Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

# A Note on the Relation Between Factor Analytic and Item Response Theory Models

Akihito Kamata & Daniel J. Bauer

**To cite this article:** Akihito Kamata & Daniel J. Bauer (2008) A Note on the Relation Between Factor Analytic and Item Response Theory Models, Structural Equation Modeling: A Multidisciplinary Journal, 15:1, 136-153, DOI: 10.1080/10705510701758406

**To link to this article:** https://doi.org/10.1080/10705510701758406

*Structural Equation Modeling*, 15:136–153, 2008

ISSN: 1070-5511 print/1532-8007 online
DOI: 10.1080/10705510701758406

Psychology Press
Taylor & Francis Group

## TEACHER'S CORNER

# A Note on the Relation Between Factor Analytic and Item Response Theory Models

Akihito Kamata
*Department of Educational Psychology & Learning Systems*
*Florida State University*

Daniel J. Bauer
*Department of Psychology*
*University of North Carolina, Chapel Hill*

The relations among several alternative parameterizations of the binary factor analysis model and the 2-parameter item response theory model are discussed. It is pointed out that different parameterizations of factor analysis model parameters can be transformed into item response model theory parameters, and general formulas are provided. Illustrative data analysis is provided to demonstrate the transformations.

The purpose of this article is to discuss the relations among several alternative parameterizations of the binary factor analysis (FA) model and the two-parameter item response theory (IRT) model. The relationship between binary FA and the two-parameter IRT model was clarified by Takane and de Leeuw (1987) and McDonald (1999), among others; however, this relation has only been described

Correspondence should be addressed to Akihito Kamata, Department of Educational Psychology & Learning Systems, Florida State University, Tallahassee, FL 32306-4453. E-mail: kamata@coe.fsu.edu

for a subset of the common parameterizations of the binary FA. Several different parameterizations of the binary FA model have been described and implemented in recent years, as noted by Millsap and Yun-Tein (2004) and Muthén and Asparouhov (2002). This article summarizes four different parameterizations of a unidimensional FA model with binary outcome variables. Then, the link between each of these parameterizations and the two-parameter IRT model for dichotomous items is clarified. This article also provides a general conversion formula to convert FA model parameters to IRT parameters under other possible parameterizations. The reader is referred to articles, such as Muthén (1978), Goldstein and Wood (1989), and Brown (2006) for general descriptions of the item response model as an FA model for binary outcome variables.

## FACTOR ANALYTIC MODEL FOR BINARY VARIABLES

The binary FA model developed through analogy to the continuous factor model. The key assumption to the binary FA is that there is a continuous underlying latent response, denoted $y_i^*$, that is an additive combination of the common factor and item-specific residual. A one-factor model for the latent response variable can thus be written as

$$y_i^* = \nu_i + \lambda_i \xi + \varepsilon_i, \tag{1a}$$

where $\nu_i$ is the intercept, $\lambda_i$ is the factor loading, the latent factor score for a particular person is $\xi$, and $\varepsilon_i$ is the residual for item $i$ (for compactness, no person subscript is included). In addition, the residuals $\varepsilon_i$ are typically assumed to be normally distributed, but a logistic distribution can also be considered.

A threshold model is then added to the linear latent response model to accommodate the dichotomous nature of the observed response, $y_i$:

$$y_i = \begin{cases} 1 \text{ if } y_i^* \geq \tau_i \\ 0 \text{ if } y_i^* < \tau_i. \end{cases} \tag{1b}$$

Here, $\tau_i$ is the threshold for item $i$. The intercepts $\nu_i$ and thresholds $\tau_i$ are not jointly identified, so the intercepts are typically dropped from the model (i.e., assumed to be zero). We thus do not consider the intercepts further here. Parameterizations for this model can be classified by the scaling of (a) the continuous latent response variables $y_i^*$, and (b) the continuous common factor $\xi$.

### Scaling the Latent Response Variables

In one parameterization, the variance of $y_i^*$ is constrained to be 1.0 for all items. Accordingly, the residual variance of $y_i^*$, $V(\varepsilon_i)$ or $V(y_i^*|\xi)$, is obtained as the remainder $V(\varepsilon_i) = 1 - \lambda_i^2 V(\xi)$. This parameterization has its origin in a particular method of estimating the binary FA, which involves the use of tetrachoric correlations. The tetrachoric correlation matrix can be thought of as a covariance matrix between underlying latent response variables with unit variance, hence the assumption of unit variance for $y_i^*$ was a natural choice. This parameterization is common in binary FA (Millsap & Yun-Tein, 2004; Muthén & Asparouhov, 2002). Because the marginal distribution of the continuous latent trait score $(y_i^*)$ is standardized, we refer to this parameterization as the marginal parameterization.

Another convention is to constrain the residuals $\varepsilon_i$ to have unit variance. The variances of $y_i^*$ are obtained as the sum of the residual variance plus the variance due to the common factor, or $V(y_i^*) = \lambda_i^2 V(\xi) + 1$. This parameterization is similar to a traditional probit regression model, where the conditional variance of the latent response is assumed to be 1. This parameterization is less commonly used with binary FA but is closer to the conventional two-parameter IRT, as we discuss shortly. Because the conditional distribution of the continuous latent trait score $(y_i^*|\xi)$ is standardized, we refer to this parameterization as the conditional parameterization.

## Scaling the Common Factor

The conditional and marginal parameterizations scale the latent response distribution, but a choice must also be made for how to scale the common factor. Two common scaling conventions are to choose a reference indicator or to standardize the common factor. In the former approach, the threshold and factor loading of one item are fixed (e.g., $\tau_1 = 0$ and $\lambda_1 = 1$). These constraints allow the mean and the variance of $\xi$ to be freely estimated. In the latter approach, the mean and the variance of $\xi$ are constrained (e.g., $E(\xi) = 0$ and $V(\xi) = 1$). Consequently, all $\lambda_i$ and $\tau_i$ are freely estimated. These same choices must be made in the continuous FA model, with the only difference that intercepts (not thresholds) are involved.

## Four Parameterizations

By the combination of the two types of scaling choices already discussed, four different parameterizations are obtained and summarized in Table 1. Note that other parameterizations are also possible, but these four alternatives seem to describe the majority of applications. In Figure 1, we attempt to clarify the nature of these four alternatives. Equation 1a implies a linear relation between $\xi$ and $y_i^*$, plotted as a regression line in Figure 1. The conditional distribution of $y_i^*|\xi$ (or $\varepsilon_i$) is also shown in the plot. Note that, by Equation 1b, if $y_i^*$ exceeds the threshold $\tau_i$, then the binary variable $y_i = 1$, otherwise it equals 0. This

TABLE 1
Assumptions for Four Factor Analysis Parameterizations

| | *Reference Indicator* | *Standardized Factor* |
|---|---|---|
| Marginal | $\lambda_1 = 1, \tau_1 = 0$<br>$V(y^*) = 1$ | $E(\xi) = 0, V(\xi) = 1$<br>$V(y^*) = 1$ |
| Conditional | $\lambda_1 = 1, \tau_1 = 0$<br>$V(\varepsilon) = 1$ | $E(\xi) = 0, V(\xi) = 1$<br>$V(\varepsilon) = 1$ |

is shown in Figure 1 by shading the region of the conditional distribution of $y_i^*|\xi$ exceeding $\tau_i$: The shaded area under the curve gives the probability that $y_i = 1$. To clarify the nature of the different parameterizations, note that both $\xi$ ($x$-axis) and $y_i^*$ ($y$-axis) are latent variables, so they have no natural scale. We must therefore fix the scale of both variables in some sensible fashion, and whatever choice we make will influence the values of the threshold and factor loading. One choice would be to select a convenient distribution for $y_i^*|\xi$. If we fix the variance of this conditional distribution (at 1, for instance, as in the probit model), then we arrive at the two conditional parameterizations. Alternatively, we might choose a convenient distribution for the marginal distribution of $y_i^*$, shown at the right of the figure. If we fix the variance of the marginal distribution (e.g., at 1), then we arrive at the two marginal parameterizations. The key is that

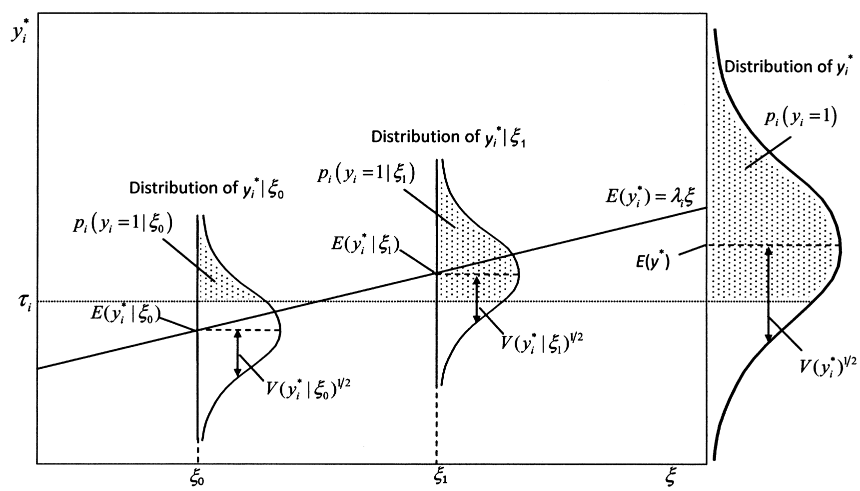


FIGURE 1 Graphical representations of key quantities in factor analysis models with binary variables.

we must fix one or the other, the conditional variance or the total variance, to set the scale of $y_i^*$. In both cases, we fix the location of $y_i^*$ by setting the intercept of the regression line to zero, making it possible to estimate the threshold.

Similarly, we must set the scale of the latent factor. One choice would be to select a convenient distribution for $\xi$, such as a standard normal distribution with mean 0 and variance 1. This choice leads to the two standardized parameterizations. Alternatively, we could fix the regression line (threshold and factor loading) for one item, which would then make it possible to estimate the mean and variance of the factor. This would lead to the two reference parameterizations. In some sense, the choice is whether to fix the *x*-axis and estimate the regression line, or fix the regression line and estimate the scale and location for the *x*-axis. The same probability model is obtained for $y_i$ in either case, but the estimates will again differ depending on the choice that is made.

We now consider how these parameterizations relate to the two-parameter item response model.

## RELATION OF BINARY FA TO TWO-PARAMETER IRT

The two-parameter IRT model can be written as

$$p_i(y_i = 1|\xi) = f(\alpha_i \xi + \beta_i),$$

where $\alpha_i$ is the slope parameter and $\beta_i$ is the intercept parameter for item $i$, $\xi$ is the latent trait score for a specific person, and $f$ is a cumulative distribution function (CDF), chosen to be either a normal or logistic CDF.[1] Previously, Takane and de Leeuw (1987) derived the relation between parameters from the standardized FA and IRT models; that is, when assuming 0 mean and unit variance for the latent factor. For unidimensional models of the kind considered here, their formulas reduce to

$$\alpha_i = \frac{\lambda_i}{q_i} \quad \text{and} \quad \beta_i = \frac{-\tau_i}{q_i},$$

where $q_i$ is $V(\varepsilon_i)^{1/2}$. Although not explicitly stated by Takane and de Leeuw, these formulas make clear that the conditional–standardized binary factor model parameterization is in fact equivalent to the IRT model, except for the reversal

[1] Parameters in a two-parameter IRT model are typically reported in the form of $f[\alpha_i(\xi - \delta_i)]$. In this case, $\delta_i = -\beta_i/\alpha_i$, and $\delta_i$ is referred to as the difficulty parameter, and $\alpha_i$ is the discrimination parameter. In this article, we deal with $\alpha_i$ and $\beta_i$ parameters directly as some other IRT literature does, including Zimowski, Muraki, Mislevy, and Bock (1999), Baker and Kim (2004), and Takane and de Leeuw (1987). In fact, BILOG-MG (Zimowski et al., 1999) estimates $\beta_i$ from data and it is transformed into $\delta_i$ when estimates are reported.

of sign for the threshold parameter $\tau_i$ relative to $\beta_i$ (given that $q_i = 1$ in this parameterization).[2] Under the marginal–standardized parameterization, $q_i = [1 - \lambda_i^2]^{1/2}$, paralleling formulas given in a number of other references (e.g., McDonald, 1999, p. 259). These results are very useful for understanding the relation between FA and IRT models. Takane and de Leeuw's derivation is not, however, directly applicable to factor models using the reference indicator convention for scaling the latent factor. We now provide more general formulas applicable to all of the binary FA parameterizations discussed already, as well as other possible parameterizations.

## TRANSFORMATION FORMULAS FOR THE FOUR PARAMETERIZATIONS

To relate the parameter values under different parameterizations, we must simply recognize that these parameterizations involve only changes in the scales of the variables in the model. We begin by considering a transformation of scale for the latent response variables, $y_i^*$. Suppose that we wish to rescale the standard deviation of $y_i^*$ (or equivalently $\varepsilon_i$) for item $i$ by a factor of $1/q_i$. To do so, we need only multiply both sides of Equations 1a and 1b by this scaling factor, yielding:

$$\frac{y_i^*}{q_i} = \frac{\nu_i}{q_i} + \frac{\lambda_i}{q_i}\xi + \frac{\varepsilon_i}{q_i},$$

and

$$y_i = \begin{cases} 1 \text{ if } \dfrac{y_i^*}{q_i} \geq \dfrac{\tau_i}{q_i} \\[2ex] 0 \text{ if } \dfrac{y_i^*}{q_i} < \dfrac{\tau_i}{q_i}. \end{cases}$$

Note that if we choose $q_i = V(y_i^*)^{1/2}$, we are standardizing the latent response variables (e.g., moving to a marginal parameterization). If we choose $q_i = V(\varepsilon_i)^{1/2}$, we are standardizing the residuals (e.g., moving to a conditional or IRT parameterization). Regardless, the intercepts, thresholds, and factor loadings

[2]Yet another possible parameterization of the binary factor model is to set all of the thresholds to 0 and estimate the intercepts of the items (rather than the other way around). These intercepts would have the same values as the previously estimated thresholds, but would be of opposite sign. Thus, a standardized conditional parameterization with 0 thresholds and estimated intercepts would exactly mimic the two-parameter IRT. In this case, the estimated intercepts would equal the $\beta_i$ parameters.

under the new scaling are equal to the values under the old scaling divided by the value $q_i$.

Now we consider an additional transformation of scale of the common factor, $\xi$. Suppose that we wish to recenter the common factor by subtracting the constant $r$, resulting in $\xi - r$. To accomplish this, we must subtract $\lambda_i / q_i$ from both sides of the equation, yielding

$$\frac{y_i^*}{q_i} - \frac{\lambda_i r}{q_i} = \frac{\nu_i}{q_i} + \frac{\lambda_i}{q_i}\xi + \frac{\varepsilon_i}{q_i} - \frac{\lambda_i r}{q_i},$$

and simplifying to the desired result

$$\frac{y_i^* - \lambda_i r}{q_i} = \frac{\nu_i}{q_i} + \frac{\lambda_i}{q_i}(\xi - r) + \frac{\varepsilon_i}{q_i}.$$

The same constant must of course be subtracted in the threshold model, yielding

$$y_i = \begin{cases} 1 \text{ if } \dfrac{y_i^* - \lambda_i r}{q_i} \geq \dfrac{\tau_i - \lambda_i r}{q_i} \\[2ex] 0 \text{ if } \dfrac{y_i^* - \lambda_i r}{q_i} < \dfrac{\tau_i - \lambda_i r}{q_i} \end{cases}.$$

Hence if the location of the common factor is changed by the subtraction of a constant $r$ then the thresholds will change by subtraction of $\lambda_i r$. Additionally, however, we might wish to change the standard deviation of $\xi$ by, say, a factor of $1/s$. The equations for the model will now change to

$$\frac{y_i^* - \lambda_i r}{q_i} = \frac{\nu_i}{q_i} + \frac{\lambda_i s}{q_i}\left(\frac{\xi - r}{s}\right) + \frac{\varepsilon_i}{q_i}.$$

Note that the scaling factor $s$ for the common factor is absorbed solely by a change in the factor loadings (multiplication by $s$), requiring no further modification of the threshold model. If $s$ is chosen to be $s = V(\xi)^{1/2}$, we can move from a reference indicator parameterization to a standardized factor parameterization, or IRT.

These transformation equations are sufficiently general to allow for conversion of parameter values from any binary factor analysis parameterization to any other binary factor analysis parameterization, or to the two-parameter IRT. Here we demonstrate the latter, showing how each binary factor analysis parameterization relates to the IRT model parameterization. To do so, we draw on the previously noted relation between the IRT and standardized conditional parameterization of the binary factor analysis. Setting $q_i = V(\varepsilon_i)^{1/2}$, $r = E(\xi)$, and $s = V(\xi)^{1/2}$ in the preceding equations, we obtain the following general transformation formulas:

$$\alpha_i = \frac{\lambda_i V(\xi)^{1/2}}{V(\varepsilon_i)^{1/2}}, \tag{2}$$

and

$$\beta_i = \frac{-[\tau_i - \lambda_i E(\xi)]}{V(\varepsilon_i)^{1/2}}. \tag{3}$$

As previously shown in Table 1, the four parameterizations differ only in whether specific values are assigned to $V(\varepsilon_i)^{1/2}$, $E(\xi)$, and $V(\xi)^{1/2}$, or these parameters are estimated in the model.[3]

We can then write four sets of conversion formulas for the four different parameterizations by substituting the appropriate values for these parameters into Equations 2 and 3. Results of substitutions are summarized in Table 2.

## ILLUSTRATIVE ANALYSIS

For illustration purposes, a unidimensional FA model, alternately employing each of the four parameterizations, was fit to the LSAT6 data (Bock & Aitkin, 1981). The data consist of five dichotomous items completed by 1,000 examinees. A summary of the data is shown in Table 3. PROC NLMIXED in SAS was used to fit the factor model. Although typical uses of the NLMIXED procedure are for fitting nonlinear mixed models, such as growth models and multilevel models, it has been shown that latent variable models with categorical outcome variables can also be fit with the procedure (e.g., Rijmen, Tuerlinckx, De Boeck, & Kuppens, 2003). Binary factor analysis models can, of course, also be fit by other software programs (e.g., M*plus*, LISREL, and EQS); however, NLMIXED was chosen here because it makes more transparent how the scaling is done and provides great flexibility regarding scaling choices. To make the results comparable to a two-parameter logistic IRT model, we used the logistic distribution for $\varepsilon_i$ in Equation 1a with each parameterization of the factor model.[4] SAS code for each of the four parameterizations is presented and documented in the Appendix.

[3]We believe these equations will also hold for ordinal factor analysis models relative to the graded response model, with the exception that more threshold parameters will be required.

[4]For FA, a normal distribution is more conventional; however we wished to compare these results to those obtained from BILOG, which uses a logistic. Results from the normal model also matched one another and are available from the first author on request. Note that results using a logistic versus a normal distribution will not match one another precisely, given the different distributions, although comparisons can be made with appropriate scaling corrections.

TABLE 2
Conversion Formulas for Four Factor Analysis Parameterizations

| | *Reference Indicator* | *Standardized Factor* |
|---|---|---|
| Marginal | $\alpha_i = \dfrac{\lambda_i V(\xi)^{1/2}}{[1 - \lambda_i^2 V(\xi)]^{1/2}}$ | $\alpha_i = \dfrac{\lambda_i}{(1 - \lambda_i^2)^{1/2}}$ |
| | $\beta_i = \dfrac{-[\tau_i - \lambda_i E(\xi)]}{[1 - \lambda_i^2 V(\xi)]^{1/2}}$ | $\beta_i = \dfrac{-\tau_i}{(1 - \lambda_i^2)^{1/2}}$ |
| Conditional | $\alpha_i = \lambda_i V(\xi)^{1/2}$ | $\alpha_i = \lambda_i$ |
| | $\beta_i = -[\tau_i - \lambda_i E(\xi)]$ | $\beta_i = -\tau_i$ |

TABLE 3
LSAT6 Data Summary

| Pattern | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Frequency |
|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 3 |
| 2 | 0 | 0 | 0 | 0 | 1 | 6 |
| 3 | 0 | 0 | 0 | 1 | 0 | 2 |
| 4 | 0 | 0 | 0 | 1 | 1 | 11 |
| 5 | 0 | 0 | 1 | 0 | 0 | 1 |
| 6 | 0 | 0 | 1 | 0 | 1 | 1 |
| 7 | 0 | 0 | 1 | 1 | 0 | 3 |
| 8 | 0 | 0 | 1 | 1 | 1 | 4 |
| 9 | 0 | 1 | 0 | 0 | 0 | 1 |
| 10 | 0 | 1 | 0 | 0 | 1 | 8 |
| 11 | 0 | 1 | 0 | 1 | 1 | 16 |
| 12 | 0 | 1 | 1 | 0 | 1 | 3 |
| 13 | 0 | 1 | 1 | 1 | 0 | 2 |
| 14 | 0 | 1 | 1 | 1 | 1 | 15 |
| 15 | 1 | 0 | 0 | 0 | 0 | 10 |
| 16 | 1 | 0 | 0 | 0 | 1 | 29 |
| 17 | 1 | 0 | 0 | 1 | 0 | 14 |
| 18 | 1 | 0 | 0 | 1 | 1 | 81 |
| 19 | 1 | 0 | 1 | 0 | 0 | 3 |
| 20 | 1 | 0 | 1 | 0 | 1 | 28 |
| 21 | 1 | 0 | 1 | 1 | 0 | 15 |
| 22 | 1 | 0 | 1 | 1 | 1 | 80 |
| 23 | 1 | 1 | 0 | 0 | 0 | 16 |
| 24 | 1 | 1 | 0 | 0 | 1 | 56 |
| 25 | 1 | 1 | 0 | 1 | 0 | 21 |
| 26 | 1 | 1 | 0 | 1 | 1 | 173 |
| 27 | 1 | 1 | 1 | 0 | 0 | 11 |
| 28 | 1 | 1 | 1 | 0 | 1 | 61 |
| 29 | 1 | 1 | 1 | 1 | 0 | 28 |
| 30 | 1 | 1 | 1 | 1 | 1 | 298 |
| *M* | .920 | .709 | .553 | .763 | .870 | |

TABLE 4
Factor Analysis Model Parameter Estimates for LSAT6 Data

| | *Conditional* | | *Marginal* | |
|---|---|---|---|---|
| | *Reference* | *Standardized* | *Reference* | *Standardized* |
| $\lambda_1$ | 1 | .8257 | 1 | .6367 |
| | | (.2581) | | (.1184) |
| $\lambda_2$ | .8754 | .7227 | .9200 | .5858 |
| | (.3627) | (.1867) | (.2367) | (.0994) |
| $\lambda_3$ | 1.0790 | .8909 | 1.0448 | .6652 |
| | (.4351) | (.2328) | (.2443) | (.0969) |
| $\lambda_4$ | .8337 | .6884 | .8906 | .5670 |
| | (.3673) | (.1851) | (.2480) | (.1035) |
| $\lambda_5$ | .7956 | .6569 | .8623 | .5490 |
| | (.3806) | (.2099) | (.2678) | (.1226) |
| $\tau_1$ | 0 | −2.7732 | 0 | −2.1385 |
| | | (.2057) | | (.1749) |
| $\tau_2$ | 1.4374 | −.9902 | 1.1649 | −.8025 |
| | (.8740) | (.0900) | (.6410) | (.0703) |
| $\tau_3$ | 2.7431 | −.2491 | 2.0482 | −.1860 |
| | (1.0769) | (.0763) | (.6618) | (.0563) |
| $\tau_4$ | 1.0273 | −1.2848 | .8462 | −1.0583 |
| | (.8790) | (.0990) | (.6720) | (.0796) |
| $\tau_5$ | .1530 | −2.0533 | .1279 | −1.7162 |
| | (.8978) | (.1354) | (.7409) | (.1207) |
| $\mu_\xi$ | 2.7732 | 0 | 2.1385 | 0 |
| | (.2057) | | (.1749) | |
| var($\xi$) | 0.6817 | 1 | .4054 | 1 |
| | (.4262) | | (.1507) | |

*Note.* Values in parentheses are standard errors.

Additionally, the two-parameter logistic IRT model was fit directly by BILOG-MG software (Zimowski et al., 1999). Parameters were estimated by maximizing the marginal likelihood by numerical integration with 100 quadrature points.[5]

The results obtained by fitting each parameterization of the factor model are summarized in Tables 4 and 5. The results showed that the transformation formulas worked perfectly for the factor model parameterizations. The largest

[5]Marginal maximum likelihood estimation (MMLE) with numerical integration has been commonly used in IRT models. Although MMLE is also becoming increasingly popular for binary FA models, weighted-least-square-based approaches still are more commonly used for binary FA models. However, when parameters are estimated with the same principal (MMLE for both IRT and binary FA models, in this case), the relation between the parameter estimates from the two can be demonstrated in a more direct manner. Also, we believe that MMLE should be a preferred choice for binary FA models whenever it is computationally feasible.

TABLE 5
Transformed Item Response Theory Model Parameter Estimates and Direct Estimates of IRT Parameters

| | *Factor Analysis* | | | | |
|---|---|---|---|---|---|
| | *Conditional* | | *Marginal* | | |
| | *Reference* | *Standardized* | *Reference* | *Standardized* | *Direct IRT* |
| $\alpha_1$ | 0.826 | 0.826 | 0.826 | 0.826 | 0.826 |
| $\alpha_2$ | 0.723 | 0.723 | 0.723 | 0.723 | 0.723 |
| $\alpha_3$ | 0.891 | 0.891 | 0.891 | 0.891 | 0.891 |
| $\alpha_4$ | 0.688 | 0.688 | 0.688 | 0.688 | 0.689 |
| $\alpha_5$ | 0.657 | 0.657 | 0.657 | 0.657 | 0.657 |
| $\beta_1$ | 2.773 | 2.773 | 2.773 | 2.773 | 2.774 |
| $\beta_2$ | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 |
| $\beta_3$ | 0.249 | 0.249 | 0.249 | 0.249 | 0.249 |
| $\beta_4$ | 1.285 | 1.285 | 1.285 | 1.285 | 1.285 |
| $\beta_5$ | 2.053 | 2.053 | 2.053 | 2.053 | 2.054 |

difference among the four parameterizations was at the fourth decimal place. When they were rounded to three decimal places, the values matched perfectly. Also, the transformed parameter estimates from the FA model were very close to the direct estimates from the IRT model by BILOG. Only three parameters ($\alpha_4$, $\beta_1$, and $\beta_5$) showed differences of .001. These differences were from the differences at the fourth decimal place and likely reflect only differences in optimization algorithms between PROC NLMIXED and BILOG.

In considering the untransformed estimates in Table 4, it is interesting to see that our inferences and interpretations might change depending on what scaling conventions are employed. For instance, in the conditional–reference parameterization, the variance of the latent ability factor is not statistically different from zero. In the marginal-reference parameterization, it does differ from zero, and in the two standardized parameterizations, no such test is available as the variance is fixed at 1. Similarly, the inferences for the threshold parameters differ depending on the scaling convention. For example, for the two reference indicator parameterizations, our tests are with respect to the threshold for the reference indicator. That is, we are testing whether items have comparable difficulty to the reference indicator. For the standardized parameterizations, the thresholds instead represent points on the centered distributions of $y_i^*$. These differences in interpretation all reflect the fact that although the models are equivalent, the parameter estimates are on different scales. Despite the difference in scales, the null hypothesis continues to be zero, and this can lead to some results being significant in one parameterization but not another. This is not problematic, as the estimates also have different interpretations in the different parameterizations. In general, a parameterization should be chosen that leads straightforwardly to the desired hypothesis tests.

## CONCLUSIONS

This article summarized four different parameterizations for a unidimensional FA model with binary variables. It was pointed out that one of the four parameterizations corresponds to the parameterization commonly used for an IRT model. General transformation formulas were then presented to transform FA parameters onto different scales, including into IRT parameters. Illustrative analysis demonstrated that the transformation formulas worked perfectly.

The choice between parameterizations is arbitrary—the results obtained under one specification can be translated directly into the results under the others. However, which parameterization one chooses does influence interpretations and can have implications when extending to more complex models. For instance, one advantage of the marginal parameterization is that the total variance of $y_i^*$ is held constant as new predictors (either latent factors or item covariates) are added to the equation for that item. As the explained variance in $y_i^*$ goes up, the residual variance, $V(\varepsilon_i)$, goes down, just as in a typical linear model. In contrast, if new predictors are added to the equation in the conditional parameterization, the residual variance, $V(\varepsilon_i)$, continues to be held constant and as the explained variance goes up, the total variance of $y_i^*$ must also go up. As a consequence, all of the coefficients of earlier predictors (e.g., factor loadings) must be adjusted to the new scale (regardless of whether the old and new predictors, e.g., factors, are correlated). These adjustments due to the changing scale of $y_i^*$ are nonintuitive for researchers more familiar with linear models and can make interpretation of the coefficients more difficult. On the other hand, the conditional parameterization has other advantages. As discussed by Millsap and Yun-Tein (2004) and Muthén and Asparouhov (2002), the conditional parameterization might be preferable when extending the binary FA to make comparisons across multiple groups or over multiple time points. The reason is that in the conditional parameterization, $V(\varepsilon_i)$ is a parameter in the model that is directly accessible to the researcher, whereas in the marginal parameterization it is a remainder involving several other model parameters. Having direct access to $V(\varepsilon_i)$ can be important if there is reason to believe this variance might differ over groups or over time, in which case it can be standardized in one group or time and estimated in another (provided that certain other constraints are implemented to identify the model).

Regarding the choice between a reference indicator and the standardized factor, standardizing the factor is a simple and elegant approach that many factor analysts use regularly. Further, when used in concert with the marginal parameterization and a normal distribution for $\varepsilon_i$, the thresholds fall on a standard normal curve, easing the translation of these estimates into marginal proportions for the observed responses. Other factor analysts, however, might prefer to use a reference indicator so that the latent factor will be in the same metric as the chosen indicator. Although this is a compelling rationale when the indicators are continuous, for binary items the metric of the latent response variables is ultimately arbitrary (e.g., depending on the choice of conditional or marginal parameterizations), so the reference indicator approach does not seem to provide a similar interpretational advantage in this case. One advantage of the reference indicator approach is that the mean and variance of the latent factor can be estimated, which may facilitate across-group or over-time comparisons. Another possible advantage is that if there is a particular item that one wishes to compare to other items, this item can be chosen as the referent. The threshold parameters estimated for the remaining items will then reflect differences in difficulty or severity relative to the referent item.

There are several other reasons we believe that the results provided in this article are important. First, they extend the long-running discussion on the relation between FA and IRT. Recognizing the different parameterizations that are possible (and are implemented) with the binary FA and how they relate to IRT might help to avoid confusion on the part of researchers using the previously published transformation formulas, which apply only to a subset of parameterizations. Second, if results are to be synthesized across different studies, it must be possible to put them on the same scale. With the formulas provided, studies employing different parameterizations can easily be compared. Further, some studies might include both measurement and structural models (as in structural equation modeling with binary outcomes), but the measurement model parameters can still be related to the IRT model using the formulas given here. Last, users of IRT might not recognize that there are a number of alternative scaling conventions other than assuming standardized residuals and a standardized common factor. These parameterizations might prove useful in IRT for the same reasons that they were developed in binary FA, either by providing different and useful interpretations or by facilitating alternative extensions of the model. It is our hope that this article raises researchers' awareness on scaling issues for binary FA models, as well as models that include binary FA submodels.

## ACKNOWLEDGMENT

This work was originally conducted when the authors participated in the 2004–2005 program on Latent Variable Models in the Social Sciences (LVSS) at the Statistical and Applied Mathematical Sciences Institute (SAMSI).

## REFERENCES


Baker, F. B., & Kim, S. H. (2004). Item response theory: Parameter estimation techniques (2nd ed.). New York: Marcel Dekker.

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

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.

Goldstein, H., & Wood, R. (1989). Five decades of item response modeling. *British Journal of Mathematical and Statistical Psychology, 42,* 139–167.

McDonald, R. P. (1999). *Test theory: Unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39,* 479–515.

Muthén, B. O. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43,* 551–560.

Muthén, B. O., & Asparouhov, T. (2002). *Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus* (Mplus Web Note No. 4). Retrieved April 28, 2005, from http://www.statmodel.com/mplus/examples/webnote.html

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods, 8,* 185–205.

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52,* 393–408.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1999). *BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items*. Chicago: Scientific Software International.


## APPENDIX
## SAS NLMIXED CODES

Here we present the SAS syntax we used to fit the binary FA model with each of the four parameterizations. To fit the model in SAS using the NLMIXED procedure, a data file was constructed that contained five rows of data for each person, corresponding to the five items. Variables included a person identifier (`id`), the item response (`y`), and five item indicator variables (`d1-d5`). The indicator variable was scored 1 if the item response was for that item; otherwise, it was scored 0 (e.g., `d1=1` when `y` is a response to Item 1, `d2=1` when `y` is a response to Item 2, etc.). Accordingly, the data file contained seven variables with 5,000 entries (five items for 1,000 examinees).

The NLMIXED code for each parameterization, given next, consists of five main parts. The first part is the line that begins with the main procedure statement `proc nlmixed`. It is followed by option specifications, including the data (`data=lsat6`), quadrature method (nonadaptive, or `noad`), the minimum number of iterations (`miniter`), convergence criterion (`gconv`), and the number of quadrature points (`qpoints`). These specifications were kept the same for all four parameterizations.

The second part of the code begins with `parms` statement and defines the parameters to be estimated and their starting values. Notice that in the two reference parameterizations `L1` ($\lambda_1$) and `tau1` ($\tau_1$) are not estimated (later they are constrained to be 1 and 0, respectively), allowing `kappa` (the mean of $\xi$) and `phi` (the variance of $\xi$) to be estimated. In contrast, in the two standardized parameterizations, all factor loadings $\lambda$ (`L1-L5`) and thresholds $\tau$ (`tau1-tau5`) are estimated, but the mean and variance of $\xi$ (`kappa` and `phi`) are constrained.

The third part of the code begins with the program statement `y_star=` and defines the model for $y^*$. Here, `d1-d5` are the indicator variables for the five items in the data file and control which parameters are operative for which items. Notice that `y_star` is defined as $\lambda$ (`L1-L5`) $\times$ $\xi$ (`ksi`), where $\lambda$ for item 1 (`L1`) is constrained to be 1 for the reference parameterization.

In the fourth part of the code, the probability model for $y = 1$ is defined. For the two conditional parameterizations, the model is first expressed in the form of `logit = y_star-tau` for each item. Note that `tau1` is constrained to be 0 in the conditional-reference parameterization. In contrast, in the two marginal parameterizations, `y_star-tau` is divided by the standard deviation of $\varepsilon$, which is $[1-\lambda^2 \operatorname{var}(\xi)]^{1/2}$ (again `tau1` is constrained to be 0 in the reference parameterization). This expression appears as `sqrt(1-phi*L2**2)`. However, notice that `L1=1` in the marginal–reference parameterization, and `phi` (variance of $\xi$) is constrained to be 1 in the marginal–standardized parameterization. After defining `logit` in this way, the probability that $y = 1$ is defined by the cumulative standard logistic distribution (`p=exp(logit)/(1+exp(logit))`).

Finally, the fifth part of the code defines the distribution of the observed responses and the latent factor (random effect). Here, the observed item response `y` is specified as a Bernoulli distributed dichotomous variable by the `binary` keyword of the `model` statement. The distribution of the latent factor (random effect) and its parameters are specified in the `random` statement. In our model, the only random effect is $\xi$ (`ksi`), which is specified as having the mean of `kappa` and the variance of `phi`. However, `kappa` and `phi` are constrained to be 0 and 1, respectively, in the two standardized parameterizations. Also, the option `subject=id` indicates that the clustering indicator variable is labeled as `id` in the data. Note that the data must be sorted on the subject ID variable prior to running the NLMIXED code.

## Conditional–Reference Parameterization

```
proc nlmixed data=lsat6 noad miniter=20 gconv=.000000000001
     qpoints=100;
*defining parameters to be estimated;
parms
```

```
kappa=.5 phi=2
L2=1 L3=1 L4=1 L5=1
tau2=0 tau3=0 tau4=0 tau5=0;
*model for y_star;
y_star =
d1*(1*ksi) +
d2*(L2*ksi) +
d3*(L3*ksi) +
d4*(L4*ksi) +
d5*(L5*ksi);
*model for probability binary y = 1;
logit=
d1*(y_star- 0)+
d2*(y_star-tau2)+
d3*(y_star-tau3)+
d4*(y_star-tau4)+
d5*(y_star-tau5);
p=exp(logit)/(1+exp(logit));
model y ~ binary(p);
*specification of random effect;
random ksi ~ normal(kappa,phi) subject=id;
run;
```

### Conditional–Standardized Parameterization

```
proc nlmixed data=lsat6 noad miniter=20 gconv=.000000000001
     qpoints=100;

*defining parameters to be estimated;
parms
L1=1 L2=1 L3=1 L4=1 L5=1
tau1=0 tau2=0 tau3=0 tau4=0 tau5=0;
*model for y_star;
y_star =
d1*(L1*ksi) +
d2*(L2*ksi) +
d3*(L3*ksi) +
d4*(L4*ksi) +
d5*(L5*ksi);
*model for probability binary y = 1;
logit =
```

```
d1*(y_star-tau1)+
d2*(y_star-tau2)+
d3*(y_star-tau3)+
d4*(y_star-tau4)+
d5*(y_star-tau5);
p=exp(logit)/(1+exp(logit));
model y ~ binary(p);
*specification of random effect;
random ksi ~ normal(0,1) subject=id;
run;
```

### Marginal–Reference Parameterization

```
proc nlmixed data=lsat6 noad miniter=20 gconv=.000000000001
     qpoints=100;

*defining parameters to be estimated;
parms
L2=.9 L3=.9 L4=.9 L5=.9
tau2=0 tau3=0 tau4=0 tau5=0
kappa=0.2 phi=0.5;
*model for y_star;
y_star =
d1*(1*ksi) +
d2*(L2*ksi) +
d3*(L3*ksi) +
d4*(L4*ksi) +
d5*(L5*ksi);
*model for probability binary y = 1;
logit =
d1*(y_star- 0)/sqrt(1-phi* 1**2)+
d2*(y_star-tau2)/sqrt(1-phi*L2**2)+
d3*(y_star-tau3)/sqrt(1-phi*L3**2)+
d4*(y_star-tau4)/sqrt(1-phi*L4**2)+
d5*(y_star-tau5)/sqrt(1-phi*L5**2);
p=exp(logit)/(1+exp(logit));
model y ~ binary(p);
*specification of random effect;
random ksi ~ normal(kappa,phi) subject=id;
run;
```

### Marginal–Standardized Parameterization

```
proc nlmixed data=lsat6 noad miniter=20 gconv=.000000000001
     qpoints=100;
*defining parameters to be estimated;
parms
L1=.5 L2=.5 L3=.5 L4=.5 L5=.5
tau1=0 tau2=0 tau3=0 tau4=0 tau5=0;
*model for y_star;
y_star =
d1*(L1*ksi) +
d2*(L2*ksi) +
d3*(L3*ksi) +
d4*(L4*ksi) +
d5*(L5*ksi);
*model for probability binary y = 1;
logit =
d1*(y_star-tau1)/sqrt(1-L1**2)+
d2*(y_star-tau2)/sqrt(1-L2**2)+
d3*(y_star-tau3)/sqrt(1-L3**2)+
d4*(y_star-tau4)/sqrt(1-L4**2)+
d5*(y_star-tau5)/sqrt(1-L5**2);
p=exp(logit)/(1+exp(logit));
model y ~ binary(p);
*specification of random effect;
random ksi ~ normal(0,1) subject=id;
run;
```