IRT in Mplus

Tihomir Asparouhov, Bengt Muthén
Version 3

June 25, 2020

In this note we describe several of the IRT modeling features implemented in Mplus, namely the the item characteristic curves, the item information curves, the total information curve, item difficulty parameter and item discrimination parameter. Different estimators and parameterizations are considered. We also provide details on the Mplus implementation of the Partial Credit Model (PCM), the 3PL-Guessing model and the 4PL Guessing with upper asymptote model.

1 ICC curves

1.1 Logit Link, ML/MLR/MLF Estimators

Let U_i be a categorical indicator for a latent factor f in the presence of a categorical latent class variable C. The item characteristic curves (ICC) for the item U_i, given that $C = k$ are computed as follows using the logistic model. If the category j is the first category

$$P_{ijk}(f) = P(U_i = j|f, C = k) = \frac{1}{1 + Exp(-\tau_{ijk} + \lambda_{ik} f)}. \tag{1}$$

If the category j is the last category

$$P_{ijk}(f) = P(U_i = j|f, C = k) = 1 - \frac{1}{1 + Exp(-\tau_{ij-1k} + \lambda_{ik} f)}. \tag{2}$$

If the category j is a middle category

$$P_{ijk}(f) = P(U_i = j|f, C = k) =$$

$$\frac{1}{1+Exp(-\tau_{ijk}+\lambda_{ik}f)}-\frac{1}{1+Exp(-\tau_{ij-1k}+\lambda_{ik}f)}. \quad (3)$$

In the presence of other covariates/other latent variables X the formulas are modified as follows. If the category j is the first category

$$P_{ijk}(f)=P(U_i=j|f,C=k,X=x)=\frac{1}{1+Exp(-\tau_{ijk}+\lambda_{ik}f+\beta_{ik}x)}. \quad (4)$$

If the category j is the last category

$$P_{ijk}(f)=P(U_i=j|f,C=k,X=x)=$$

$$1-\frac{1}{1+Exp(-\tau_{ij-1k}+\lambda_{ik}f+\beta_{ik}x)}. \quad (5)$$

If the category j is a middle category

$$P_{ijk}(f)=P(U_i=j|f,C=k,X=x)=$$

$$\frac{1}{1+Exp(-\tau_{ijk}+\lambda_{ik}f+\beta_{ik}x)}-\frac{1}{1+Exp(-\tau_{ij-1k}+\lambda_{ik}f+\beta_{ik}x)}. \quad (6)$$

1.2 Probit Link, ML/MLR/MLF Estimators

Let Φ be the standard normal cumulative distribution function. The ICC curves are given as follows. If the category j is the first category

$$P_{ijk}(f)=P(U_i=j|f,C=k,X=x)=\Phi(\tau_{ijk}-\lambda_{ik}f-\beta_{ik}x). \quad (7)$$

If the category j is the last category

$$P_{ijk}(f)=P(U_i=j|f,C=k,X=x)=$$

$$1-\Phi(\tau_{ij-1k}-\lambda_{ik}f-\beta_{ik}x). \quad (8)$$

If the category j is a middle category

$$P_{ijk}(f)=P(U_i=j|f,C=k,X=x)=$$

$$\Phi(\tau_{ijk}-\lambda_{ik}f-\beta_{ik}x)-\Phi(\tau_{ij-1k}-\lambda_{ik}f-\beta_{ik}x). \quad (9)$$

1.3 Probit Link, WLS/WLSM/WLSMV/ULS Estimators, Theta Parametrization

In this situation the model does not include latent categorical variable C however multiple group models are included. Let G denote the group variable. With the Theta parametrization the residual parameter θ_{ik} is an actual parameter in the model. For basic models this parameter is fixed to 1 since it will not be identified without model restrictions, however for multiple group and growth models the parameter could be identified. If these parameters are not printed in the results section that means that they are fixed to 1. The ICC curves are given as follows. If the category j is the first category

$$P_{ijk}(f) = P(U_i = j|f, G = k, X = x) = \Phi\left(\frac{\tau_{ijk} - \lambda_{ik} f - \beta_{ik} x}{\sqrt{\theta_{ik}}}\right). \tag{10}$$

If the category j is the last category

$$P_{ijk}(f) = P(U_i = j|f, G = k, X = x) =$$
$$1 - \Phi\left(\frac{\tau_{ij-1k} - \lambda_{ik} f - \beta_{ik} x}{\sqrt{\theta_{ik}}}\right). \tag{11}$$

If the category j is a middle category

$$P_{ijk}(f) = P(U_i = j|f, G = k, X = x) =$$
$$\Phi\left(\frac{\tau_{ijk} - \lambda_{ik} f - \beta_{ik} x}{\sqrt{\theta_{ik}}}\right) - \Phi\left(\frac{\tau_{ij-1k} - \lambda_{ik} f - \beta_{ik} x}{\sqrt{\theta_{ik}}}\right). \tag{12}$$

1.4 Probit Link, WLS/WLSM/WLSMV/ULS Estimators, Delta Parametrization

With the Delta parametrization the θ_{ik} are not actual parameters but are dependent parameters that are obtained from the following equation

$$\theta_{ik} = \Delta_{ik}^{-2} - Var(\lambda_{ik} f)$$

where Δ_{ik} are actual parameters that can be either free or fixed. Again the Δ_{ik} are typically not identifiable and are fixed to 1, however in growth and multiple group models the parameter can be free and identified. When the Δ_{ik} parameters are not present in the results, they are fixed to 1. The θ_{ik} parameters are always reported in the results section and are typically smaller than 1. For example when the Δ_{ik} parameters are fixed to 1 the θ_{ik} are smaller than 1. The ICC curves are given as in the previous section.

2 IIC curves

The item information curves (IIC) for a categorical indicator U_i and a latent factor f in class $C = k$ (or group k) is computed as in Samejima (1974). Define for $1 \leq j \leq l-1$

$$Q_{ijk} = \sum_{r=1}^{j} P_{irk}. \tag{13}$$

and $Q_{i0k} = 0$, $Q_{ilk} = 1$. The IIC is defined as follows

$$I_{ik}(f) = \sum_{r=1}^{l} \frac{(\partial P_{irk}/\partial f)^2}{P_{irk}}. \tag{14}$$

For the ML/MLF/MLR estimators with the logit link functions the IIC curve is given by

$$I_{ik}(f) = \lambda_{ik}^2 \sum_{r=1}^{l} \frac{(Q_{irk}(1-Q_{irk}) - Q_{i,r-1,k}(1-Q_{i,r-1,k}))^2}{P_{irk}}. \tag{15}$$

For binary items the above formula reduces to

$$I_{ik}(f) = \lambda_{ik}^2 P_{i1k}(1-P_{i1k}). \tag{16}$$

For the ML/MLF/MLR estimators with the probit link functions, we use the logit to probit approximation and give the IIC curve by

$$I_{ik}(f) = 3.29 \cdot \lambda_{ik}^2 \sum_{r=1}^{l} \frac{(Q_{irk}(1-Q_{irk}) - Q_{i,r-1,k}(1-Q_{i,r-1,k}))^2}{P_{irk}}. \tag{17}$$

For the WLS/WLSM/WLSMV/ULS estimators with the probit link functions and either theta or delta parametrization the IIC curve is given by

$$I_{ik}(f) = 3.29 \cdot \frac{\lambda_{ik}^2}{\theta_{ik}} \sum_{r=1}^{l} \frac{(Q_{irk}(1-Q_{irk}) - Q_{i,r-1,k}(1-Q_{i,r-1,k}))^2}{P_{irk}}. \tag{18}$$

The total information function is obtained by adding all item information functions and the prior information

$$I_k(f) = 1/\psi + \sum_i I_{ik}(f). \tag{19}$$

where ψ is the variance of the factor f. The term $1/\psi$ is minus the second derivative of the log-likelihood of the prior. The meaning of prior here is simply the part of the likelihood that specifies the factor as a $N(0,1)$ latent variable or more generally as a $N(\alpha, \psi)$ latent variable. The above formulas are obtained using the logit link function and are exact in this case. They are simply an approximation for the probit link function. The constant 3.29 used with the probit link function is simply the $\pi^2/3$ constant which is needed to adjust the scale of the loadings.

The information function $I(f)$ can be used to calculate approximate standard errors for the factor score estimates

$$SE(f) = \frac{1}{\sqrt{I(f)}}.$$

This is because $I(f)$ is the expected information function. To obtain the standard errors for the factor score in the presence of missing data the total information function in the above formula is replaced by the sum of the IIC for the indicators that are present for that observation. These factor score standard errors can also be obtained in Mplus directly using the SAVEDATA command and the ML/MLF/MLR estimators with numerical integration from the estimated posterior distribution of the factor. The two methods differ to some extent but generally yield approximately the same results. A third method for computing the factor score standard errors is with the Bayes estimator where the posterior distribution for each factor can also be estimated.

3 IRT Parameterization

For binary items with a single factor we provide the parameter estimates also in the traditional IRT scale. Let the factor mean be α and the factor variance be ψ. Thus $f = \alpha + \sqrt{\psi}\theta$ where θ is the IRT standard normal latent variable with mean 0 and standard deviation 1. For the ML/MLF/MLR estimators with the logit link function

$$P(U_i = 1|f) = \frac{1}{1 + Exp(\tau_{ik} - \lambda_{ik} f)} = \frac{1}{1 + Exp(-a_{ik}(\theta - b_{ik}))} \tag{20}$$

where a_{ik} is the item discrimination parameter and b_{ik} is the item difficulty parameter. These parameters are computed as follows

$$a_{ik} = \lambda_{ik}\sqrt{\psi} \tag{21}$$

$$b_{ik} = \frac{\tau_{ik} - \lambda_{ik}\alpha}{\lambda_{ik}\sqrt{\psi}}. \tag{22}$$

For the other estimations, links and parametrization the IRT parametrization is obtained by the same approach. The resulting formulas for b_{ik} is the same as (22), while the parameter a_{ik} is obtained as follows. For the ML/MLF/MLR estimators with the probit link function

$$a_{ik} = \lambda_{ik}\sqrt{\psi}. \tag{23}$$

For the WLS/WLSM/WLSMV/ULS estimators with the probit link functions and theta parametrization

$$a_{ik} = \frac{\lambda_{ik}\sqrt{\psi}}{\sqrt{\theta_{ik}}} \tag{24}$$

For the WLS/WLSM/WLSMV/ULS estimators with the probit link functions and delta parametrization

$$a_{ik} = \frac{1}{\sqrt{\Delta_{ik}^{-2}\lambda_{ik}^{-2}\psi^{-1} - 1}} \tag{25}$$

The standard errors of these parameters are computed by the delta method.

Note that this parameterization will be computed precisely when all factor indicators are binary and there is just one factor. The parameterization is not computed if there are negative residual variances, zero loadings, negative factor variance, additional structural relationships between the indicators such as direct regressions between the indicators, or when there are predictor variables in the model. The parametrization is also not available for montecarlo simulations, multiple imputations, or two-level models.

4 The Partial Credit Model (PCM)

Mplus supports both the Partial Credit Model (PCM) and the Generalized Partial Credit Model (GPCM). These models are described in chapters 7 and 8 of van der Linden (2016).

Suppose U is an ordered categorical variable and X is a vector of observed or latent predictors. Suppose that U takes m possible observed values $0, ..., m-1$. The partial credit model implemented in Mplus is a general model that accommodates any number of observed and latent predictors as well as mixture and two-level models, such as random intercept and slope models. The model is given by the following equation. For $k = 0, ..., m-1$

$$P(U = k|X) = \frac{Exp(\sum_{i=1}^{k}(\beta X - \tau_i))}{\sum_{g=0}^{m-1} Exp(\sum_{i=1}^{g}(\beta X - \tau_i))} \tag{26}$$

where β is a set of regression coefficients and τ_i are a set of threshold values. For identifiability purposes $\tau_0 = 0$. The above model can also be rewritten as

$$P(U = k|X) = \frac{Exp(k\beta X - \sum_{i=1}^{k} \tau_i)}{\sum_{g=0}^{m-1} Exp(g\beta X - \sum_{i=1}^{g} \tau_i))} \tag{27}$$

It is shown in Huggins-Manley and Algina (2015) that this model is a special case of the multinominal regression model implemented in Mplus. In the case where the variable U has only 2 categories the model becomes equivalent to the logistic regression model. The above PCM model is an alternative model and a competing model to the logistic regression model described in Section 1.1 and the probit regression model described in Section 1.2. It has the special property that the log-odds is linear in terms of the predictors X, that is, $\log(P(U = i|X)/P(U = j|X))$ is a linear function in terms of X. This property does not hold for either of the logistic or probit regression models. The regression coefficients in the PCM model are not equal to the probit or logit regression coefficients but they behave similarly, that is, positive regression coefficient β implies that higher vales of X leads to higher values of U. The logit, the probit and the PCM regression use the same number of parameters. The three models can be compared using the BIC criterion.

In the special case when there is a single factor η measured by several PCM items U_j, taking values from 0 to $m_j - 1$, Mplus also computes the model parameters in the standard IRT metric.

The Generalized Partial Credit Model (GPCM) is given by the following equation

$$P(U_j = k|\eta) = \frac{Exp(\sum_{i=1}^{k} \lambda_j(\eta - b_j + d_{i,j}))}{\sum_{g=0}^{m-1} Exp(\sum_{i=1}^{g} \lambda_j(\eta - b_j + d_{i,j})))} \tag{28}$$

where the ability factor η is assumed to have a standard normal distribution with mean zero and variance 1. The parameters $d_{i,j}$ are constrained for identification purposes by the following equations

$$d_{0,j} = 0 \tag{29}$$

$$\sum_{i=1}^{m_j-1} d_{i,j} = 0 \tag{30}$$

The parameter λ_j is the *Item Discrimination* parameter, the parameter b_j is the *Item Location*, and the parameters $d_{i,j}$ are the *Item Categories*. All of these parameters can be found in the IRT PARAMETERIZATION section of the Mplus output. Note that model (28) is a reparameterization of the general model (26). To obtain the GPCM parameters, Mplus first computes the model estimates for the general model

$$P(U_j = k|\eta) = \frac{Exp(\sum_{i=1}^{k}(\beta_j\eta - \tau_{i,j}))}{\sum_{g=0}^{m_j-1} Exp(\sum_{i=1}^{g}(\beta_j\eta - \tau_{i,j}))}. \tag{31}$$

Suppose that the estimates for the mean and the variance of the latent variable η are α and ψ. The GPCM reparametrization is derived by

$$\lambda_j = \beta_j\sqrt{\psi} \tag{32}$$

$$b_j = \sum_{i=1}^{m_j-1} (\tau_{i,j} - \beta_j\alpha)/((m_j - 1)\lambda_j) \tag{33}$$

$$d_{i,j} = b_j - \tau_{i,j}/\lambda_j. \tag{34}$$

The item information function for the GPCM is computed as in Reckase (2009)

$$I(U_j, \eta) = \lambda_j^2 \sum_{k=1}^{m_j-1} (k - E(U_j|\eta))^2 P(U_j = k|\eta). \tag{35}$$

5 The Guessing (3PL) and the Guessing with Upper Asymptote Models (4PL)

Suppose U is a binary variable, with possible values 0 and 1, and X is a set of observed or latent predictors. The 3PL Guessing model implemented in

Mplus is given by the following equation

$$P(U = 0|X) = \frac{1}{1 + Exp(-\tau_2)} \frac{1}{1 + Exp(-\tau_1 + \beta X)}. \tag{36}$$

The guessing parameter is

$$c = 1 - \frac{1}{1 + Exp(-\tau_2)}. \tag{37}$$

An equivalent way to write the above model is

$$P(U = 1|X) = c + \frac{1 - c}{1 + Exp(\tau_1 - \beta X)}. \tag{38}$$

The 3PL model has 2 threshold parameters τ_1 and τ_2 as well as the regression coefficients β. If $\tau_2 \geq 15$ then $c = 0$ and the 3PL model essentially becomes equivalent to the logistic regression model. This model arises naturally in modeling testing items that are multiple choice, with no penalty for guessing. With probability c an item is answered correctly due to guessing irrespective of the person's ability or the difficulty of the item. The model is particularly useful when modeling a difficult item and the item is answered correctly even though the person's ability indicates that it is unlikely that this has happened due to their ability. The guessing parameter c is also referred to as a lower asymptote for the item. Even if nobody in the population is able to solve correctly the test item, a minimum of c percentage of the population is expected to answer the item correctly due to guessing.

The 4PL model is a further generalization of the guessing 3PL model and provides an upper asymptote for the item, entirely symmetric to the lower asymptote. The model is given by the following equations

$$c = 1 - \frac{1}{1 + Exp(-\tau_2)} \tag{39}$$

$$d = 1 - \frac{1}{1 + Exp(-\tau_3)} \tag{40}$$

$$P(U = 0|X) = 1 - c - (d - c)\frac{1}{1 + Exp(\tau_1 - \beta X)} = \tag{41}$$

$$1 - d + (d - c)\frac{1}{1 + Exp(-\tau_1 + \beta X)} \tag{42}$$

$$P(U=1|X) = c + (d-c)\frac{1}{1+Exp(\tau_1 - \beta X)} = \tag{43}$$

$$d - (d-c)\frac{1}{1+Exp(-\tau_1 + \beta X)} \tag{44}$$

The 4PL model has 3 threshold parameters τ_1, τ_2 and τ_3 as well as the regression coefficients β. If $\tau_3 \leq -15$, then $d = 1$ and the 4PL model becomes equivalent to the 3PL model. A special case when only the upper asymptote is present is of interest as well, see Loken and Rulison (2010) for an application where the upper asymptote is the more important asymptote. In that case, $\tau_2 \geq 15$, $c = 0$, and the variable U has only upper asymptote. Then the distribution of $1-U$ is described by the 3PL model. This symmetry between the asymptotes is useful in understanding the properties of the model and to provide proper interpretation for the parameters. Items that are worded in a tricky way may cause incorrect answer due to carelessness even when the item is within the person's ability range. Such items can be fitted better by the 4PL model because that model allows for incorrect answer even when the person's ability would align more with a correct answer.

The above 3PL and 4PL models allow for multiple observed and unobserved predictors. When there is only one latent variable predictor η the results are also reported in the standard IRT parametrization scale. The parameters β and τ_1 are reparameterized the same way as in the 2PL model in equation (20), i.e., by computing the item discrimination parameter and the item difficulty parameter as in equations (21) and (22). The threshold parameter τ_2 is reparameterized as the guessing parameter c and the threshold τ_3 is reparameterized as the upper asymptote parameter d. The item information function is computed as in Magis (2013)

$$I(U_j, \eta) = \lambda_j^2 \frac{(P(\eta) - c)^2 (d - P(\eta))^2}{(d-c)^2 P(\eta)(1 - P(\eta))} \tag{45}$$

where $P(\eta) = P(U = 1|\eta)$. Note that if $c = 0$ and $d = 1$ the formula reduces down to formula (16) for the 2PL model.

The 3PL model is equivalent to a 2-class mixture model where in the first class the outcome follows a logistic regression while in the second class the outcome is always 1. Similarly the the 4PL model is equivalent to a 3-class mixture model where in the first class the outcome follows a logistic regression while in the second class the outcome is always 1 and in the third class the outcome is always 0. Note however that the latent class variable is defined for each item rather than for each person as in a typical mixture model, i.e., the model can be viewed as a grade of membership model, see Asparouhov and Muthén (2008) for example. Due to this connection with Mixture models the 3PL and 4PL likelihoods can be prone to multiple local optima and it may be necessary in certain situations to explore random starting values in the optimization using the STARTS option of the ANALYSIS command in Mplus.

The guessing parameter c and the upper asymptote parameter d are not identified when there is no observed or latent covariate in the model. Even in the presence of an observed covariate sufficient sample size is needed to obtain good parameter estimates. Often the parameter c will converge to 0 and the parameter d will converge to 1 which essentially reduces the 4PL model to the standard 2PL model. A sample size of $N = 5000$ might be needed to estimate well a model with one asymptote and $N = 20000$ might be needed to estimate well a model with both asymptotes. For an unobserved covariate such as an ability factor model measured by 4PL items, around 10 items or more are needed to measure the ability factor well enough to be able to replicate the identifiability of the 3PL and 4PL models with observed covariate.

It is possible to add priors for the model parameters to improve identifiability of the 3PL and 4PL models or to incorporate additional information in the model that is not a part of the observed data. The maximum-likelihood estimation in the presence of priors simply maximizes the combined likelihood

$$L = P(data|\theta) \cdot Prior(\theta) \tag{46}$$

where θ represents the vector of model parameters. Priors can be given for all or only for some of the model parameters. To improve identifiability however for the 3PL and 4PL models, priors might be needed for all model parameters.

The priors in this maximum-likelihood based estimation are similar to the priors in the Bayesian estimation and are specified in Mplus via the MODEL PRIORS command. Only normal priors are implemented in Mplus for the 3PL and 4PL models. The priors are specified for the native parametrization, meaning that to specify a prior for the guessing parameter c, one has to specify a prior for the second threshold parameter τ_2. Using the delta method one can use the following approximation to obtain the appropriate prior for τ_2. Suppose that a prior is desired for c with mean m and variance v. One can approximately accomplish this by specifying a prior for τ_2 with mean $-\log(m/(1-m))$ and variance $v/(m(1-m))^2$. Similarly if a prior is desired for d with mean m and variance v, a prior is specified for τ_3 with mean $\log(m/(1-m))$ and variance $v/(m(1-m))^2$.

The likelihood of the 3PL and the 4PL models can be fairly flat even for larger sample sizes. That means that the priors may actually have unusually large effect on the parameter estimates. There are four consequences of that fact. First, sensitivity of the priors should be examined. Second, priors should be selected as meaningfully as possible rather than as mathematical artifacts. Third, it should be noted that even when parameter estimates differ quite substantially due to different prior specification, the estimated models might not differ substantially. One should examine TECH10 for the estimated contingency tables, Pearson chi-square, univariate and bivariate contingency tables and univariate and bivariate Pearson chi-squares. Fourth, due to the flatness of the likelihood priors may become unintentionally too informative. The TECH10 results should be compared for models with priors and models without priors to make sure that the final estimated model is not pulled away from the data too severely by the prior specifications. Thus priors should generally be set as weakly informative priors.

The main advantage of using priors with the ML estimation is the ability of the priors to stabilize the estimation, minimize MSE of the estimates and improve coverage. Consider the following univariate example with one binary outcome and one observed standard normal covariate in a guessing 3PL model. The three parameter in the model are set as follows $\tau_1 = 0$, $\tau_2 = 1$ which yields a guessing parameter $c = 0.27$, and $\beta = 1$. We estimate the model with and without prior. Prior is set only for τ_2. We used a normal prior with mean 1 and variance 1. Table 1 reports the bias, coverage and MSE for the model parameters for 3 sample sizes $N = 500, 1000$ and 5000. The results are based on a simulation study with 100 replications. This simulation shows that the main effect of the prior is on the MSE, which is consistently better when the prior is included. In small sample size it appears also that the prior improves the coverage as well.

Table 1: Comparing ML estimation with and without priors. Absolute Bias/Coverage/MSE

N	Parameter	With Prior	Without Prior
500	c	.01/.85/.006	.01/.77/.020
1000	c	.00/.89/.004	.02/.90/.012
5000	c	.00/.91/.002	.00/.95/.003
500	τ_1	.07/.90/.11	.05/.84/.23
1000	τ_1	.03/.91/.06	.01/.91/.11
5000	τ_1	.02/.92/.02	.01/.96/.03
500	β	.10/.98/.08	.11/.91/.11
1000	β	.03/.97/.03	.02/.93/.04
5000	β	.00/.94/.01	.00/.96/.01

The following references can be used for additional information on the IRT models.

References

[1] Asparouhov, T. & Muthén, B. (2008). Multilevel mixture models. In Hancock, G. R., & Samuelsen, K. M. (Eds.), Advances in latent variable mixture models, pp. 27-51. Charlotte, NC: Information Age Publishing

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

[3] Bock, R.D. (1997). A brief history of item response theory. Educational Measurement: Issues and Practice, 16, 21-33.

[4] Cheng Y. & Liu C. (2015). The Effect of Upper and Lower Asymptotes of IRT Models on Computerized Adaptive Testing. Applied Psychological Measurement, 39, 551-565.

[5] du Toit, M. (2003). IRT from SSI. Lincolnwood, IL: Scientific Software International, Inc. (BILOG, MULTILOG, PARSCALE, TESTFACT)

[6] Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Mahwah, NJ: Erlbaum.

[7] Hambleton, R.K. & Swaminathan, H. (1985). Item response theory. Boston: Kluwer-Nijhoff.

[8] Huggins-Manley A. C. & Algina J. (2015). The Partial Credit Model and Generalized Partial Credit Model as Constrained Nominal Response Models, With Applications in Mplus. Structural Equation Modeling: A Multidisciplinary Journal, 22, 308-318.

[9] Loken E. & Rulison K. L. (2010). Estimation of a four-parameter item response theory model. British Journal of Mathematical and Statistical Psychology, 63, 509-525.

[10] MacIntosh, R. & Hashim, S. (2003). Variance estimation for converting MIMIC model parameters to IRT parameters in DIF analysis. Applied Psychological Measurement, 27, 372-379.

[11] Magis D. (2013) A Note on the Item Information Function of the Four-Parameter Logistic Model. Applied Psychological Measurement, 37, 304-315.

[12] Muthén, B. (1985). A method for studying the homogeneity of test items with respect to other relevant variables. Journal of Educational Statistics, 10, 121-132.

[13] Muthén, B. (1988). Some uses of structural equation modeling in validity studies: Extending IRT to external variables. In H. Wainer & H. Braun (Eds.), Test Validity (pp. 213-238). Hillsdale, NJ: Erlbaum Associates.

[14] Muthén, B. (1989). Using item-specific instructional information in achievement modeling. Psychometrika, 54, 385-396.

[15] Muthén, B. (1994). Instructionally sensitive psychometrics: Applications to the Second International Mathematics Study. In I. Westbury, C. Ethington, L. Sosniak & D. Baker (Eds.), In search of more effective mathematics education: Examining data from the IEA second international mathematics study (pp. 293-324). Norwood, NJ: Ablex.

[16] Muthén, B. & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. Mplus Web Note #4 (www.statmodel.com).

[17] Muthén, B., Kao, Chih-Fen & Burstein, L. (1991). Instructional sensitivity in mathematics achievement test items: Applications of a new IRT-based detection technique. Journal of Educational Measurement, 28, 1-22.

[18] Muthén, B. & Lehman, J. (1985). Multiple-group IRT modeling: Applications to item bias analysis. Journal of Educational Statistics, 10, 133-142.

[19] Reckase, M. (2009). Multidimensional item response theory. New York: Springer.

[20] Samejima, F. (1972) A general model for free-response data. Psychometrika Monograph Supplement, No.18

[21] Samejima, F. (1974). Normal ogive model on the continuous response level in the multidimensional latent space. Psychometrika, 39, 111-121.

[22] Takane, Y. & DeLeeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. Psychometrika, 52, 393-408.

[23] Van der Linden, W.J. (2016). Handbook of item response theory. Volume one. Models. Boca Raton, FL: CRC Press.