

# UCLA

## UCLA Previously Published Works

### Title

An Explanatory Multidimensional Random Item Effects Rating Scale Model.

### Permalink

<https://escholarship.org/uc/item/9r7122hv>

### Journal

Educational and Psychological Measurement, 83(6)

### Authors

Huang, Sijia

Luo, Jinwen

Cai, Li

### Publication Date

2023-12-01

### DOI

10.1177/00131644221140906

Peer reviewed

# An Explanatory Multidimensional Random Item Effects Rating Scale Model

Educational and Psychological  
Measurement  
2023, Vol. 83(6) 1229–1248  
© The Author(s) 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/00131644221140906  
journals.sagepub.com/home/epm



Sijia Huang<sup>1</sup> , Jinwen (Jevan) Luo<sup>2</sup> and Li Cai<sup>2</sup>

## Abstract

Random item effects item response theory (IRT) models, which treat both person and item effects as random, have received much attention for more than a decade. The random item effects approach has several advantages in many practical settings. The present study introduced an explanatory multidimensional random item effects rating scale model. The proposed model was formulated under a novel parameterization of the nominal response model (NRM), and allows for flexible inclusion of person-related and item-related covariates (e.g., person characteristics and item features) to study their impacts on the person and item latent variables. A new variant of the Metropolis-Hastings Robbins-Monro (MH-RM) algorithm designed for latent variable models with crossed random effects was applied to obtain parameter estimates for the proposed model. A preliminary simulation study was conducted to evaluate the performance of the MH-RM algorithm for estimating the proposed model. Results indicated that the model parameters were well recovered. An empirical data set was analyzed to further illustrate the usage of the proposed model.

## Keywords

item response theory, random item effects model, rating scale model, explanatory item response theory model

---

<sup>1</sup>Indiana University Bloomington, USA

<sup>2</sup>University of California, Los Angeles, USA

## Corresponding Author:

Sijia Huang, School of Education, Indiana University Bloomington, 201 North Rose Avenue, Bloomington, IN 47405-7000, USA.

Email: sijhuang@iu.edu

## Introduction

Likert-type items are ubiquitously utilized in education and psychology to measure traits that cannot be observed directly. Item response theory (IRT) models, such as the rating scale model (RSM; Andrich, 1978a, 1978b) and the partial credit model (PCM; Masters, 1982), are routinely applied in practice to analyze item-level responses with more than two categories. These polytomous IRT models depict the relationships between observed item responses, person latent traits (e.g., attitude, and academic proficiency), and item properties (e.g., item location). For example, the RSM specifies the log-odds of conditional probabilities that the response of a person  $p$  ( $p = 1, \dots, P$ ) to an item  $i$  ( $i = 1, \dots, I$ ), denoted by  $y_{pi}$ , falls into category  $k$  ( $k = 0, 1, \dots, K - 1$ ) over  $k - 1$  as

$$\log \left[ \frac{P(y_{pi} = k | \theta_p)}{P(y_{pi} = k-1 | \theta_p)} \right] = \theta_p - (\delta_i + \tau_k), \quad (1)$$

where  $\theta_p$  is a theoretical person latent variable and is often assumed to follow a normal distribution,  $\theta_p \sim N(0, \sigma^2)$ . Depending on the parameterization of the model,  $\sigma^2$  can be either estimated from the data or fixed at one. The RSM assumes that the threshold structure is the same across all items in a test so that  $\delta_i$  represents the overall location of item  $i$ , and  $\tau_k$  is the threshold parameter for category  $k$  from category  $k - 1$ . The RSM, like most IRT models, assumes the person effects (i.e.,  $\theta_p$ ) are random. In other words, the persons who respond to items are viewed as the sample of a larger person population. On the other hand, item effects/parameters (i.e.,  $\delta_i$ ) are often assumed fixed in using marginal maximum likelihood (MML) estimation (Andrich, 1978a; De Boeck, 2008).

Random item effects IRT models, which consider that item effects/parameters random, have received much attention for more than a decade (e.g., De Boeck, 2008; Janssen et al., 2000; Van den Noortgate et al., 2003). Theoretically, it is reasonable that an IRT model treats item parameters as random variables. As the generalizability theory (G-theory; Shavelson & Webb, 1991) states, items in a measurement scale can be viewed as a random sample of an *item universe* (i.e., the population of items). Therefore, distributional assumptions regarding item effects need to be made if generalized conclusions are to be drawn for more items (Briggs & Wilson, 2007). IRT models are also recognized as special cases of the generalized linear mixed model (GLMM; Chalmers, 2015; De Boeck & Wilson, 2004; Rijmen et al., 2003) so that random item effects IRT models become natural extensions of conventional IRT models.

The random item effects approach also has substantive advantages in many practical settings. For example, in large-scale international survey studies, different nations would use a common measurement scale but it is not uncommon to observe that item characteristics vary across nations (e.g., De Jong et al., 2007; De Jong & Steenkamp, 2010; Fox & Verhagen, 2010; Rijmen & Jeon, 2013). In this context, random item effects IRT models can serve as a general tool for studying the issue of measurement

invariance (Meredith, 1993), which is also known as *differential item functioning* (DIF; Holland & Wainer, 2012), across nations (De Boeck, 2008; Muthén & Asparouhov, 2018). Automated item generation, where hundreds of items are created as random “clones” from item families (Geerlings et al., 2011; Glas & van der Linden, 2003), is another area where the random item effects IRT models apply. The random item effects perspective allows understanding the heterogeneity within item families, identifying problematic items, and empowering test assembly. In addition, random item effects IRT models can be also desired in scenarios where the number of persons for item calibration is relatively small. This is because they generally require estimating fewer parameters than their fixed item effects counterparts.

Based on what item effects are random, the random item effects IRT models can be categorized into two types. In the first type, the item effects are random over items; while in the second type, the item effects are made random over persons or clusters. A well-known example of the first type is the random item effects Rasch model introduced in Van den Noortgate et al. (2003). While considering both person and item effects random, the binary item responses are cross-classified by persons and items. In other words, the random item effects Rasch model is essentially a generalized linear model with crossed random effects. The logit that a person  $p$  correctly answers a dichotomous item  $i$  is specified as

$$\text{logit}[P(y_{pi} = 1)] = u_p^{\text{person}} + u_i^{\text{item}} + \beta_0, \quad (2)$$

with  $u_p^{\text{person}}$  represents person  $p$ 's latent variable, and  $u_i^{\text{item}}$  is item  $i$ 's random effect and represents its easiness level. Both  $u_p^{\text{person}}$  and  $u_i^{\text{item}}$  are assumed to be normally distributed,  $u_p^{\text{person}} \sim N(0, \sigma_{\text{person}}^2)$  and  $u_i^{\text{item}} \sim N(0, \sigma_{\text{item}}^2)$ .  $\beta_0$  represents the logit of the probability that an *average* person correctly answers an *average* item, that is,  $u_p^{\text{person}} = u_i^{\text{item}} = 0$ .

The basic model presented in Equation 2 was extended to the two-parameter logistic (2PL) model (e.g., Janssen et al., 2000; Van den Noortgate et al., 2003), the three-parameter logistic (3PL; e.g., Johnson & Sinharay, 2005; Van den Noortgate et al., 2003) model, and generalized partial credit (GPC; e.g., Johnson & Sinharay, 2005) model. Janssen et al. (2000) proposed a hierarchical IRT model for dichotomously scored items in criterion-referenced measurement. The hierarchical IRT model assumes the difficulty and discrimination parameters of items are both random effects and are drawn from certain normal distributions. Effects of items that measure the same criteria would share a common mean and variance. Johnson and Sinharay (2005) considered the 3PL (Birnbaum, 1968) and the GPC (Muraki, 1992) models so that the difficulty, discrimination, asymptote (for the 3PL model), and step (for the GPC model) parameters of items are all treated as random effects.

The second type of random item effects IRT models allows the item effects to be random over persons or clusters (e.g., countries), and are often adopted in cross-national studies (e.g., De Jong et al., 2007; De Jong & Steenkamp, 2010; Rijmen &

Jeon, 2013). For an item  $i$  with two score categories, the logit of the probability that a person  $p$  from country  $j$  ( $j = 1, \dots, J$ ) correctly answers it is

$$\text{logit}[P(y_{p(j)i} = 1)] = \theta_{p(j)} + \beta_{ij}, \quad (3)$$

where  $\theta_{p(j)}$  presents this person's latent variable level and is sampled from a distribution with its country average  $\theta_j$  and variance  $\sigma_j^2$ ,  $\theta_{p(j)} \sim N(\theta_j, \sigma_j^2)$ . The country average  $\theta_j$  is then viewed to be drawn from a normal distribution with the grand average  $\theta$  and variance  $\tau^2$ . The item easiness parameter  $\beta_{ij}$  is also assumed to be country-specific,  $\beta_{ij} \sim N(\beta_i, \sigma_b^2)$ . The 2PL model and graded response model (GRM; Samejima, 1969) versions of the model in Equation 3 have also been developed (De Jong et al., 2007; De Jong & Steenkamp, 2010), in which each country has its own discrimination and threshold parameters. Wang et al. (2006) proposed a random item effects rating scale model (RERSM), where item thresholds are treated as random effects to account for randomness in subjective judgment across persons. The variation of thresholds reflects the magnitude of the randomness. This RERSM was further extended to incorporate item-specific discrimination parameters (Wang & Wu, 2011) and to accommodate multidimensional and multilevel data (Wang & Qiu, 2013).

An issue that has been of substantive interest and needs more investigation with random item effects IRT models is how person characteristics (e.g., gender) and item features (e.g., if an item is reservedly worded) explain differences in the person latent variable(s) and the item properties (e.g., item location). De Boeck and Wilson (2004) introduced the explanatory item response models (EIRM) to simultaneously model the impacts of person-related and item-related covariates and the item response process. A widely-used example of EIRM is Fischer's (1973) linear logistic test model (LLTM), which specifies the item location parameter as a weighted sum of multiple item features. However, in practice, the person latent variable and the item location parameters may not be fully explained by the covariates considered, leaving room for the inclusion of random residuals. Several explanatory random item effects models have been proposed for dichotomous data, including the explanatory multidimensional multilevel random item response model (EMMRIRM; Cho et al., 2013) and additive multilevel item structure (AMIS; Cho et al., 2014), while the models have not yet been extended to polytomous response data.

The present study focuses on the first type of random item effects model (i.e., models whose random effects vary over items.) To extend the utility of the random item effects approach and EIRM to more contexts, this study introduces an explanatory multidimensional random item effects RSM for polytomous items. The proposed model considers both person and item effects as random and allows the impacts of person-related and item-related covariates to be studied simultaneously. The proposed model is formulated under a novel parameterization of the nominal response model (NRM; Bock, 1972, 1997; Thissen & Cai, 2016; Thissen et al., 2010). The new parameterization unifies a series of *divided-by-total* (Thissen & Steinberg, 1986) polytomous IRT models, allows for straightforward multidimensional extensions of these models, and has been implemented in widely-used IRT software packages, including

flexMIRT (Cai, 2017), IRTPRO (Cai et al., 2011), and OpenMx/ rpf (Pritikin & Falk, 2020). In addition, as Falk (2021) noted, the new parameterization facilitates Monte Carlo simulation studies since it allows for easy simulating a greater variety of category response functions that are reasonable. With the original parameterization, varying one of the item parameters may make other parameters unrealistic and lead to impractical category response functions. To estimate the proposed model, a new variant of the Metropolis-Hastings Robbins-Monro (MH-RM) algorithm designed for estimating latent variable models with crossed random effects (Cai, 2008, 2010a, 2010b; Chung & Cai, 2021; Huang, 2021) is applied.

The remainder of this article is organized as follows. The “Nominal Response Model” section presents the fixed item effects RSM as a constrained case of the NRM based on a novel parameterization (Thissen & Cai, 2016; Thissen et al., 2010). The “Proposed Modeling Approach” section proposes an explanatory multidimensional random item effects RSM and illustrates the new variant of the MH-RM algorithm for parameter estimation. The “Simulation Study” section presents a preliminary simulation study for evaluating the performance of the MH-RM algorithm for estimating the proposed model. The “Empirical Example” section analyzes an empirical data set and uses the proposed model to answer research questions that are of substantive interest. The “Conclusion and Discussion” section summarizes this study and discusses future research directions.

## Nominal Response Model

In this section, the parameterization of the NRM proposed by Thissen et al. (2010) and Thissen and Cai (2016) is briefly introduced to facilitate the understanding of the proposed model.

### A Novel Parameterization of NRM

The NRM was originally designed for item responses with no pre-determined orders (Bock, 1972, 1997). Thissen et al. (2010) and Thissen and Cai (2016) presented a novel parameterization of the NRM and showed that the NRM is best treated as a template for polytomous IRT models. Adopting the new parameterization, an NRM specifies the probability that person  $p$ 's response to item  $i$  is in category  $k$  ( $k = 0, 1, \dots, K - 1$ ) as

$$P(y_{pi} = k | \theta_p; a_i^*, \mathbf{s}_i, \mathbf{c}_i) = \frac{\exp(z_{pik})}{\sum_{m=0}^{k-1} \exp(z_{pim})}. \tag{4}$$

In Equation 4,  $z_{pik}$  is a linear predictor:

$$z_{pik} = a_i^* s_{i(k+1)} \theta_p + c_{i(k+1)}. \tag{5}$$

in which  $a_i^*$  represents the overall slope for item  $i$ ,  $s_{i(k+1)}$  is the scoring function for category  $k$ , and  $c_{i(k+1)}$  is the intercept parameter for the  $k$  th category of item  $i$ .  $\theta_p$  is the person latent variable. The denominator of Equation 4 is the sum of exponentials of linear predictors for all response categories, and  $m$  is an indicator of the response category. For identification purposes, restrictions on the scoring functions  $s_i$  and the intercepts  $\mathbf{c}_i$  need to be imposed. These restrictions are implemented through reparameterizing  $s_i$  and  $\mathbf{c}_i$  into two new vectors, namely,  $\boldsymbol{\alpha}_i$  and  $\boldsymbol{\gamma}_i$ , via a contrast matrix  $\mathbf{T}$ :

$$\mathbf{s}_i = \mathbf{T}\boldsymbol{\alpha}_i \text{ and } \mathbf{c}_i = \mathbf{T}\boldsymbol{\gamma}_i. \quad (6)$$

As pointed out in Thissen et al. (2010) and Falk (2021), elements in  $\boldsymbol{\gamma}_i$  are interpretable. The first element of  $\boldsymbol{\gamma}_i$  reflects the overall location of item  $i$ , while the remaining elements of  $\boldsymbol{\gamma}_i$  parameterizes the spacing among the crossover points of item characteristic curves.

A major benefit of adopting this novel parameterization is that it allows for straightforward multidimensional generalizations of the NRM and its special cases by expanding the scalar latent variable and the associated overall slope into vectors.

### RSM as a Constrained NRM

As shown in Thissen et al. (2010) and Thissen and Cai (2016), several well-known *divided-by-total* (Thissen & Steinberg, 1986) polytomous IRT models, including the RSM, the PCM, and the GPC model (Muraki, 1992), can all be viewed as constrained cases of the full-rank NRM. To formulate an RSM under the new parameterization, three constraints need to be imposed: (a) the overall slopes are restricted to be equal across items,  $a_1^* = \dots = a_i^* = \dots = a_I^*$ , (b)  $\boldsymbol{\alpha}_i$  is fixed to  $(1, 0, \dots, 0)'$ , which effectively makes the scoring function values  $(0, 1, 2, \dots, K - 1)'$ , and (c) the second to the last elements of  $\boldsymbol{\gamma}_i$  are constrained to be equal across items.

If the person latent variable  $\theta_p$  is assumed unidimensional and follows a standard normal distribution, free parameters in an RSM under the new parameterization include: (a) an overall slope parameter  $a^*$ , which also reflects the variation of the person latent variable, (b) the first element of the  $\boldsymbol{\gamma}_i$  vector for each item,  $\gamma_{i1}$  for  $i = 1, \dots, I$ , and (c) the second to the last elements of  $\boldsymbol{\gamma}_i$ , which are constrained equal across items. Once estimates of these free parameters are obtained, parameters under the original parameterization of the RSM can be derived. When the linear-Fourier basis contrast matrix is used for the reparameterization, the location parameters in Equation 1 can be computed as

$$\delta_i = \frac{-c_K}{a^*(K-1)} = \frac{-\gamma_{i1}}{a^*}. \quad (7)$$

The threshold parameters are

$$\tau_k = \frac{c_K}{K-1} - \frac{c_k - c_{k-1}}{a^*}. \quad (8)$$

## Proposed Modeling Approach

In this section, an explanatory multidimensional random item effects RSM based on the new parameterization is proposed. The measurement model, which specifies the probability of selecting a response option as a function of the person and item latent variables, is introduced first. The person latent variables represent the latent constructs that a test aims to measure. The level of the item latent variable reflects the deviation of an item from the average item location. Then the structural models are presented, which connect the person-related and item-related covariates with person and item latent variables through regression equations.

### Measurement Model

Adopting the novel parameterization, the probability that person  $p$ 's response to item  $i$  falls into category  $k$  ( $k=0, 1, \dots, K-1$ ) is specified as

$$P(y_{pi} = k | \boldsymbol{\theta}_p, \delta_i; \mathbf{a}^{\text{person}}, a^{\text{item}}, \mathbf{S}^{\text{person}}, \mathbf{s}^{\text{item}}, \mathbf{c}) = \frac{\exp(z_{pik})}{\sum_{m=0}^{K-1} \exp(z_{pim})}, \quad (9)$$

where the linear predictor is

$$z_{pik} = [\mathbf{a}^{\text{person}} \circ \mathbf{s}_{k+1}^{\text{person}}]^T \boldsymbol{\theta}_p + a^{\text{item}} s_{k+1}^{\text{item}} \delta_i + c_{k+1}. \quad (10)$$

The item responses are cross-classified by persons and items. Therefore, like the random item effects model presented in Equation 2, the first two terms of Equation 10 capture the person and item random effects, respectively. Specifically, in Equations 9 and 10,  $\mathbf{a}^{\text{person}}$  is a slope vector that corresponds to  $\boldsymbol{\theta}_p$ , the latent variables that vary over persons.  $s_{k+1}^{\text{person}}$  is the  $(k+1)$  th column of the scoring function matrix  $\mathbf{S}^{\text{person}}$ , which is of size  $D \times K$ . Each column of the  $\mathbf{S}^{\text{person}}$  matrix corresponds to a category and each row of it corresponds to a latent dimension. The scoring function values are allowed to vary across latent dimensions. The symbol  $\circ$  denotes the Schur product.  $a^{\text{item}}$  and  $s_{k+1}^{\text{item}}$  are respectively the slope and scoring function value associated with item latent variable  $\delta_i$ , which varies over items and captures item  $i$ 's deviation from the average item location.  $c_{k+1}$  is the intercept parameter.

Restrictions on the scoring function matrix associated with person latent variables  $\mathbf{S}^{\text{person}}$ , scoring function values that correspond to item latent variable  $\mathbf{s}^{\text{item}}$ , and the intercept parameter vector  $\mathbf{c}$  are implemented through reparametrizing them as

$$\mathbf{s}_k^{\text{person}} = \mathbf{T} \boldsymbol{\alpha}_k^{\text{person}}, \mathbf{s}^{\text{item}} = \mathbf{T} \boldsymbol{\alpha}^{\text{item}}, \text{ and } \mathbf{c} = \mathbf{T} \boldsymbol{\gamma}, \quad (11)$$

where  $\boldsymbol{\alpha}_k^{\text{person}}$ ,  $\boldsymbol{\alpha}^{\text{item}}$ , and  $\boldsymbol{\gamma}$  are vectors of size  $K-1$ . In Equation 11,  $\mathbf{T}$  represents a  $K \times (K-1)$  contrast matrix and can take a linear Fourier-basis form:



$$T = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & f_{22} & \dots & f_{2(K-1)} \\ 2 & f_{32} & \dots & f_{3(K-1)} \\ \vdots & \vdots & \ddots & \vdots \\ K-1 & 0 & \dots & 0 \end{bmatrix},$$

in which

$$f_{mm'} = \sin \left[ \frac{\pi(m' - 1)(m - 1)}{(K - 1)} \right],$$

so that the second to last columns of  $\mathbf{T}$  are mutually orthogonal.

To formulate the proposed model as an RSM type model,  $\alpha_k^{\text{person}}$  and  $\alpha^{\text{item}}$  are valued  $(1, 0, \dots, 0)'$  so that the scoring function values effectively become  $(0, 1, 2, \dots, K - 1)'$ . Note that in Equation 11, the  $\mathbf{c}$  and  $\boldsymbol{\gamma}$  vector has no item indicator, meaning that  $\boldsymbol{\gamma}$  is constrained to be equal across items. In addition, the first element of  $\boldsymbol{\gamma}$  is fixed to zero so that the average location of all items is 0.

### Structural Model

Two regression equations for the person and item latent variables are defined, respectively, incorporating person-related and item-related covariates:

$$\boldsymbol{\theta}_p = \mathbf{B}\mathbf{x}_p + \boldsymbol{\varepsilon}_p, \quad (12)$$

$$\delta_i = \mathbf{w}_i'\boldsymbol{\lambda} + \xi_i. \quad (13)$$

Equation 12 models the relationships between person-related covariates and the person latent variables.  $\boldsymbol{\theta}_p$  is a vector of size  $D$  and varies over persons.  $\mathbf{x}_p$  is a vector that consists of  $m$  person-related covariates (e.g., gender),  $\mathbf{B}$  is a  $D \times m$  regression coefficient matrix,  $\boldsymbol{\varepsilon}_p$  is a  $D$ -dimensional vector of person random effects, and is assumed to follow a multivariate normal distribution,  $\boldsymbol{\varepsilon}_p \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ . Equation 13 is for the item latent variable  $\delta_i$ , which represents the deviation of item  $i$ 's location from the average location.  $\mathbf{w}_i$  is the item covariate vector (e.g., reversely worded) and is of size  $n$ , and  $\boldsymbol{\lambda}$  is the corresponding regression coefficient vector.  $\xi_i$  represents item  $i$ 's random effects, and is assumed to be normally distributed.

### Identification Constraints

To identify the proposed model, a few constraints need to be imposed. Specifically, the person and item random effects are let to have unit variances. In other words, diagonal elements of  $\boldsymbol{\Sigma}$  are all ones, and  $\xi_i$  follows a standard normal distribution. As a result, values of the associated slope parameters  $\mathbf{a}^{\text{person}}$  and  $a^{\text{item}}$  effectively reflect the variations in the person latent traits and item locations.

Free parameters in the proposed model include elements of the regression coefficient matrix  $\mathbf{B}$ , elements of the regression coefficient vector  $\lambda$ , off-diagonal elements of  $\Sigma$ ,  $\mathbf{a}^{\text{person}}$ ,  $a^{\text{item}}$  and the second to the  $(K - 1)$  th elements of  $\gamma$ . Compared with the conventional RSM, the proposed model generally requires estimating fewer parameters, which can be desired in practical settings. For example, assuming no person-related and item-related covariates are considered, for a unidimensional test consisting of  $I$  items with  $K$  categories, a conventional RSM requires estimating  $I + K - 1$  parameters ( $I$  location parameters and  $K - 1$  thresholds), while the number of free parameters in a random item effects RSM is  $K$  (2 slope parameters associated with the person and item latent variables and  $K - 2$  elements of  $\gamma$ ).

### Estimation

A variant of the MH-RM algorithm (Cai, 2008, 2010a, 2010b; Chung & Cai, 2021; Huang, 2021) that implements a new sampling strategy for estimating latent variable models with crossed random effects is applied to estimate the explanatory multidimensional random item effects RSM introduced above. The MH-RM algorithm produces the maximum likelihood estimates (MLEs) of model parameters and is preferred in high-dimensional IRT settings (e.g., Chung & Cai, 2021; Falk & Cai, 2016; Monroe & Cai, 2014; Yang & Cai, 2014). The MH-RM algorithm makes use of the idea of data augmentation and combines the MH (Hastings, 1970; Metropolis et al., 1953) sampler with the RM (Robbins & Monro, 1951) Stochastic Approximation (SA) algorithm.

The MH-RM algorithm has two strong motivations: (a) Fisher's (1925) identity, which states that the conditional expectation of the gradient of the complete data log-likelihood is equal to the gradient of the observed data log-likelihood, and (b) the RM algorithm, which is a root-finding algorithm designed for functions that are corrupted by noise. To illustrate how the two motivations facilitate the MLE, the iterative scheme of the standard MH-RM algorithm is outlined below first. Following the general scheme, the sampling strategy applied in the new variant of the algorithm to aid the estimation of crossed random person and random item effects is introduced. Readers that are interested in more technical details are referred to Cai (2008, 2010a, 2010b), Chung and Cai (2021), and Huang (2021).

Each iteration of the MH-RM algorithm consists of three steps: *stochastic imputation*, *stochastic approximation*, and *Robbins-Monro update*. With the MH-RM algorithm, the person and item latent variables/random effects are treated as *missing data* (Dempster et al., 1977). The missing data and the *observed data* (i.e., item response and covariates) form the *complete data*. In the first *stochastic imputation* step of each iteration, the MH sampler is applied to impute missing data (i.e., person and item random effects) so that the complete data are formed. In the *stochastic approximation* step, made possible by Fisher's identity, the gradient vector and information matrix of the complete data log-likelihood are evaluated as an approximation of the observed data log-likelihood, which is more difficult to evaluate. In the *Robbins-Monro update*

step, the RM algorithm is applied to update the model parameters, accounting for the missing data (i.e., noise) introduced in the *stochastic imputation* step.

In the *stochastic imputation* step of each iteration, a sampling strategy that couples the Metropolis-within-Gibbs algorithm with the alternating imputation posterior (AIP) algorithm (Cho & Rabe-Hesketh, 2011) is adopted. In this process, values of the person and item latent variables/random effects are simulated in alternation. Specifically, the person latent variables/random effects are imputed first, fixing item latent variables/random effects to the imputed values from the previous iteration. Direct simulating latent variables/random effects for persons from their conditional densities are not feasible. Therefore, a Gibbs sampler is constructed and combined with the MH algorithm. Then the item latent variables/random effects are simulated in the same manner, fixing the person latent variables/random effects to the imputed values obtained in the current iteration.

## Simulation Study

To evaluate the performance of the MH-RM algorithm in terms of estimating the proposed explanatory multidimensional random item effects RSM, a simulation study was conducted. Data were simulated and analyzed with the popular IRT software flexMIRT (Cai, 2017).

### Simulation Design

Item responses were simulated based on a five-category explanatory random item effects RSM presented in Equations 9, 10, and 12. For illustration purposes, the person latent variable  $\theta_p$  was assumed to be unidimensional and was predicted by one person-related continuous covariate. The simulated value for the regression coefficient was 0.7. The item latent variable  $\delta_i$  was unidimensional and was predicted by one item-related covariate. The data-generating value of the regression coefficient was 0.3. Generating values of the second to the last elements of  $\boldsymbol{\gamma}$  were 1.2,  $-0.4$ , and 0.1 respectively. A linear Fourier-basis contrast matrix was applied. Note that the full generalization of the proposed model allows for accommodating multidimensional person latent variables and multiple person-related and item-related covariates that are of different types.

Manipulated factors considered in this simulation study were: (a) the number of persons, (b) the number of items, and (b) the variances of persons and item random effects. Specifically, the numbers of persons generated were 500 and 1,000, which aimed to reflect a relatively small and large sample size. The numbers of items simulated were 100 and 200. The numbers of persons and items were chosen to mimic real-world scenarios where a large number of persons are employed to evaluate an item pool that consists of a relatively small number of items. Another consideration is that, as mentioned, the random item effects approach can be ideal in situations where

the sample size is too small to produce stable parameter estimates if conventional fixed item effects IRT models are adopted.

Three combinations of variances of person and item random effects were simulated, which were  $a^{person} = a^{item} = 1$ ,  $a^{person} = 2$  and  $a^{item} = 1$ , and  $a^{person} = 1$  and  $a^{item} = 2$ . The three combinations led to conditions with equal person and item random effects variances, conditions with larger person random effects variance, and conditions with larger item random effects variance, respectively.

A total of  $2 \times 2 \times 3 = 12$  conditions were simulated. For each condition, 100 replications were simulated.

### Evaluation Criteria

The bias and root mean square error (RMSE) were used to evaluate if the model parameters were well recovered. Specifically, for a generic model parameter  $\omega$ , the bias and RMSE were computed as

$$Bias(\hat{\omega}) = \frac{\sum_{r=1}^R (\hat{\omega} - \omega)}{R}, \tag{14}$$

$$RMSE(\hat{\omega}) = \sqrt{\frac{\sum_{r=1}^R (\hat{\omega} - \omega)^2}{R}}, \tag{15}$$

where  $R$  is the number of converged replications.

### Simulation Results

Table 1 summarizes the biases for parameter estimates of the proposed model under various simulation conditions. As shown in the table, the slope parameter ( $a^{person}$ ) and the regression coefficient ( $\beta$ ) associated with the person latent variable and elements of the  $\gamma$  vector were very well recovered. The biases for these parameters were all below .01 with only two exceptions, which was the  $a^{person}$  estimate in the conditions that combined 200 items and a larger person variance.

The slope parameter ( $a^{item}$ ) and regression coefficient ( $\lambda$ ) associated with the item latent variable were well recovered but slightly worse than  $a^{person}$ ,  $\beta$ , and elements of  $\gamma$ . When the number of persons was relatively small, the absolute values of biases were all below .03. For conditions with a relatively larger number of persons, the absolute values of biases of  $a^{item}$  ranged from .032 to .072, and the absolute values of biases of  $\lambda$  ranged from .017 to .029. The relatively large biases of  $a^{item}$  and  $\lambda$  could be attributed to the relatively small numbers of items compared with persons.

Table 2 shows RMSEs for parameter estimates of the explanatory random item effects RSM under all simulation conditions. Holding the number of persons constant, the RMSEs for conditions with 100 items were slightly larger than those for conditions with 200 items. For all equal variance and larger person variance conditions, the

**Table 1.** Biases for Parameter Estimates of the Random Item Effects Model.

Number of person	Number of item	$\sigma^{\text{person}}$	$\sigma^{\text{item}}$	$\gamma_2$	$\gamma_3$	$\gamma_4$	$\beta$	$\lambda$
Equal variance conditions								
500	100	-.001	.000	.002	.001	-.001	-.004	-.001
	200	-.001	.002	.001	.001	-.001	-.004	-.002
1,000	100	-.005	.032	-.001	.001	.000	.000	-.027
	200	-.006	.053	.000	.000	.000	-.004	-.029
Larger person variance conditions								
500	100	.002	.008	-.004	.000	.000	-.003	-.012
	200	-.013	.002	-.002	.000	-.001	.005	-.006
1,000	100	-.003	.072	-.001	.000	-.001	-.003	-.030
	200	-.025	.066	.000	.000	.000	.008	-.044
Larger item variance conditions								
500	100	.001	-.015	.001	.001	.000	-.001	.028
	200	-.005	.003	-.002	.000	-.001	-.003	-.006
1,000	100	-.006	.056	-.002	.002	.000	.003	-.017
	200	-.006	.063	.000	.001	.000	.002	-.019

**Table 2.** RMSEs for Parameter Estimates of the Random Item Effects Model.

Number of person	Number of item	$\sigma^{\text{person}}$	$\sigma^{\text{item}}$	$\gamma_2$	$\gamma_3$	$\gamma_4$	$\beta$	$\lambda$
Equal variance conditions								
500	100	.037	.076	.019	.008	.009	.049	.115
	200	.035	.053	.015	.006	.006	.050	.086
1,000	100	.024	.084	.014	.006	.005	.031	.113
	200	.025	.074	.010	.004	.004	.036	.074
Larger person variance conditions								
500	100	.067	.085	.022	.011	.010	.052	.100
	200	.058	.049	.014	.007	.007	.044	.072
1,000	100	.052	.125	.012	.006	.006	.031	.104
	200	.053	.089	.010	.005	.004	.038	.087
Larger item variance conditions								
500	100	.036	.163	.023	.011	.009	.048	.139
	200	.038	.105	.016	.007	.007	.045	.075
1,000	100	.024	.177	.014	.008	.007	.030	.109
	200	.023	.134	.011	.005	.005	.032	.077

Note. RMSE = root mean square error.

RMSEs for all parameters were all below .1 with four exceptions. For larger item variance conditions, the RMSEs for the slope parameter ( $\sigma^{\text{person}}$ ) and the regression coefficient ( $\beta$ ) associated with the person latent variable and elements of the  $\gamma$  vector were all below .05. Similar to the biases, the RMSEs of the slope parameter ( $\sigma^{\text{item}}$ )

associated with the item latent variable were relatively large, ranging from .105 to .177. The RMSEs of the regression coefficient ( $\lambda$ ) associated with item latent variables ranged from 0.075 to 0.139.

## Empirical Example

### Data and Analysis

The proposed explanatory multidimensional random item effects RSM approach was applied to an empirical data set presented in Zhang et al. (2016). The data included responses of 312 persons to the abbreviated 18-item *Need for Cognition* (NFC) scale (Cacioppo & Petty, 1982). The sample consisted of 252 females and 59 males. The mean and standard deviation of the age were 19.95 and 2.72, respectively. The NFC scale measures the tendency of an individual to engage in and enjoy thinking. Each item of the NFC scale describes a cognitive activity (e.g., “*I would prefer complex to simple problems.*”) and has nine categories. Among the 18 items, nine are reversely worded (e.g., “*Thinking is not my idea of fun*”). Two person-related covariates, including gender and age, and one item-related covariate that indicates if an item is reversely worded were available.

Three research questions that are of substantive interest were asked:

**Research Question 1:** Compared with the variation in persons’ tendency to engage in and enjoy thinking, how do the items in the abbreviated 18-item NFC scale vary in their location parameters?

**Research Question 2:** If gender and age can predict a person’s tendency to engage in and enjoy thinking?

**Research Question 3:** If the item type (reverse worded vs. positively worded items) has an impact on the item location?

Based on the frequencies of item responses (shown in Table 3), categories 1 and 2 were collapsed into one category due to the low frequency of category 1, and categories 8 and 9 were combined as well due to the low frequency of category 9. Thus, a seven-category explanatory random item effects RSM, which included a unidimensional person latent variable and a unidimensional item latent variable, was applied to fit the data. In addition, persons’ gender and their standardized age were incorporated as predictors of the person latent variable. The item type indicator was utilized as the predictor of item latent variable. The analysis was conducted using flexMIRT (Cai, 2017). The syntax is presented in the appendix.

## Results

The estimates of the slopes associated with person and item latent variables were .44 (.02) and .31 (.01), respectively. These estimates indicated that compared with the

**Table 3.** Item Response Frequencies.

Item	Response categories								
	1	2	3	4	5	6	7	8	9
1	9	36	69	51	44	33	33	24	13
2	15	45	80	59	28	42	26	12	3
3	11	39	61	55	48	32	38	17	8
4	7	36	62	49	44	49	32	25	6
5	15	55	63	52	42	44	27	8	4
6	12	24	55	57	50	50	32	17	11
7	7	24	37	44	27	60	60	35	14
8	7	20	33	33	50	65	57	31	12
9	5	9	28	40	35	83	64	30	13
10	23	38	81	66	62	20	16	1	1
11	18	51	74	70	55	20	16	2	2
12	21	48	57	73	42	26	23	11	7
13	5	29	55	50	70	53	24	19	6
14	13	41	51	75	57	31	28	11	4
15	17	40	54	75	62	33	18	8	3
16	1	24	47	49	35	54	48	32	21
17	17	43	48	57	38	53	26	21	8
18	30	48	65	69	53	23	14	6	3

item locations, the persons had more variation in the tendency of engaging in and enjoying thinking. Estimates of the second to the last elements of the  $\gamma$  and their standard errors were 1.08 (.05), 0.07 (.03), 0.07 (.02), 0.02 (.02), and  $-0.03$  (.02), respectively.

Estimates of the two regression coefficients associated with gender and age were .07 (.09) and .04 (.06). Neither of the two estimates was significant, indicating that there was no difference between the female and male groups in their tendency to engage in and enjoy thinking, and age is not a significant predictor of the tendency. The estimated regression coefficient for the item type is  $-.17$  (.35). This insignificant estimate indicates that if an item is reversely worded has no impact on its location.

## Conclusion and Discussion

This study contributes to the IRT literature by introducing an explanatory multidimensional random item effects RSM for polytomous items. The proposed model adopted a novel parameterization of the NRM, which facilitates extending commonly known polytomous IRT models to their multidimensional versions. The proposed model allows for studying the relationships between covariates that are of interest

with latent variables, such as the relationship between gender and the person's personality, and the relationship between item features and item locations. A variant of the MH-RM algorithm designed for estimating latent variable models with crossed random effects was used to estimate the proposed model. A simulation study was conducted to evaluate the performance of the MH-RM algorithm in terms of parameter recovery with a popular IRT software flexMIRT (Cai, 2017). The simulation results indicated that the model parameter estimates were well recovered. The slope parameter and regression coefficient associated with the person latent variable and elements of  $\gamma$  have smaller biases and RMSEs than the item latent variable slope and regression coefficient. This could be due to the fact that in the simulation conditions, which aimed to mimic real-world scenarios, the number of items were much smaller than the number of persons. The proposed model was applied to an empirical data set to further demonstrate how the proposed model can be used to address substantive research questions.

The proposed explanatory multidimensional random item effects RSM considers item locations random over items and can be applied in many practical settings, such as the automatic item generation. As illustrated in the "Proposed Modeling Approach" section, the proposed random item effects RSM requires estimating a smaller number of model parameters (compared with the fixed item effects RSM). Therefore, it can also be desired in scenarios where the number of persons is relatively small to obtain stable estimates of model parameters.

The present study enhances the idea that the full-rank NRM is best treated as a flexible template and promotes the understanding of the novel parameterization proposed by Thissen et al. (2010) and Thissen and Cai (2016). The flexibility of the full-rank NRM template allows for straightforward extensions of unidimensional polytomous IRT models to multidimensional models. It also makes it possible to *design* models for specific purposes by imposing constraints on model parameters. In addition, this paper facilitates the interpretations of outputs of popular IRT software packages that implement this parameterization.

However, the present study evaluated the proposed model in a unidimensional person latent variable setting and under a relatively small number of conditions. Therefore, it is recommended that the proposed model be evaluated in a more general multidimensional person latent variable case, and under a broader range of simulation conditions, such as different sets of item parameters, multiple combinations of numbers of persons and items, and multiple person-related and item-related covariates, to facilitate decision-making in empirical studies. In addition, the present study focused on the RSM and developed its random item effects counterpart. Future studies can adopt this novel parameterization of the NRM to extend other polytomous IRT models (such as the GPC model) to their random item effects versions.



## Appendix

Sample flexMIRT syntax for estimating the explanatory multidimensional random item effects RSM

```

<Project >
  Title = "Empirical Data Analysis";
  Description = "Zhang et al., 2016 Data";

<Options >
  Mode = Calibration;
  Rndseed = 1234;
  Algorithm = MHRM; // The MH-RM algorithm is applied to estimate the model.
  Processors = 4;
  ProposalStd = 2.2;
  ProposalStd2 = 2.4;
  InitGain = 0.1;
  Stage1 = 20000;
  Stage2 = 1000;
  SavePRM = Yes;
  SaveMCO = Yes;
  SaveSCO = Yes;
  Score = EAP;

<Groups >
%Gr%
  File = "LongData.dat";
  Varnames = id, gender, age, item, res, rw; // Variable names in the data set
  Select = res; // The variable res consists of item responses.
  Dimensions = 2; // The proposed model included a person dimension and an item dimension.
  Between = 1; // Item responses are cross-classified by the person and item dimensions.
  Cluster = id; // The variable id indicates persons.
  Block = item; // The variable item indicates items.
  Code(res) = (1,2,3,4,5,6,7,8,9),(0,0,1,2,3,4,5,6,6); // The nine-category items are recoded.
  Ncats(res) = 7;
  Model(res) = Nominal(7); // The model is specified as an NRM with seven categories.
  Crossed = Yes;
  Covariates = gender, age, rw; // Two covariates are incorporated in the model.
  L2covariates = 2; // The first two covariates are to be regressed to the person latent variable.

<Constraints > // Constraints are imposed to the full-rank NRM.
  Fix Gr, (res), ScoringFn; // The alpha vector/scoring functions are fixed.
  Fix Gr, (res), Intercept(1); // The first element of gamma vector is fixed to zero.
  Value Gr, (res), Intercept(2), 0.2; // Assign a starting value to the 2nd element of gamma.

```

Value Gr, (res), Intercept(3), 0.2;  
Value Gr, (res), Intercept(4), 0.2;  
Value Gr, (res), Intercept(5), 0.2;  
Value Gr, (res), Intercept(6), 0.2;  
Free Beta(1,1); // Allow the regression coefficient to be free estimated.  
Value Beta(1,1), 0.2;  
Free Beta(1,2);  
Value Beta(1,2), 0.2;  
Free Beta(2,3);  
Value Beta(2,3), 0.2;


### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

Sijia Huang  <https://orcid.org/0000-0002-1504-3965>

### References

- Andrich, D. (1978a). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2*, 581–594.
- Andrich, D. (1978b). A rating formulation for ordered response categories. *Psychometrika, 43*, 561–573.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–472). Addison-Wesley.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29–51.
- Bock, R. D. (1997). The nominal categories model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 33–50). Springer.
- Briggs, D. C., & Wilson, M. (2007). Generalizability in item response modeling. *Journal of Educational Measurement, 44*, 131–155.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116–131.
- Cai, L. (2008). *A Metropolis-Hastings Robbins-Monro algorithm for maximum likelihood nonlinear latent structure analysis with a comprehensive measurement model*. The University of North Carolina at Chapel Hill.

- Cai, L. (2010a). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro Algorithm. *Psychometrika*, *75*, 33–57.
- Cai, L. (2010b). Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, *35*, 307–335.
- Cai, L. (2017). *flexMIRT: Flexible multilevel multidimensional item analysis and test scoring. (version 3.51)* [Computer software]. Vector Psychometric Group.
- Cai, L., Du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling*. Scientific Software International.
- Chalmers, R. P. (2015). Extended mixed-effects item response models with the MH-RM algorithm. *Journal of Educational Measurement*, *52*, 200–222.
- Cho, S.-J., De Boeck, P., Embretson, S., & Rabe-Hesketh, S. (2014). Additive multilevel item structure models with random residuals: Item modeling for explanation and item generation. *Psychometrika*, *79*, 84–104.
- Cho, S.-J., Gilbert, J. K., & Goodwin, A. P. (2013). Explanatory multidimensional multilevel random item response model: An application to simultaneous investigation of word and person contributions to multidimensional lexical representations. *Psychometrika*, *78*, 830–855.
- Cho, S.-J., & Rabe-Hesketh, S. (2011). Alternating imputation posterior estimation of models with crossed random effects. *Computational Statistics & Data Analysis*, *55*, 12–25.
- Chung, S., & Cai, L. (2021). Cross-classified random effects modeling for moderated item calibration. *Journal of Educational and Behavioral Statistics*, *46*, 651–681.
- De Boeck, P. (2008). Random item IRT models. *Psychometrika*, *73*, 533–559.
- De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer.
- De Jong, M. G., & Steenkamp, J.-B. E. (2010). Finite mixture multilevel multidimensional ordinal IRT models for large scale cross-cultural research. *Psychometrika*, *75*, 3–32.
- De Jong, M. G., Steenkamp, J.-B. E., & Fox, J.-P. (2007). Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. *Journal of Consumer Research*, *34*, 260–278.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, *39*, 1–22.
- Falk, C. F. (2021). A note on the interpretation and simulation of reparameterized intercepts in constrained versions of the nominal response model. *The Quantitative Methods for Psychology*, *17*, 345–354.
- Falk, C. F., & Cai, L. (2016). A flexible full-information approach to the modeling of response styles. *Psychological Methods*, *21*, 328–347.
- Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, *37*, 359–374.
- Fisher, R. A. (1925). Theory of statistical estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, *22*(5), 700–725.
- Fox, J.-P., & Verhagen, J. (2010). Random item effects modeling for cross-national survey data. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 467–488). Routledge.
- Geerlings, H., Glas, C. A., & Van Der Linden, W. J. (2011). Modeling rule-based item generation. *Psychometrika*, *76*, 337–359.
- Glas, C. A., & van der Linden, W. J. (2003). Computerized adaptive testing with item cloning. *Applied Psychological Measurement*, *27*, 247–261.

- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109.
- Holland, P. W., & Wainer, H. (2012). *Differential item functioning*. Routledge.
- Huang, S. (2021). *Estimation of cross-classified multilevel item response theory models with Metropolis-Hastings Robbins-Monro Algorithm*. University of California, Los Angeles.
- Janssen, R., Tuerlinckx, F., Meulders, M., & De Boeck, P. (2000). A hierarchical IRT model for criterion-referenced measurement. *Journal of Educational and Behavioral Statistics*, *25*, 285–306.
- Johnson, M. S., & Sinharay, S. (2005). Calibration of polytomous item families using Bayesian hierarchical modeling. *Applied Psychological Measurement*, *29*, 369–400.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*, 1087–1092.
- Monroe, S., & Cai, L. (2014). Estimation of a Ramsay-curve item response theory model by the Metropolis–Hastings Robbins–Monro algorithm. *Educational and Psychological Measurement*, *74*, 343–369.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*, 159–176.
- Muthén, B., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: Alignment and random effects. *Sociological Methods & Research*, *47*, 637–664.
- Pritikin, J. N., & Falk, C. F. (2020). OpenMx: A modular research environment for item response theory method development. *Applied Psychological Measurement*, *44*, 561–562.
- Rijmen, F., & Jeon, M. (2013). Fitting an item response theory model with random item effects across groups by a variational approximation method. *Annals of Operations Research*, *206*, 647–662.
- Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods*, *8*, 185–205.
- Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, *22*(3), 400–407.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika* *34*, 1–97.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. SAGE.
- Thissen, D., & Cai, L. (2016). Nominal categories models. In W. J. van der Linden (Ed.), *Handbook of item response theory (Vol. 1)*, pp. 51–73. Chapman and Hall/CRC.
- Thissen, D., Cai, L., & Bock, R. D. (2010). The nominal categories item response model. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models: Development and applications* (pp. 43–75). Taylor & Francis.
- Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, *49*, 501–519.
- Van den Noortgate, W., De Boeck, P., & Meulders, M. (2003). Cross-classification multilevel logistic models in psychometrics. *Journal of Educational and Behavioral Statistics*, *28*, 369–386.

- Wang, W.-C., & Qiu, X.-L. (2013). A multidimensional and multilevel extension of a random-effect approach to subjective judgment in rating scales. *Multivariate Behavioral Research, 48*, 398–427.
- Wang, W.-C., Wilson, M., & Shih, C. L. (2006). Modeling randomness in judging rating scales with a random-effects rating scale model. *Journal of Educational Measurement, 43*, 335–353.
- Wang, W.-C., & Wu, S. L. (2011). The random-effect generalized rating scale model. *Journal of Educational Measurement, 48*, 441–456.
- Yang, J. S., & Cai, L. (2014). Estimation of contextual effects through nonlinear multilevel latent variable modeling with a Metropolis–Hastings Robbins–Monro algorithm. *Journal of Educational and Behavioral Statistics, 39*, 550–582.
- Zhang, X., Noor, R., & Savalei, V. (2016). Examining the effect of reverse worded items on the factor structure of the need for cognition scale. *PLOS ONE, 11*, 1–15.