Test Equating under the Multiple–Choice Model

Jee–Seon Kim

University of Illinois at Urbana–Champaign

Bradley A. Hanson

ACT, Inc.

May 12, 2000

Abstract

This paper presents a characteristic curve procedure for computing transformations of the item response theory (IRT) ability scale assuming the multiple–choice model (Thissen & Steinberg, 1984). The multiple–choice model provides a realistic and informative approach to analyzing multiple–choice items in two important ways. First, the probability of guessing is a decreasing function of ability rather than a constant across different ability levels as in the three–parameter logistic model. Second, the model utilizes information from incorrect answers as well as from correct answers. The multiple–choice model includes many well–known IRT models as special cases, such as Bock's (1972) nominal response model. Formulas needed to implement a characteristic curve method for scale transformation are presented for the multiple–choice model. Two moment methods of estimating a scale transformation for the multiple–choice model (the mean/mean and mean/sigma methods) are also presented. The use of the characteristic curve method for the multiple–choice model is illustrated in an example equating ACT mathematics tests, and is compared to the results from the mean/mean and mean/sigma methods. In the process of deriving the scale transformation procedure for the multiple–choice model, corrections were made in some of the formulas presented by Baker (1993) for computing a scale transformation for the nominal response model.

Index terms: equating, item response theory, multiple–choice model, IRT scale transformation, quadratic loss function.

To make scores and item parameters on two item response theory (IRT) scales comparable, a transformation is required from one metric to another. Equating tests under item response theory involves estimating the slope and intercept of a linear transformation that can be used to transform the item parameter estimates for the tests to a target metric. The procedures described in this paper can be used with data from a common–item nonequivalent groups equating design. In such a design two forms of a test are taken by separate samples of examinees from two populations. The two forms have a set of items in common. If item parameter estimates for the two forms based on an IRT model are separately estimated, then the item parameter estimates for the two forms will be on different IRT scales. The two sets of item parameter estimates for the common items are used to compute a slope and intercept that can be used to transform the item parameter estimates for one form to the same scale as the item parameter estimates for the other form.

Stocking and Lord's (1983) and Haebara's (1980) procedures provide estimates of a scale transformation using data from dichotomously scored items collected under a common item design for binary IRT models such as the Rasch, two–parameter logistic, and three–parameter logistic models. The procedures developed by Stocking and Lord (1983) and Haebara (1980) are characteristic curve transformation methods (Kolen & Brennan, 1995). Another set of procedures using moments of the distributions of item parameter estimates are available for computing scale transformations, although the general conclusion from research comparing characteristic curve and moment methods is that the characteristic curve methods produce more accurate results (Kolen & Brennan, 1995).

The nominal response (Bock, 1972), graded response (Samejima, 1969), and multiple–choice models (Thissen & Steinberg, 1984, 1997) can be used to model response curves for all alternatives of a multiple–choice item. Of these models the multiple–choice model is most appropriate for multiple–choice items because it allows for the possibility of non-zero response probabilities for all alternatives over the entire range of proficiency. It is reasonable to have non-zero response probabilities for all alternatives if it is possible some examinees would choose a response by guessing (the response to the item does not depend on the ability of the examinee).

Baker (1992, 1993) presented a characteristic curve scale transformation procedure for Samejima's (1969) graded response model and Bock's (1972) nominal response model. This study presents a characteristic curve scale transformation procedure for the multiple–choice model.

The remainder of this paper is structured as follows. First, the multiple–choice model is presented in the form given by Thissen & Steinberg (1984), and an alternative formulation is presented based on a latent response strategy interpretation. The reformulation implicitly denotes the relationship between the multiple–choice model and the nominal response model. Next, a characteristic curve scale transformation procedure and essential derivations for the transformation procedure under the multiple–choice model are presented. The following section revisits Baker's (1993) test equating under the nominal response model. It is shown that his formulas are not satisfied in general, and revised formulas are presented. In the following section, two moment methods of estimating a scale transformation for the multiple–choice model (the mean/sigma and mean/mean methods) are presented. An example of applying the scale transformations discussed in the paper is given using data from an ACT mathematics test. The paper concludes by summarizing advantageous features of ICC equating under the multiple–choice model.

Multiple–Choice Model

For multiple–choice model (Thissen & Steinberg, 1984) the probability of responding in response category k of item j as a function of proficiency (the item category response function) is

$$\mathbf{P}(\theta_i \mid a_{jk}, b_{jk}, d_{jk}) = \frac{\exp\left[a_{jk}\ \theta_i + b_{jk}\right] + d_{jk} \exp\left[a_{j0}\ \theta_i + b_{j0}\right]}{\sum_{h=0}^{m_j} \exp\left[a_{jh}\ \theta_i + b_{jh}\right]}, \tag{1}$$

where

$j = 1, 2, \ldots, J$ indexes the test items

$k = 1, 2, \ldots, m_j$ indexes the response categories for item j.

For identification, three constraints are imposed on the parameters in the multiple–choice model:

$$\sum_{k=0}^{m_j} a_{jk} = 0, \tag{2}$$

$$\sum_{k=0}^{m_j} b_{jk} = 0, \tag{3}$$

and

$$\sum_{k=1}^{m_j} d_{jk} = 1. \tag{4}$$

The multiple–choice model given in Equation 1 can also be written as

$$\mathbf{P}(\theta_i \mid a_{jk}, b_{jk}, d_{jk}) \;=\; \mathbf{C}_{j0}(\theta_i)\; d_{jk} \;+\; [1 - \mathbf{C}_{j0}(\theta_i)]\, \frac{\exp\left[a_{jk}\;\theta_i + b_{jk}\right]}{\sum_{h=1}^{m_j} \exp\left[a_{jh}\;\theta_i + b_{jh}\right]}\,, \tag{5}$$

where

$$\mathbf{C}_{j0}(\theta_i) = \frac{\exp\left[a_{j0}\;\theta_i + b_{j0}\right]}{\sum_{h=0}^{m_j} \exp\left[a_{jh}\;\theta_i + b_{jh}\right]}. \tag{6}$$

One way of interpreting the multiple–choice model using Equation 5 is to suppose there are two latent response strategies. For the first response strategy the probability of response k to item j is given by the nominal response model (Bock, 1972):

$$\mathbf{P}(\theta_i \mid a_{jk}, b_{jk}) = \frac{\exp\left[a_{jk}\;\theta_i + b_{jk}\right]}{\sum_{h=1}^{m_j} \exp\left[a_{jh}\;\theta_i + b_{jh}\right]}\,. \tag{7}$$

For the second response strategy the probability of response k to item j is d_{jk}, which does not depend on the latent proficiency. The first and second response strategies could be called non–guessing and guessing strategies, respectively. If the probability of an examinee with latent proficiency θ_i using the guessing strategy is $\mathbf{C}_{j0}(\theta_i)$, and the probability of the examinee using the non–guessing strategy is $1 - \mathbf{C}_{j0}(\theta_i)$, then the marginal probability over both response strategies of an examinee with latent proficiency θ_i giving response k to item j is given by Equation 5.

The multiple–choice model combines aspects of the three–parameter logistic (3PL) model (Lord, 1980) and the nominal response model (Bock, 1972), overcoming some drawbacks of these models in analyzing multiple–choice items. Unlike the 3PL model, the multiple–choice model maintains multiple category information without dichotomizing responses into binary scores. Like the 3PL model, the multiple–choice model allows for guessing by not requiring the lower asymptote of item category response functions for incorrect responses to be zero, although the guessing probability $\mathbf{C}_{j0}(\theta_i)$ in the multiple–choice model is constructed as a function of ability, not held constant over θ as in the 3PL model.

The nominal response model has been widely used in choice/preference data but is not as appropriate for multiple–choice items due to the fact that as the latent proficiency approaches negative infinity there will be one response for which the response function approaches one, and the response functions associated with all other responses will approach zero. This is not consistent with the possibility that examinees with low proficiencies could choose any of the responses by guessing.

For some applications of the multiple–choice model, equating is necessary. For example, equating would be needed in order to produce comparable scores from two different test forms. The purpose of this paper is to present a characteristic curve equating procedure for the multiple–choice model.

IRT Scale Transformation

Item parameter estimates for two different tests estimated using samples from different populations are on different IRT scales, say θ for a target scale and θ' for a current scale. The two scales are related by a linear transformation. A method of estimating the transformation from the current to the target scale involves finding the slope and intercept of a linear scale transformation function so as to minimize the discrepancy between the characteristic curves using parameter estimates on the target scale and characteristic curves using parameter estimates on the current scale *transformed* by this linear function.

The linear transformation from θ_i' to θ_i is given as

$$\theta_i = \mathrm{S}\ \theta_i' + \mathrm{I}\,, \tag{8}$$

A scale transformation is computed using two sets of item parameter estimates on a group of common items. The two sets of parameter estimates are obtained using samples from different populations. One set of item parameter estimates $(\hat{a}_{jk}, \hat{b}_{jk}, \hat{d}_{jk})$ are on the target scale, and the other set of item parameter estimates $(\hat{a}'_{jk}, \hat{b}'_{jk}, \hat{d}'_{jk})$ are on the current scale.

Assuming two tests used to obtain the two sets of item parameter estimates for the common items are developed under the multiple–choice model, transformations of item parameter estimates on a current θ' scale to the target θ scale using a slope and intercept S and I of the scale transformation in Equation 8 are

$$\hat{a}^*_{jk} = \frac{\hat{a}'_{jk}}{\mathrm{S}} \tag{9}$$

$$\hat{b}^*_{jk} = \hat{b}'_{jk} - \frac{\hat{a}'_{jk}}{\mathrm{S}}\mathrm{I} = \hat{b}'_{jk} - \hat{a}^*_{jk}\ \mathrm{I} \tag{10}$$

$$\hat{d}^*_{jk} = \hat{d}'_{jk}$$

The parameters d_{jk} are not affected by scale transformations. Note that since the two sets of parameter estimates contain sampling error it will not be the case that $\hat{a}^*_{jk} = \hat{a}_{jk}$, and $\hat{b}^*_{jk} = \hat{b}_{jk}$, for all j and k, even if S and I are the slope and intercept of the true scale transformation from the current to the target scale.

The characteristic curve method of estimating a scale transformation finds values of S and I that minimize the squared difference in the item category response functions using the parameter estimates on the target scale and the parameter estimates on the current scale transformed to the target scale:

$$\mathbf{F} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{n} \sum_{k=1}^{m_j} \left[\mathbf{P}(\theta_i \mid \hat{a}_{jk}, \hat{b}_{jk}, \hat{d}_{jk}) - \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk}) \right]^2 , \tag{11}$$

where

$i = 1, 2, \ldots, N$ indexes N arbitrary points values of θ on the target scale,

$j = 1, 2, \ldots, n$ indexes the anchor items common to the two tests, where $n \leq J$,

$k = 1, 2, \ldots, m_j$ indexes the response categories for item j, and

$M = \sum_{j=1}^{n} m_j$.

Finding values of S and I which minimize the loss function in Equation 11 requires gradients of the loss function with respect to S and I:

$$\frac{d\mathbf{F}}{d\mathrm{S}} = -\frac{2}{NM} \sum_{i=1}^{N} \sum_{j=1}^{n} \sum_{k=1}^{m_j} \left[\mathbf{P}(\theta_i \mid \hat{a}_{jk}, \hat{b}_{jk}, \hat{d}_{jk}) - \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk}) \right] \left(\frac{d}{d\mathrm{S}} \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk}) \right) ,$$

and

$$\frac{d\mathbf{F}}{d\mathrm{I}} = -\frac{2}{NM} \sum_{i=1}^{N} \sum_{j=1}^{n} \sum_{k=1}^{m_j} \left[\mathbf{P}(\theta_i \mid \hat{a}_{jk}, \hat{b}_{jk}, \hat{d}_{jk}) - \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk}) \right] \left(\frac{d}{d\mathrm{I}} \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk}) \right) .$$

Let u_{jk} and v_j be

$$u_{jk} = \exp\left[\hat{a}^*_{jk} \ \theta_i + \hat{b}^*_{jk} \right] + \hat{d}^*_{jk} \exp\left[\hat{a}^*_{j0} \ \theta_i + \hat{b}^*_{j0} \right]$$

and

$$v_j = \sum_{h=0}^{m_j} \exp\left[\hat{a}^*_{jh} \ \theta_i + \hat{b}^*_{jh} \right] .$$

Let λ_{jk} be

$$\lambda_{jk} = \exp\left[\hat{a}^*_{jk} \ \theta_i + \hat{b}^*_{jk} \right] .$$

Then

$$u_{jk} = \lambda_{jk} + \hat{d}^*_{jk}\lambda_{j0},$$

$$v_j = \sum_{h=0}^{m_j} \lambda_{jh},$$

and

$$\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk}) = \frac{u_{jk}}{v_j} = \frac{\lambda_{jk} + \hat{d}^*_{jk}\lambda_{j0}}{\sum_{h=0}^{m_j} \lambda_{jh}}.$$

The derivatives of $\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk})$ with respect to I and S are

$$\frac{d}{d\mathrm{I}}\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk}) = \frac{1}{v_j}\frac{du_{jk}}{d\mathrm{I}} - \frac{u_{jk}}{v_j^2}\frac{dv_j}{d\mathrm{I}}$$

and

$$\frac{d}{d\mathrm{S}}\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk}) = \frac{1}{v_j}\frac{du_{jk}}{d\mathrm{S}} - \frac{u_{jk}}{v_j^2}\frac{dv_j}{d\mathrm{S}}$$

By Equations 9 & 10,

$$\lambda_{jk} = \exp\left[\hat{a}^*_{jk}\ \theta_i + \hat{b}^*_{jk}\right] = \exp\left[\frac{\hat{a}'_{jk}}{\mathrm{S}}\ \theta_i + \hat{b}'_{jk} - \frac{\hat{a}'_{jk}}{\mathrm{S}}\ \mathrm{I}\ \right]$$

The derivative of $\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk})$ with respect to I is

$$\begin{aligned}\frac{d}{d\mathrm{I}}\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk}) &= \frac{1}{v_j}\left[\lambda_{jk}\left(-\frac{\hat{a}'_{jk}}{\mathrm{S}}\right) + \hat{d}^*_{jk}\ \lambda_{j0}\left(-\frac{\hat{a}'_{j0}}{\mathrm{S}}\right)\right] - \ \frac{u_{jk}}{v_j}\ \frac{1}{v_j}\ \sum_{h=0}^{m_j}\lambda_{jh}\ \left(-\frac{\hat{a}'_{jh}}{\mathrm{S}}\right) \\ &= -\frac{1}{\sum_{h=0}^{m_j}\lambda_{jh}}\ \left[\ \lambda_{jk}\ \hat{a}^*_{jk} + \hat{d}^*_{jk}\ \lambda_{j0}\ \hat{a}^*_{j0}\ -\ \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk})\ \sum_{h=0}^{m_j}\lambda_{jh}\ \hat{a}^*_{jh}\ \right]\end{aligned}$$

Similarly, the derivative of $\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk})$ with respect to S is

$$\frac{d}{d\mathrm{S}}\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk}) \ = \ \frac{1}{v_j}\left[\lambda_{jk}\left(\frac{-\hat{a}'_{jk}\theta_i + \hat{a}'_{jk}\mathrm{I}}{\mathrm{S}^2}\right) + \hat{d}^*_{jk}\ \lambda_{j0}\left(\frac{-\hat{a}'_{j0}\theta_i + \hat{a}'_{j0}\mathrm{I}}{\mathrm{S}^2}\right)\right]$$

$$
\begin{aligned}
&\quad - \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk}) \frac{1}{v_j} \sum_{h=0}^{m_j} \lambda_{jh} \left(\frac{-\hat{a}'_{jk}\theta_i + \hat{a}'_{jk}\mathrm{I}}{\mathrm{S}^2} \right) \\
&= \frac{1}{v_j} \left[\lambda_{jk} \left(\frac{\hat{a}'_{jk}}{\mathrm{S}} \right) \left(\frac{\mathrm{I} - \theta_i}{\mathrm{S}} \right) + \hat{d}^*_{jk} \ \lambda_{j0} \left(\frac{\hat{a}'_{j0}}{\mathrm{S}} \right) \left(\frac{\mathrm{I} - \theta_i}{\mathrm{S}} \right) \right] \\
&\quad - \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk}) \frac{1}{v_j} \sum_{h=0}^{m_j} \lambda_{jh} \left(\frac{\hat{a}'_{jh}}{\mathrm{S}} \right) \left(\frac{\mathrm{I} - \theta_i}{\mathrm{S}} \right)
\end{aligned}
$$

$$
= \frac{1}{\sum_{h=0}^{m_j} \lambda_{jh}} \left(\frac{\mathrm{I} - \theta_i}{\mathrm{S}} \right) \left[\lambda_{jk} \ \hat{a}^*_{jk} + \hat{d}^*_{jk} \ \lambda_{j0} \ \hat{a}^*_{j0} - \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}, \hat{d}^*_{jk}) \ \sum_{h=0}^{m_j} \lambda_{jh} \ \hat{a}^*_{jh} \right]
$$

Revised Formulas for the Nominal Response Model

Baker (1993) presented a characteristic curve scale transformation procedure for the nominal response model. In the course of deriving the gradients presented in the previous section for the multiple–choice model errors were discovered in formulas presented in Baker (1993). This section presents revised formulas for the gradients used in the characteristic curve scale transformation for the nominal response model presented by Baker (1993).

The nominal response model (Bock, 1972) is given as follows:

$$
\mathbf{P}(\theta_i \mid a_{jk}, b_{jk}) = \frac{\exp\left[a_{jk} \ \theta_i + b_{jk}\right]}{\sum_{h=1}^{m_j} \exp\left[a_{jh} \ \theta_i + b_{jh}\right]} \tag{12}
$$

where

$j = 1, 2, \ldots, J$ indexes the test items,

$k = 1, 2, \ldots, m_j$ indexes the items' response categories.

There are two constraints imposed for model identification:

$$
\sum_{k=1}^{m_j} a_{jk} = 0
$$

and

$$\sum_{k=1}^{m_j} b_{jk} = 0.$$

The transformation of parameter estimates on the current scale to the target scale is

$$\hat{a}^*_{jk} = \frac{\hat{a}'_{jk}}{\mathrm{S}},$$

and

$$\hat{b}^*_{jk} = \hat{b}'_{jk} - \frac{\hat{a}'_{jk}}{\mathrm{S}}\ \mathrm{I} = \hat{b}'_{jk} - \hat{a}^*_{jk}\ \mathrm{I}.$$

The loss function is analogous to that for the multiple–choice model presented in Equation 11:

$$\mathbf{F} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{n} \sum_{k=1}^{m_j} \left[\mathbf{P}(\theta_i \mid \hat{a}_{jk}, \hat{b}_{jk}) - \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})\right]^2 .$$

Yet, the gradients of the loss with respect to I and S are slightly different:

$$\frac{d\mathbf{F}}{d\mathrm{S}} = -\frac{2}{NM} \sum_{i=1}^{N} \sum_{j=1}^{n} \sum_{k=1}^{m_j} \left[\mathbf{P}(\theta_i \mid \hat{a}_{jk}, \hat{b}_{jk}) - \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})\right] \left[\frac{d}{d\mathrm{S}}\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})\right],$$

and

$$\frac{d\mathbf{F}}{d\mathrm{I}} = -\frac{2}{NM} \sum_{i=1}^{N} \sum_{j=1}^{n} \sum_{k=1}^{m_j} \left[\mathbf{P}(\theta_i \mid \hat{a}_{jk}, \hat{b}_{jk}) - \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})\right] \left[\frac{d}{d\mathrm{I}}\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})\right].$$

Let u_k and v be

$$u_k = \exp\left[\hat{a}^*_{jk}\ \theta_i + \hat{b}^*_{jk}\right]$$

and

$$v = \sum_{h=1}^{m_j} \exp\left[\hat{a}^*_{jh}\ \theta_i + \hat{b}^*_{jh}\right] = \sum_{h=1}^{m_j} u_h,$$

then

$$\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}) = \frac{u_k}{v}.$$

The derivatives of $\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})$ with respect to I and S are

$$\frac{d}{d\mathrm{I}}\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}) = \frac{1}{v}\frac{du_k}{d\mathrm{I}} - \frac{u_k}{v^2}\frac{dv}{d\mathrm{I}}$$

and

$$\frac{d}{d\mathrm{S}}\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}) = \frac{1}{v}\frac{du_k}{d\mathrm{S}} - \frac{u_k}{v^2}\frac{dv}{d\mathrm{S}}$$

By Equations 9 & 10,

$$\exp\left[\hat{a}^*_{jk}\ \theta_i + \hat{b}^*_{jk}\right] = \exp\left[\frac{\hat{a}'_{jk}}{\mathrm{S}}\ \theta_i + \left(\hat{b}'_{jk} - \frac{\hat{a}'_{jk}}{\mathrm{S}}\right)\ \mathrm{I}\ \right]$$

The derivative of $\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})$ with respect to I is

$$\begin{aligned}
\frac{d}{d\mathrm{I}}\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}) &= \frac{1}{v}\ u_k\ \left(-\frac{\hat{a}'_{jk}}{\mathrm{S}}\right) \quad - \quad \frac{u_k}{v}\ \frac{1}{v}\ \left[\ \sum_{h=1}^{m_j} u_h\ \left(-\frac{\hat{a}'_{jh}}{\mathrm{S}}\right)\ \right] \\
&= \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})\ (-\hat{a}^*_{jk}) \quad - \quad \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})\ \sum_{h=1}^{m_j}(-\hat{a}^*_{jh})\ \frac{u_h}{v} \\
&= \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})\ \left[\ (-\hat{a}^*_{jk}) \quad + \quad \sum_{h=1}^{m_j}(\hat{a}^*_{jh})\ \mathbf{P}(\theta_i \mid \hat{a}^*_{jh}, \hat{b}^*_{jh})\ \right]. \qquad (13)
\end{aligned}$$

Similarly, the derivative of $\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})$ with respect to S is

$$\frac{d}{d\mathrm{S}}\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}) = \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})\left(\frac{\mathrm{I}-\theta_i}{\mathrm{S}}\right)\ \left[\ \hat{a}^*_{jk} \quad - \quad \sum_{h=1}^{m_j}\hat{a}^*_{jh}\ \mathbf{P}(\theta_i \mid \hat{a}^*_{jh}, \hat{b}^*_{jh})\ \right] \qquad (14)$$

Baker (1993) provided a procedure for computing a scale transformation under the nominal response model and derived gradients that are not equivalent to Equations 13 and 14. In his paper, the derivative of $\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})$ with respect to I is numbered as Equation 43 and the derivative of $\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})$ with respect to I is numbered as Equation 42.

$$\frac{d}{d\mathrm{I}}\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}) = (-\hat{a}^*_{jk})\ \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})\ \left[1 - \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})\right] \qquad (15)$$

$$\frac{d}{d\mathrm{S}}\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}) = \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}) \ \left[1 \ - \ \mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk}) \right] \ \hat{a}^*_{jk} \ \left(\frac{\mathrm{I} - \theta_i}{\mathrm{S}} \right) \tag{16}$$

The discrepancy between Baker's formulas and the formulas presented in this paper lies in a strong assumption implied by Baker's formulas. Baker's gradients match the gradients presented in Equations 13 and 14 if an additional condition is satisfied.

Bakers formulas holds if and only if

$$\sum_{h=1,\ h\neq k}^{m_j} \hat{a}^*_{jh} \ \mathbf{P}(\theta_i \mid \hat{a}^*_{jh}, \hat{b}^*_{jh}) = 0. \tag{17}$$

However, as shown in the appendix, the condition is prone to break down and is typically not assumed in practice.

Mean/Mean and Mean/Sigma Methods

An alternative to the characteristic curve method of scale transformation is to estimate the scale transformation using moments of the estimated item parameters. Examples of moment methods of estimating a scale transformation are the mean/mean and mean/sigma methods for the 3PL model (Kolen & Brennan, 1995). The mean/sigma method was first described by Marco (1977), and the mean/mean method was first described by Loyd and Hoover (1980). Mean/mean and mean/sigma methods for the multiple–choice model analogous to the mean/mean and mean/sigma methods for the 3PL model are presented in this section.

To implement the mean/mean and mean/sigma methods a reparameterization of the multiple–choice model as given in Equation 1 is used. The reparameterization of the multiple–choice model consists of using the parameters t_{jk} in place of the parameters b_{jk}, where

$t_{jk} = -b_{jk}/a_{jk}$. With this reparameterization the model in Equation 1 can be written as

$$\mathbf{P}(\theta_i \mid a_{jk}, t_{jk}, d_{jk}) = \frac{\exp\left[a_{jk}(\theta_i - t_{jk})\right] + d_{jk}\exp\left[a_{j0}(\theta_i - t_{j0})\right]}{\sum_{h=0}^{m_j}\exp\left[a_{jh}(\theta_i - t_{jh})\right]}$$

If a_{jk} and t_{jk} are parameters for a set of items on the target scale and a'_{jk} and t'_{jk} are the corresponding values of these parameters on the current scale then

$$a_{jk} = \frac{a'_{jk}}{\mathrm{S}}\,, \tag{18}$$

and

$$t_{jk} = \mathrm{S}\; t'_{jk} + \mathrm{I}\,. \tag{19}$$

Both the mean/mean and mean/sigma methods use two sets of item parameter estimates for a set of n common items. One set of parameter estimates for the common items is on the target scale ($\hat{a}_{jk}$, $\hat{t}_{jk}$), and one set of item parameter estimates on the current scale ($\hat{a}'_{jk}$, $\hat{t}'_{jk}$). To obtain an estimate of S the mean/mean method uses the mean of all the $\hat{a}_{jk}$

$$\hat{\mu}_a = \frac{1}{n+M}\sum_{j=1}^{n}\sum_{k=0}^{m_j}\hat{a}_{jk}\,, \tag{20}$$

and the mean of all the $\hat{a}'_{jk}$

$$\hat{\mu}_{a'} = \frac{1}{n+M}\sum_{j=1}^{n}\sum_{k=0}^{m_j}\hat{a}'_{jk}\,. \tag{21}$$

Using Equation 18

$$\hat{\mu}_a = \frac{\hat{\mu}_{a'}}{\mathrm{S}}\,. \tag{22}$$

The mean/mean estimate of S ($\hat{\mathrm{S}}^{mm}$) is obtained by solving Equation 22 for S

$$\hat{\mathrm{S}}^{mm} = \frac{\hat{\mu}_{a'}}{\hat{\mu}_a}\,. \tag{23}$$

Similarly, using Equation 19

$$\hat{\mu}_t = \mathrm{S}\ \hat{\mu}_{t'} + \mathrm{I}\,, \tag{24}$$

where $\hat{\mu}_t$ and $\hat{\mu}_{t'}$ are the means of $\hat{t}_{jk}$ and $\hat{t}'_{jk}$ analogous to $\hat{\mu}_a$ and $\hat{\mu}_{a'}$ in Equations 20 and 21. Using Equation 24, and the mean/mean estimate of S given in Equation 23, a mean/mean estimate of I can be obtained

$$\begin{aligned}\hat{\mathrm{I}}^{mm} &= \hat{\mu}_t - \hat{\mathrm{S}}^{mm}\hat{\mu}_{t'} \\ &= \hat{\mu}_t - \frac{\hat{\mu}_{a'}}{\hat{\mu}_a}\hat{\mu}_{t'}\,. \end{aligned} \tag{25}$$

The mean/sigma estimate of S uses the standard deviation of $\hat{t}$ and $\hat{t}'$ ($\hat{\sigma}_t$ and $\hat{\sigma}_{t'}$), where

$$\hat{\sigma}_t^2 = \frac{1}{n+M}\sum_{j=1}^{n}\sum_{k=0}^{m_j}(\hat{t}_{jk} - \hat{\mu}_t)^2 \tag{26}$$

$$\hat{\sigma}_{t'}^2 = \frac{1}{n+M}\sum_{j=1}^{n}\sum_{k=0}^{m_j}(\hat{t}'_{jk} - \hat{\mu}_{t'})^2 \tag{27}$$

From Equation 19

$$\hat{\sigma}_t = \mathrm{S}\ \hat{\sigma}_{t'}\,. \tag{28}$$

The mean/sigma estimate of S is obtained by solving Equation 28 for S

$$\hat{\mathrm{S}}^{ms} = \frac{\hat{\sigma}_t}{\hat{\sigma}_{t'}}\,. \tag{29}$$

The mean/sigma estimate of I is obtained in the same manner as the mean/mean estimate of I by using the means of $\hat{t}'_{jk}$ and $\hat{t}_{jk}$

$$\begin{aligned}\hat{\mathrm{I}}^{ms} &= \hat{\mu}_t - \hat{\mathrm{S}}^{ms}\hat{\mu}_{t'} \\ &= \hat{\mu}_t - \frac{\hat{\sigma}_t}{\hat{\sigma}_{t'}}\hat{\mu}_{t'} \end{aligned} \tag{30}$$

Example: ACT mathematics tests 1997 vs. 1998

An example applying the scale transformation methods derived in this paper is presented using a form of the ACT Assessment math test administered in October, 1997 and again in October, 1998. The ACT Assessment math test contains 60 five-alternative multiple choice items. The groups administered the forms in 1997 and 1998 are not equivalent groups, so when item parameter estimates are obtained separately using the 1997 and 1998 data a scale transformation is needed to put the parameter estimates on the same scale.

Parameter estimates for the multiple–choice model were computed using MULTILOG (Thissen, 1991) for both the 1997 and 1998 data, separately. Due to limitations in MULTILOG the 60 items needed to be split into two sets of 30 items (even and odd items), and separate MULTILOG runs used on each set of items for both the 1997 and 1998 data.

The scale transformation of the 1998 item parameter estimates to put them on the scale of the 1997 item parameter estimates was computed using the characteristic curve method described in this paper. Routines described in Dennis and Schnabel (1996) were used to find values of the slope and intercept that minimized the loss function given in Equation 11. These routines implemented a quasi-Newton minimization procedure using a finite-difference approximation of the Hessian matrix. The programs used in this paper are available at http://www.b-a-h.com/software/mcmequate/.

The estimated intercept and slope of the scale transformation were -0.059245 and 1.181568, respectively. Plots of the two sets of item parameter estimates for the a–parameters, b–parameters, and d–parameters are presented in Figure 1. In the plots of the a–parameters and b–parameters there are 360 points representing the six a–parameters or b–parameters for each of 60 items. In the plots of the d–parameters there are 300 points representing the five d–parameters on each of 60 items. The horizontal coordinate of each point is the parameter estimate from the 1998 data (on the current scale) and the vertical coordinate of each point is the parameter estimate from the 1997 data (on the target scale). In the plot of the a–parameters the line converting the a–parameters on the current scale to the target scale is presented. Such a line is not presented for the b–parameters because the transformation of the b–parameters from the current to the target scale is a function of both the a–parameters and b–parameters on the current scale, not the b–parameters alone.

Plots of threshold parameters ($-b/a$) used in the mean/mean and mean/sigma methods are given in Figure 2. The top plot in Figure 2 contains points for all 360 threshold parameters. The bottom plot in Figure 2 removes outlying points corresponding to extreme values of the threshold parameters. For the threshold parameters, there is a linear function which converts the parameters on the current scale to the target scale which is presented in Figure 2.

The plot in Figure 2 indicates that trying to compute estimates of the scale transformation using the mean/mean or mean/sigma methods, which use the mean and/or standard deviation of threshold parameters, would be problematic. The mean and standard deviation of the threshold parameters will be highly distorted by the outlying observations shown in Figure 2. Consequently, for these data it is not feasible to compute the scale transformation using the mean/mean and mean/sigma methods.

The nominal response model was also fit to these data using MULTILOG. Again, two separate MULTILOG runs were used on two subsets of 30 items (even and odd items). Scale transformations for transforming the 1998 item parameter estimates to the scale of the 1997 item parameter estimates were computed using the revision of Baker's formulas presented in this paper. The same procedures for minimizing the loss function as those described for estimating the scale transformation for the multiple–choice model were used. The estimated intercept and slope of the scale transformation were 0.018253 and 1.054735, respectively. Figures 3 and 4 present plots analogous to those in Figures 1 and 2, respectively, for the parameter estimates of the nominal response model. There are 300 points presented in each plot corresponding to five category parameters for each of 60 items.

An estimated scale transformation for the nominal response model was also computed using Baker's original formulas for the gradient. The intercept and slope of the scale transformation computed using Baker's formulas were 0.019187 and 1.055032, which are almost the same as those computed using the corrected formulas in this paper. Note that in this case an incorrect gradient was used in the minimization procedure. The minimization routine still ran, but the procedure ended with a warning that iterations were stopped due to an inability to find further feasible values of the slope and intercept rather than due to the convergence criterion being met. This warning resulted from the fact that the gradients given to the minimization procedure were not correct. Still, in this case the minimization procedure ran and produced a transformation almost identical to the transformation computed using the correct gradients. It could be that in other cases the use of Baker's gradient formulas may produce much less accurate results than using the correct formulas, or could result in the minimization procedure completely failing.

Conclusion

The multiple–choice model is useful for modeling responses to all alternatives of multiple choice items. This paper presents a characteristic curve method of scale transformation for the multiple–choice model. In deriving the characteristic curve procedure for the multiple–choice model it was discovered that the gradients presented by Baker (1993) for a nominal response model characteristic curve method were incorrect. The correct gradients needed for implementing the characteristic curve method for the nominal response model are presented.

Moments methods of scale transformation (the mean/mean and mean/sigma methods) were also presented for the multiple–choice model. Outlying observations made it infeasible to compute the mean/mean and mean/sigma scale transformations in the example presented. Previous research comparing the performance of characteristic curve methods and moment methods for other models has indicated that there is less error in the characteristic curve methods than in moment methods (Kolen & Brennan, 1995). It is recommended that the characteristic curve method of scale transformation for the multiple–choice model be preferred to the mean/mean and mean/sigma methods, or other moment methods.

References

Baker, F. B. (1992). Equating tests under the graded response model. *Applied Psychological Measurement, 16*, 87–96.

Baker, F. B. (1993). Equating tests under the nominal response model. *Applied Psychological Measurement. 17*, 239–251.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29–51.

Dennis, J. E., & Schnabel, R. B. (1996). *Numerical methods for unconstrained optimization and nonlinear equations.* Philadelphia: Society for Industrial and Applied Mathematics.

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

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

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

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.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17.

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 International.

Thissen, D., & Steinberg, L. (1984). A response model for multiple choice items. *Psychometrika, 49*, 501–519.

Thissen, D., & Steinberg, L. (1997). A response model for multiple–choice items. In W. J. van der Linden and R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 51–65). New York: Springer–Verlag.

Acknowledgment:

We would like to thank Deborah J. Harris for her support and Ulf Böckenholt and Rod McDonald for their helpful comments on this research. We would also like to thank ACT for making available the data set used in the reported application. Corresponding concerning this article should be addressed to Jee–Seon Kim, Department of Psychology, University of Illinois at Urbana–Champaign, 603 East Daniel Street, Champaign, IL 61820 or Bradley A. Hanson, ACT, Inc., 2255 North Dubuque Road, P.O. Box 168, Iowa City, IA 52243–0168. Electronic mail may be addressed to jskim@s.psych.uiuc.edu or hanson@act.org.

Appendix

Is Equation 17: $\sum_{h=1,\ h\neq k}^{m_j} \hat{a}^*_{jh}\ \mathbf{P}(\theta_i \mid \hat{a}^*_{jh}, \hat{b}^*_{jh}) \equiv 0$ always true?

For an item j,

assume the number of response options, m_j is 3 and

let the estimated slopes of three categories, $\hat{a}^*_{jh}$, be .5, $-.25$, $-.25$ and

the estimated intercepts of three categories, $\hat{b}^*_{jh}$, be $-.1$, 0, 1.

$\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})$ are probabilities, thus nonnegative.

For $k = 1$, therefore,

$$\sum_{h=1,\ h\neq k}^{m_j} \hat{a}^*_{jh}\ \mathbf{P}(\theta_i \mid \hat{a}^*_{hk}, \hat{b}^*_{hk}) = -.25\ \mathbf{P}(\theta_i \mid \hat{a}^*_{2k}, \hat{b}^*_{2k}) - .25\ \mathbf{P}(\theta_i \mid \hat{a}^*_{3k}, \hat{b}^*_{3k}) \leq 0.$$

The equality holds if and only if

$$\mathbf{P}(\theta_i \mid \hat{a}^*_{jh}, \hat{b}^*_{jh}) = 0, \quad h = 2, 3.$$

Since

$$\sum_{h=1}^{m_j} \mathbf{P}(\theta_i \mid \hat{a}^*_{jh}, \hat{b}^*_{jh}) \equiv 1,$$

the equality also implies

$$\mathbf{P}(\theta_i \mid \hat{a}^*_{j1}, \hat{b}^*_{j1}) = 1.$$

However, $\mathbf{P}(\theta_i \mid \hat{a}^*_{jk}, \hat{b}^*_{jk})$ is neither 0 nor 1 unless θ reaches $-\infty$ or ∞. Considering θ ranges over a set of discrete points for estimation (e.g., -4 to 4),

$$\mathbf{P}(\theta_i \mid \hat{a}^*_{jh}, \hat{b}^*_{jh}) \neq 0 \ \text{ nor } \ 1 \ \text{ for } \ \text{any} \ \ \theta.$$

In conclusion, Equation 17 does not hold in general and Baker's equations wouldn't always provide the proper derivatives. We can find a counter example for Equation 17 from the case where $\hat{a}^*_{jh}$, $h = 1, 2, \ldots, k-1, k+1, \ldots, m_j$ have the opposite sign of $\hat{a}^*_{jk}$, as shown above.

List of figures

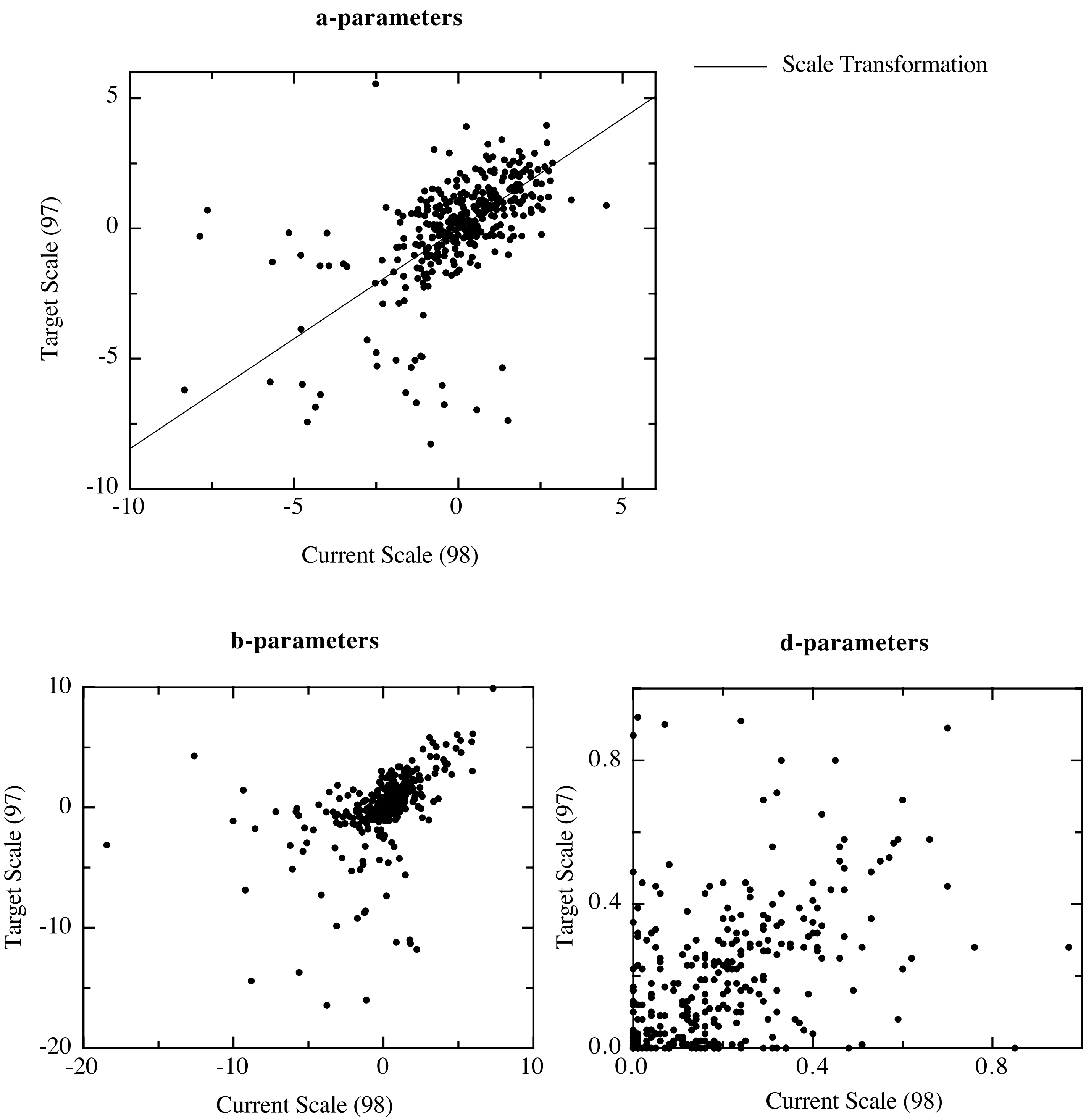


Figure 1. Multiple-Choice Model Parameter Estimates for 1997 and 1998 Data.

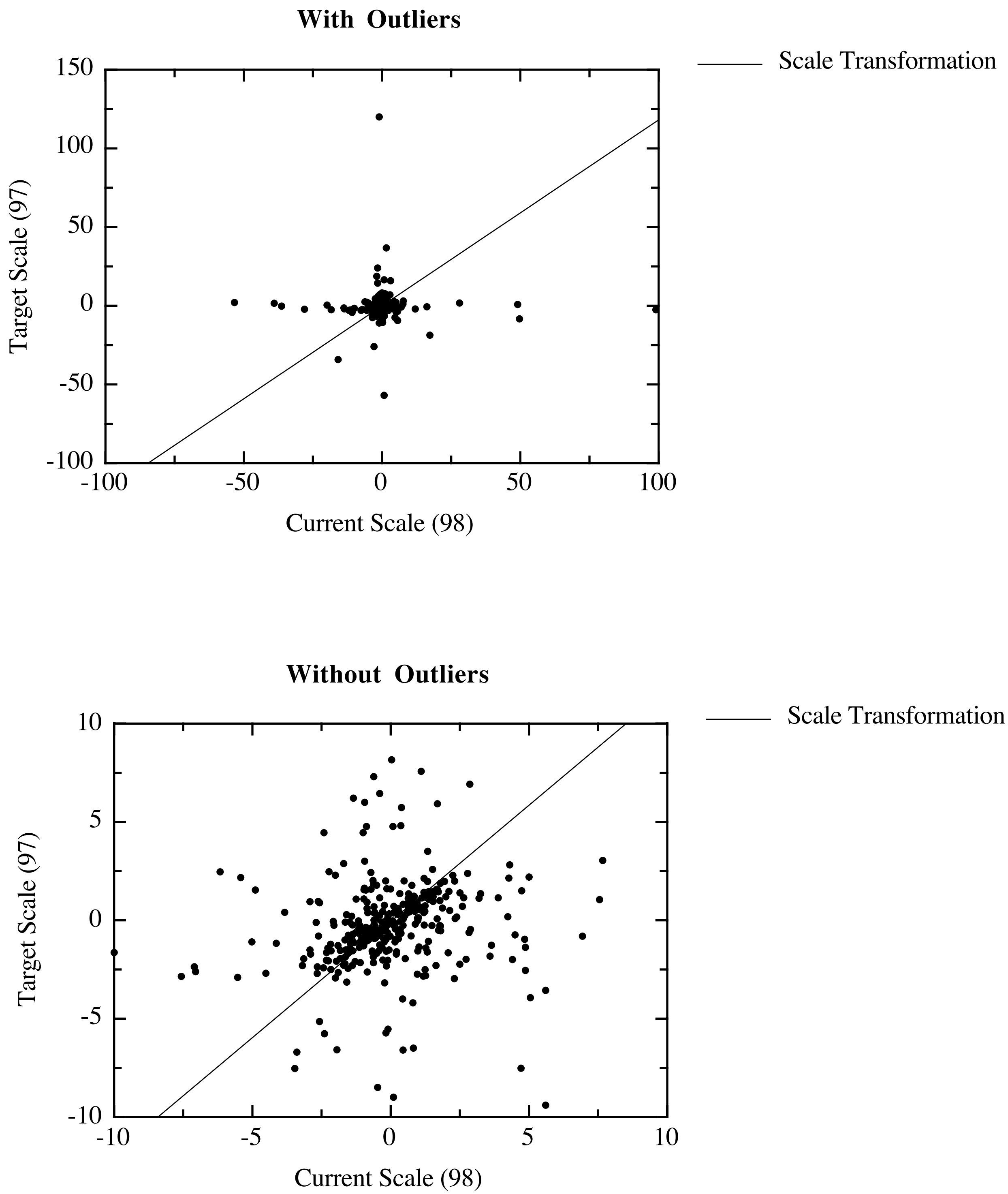
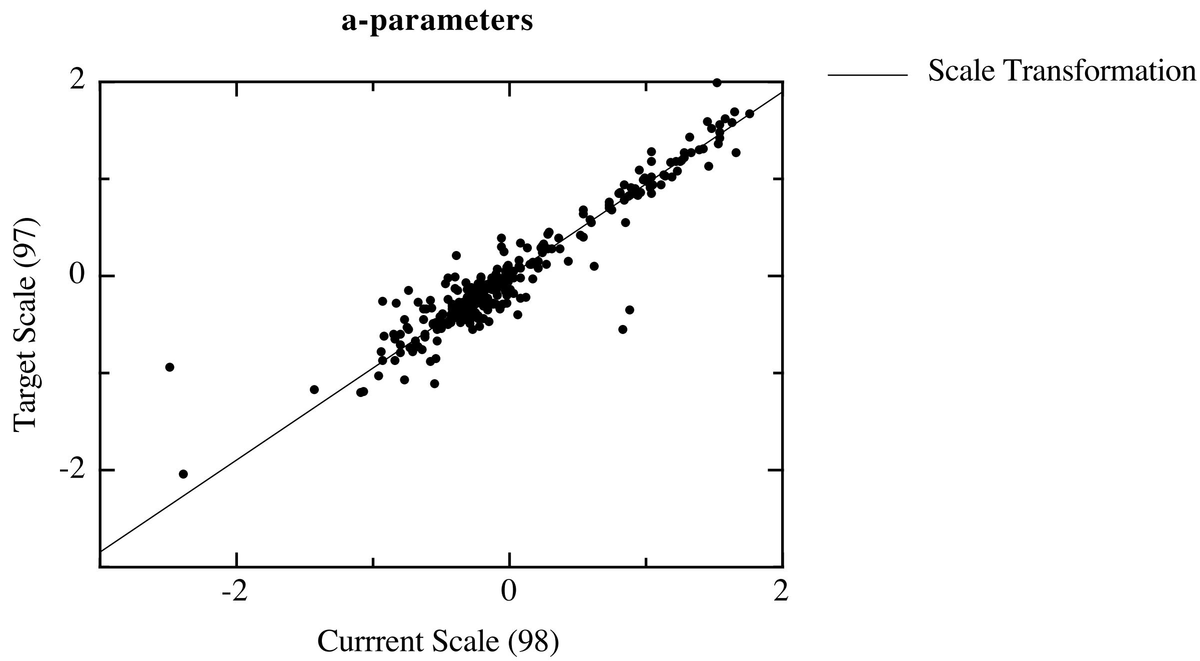


Figure 2. Multiple-Choice Model Threshold Parameter Estimates for 1997 and 1998 Data.



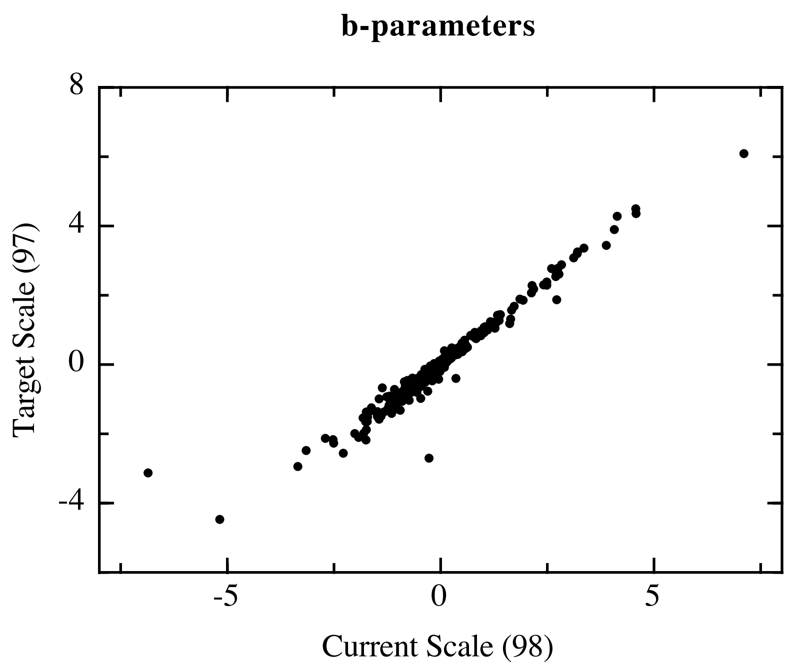


Figure 3. Nominal Response Model Parameter Estimates for 1997 and 1998 Data.

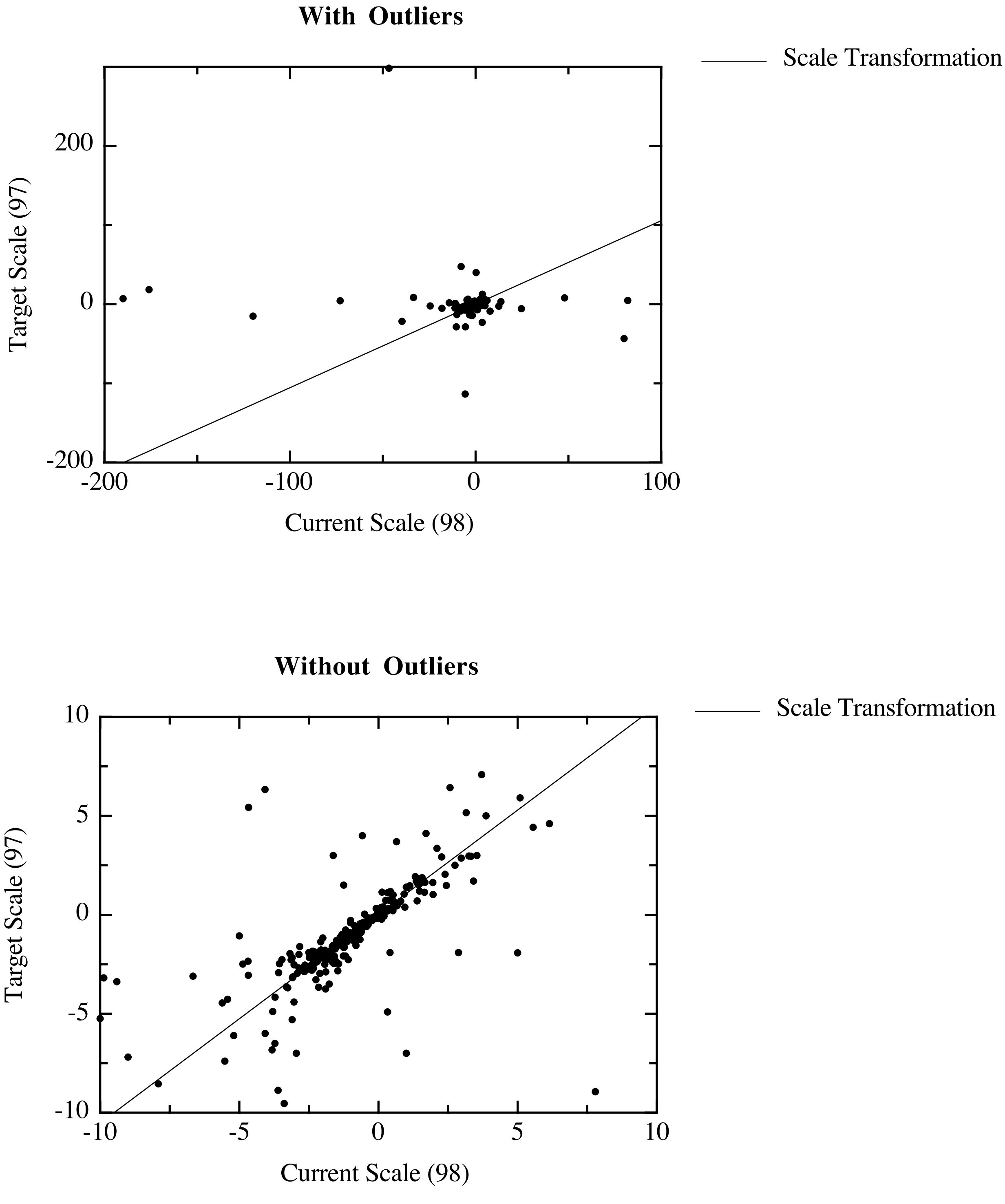


Figure 4. Nominal Response Model Threshold Parameter Estimates for 1997 and 1998 Data.