normal distribution N $(\boldsymbol{\mu}_2, \Sigma_2)$. In estimating (**a**, d) in NOHARM, $\boldsymbol{\theta}$ is assumed to have a standard normal distribution. Therefore, we first transfer the standard normal distribution $\boldsymbol{\theta}$ to $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ such that

$$\boldsymbol{\theta}_1 = \boldsymbol{B}_1\boldsymbol{\theta} + \boldsymbol{b}_1,\ \boldsymbol{\theta}_2 = \boldsymbol{B}_2\boldsymbol{\theta} + \boldsymbol{b}_2, \tag{A2}$$

where $\boldsymbol{B}_1$ satisfies $\Sigma_1 = \boldsymbol{B}_1\boldsymbol{B}_1'$, $\boldsymbol{B}_2$ satisfies $\Sigma_2 = \boldsymbol{B}_2\boldsymbol{B}_2'$, $\boldsymbol{b}_1 = \boldsymbol{\mu}_1$, and $\boldsymbol{b}_2 = \boldsymbol{\mu}_2$. $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ can be uniquely defined by the Cholesky decomposition. Resulting $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are lower triangular matrices.

Substituting $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in the probability functions P_1 and P_2, we obtain

$$p_1 = \frac{1}{1 + e^{-1.7(\mathbf{a}'\boldsymbol{B}_1\boldsymbol{\theta} + \mathbf{a}'\boldsymbol{b}_1 + d)}},\ p_2 = \frac{1}{1 + e^{-1.7(\mathbf{a}'\boldsymbol{B}_2\boldsymbol{\theta} + \mathbf{a}'\boldsymbol{b}_2 + d)}}. \tag{A3}$$

Thus, the item parameters we actually estimate are $(\mathbf{a}'\boldsymbol{B}_1, \mathbf{a}'\boldsymbol{b}_1 + d)$ and $(\mathbf{a}'\boldsymbol{B}_2, \mathbf{a}'\boldsymbol{b}_2 + d)$. Now we link the second set to the first set with the help of Equations 5 and 6 using linking coefficients $(\boldsymbol{A}, \boldsymbol{\beta})$. Thus,

$$\mathbf{a}'\boldsymbol{B}_1 = \mathbf{a}'\boldsymbol{B}_2\boldsymbol{A}^{-1},\ \mathbf{a}'\boldsymbol{b}_1 + d = \mathbf{a}'\boldsymbol{b}_2 + d - \mathbf{a}'\boldsymbol{B}_2\boldsymbol{A}^{-1}\boldsymbol{\beta}. \tag{A4}$$

Solving for $\boldsymbol{A}$ and $\boldsymbol{\beta}$,

$$\boldsymbol{A} = \boldsymbol{B}_1^{-1}\boldsymbol{B}_2,\ \boldsymbol{\beta} = \boldsymbol{B}_1^{-1}(\boldsymbol{b}_2 - \boldsymbol{b}_1). \tag{A5}$$

The estimate of $\boldsymbol{A}$ obtained from IPLINK, however, is not restricted to a lower triangular matrix. Therefore, such an estimate of $\boldsymbol{A}$ cannot be directly compared to $\boldsymbol{B}_1^{-1}\boldsymbol{B}_2$. To compare the estimate of $\boldsymbol{A}$ to a known value, the Cholesky decomposition $\boldsymbol{T}$ is obtained for the estimate of $\boldsymbol{A}\boldsymbol{A}'$ such that

$$\boldsymbol{T}\boldsymbol{T}' = \boldsymbol{A}\boldsymbol{A}' = \boldsymbol{B}_1^{-1}\boldsymbol{B}_2(\boldsymbol{B}_1^{-1}\boldsymbol{B}_2)'. \tag{A6}$$

Since $\boldsymbol{T}$, $\boldsymbol{B}_1$, and $\boldsymbol{B}_2$ are all uniquely obtained through the Cholesky decomposition, $\boldsymbol{T}$ then can be compared to $\boldsymbol{B}_1^{-1}\boldsymbol{B}_2$ (true rotation matrix).

As for the translation vector $\boldsymbol{\beta}$,

$$\begin{aligned}\boldsymbol{\beta}'\boldsymbol{\beta} &= (\boldsymbol{b}_2 - \boldsymbol{b}_1)'(\boldsymbol{B}_1^{-1})'\boldsymbol{B}_1^{-1}(\boldsymbol{b}_2 - \boldsymbol{b}_1) = (\boldsymbol{b}_2 - \boldsymbol{b}_1)'(\boldsymbol{B}_1\boldsymbol{B}_1')^{-1}(\boldsymbol{b}_2 - \boldsymbol{b}_1)\\ &= (\boldsymbol{b}_2 - \boldsymbol{b}_1)'\Sigma_1^{-1}(\boldsymbol{b}_2 - \boldsymbol{b}_1).\end{aligned} \tag{A7}$$

Here we can directly compare $\boldsymbol{\beta}'\boldsymbol{\beta}$ (a scaler value of estimated translation vector) to $(\boldsymbol{b}_2 - \boldsymbol{b}_1)'\Sigma_1^1(\boldsymbol{b}_2 - \boldsymbol{b}_1)$ (a scaler value of true translation vector), since Σ_1, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ are known.

References

Davey, T. C. (1991, June). *Some issues in linking multidimensional item calibrations.* Paper presented at the Office of Naval Research Contractors Meeting on Model-based Psychological Measurement, Princeton, NJ.

Davey, T. C., Oshima, T. C., & Lee, K. (1996). Linking multidimensional item calibrations. *Applied Psychological Measurement, 20*, 405–416.

Divgi, D. R. (1985). A minimum chi-square method for developing a common metric in item response theory. *Applied Psychological Measurement, 9*, 413–415.

Fraser, C. (1987). NOHARM: *An IBM PC computer program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory.* [Computer Program].

Center for Behavioral Studies, The University of New England, Armidale, New South Wales, Australia.

Goldfeld, S. M., Quardt, R. E., & Trotter, H. F. (1966). Maximization by quadratic hill climbing. *Econometrica, 34*, 541–551.

Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research, 22*, 144–149.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: SAGE Publications.

Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices.* New York: Springer-Verlag.

Lee, K., & Oshima, T. C. (1996). IPLINK: Multidimensional and unidimensional item parameter linking in item response theory. *Applied Psychological Measurement, 20*, 230.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

McKinley, R. L., & Reckase, M. D. (1983). *An extension of the two-parameter logistic model to the multidimensional latent space* (Research Report, ONR 83–2). Iowa City, IA: American College Testing Program.

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika, 53*, 495–502.

Raju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement, 19*, 353–368.

Stocking, M. L., & Lord, F. M. (1983) Developing a common metric in item response theory. *Applied Psychological Measurement, 7*, 201–210.

Authors

T. C. OSHIMA is Associate Professor of Research, Measurement, and Statistics in the Department of Educational Policy Studies at Georgia State University, 30 Pryor St., Suite 450, Atlanta, GA 30303-3083; *oshima@gsu.edu*. Her research interests include item response theory and differential item function.

TIM DAVEY is ETS is Analysis Leader at the Educational Testing Service, Mail stop 11-L, Rosedale Road, Princeton, NJ 08541. His research interests include psychometrics, item response theory, and computerized testing.

K. LEE is program analyst at S.I.T.A., 3100 Cumberland Circle, Suite 300, Atlantic, GA 30339. His research interests include software programming and statistics.

References

Maximization by Quadratic Hill-Climbing
Stephen M. Goldfeld; Richard E. Quandt; Hale F. Trotter
Econometrica, Vol. 34, No. 3. (Jul., 1966), pp. 541-551.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28196607%2934%3A3%3C541%3AMBQH%3E2.0.CO%3B2-K