Unfolding IRT Models for Likert-Type Items With a Don't Know Option

Applied Psychological Measurement

2016, Vol. 40(7) 517–533

© The Author(s) 2016

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0146621616664047

apm.sagepub.com

Chen-Wei Liu¹ and Wen-Chung Wang¹

Abstract

Attitude surveys are widely used in the social sciences. It has been argued that the underlying response process to attitude items may be more aligned with the ideal-point (unfolding) process than with the cumulative (dominance) process, and therefore, unfolding item response theory (IRT) models are more appropriate than dominance IRT models for these surveys. Missing data and don't know (DK) responses are common in attitude surveys, and they may not be ignorable in the likelihood for parameter estimation. Existing unfolding IRT models often treat missing data or DK as missing at random. In this study, a new class of unfolding IRT models for non-ignorable missing data and DK were developed, in which the missingness and DK were assumed to measure a hierarchy of latent traits, which may be correlated with the latent attitude that a test intended to measure. The Bayesian approach with Markov chain Monte Carlo methods was used to estimate the parameters of the new models. Simulation studies demonstrated that the parameters were recovered fairly well, and ignoring nonignorable missingness or DK resulted in poor parameter estimates. An empirical example of a religious belief scale about health was given.

Keywords

item response unfolding models, nonignorable missing data, don't know option

Social surveys often have missing data as well as nonsubstantive response such as “don't know” (DK). They are also referred to as item nonresponses because respondents do not give a substantive response such as agreement or disagreement (Dillman, Eltinge, Groves, & Little, 2002). Nonresponse data can be classified into three types—missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR; Rubin, 1976). Let z_{com} denote the complete data, which consist of observed data z_{obs} and missing data z_{mis} . Let d denote the missingness indicator ($d = 1$ if observed and $d = 0$ otherwise). If $\Pr(d|z_{\text{com}}) = \Pr(d)$, that is, the distribution of the missingness indicator does not depend on z_{obs} or z_{mis} , the data are MCAR. If $\Pr(d|z_{\text{com}}) = \Pr(d|z_{\text{obs}})$, that is, the distribution of the missingness indicator depends on the observed data, not on the missing data, the data are MAR. If $\Pr(d|z_{\text{com}}) = \Pr(d|z_{\text{mis}})$, that is, the distribution of the missingness indicator depends on the missing data, the data are

¹The Education University of Hong Kong, Hong Kong

Corresponding Author:

Wen-Chung Wang, The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong.

Email: wcwang@eduhk.hk

MNAR. The former two mechanisms can be ignored in the likelihood-based inference (Little & Rubin, 1987), so this study focuses on MNAR.

In traditional statistical analysis, only z is treated as a random variable but d is not. In the framework proposed by Rubin (1976), both d and z are conceptualized as a random variable with a joint probability density function, $\Pr(d, z)$, so the underlying parameters controlling the realized values of d and z could be inferred. $\Pr(d, z)$ can be factored as $\Pr(d|z)\Pr(z)$ or $\Pr(z|d)\Pr(d)$. The former is called the selection model, and the latter is the pattern-mixture model (Little & Rubin, 1987). The two approaches are exchangeable, but they are different in interpretation and may not produce the same parameter estimates. In this study, the framework of the selection model was used to describe the missingness mechanism.

The development of linear regression, factor analysis, and item response theory (IRT) models for nonignorable missing data has been growing in recent years (Glas & Pimentel, 2008; Holman & Glas, 2005). In the IRT framework, the joint distribution of an IRT model for observed data and another IRT model for missing data can be expressed as $[\Pr(z|\theta_1)\Pr(d|\theta_2)g(\theta_1, \theta_2|\Phi)]$, where θ_1 is target latent attitude that the survey intends to measure and it defines the observed-data mechanism $\Pr(z|\theta_1)$; θ_2 is the latent "propensity" that defines the missingness mechanism $\Pr(d|\theta_2)$; and Φ is the parameters of the joint distribution of θ_1 and θ_2 . Item parameters are omitted in this formula for notational simplicity. This "latent-variable" approach enables the combination of IRT models for both observed-data and missing-data mechanisms, and thus, both kinds of information about respondents can be utilized.

In the literature of IRT models for MNAR data on Likert-type items, both models for observed data and missing data belong to dominance IRT models, in which the item response function increases monotonically as the latent trait increases (Glas & Pimentel, 2008; Holman & Glas, 2005). For example, the IRT models in $\Pr(z|\theta_1)$ and $\Pr(d|\theta_2)$ can both be the Rasch model (Rasch, 1960). Actually, when appropriate, the IRT model in $\Pr(z|\theta_1)$ or $\Pr(d|\theta_2)$ can be an unfolding IRT model, in which the probability of endorsing (agreeing with) an item depends on the distance between the person's location parameter and the item's location parameter: The shorter the distance, the higher the probability. Thus, a two-by-two array can be created. The IRT model in $\Pr(z|\theta_1)$ can be either a dominance or an unfolding model, and the IRT model in $\Pr(d|\theta_2)$ can be either a dominance or an unfolding model, leading to four possible combinations: dominance–dominance, dominance–unfolding, unfolding–dominance, and unfolding–unfolding. The dominance–dominance case has been investigated intensively in the literature, whereas the other three types are less studied.

In some cases, the missingness mechanism can be described with unfolding models. For instance, in a survey of income, wealthy people and poor people may be less willing to reveal their income than others, such that the missingness mechanism for income becomes nonignorable and has an inverted U-shaped relation with income (Lillard, Smith, & Welch, 1986). In general, the selection of the four combinations depends on practical needs and rationales.

In addition to missing data, DK is a common nonresponse in survey. The popularity of including a DK option in survey may be due to the "nonattitude-reduction" hypothesis: Offering a DK option will prevent those respondents without clear opinions from giving misleading responses if a DK option is not provided to them (Krosnick, 2002). If respondents cannot comprehend the question well, they may yield DK responses. Thus, DK could be regarded as an indicator of the level of the respondent's knowledge to the question. A DK option is sometimes treated as a midpoint response on an ordinal scale (Coombs & Coombs, 1976) or simply as missing data following the MCAR or MAR assumption. However, DK might elicit different utility that involves a different latent trait than the intended-to-be-measured (target) latent trait (Krosnick, 2002). Take a political question for an example: "Counties that can hold referendum are more democratic than others that cannot." This question is not hard to

understand, but it apparently demands sufficient knowledge about referendum. In addition, the term “more democratic” could elicit different meanings for different respondents. If the occurrence of DK is related to opinion expressions, the DK response is no longer ignorable and should not be treated as ignorable.

The role of a middle category (e.g., neutral, undecided, no opinion, DK) in a rating scale has been investigated. The reasons of choosing a middle category could be attributed to ignorance, ambivalence, knowledge, or indecisiveness (Baka, Figgou, & Triga, 2012; Hanisch, 1992; Krosnick, 2002; Nowlis, Kahn, & Dhar, 2002). In general, putting DK in the middle or the side of a scale induces different psychological processes and thus should be treated differently (Sturgis, Roberts, & Smith, 2014).

In this study, the authors gathered the information, if any, of missing data and DK responses which could not only reveal respondents' propensity to items but also increase measurement precision of the target latent attitude. A new model was proposed to treat missingness and DK as measuring two distinct latent variables, which can be correlated with the target latent attitude. In the following sections, unfolding IRT models are described, followed by the introduction to a new class of hierarchical IRT models for missingness and DK. A series of simulation studies were conducted to investigate parameter recovery of the new models and the consequences of ignoring nonignorable missingness and DK. The results are summarized. A data set of religious belief was analyzed to illustrate implication and applications of the new models. Finally, limitations and future directions are discussed.

Unfolding IRT Models

In attitude measurement, it is argued that an ideal-point (unfolding) process may be more appropriate than a cumulative (dominance) process (Roberts, Donoghue, & Laughlin, 2000; Roberts & Laughlin, 1996). Luo (2001) proposed a class of unfolding IRT models that encompass many unfolding IRT models. According to Luo's formulation, the rating scale maps a manifest polytomous response $Z \in (0, 1, \dots, C)$ to C dichotomous variables ($Y_1, \dots, Y_k, \dots, Y_C$), where $Y_k \in (0,1)$, $k \in (1, 2, \dots, C)$, and C is the number of options minus one. The probability function for a dichotomous rating scale response is given as follows:

$$\Pr_{jik} \equiv \Pr(Y_{jik} = 1 | \theta_j, \delta_i, \rho_{ik}) = \frac{f_k(\rho_{ik})}{f_k(\rho_{ik}) + f_k(\theta_j - \delta_i)}, \quad (1)$$

where Y_{jik} is the k th dichotomous variable (0,1); θ_j is the latent trait of person j ; δ_i is the location of item i ; ρ_{ik} is the threshold parameter, which describes the range on the continuum within which the response of a person to a statement is more likely to be positive than negative; and $f(t)$ for any real t could be defined as any function that is nonnegative, monotonic in the positive domain, and symmetric about the origin (Luo, 2001). Assuming the rating processes are independent, the probability of an observed polytomous response, given the person and item parameters, can be simplified as follows:

$$\Pr(Z_{ji} = z | \theta_j, \delta_i, \rho_{ik}) = \frac{\prod_{k=1}^C \Pr_{jik}^{U_{zk}} Q_{jik}^{1-U_{zk}}}{\sum_{w=0}^C \prod_{k=1}^C \Pr_{jik}^{U_{wk}} Q_{jik}^{1-U_{wk}}}, \quad (2)$$

where the dummy variable $U_{zk} = 1$ if $z \geq k$, otherwise $U_{zk} = 0$; $U_{wk} = 1$ if $w \geq k$, otherwise $U_{wk} = 0$; $\Pr_{jik}$ is defined in Equation 1; and Q_{jik} is $1 - \Pr_{jik}$. Different forms of $f(t)$ will result in

different unfolding models (Luo, 2001). In this study, an operational function $f(t)$ for variable k was adopted:

$$f_k(t) = \frac{\cosh\left[\left(\frac{2C+1}{2} + 1 - k\right)t\right]}{\cosh\left[\left(\frac{2C+1}{2} - k\right)t\right]}, \quad (3)$$

where $\cosh(\cdot)$ denotes the hyperbolic cosine function. The resulting model is referred to as the graded unfolding model (GUM), which is applicable for both dichotomous and polytomous items (Luo, 2001). Other unfolding models can be adopted, though the GUM was used in this study for illustration.

Unfolding IRT Models for MNAR Data

The authors proposed a selection model that combines an unfolding model for Likert-type items and a dominance or unfolding model for missingness. Based on the assumption of local independence, the posterior distribution of the person parameters and item parameters conditioned on observed data is proportional to

$$\prod_{j,i} \Pr(Z_{ji} = z | \theta_j, \beta_i) \Pr(D_{ji} = d | \theta_j^{(P)}, \beta_i^{(P)}) g(\theta_j, \theta_j^{(P)} | \Phi), \quad (4)$$

where $\Pr(Z_{ji} = z | \theta_j, \beta_i)$ is an unfolding IRT model; Z_{ji} is the random variable of observable responses to item i for person j , z is the realization of Z_{ji} ; D_{ji} is the missing-data random variable with indicator d equal to 1 when z is observed, and equal to 0 when z is not observed; θ_j is the target latent attitude of person j ; β_i is a vector consisting of parameters for item i (e.g., δ_i and ρ_i); $\Pr(D_{ji} = d | \theta_j^{(P)}, \beta_i^{(P)})$ describes the missingness mechanism; $\theta_j^{(P)}$ in this context can be viewed as the latent “propensity” for person j , which determines how likely the person would respond to the items; $\beta_i^{(P)}$ is a vector consisting of parameters for item i (e.g., b_i), which determines how likely the item would be responded to; and $g(\theta_j, \theta_j^{(P)} | \Phi)$ is a joint distribution indexed by parameters Φ . When θ_j and $\theta_j^{(P)}$ are independent, $g(\theta_j, \theta_j^{(P)} | \Phi) = g_1(\theta_j)g_2(\theta_j^{(P)})$, suggesting the missingness is ignorable. Note that the distributional forms of $g_1(t)$, $g_2(t)$, and $g(t)$ are all assumed to be the same (e.g., normal distributions).

In the literature, the missingness mechanism is often assumed to follow a dominance model such as the Rasch model:

$$\Pr(D_{ji} = d | \theta_j^{(P)}, b_i) = \frac{\exp(\theta_j^{(P)} - b_i)}{1 + \exp(\theta_j^{(P)} - b_i)}, \quad (5)$$

where b_i is the location parameter of item i ; the others are defined previously. It is possible that respondents with extreme latent propensity are less likely to respond than those with moderate latent propensity, resulting in an inverted U-shaped relationship between the missingness rate and the latent propensity ($d = 0$ means missing response). Under such cases, the use of unfolding models for missingness appears justifiable (e.g., Equations 1-3). Korobko, Glas, Bosker, and Luyten (2008) proposed an IRT model to account for the choice behavior among a limited number of examination subjects, in which the probability of choosing a subject is assumed to follow unfolding models. Survey of wage is another example. Respondents with either very low or very high income may be more reluctant than others to respond to items (Lillard et al.,

1986). Prior to using IRT models for missing data in practice, it is advisable to check missing response patterns with exploratory analysis, such as correspondence analysis (Polak, Heiser, & De Rooij, 2009). If items and persons are ordered along an arch on a two-component plot, the missingness mechanism might follow an unfolding process. In addition, nonparametric diagnostic for unimodality of missingness can be applied to check whether unfolding models are appropriate (Polak, De Rooij, & Heiser, 2012).

The Hierarchy of Missingness, DK, and Substantive Responses

Besides missingness, the proposed modeling above can be extended to deal with DK responses. When missingness and DK coexist in response patterns, the latter could be regarded as measuring a construct of knowledge level toward the item content. In this study, a hierarchical mental process in responding to an item was hypothesized based on literature review. First, a respondent decides whether to give a response. If the answer is no, then the response is missing. If the answer is yes, then the respondent proceeds to the second step to decide which option to endorse. Below, the authors describe such a hierarchy among missingness, DK, and substantive responses.

It is important to ascertain whether DK shares the same mental process as the missingness mechanism so that DK can be treated as missing data. The answer is probably “no.” Missingness could result from lacking interest, purposely skipping, or refusing to answer. If respondents lack interest in the questionnaire, they would probably ignore the questions without reading them carefully. When questions are too sensitive, respondents may purposely skip them or refuse to express their opinions. When respondents decide to answer a question, the following four steps may be involved (Tourangeau, 1984): (a) interpreting the question and activating relevant attitude, (b) retrieving an overall summary of knowledge, (c) rendering a judgment from the retrieved knowledge, and (d) mapping the judgment onto one of the response options. If respondents cannot accomplish the first two steps, they would tend to choose DK (Krosnick, 2002). The third and final processes are related to substantive responses. To sum up, it seems that the missingness mechanism precedes the mental process of DK, followed by that for substantive response processes.

Figure 1 summarizes the hypothetical hierarchy of missingness, DK, and substantive responses, and their corresponding IRT models, where circles denote latent variables and squares denote observed responses. The first process involves the missingness mechanism, in which a respondent decides whether to respond to an item. Let $1 - \Pr(\theta^{(P)})$ denote the probability of missing an item for a person with a propensity of $\theta^{(P)}$. When a respondent decides to respond, with a probability of $\Pr(\theta^{(P)})$, it comes to the second process of knowledge evaluation, in which the respondent searches relevant knowledge to understand the item. If not successful, the respondent would give DK with a probability of $1 - \Pr(\theta^{(DK)})$; if not, there is a probability of $\Pr(\theta^{(DK)})$ to move to the third process of attitude mapping, in which the respondent maps his or her attitude to one of the given substantive options (e.g., strongly agree, agree, disagree, and strongly disagree) with a probability of $\Pr(z|\theta)$, where z is a realized value. The item thresholds of missingness and DK are denoted as $b^{(P)}$ and $b^{(DK)}$, and interpreted as the missingness propensity and item ambiguity, respectively.

There are altogether six options in Figure 1: missingness, DK, strongly disagree, disagree, agree, and strongly agree (denoted as A-F), which are reorganized into three pseudo items (one dichotomous item for each of missingness and DK, and one 4-point item for the substantive options). Based on the assumption of local independence, the probabilities of the six options are shown in the last column of Table 1. The posterior distribution of the person parameters and item parameters conditioned on observed data is proportional to

Figure 1. Hypothetical hierarchy of missingness, “don’t know,” and substantive responses, and their corresponding probabilities.

$$\prod_{j,i} \Pr(z_{ij} | \theta_j, \beta_i) \Pr(d_{ij} | \theta_j^{(P)}, \beta_i^{(P)}) \Pr(x_{ij} | \theta_j^{(DK)}, \beta_i^{(DK)}) g(\theta_j, \theta_j^{(P)}, \theta_j^{(DK)} | \Phi), \quad (6)$$

where x_{ji} is the DK response for person j on item i ($x = 0$ indicates observed DK response, otherwise $x = 1$), $\theta_j^{(DK)}$ is the latent “knowledge” trait for person j which determines how likely the person would yield DK response, $\beta_i^{(DK)}$ is a vector consisting of parameters for the DK of item i , and $g(\theta_j | \Phi)$ is a joint distribution indexed by parameters Φ .

These probabilities can follow either a dominance IRT model or an unfolding IRT model, where appropriate. In practice, if there are no or very few missingness or DK, the corresponding pseudo items and options can be dropped, which reduces to conventional IRT models. Furthermore, in case a different hierarchy of missingness, DK, and substantive responses is theoretically more appealing than that presented in Figure 1, one can easily revise the hierarchy to develop customized models. Other hierarchical structures have been proposed in the studies of sequential choice processes, nested choice processes, sensitivity survey, extreme responses, and health-related survey (De Boeck & Partchev, 2012; Jeon & De Boeck, 2015). The authors did not argue that the hierarchy in Figure 1 was universally appropriate; rather, they concentrated on how to model nonignorable missingness and DK given a hierarchy. The modeling can be easily adjusted if a different hierarchy is adopted (Jeon & De Boeck, 2015).

Parameter Estimation

The parameters of the newly proposed unfolding models for nonignorable missingness and DK can be estimated with the Bayesian approach with Markov chain Monte Carlo (MCMC) methods, which has been adopted to estimate the parameters of unfolding models (Wang, Liu, & Wu, 2013). In this approach, priors for the parameters should be specified. The freeware JAGS

Table 1. Formulation of Pseudo Items and Corresponding Probabilities of Missingness, “Don’t Know,” and Substantive Responses.

Option	Pseudo-Item 1	Pseudo-Item 2	4-point item	Probability
A	0	—	—	$1 - \Pr(\theta^{(P)})$
B	1	0	—	$\Pr(\theta^{(P)})[1 - \Pr(\theta^{(DK)})]$
C	1	1	0	$\Pr(\theta^{(P)})\Pr(\theta^{(DK)})\Pr(z = 0 \theta)$
D	1	1	1	$\Pr(\theta^{(P)})\Pr(\theta^{(DK)})\Pr(z = 1 \theta)$
E	1	1	2	$\Pr(\theta^{(P)})\Pr(\theta^{(DK)})\Pr(z = 2 \theta)$
F	1	1	3	$\Pr(\theta^{(P)})\Pr(\theta^{(DK)})\Pr(z = 3 \theta)$

Note. “—” indicates “not observable.”

(Plummer, 2003) was used in this study, which allows users to implement their own customized models. No effort is needed to develop parameter estimation procedures and corresponding computer programs, making the new models readily available to practitioners. To facilitate the analysis, one can adopt the jagsUI packages in R to call JAGS. An example of jagsUI syntax is provided in Online Appendix A.

Method

A series of simulations were conducted to investigate parameter recovery of the new unfolding models and to evaluate the consequences on parameter estimation when nonignorable missingness or DK existed but were ignored. Two models were investigated: the unfolding model for nonignorable missingness (UMNonM) and the unfolding model for nonignorable missingness and DK (UMNonMD). In the UMNNonM, the data consisted of observed data and missing data, whereas in the UMNNonMD, the data consisted of observed data, missing data, and DK responses. The observed data in both models followed the GUM. In the UMNNonM, the missing data followed either a dominance IRT model (the Rasch model) or an unfolding model (i.e., the dichotomous version of the GUM), and the target latent attitude and the latent propensity for missingness followed a bivariate normal distribution. In the UMNNonMD, the missing data followed a dominance IRT model (the Rasch model), and DK followed another dominance IRT model (the Rasch model), and the target latent attitude, the latent propensity, and the latent “knowledge” for DK followed a multivariate normal distribution. The authors did not intend to compare the UMNNonM and UMNNonMD; instead, they aimed to evaluate parameter recovery of each model and consequences of ignoring nonignorable missingness or DK.

After the data were simulated from the UMNNonM or UMNNonMD, the data-generating models were fit to evaluate parameter recovery, and the corresponding constrained models (denoted as UMNNonM-C and UMNNonMD-C, where the correlation between latent variables was constrained at zero, indicating the missingness or DK was treated as ignorable) were fit to evaluate the consequences on parameter estimation when the nonignorable missingness or DK was ignored.

For the UMNNonM, major independent variables included (a) magnitude of the nonignorable missingness, indexed by the correlation between the latent attitude and the latent propensity ($\text{Cor} = 0, .4, \text{ and } .8$); (b) sample size (500 and 1,000); (c) test length (10 and 20 4-point items); and (d) missingness mechanism (dominance and unfolding). A Cor as high as .8 was used in simulation studies on nonignorable missingness (Glas & Pimentel, 2008; Holman & Glas, 2005), and the sample sizes and test lengths were adopted in studies on unfolding models

(Roberts et al., 2000; Roberts & Laughlin, 1996). For the UMNonMD, major independent variables included (a) magnitude of the nonignorable missingness and DK, indexed by the correlation among the three latent variables ($\text{Cor} = 0, .4, \text{ and } .8$); (b) sample size (500 and 1,000); and (c) test length (10 and 20 four-point items). The mean and variance of each latent variable were set at 0 and 1, respectively. The larger the correlation between the latent attitude and the other two latent variables for missingness and DK, the stronger the nonignorability would be. A zero correlation suggested ignorability. All responses were simulated and coded using the R program. There were 70 replications in each condition.

When the missing data followed the Rasch model, it was the location parameter that determined the rate of missingness for an item: The larger the location parameter, the larger the rate would be. The DK responses also followed the Rasch model; thus, the larger the location parameter for DK, the higher the probability of DK would be. The item location parameters were set between -2 and 2 , with an equal increment between adjacent items. When the missingness followed the dichotomous GUM, the item location parameters were set between -2 and 2 , with an equal increment between adjacent items, and the ρ -parameters were set at 1 for all items. The substantive responses in the UMNonM and UMNonMD followed the GUM, in which the item location parameters were set between -2 and 2 , with an equal increment between adjacent items, and $\rho_0 = 1.2$, $\rho_1 = 0.7$, and $\rho_2 = 0.4$. Similar configurations were used in the literature (Roberts et al., 2000; Roberts & Laughlin, 1996; Wang et al., 2013).

To identify the UMNonM or UMNonMD, the mean of every latent variable was set at 0. The variance–covariance matrix of these latent variables was estimated. Setting appropriate initial values for the item location parameters (δ -parameters) in the GUM was crucial because the opposite sign of the δ -parameters would lead to the same likelihood due to the symmetric item characteristic curves. The signs of the δ -parameters should be determined by content experts, and the order of the δ -parameters could be approximated by using the correspondence analysis (Polak et al., 2009). Previous simulation results showed that the order of the location parameters could be recovered fairly well (Polak et al., 2009). For other parameters, the initial values were internally generated by JAGS.

The configuration of the prior distributions for parameters in JAGS was as follows. The initial values of the δ -parameters were generated from correspondence analysis (Polak et al., 2009), where the signs of the δ -parameters were treated as known in the simulation studies. In real data analysis, the signs have to be determined by content experts. In the software GGUM2004 for unidimensional unfolding models (Roberts, Fang, Cui, & Wang, 2006), the signs of starting values for item locations are determined by principal components analysis but are free to vary during the estimation procedure. In the authors' experiences, setting signs to more δ -parameters yields better estimation, especially when unfolding models are multidimensional or sample sizes are small. A positively truncated normal prior distribution was used for positive b -parameters, whereas a negatively truncated normal distribution for negative b -parameters, both with means 0 and variances 10. For items of ambivalent contents where their signs were difficult to determine, a normal prior distribution with a large variance could be used. A positively truncated normal prior distribution was used for all ρ -parameters with mean 1 and variance 10. The prior distribution for all θ -parameters was a multivariate normal distribution with mean vector zero, where the Wishart distribution was adopted for the inverse variance–covariance matrix with the scale matrix equal to the identity matrix and the degrees of freedom equal to the number of latent variables. Overall, all the prior distributions were regulated as weakly informative. The burn-in period was 5,000, followed by drawing 5,000 MCMC samples. To ensure the sampled values were generated from a stationary distribution, the authors computed the Cramér–von Mises statistic (Heidelberger & Welch, 1983) via the coda package

in R. If it was not a stationary distribution for parameters, another 5,000 MCMC samples were drawn and assessed by the Cramér–von Mises statistic until convergence.

The bias and root mean square error (RMSE) in parameter estimates across the 70 replications were computed. As there were many items in the test, the mean absolute bias (MAB) and the mean root mean square error (MRMSE) across items were also computed:

$$\text{MAB}(\hat{\beta}) = \frac{\sum_{i=1}^I \left| \frac{\sum_{r=1}^R (\hat{\beta}_{ir} - \beta_i)}{R} \right|}{I}, \quad (7)$$

$$\text{MRMSE}(\hat{\beta}) = \frac{\sum_{i=1}^I \sqrt{\frac{\sum_{r=1}^R (\hat{\beta}_{ir} - \beta_i)^2}{R}}}{I}, \quad (8)$$

where $\hat{\beta}_{ir}$ is the parameter estimate for item i at replication r ; β_i is the true value of parameter; R was the number of replications (70); and I was the total number of items when β denoted the δ - or b -parameters, and the total number of thresholds when β denoted the ρ -parameters.

It was anticipated that (a) the parameters of the UMNonM and the UMNonMD could be recovered fairly well when the data-generating models were fit; (b) when the UMNonM and the UMNonMD were fit to data where the correlation between latent variables was zero (the missingness and DK were actually ignorable), the estimated correlation would be very close to zero and the other parameters would be recovered fairly well, suggesting it did little harm to fit unnecessarily complicated models (i.e., the UMNonM and the UMNonMD) to ignorable missingness and DK; and (c) when nonignorable missingness or DK existed but were ignored by fitting the UMNonM-C and the UMNonMD-C, the parameter estimates would be poor; furthermore, the larger the correlation between latent variables (i.e., the stronger the nonignorability), the worse the parameter estimation would be.

Results

Parameter Recovery of the UMNonM

Missingness following the Rasch model. As expected, when the UMNonM was fit, the parameters were recovered fairly well. The MAB and MRMSE values for the δ -parameters of the observed data (which are of greatest interest in practice) were between 0.02 and 0.05 ($M = 0.03$), and between 0.01 and 0.15 ($M = 0.11$), respectively. Besides, when the missingness was actually ignorable (i.e., $\text{Cor} = 0$), fitting the unnecessarily complicated UMNonM yielded a correlation very close to zero ($M = 0.02$) and a good recovery for the other parameters.

In contrast, ignoring the nonignorable missing data by fitting the UMNonM-C yielded poor parameter estimates (see the lower panel). When $\text{Cor} = .8$, the MAB and MRMSE values for the δ -parameters were between 0.08 and 0.15 ($M = 0.07$), and between 0.10 and 0.20 ($M = 0.13$), respectively. When $\text{Cor} = .45$, the parameter estimation was only slightly poorer than that when the data-generating model was fit. According to the Rasch model for missingness, persons with a lower latent propensity would yield more missingness than persons with a higher latent propensity. A positive correlation between the latent attitude and the latent propensity indicated that persons with a lower latent attitude would yield more missingness than persons with a higher

latent attitude. Constraining such a correlation at zero, the UMNonM-C yielded biased estimates: The larger the correlation, the worse the parameter estimation would be.

The sample size and test length had slight impact on parameter estimation. As expected, in general the larger the sample size and the longer the test, the better the parameter estimation would be.

Missingness following the unfolding model. In general, fitting the UMNonM yielded a good parameter recovery regardless of the correlation. The MAB and MRMSE values for the δ -parameters of the observed data were between 0.02 and 0.05 ($M = 0.03$), and between 0.09 and 0.15 ($M = 0.11$), respectively. Ignoring the nonignorable missing data by fitting the UMNonM-C yielded slightly poorer parameter estimates. The MAB and MRMSE values for the δ -parameters of the observed data were between 0.02 and 0.15 ($M = 0.05$), and between 0.09 and 0.18 ($M = 0.12$), respectively. For the δ -parameters, the maximum difference between $\text{Cor} = 0$ and $\text{Cor} = .8$ was 0.12, which was similar to when missingness followed the Rasch model.

Parameter Recovery of the UMNonMD

Figures 2 and 3 show the MAB and MRMSE values, respectively, when the UMNonMD and UMNonMD-C were fit. Fitting the UMNonMD yielded MAB and MRMSE values for the δ -parameters between 0.03 and 0.08 ($M = 0.04$), and between 0.14 and 0.25 ($M = 0.17$), respectively, indicating a good parameter recovery (see the upper two panels). On the contrary, when the nonignorable missingness and DK were treated as ignorable by fitting the UMNonMD-C, both the estimates for the δ - and $b^{(\text{DK})}$ -parameters were substantially biased (see the lower two panels), but those for the $b^{(\text{P})}$ -parameters were not. For example, when $\text{Cor} = .8$, the MAB values were between 0.19 and 0.31 ($M = 0.25$) for the δ -parameters, between 0.16 and 0.22 ($M = 0.19$) for the $b^{(\text{DK})}$ -parameters, but between 0.01 and 0.02 ($M = 0.01$) for the $b^{(\text{P})}$ -parameters. It appeared that ignoring the nonignorable missingness had no impact on the $b^{(\text{P})}$ -parameters, but a serious impact on the δ -parameters. This was because there were no missing data in the first step of missingness, so the parameter estimation was not biased, a small amount of missing data in the second step of DK, and a large amount of missing data in the final step of latent attitude, so the parameter estimation was subject to the ignorance. Finally, in general the larger the sample size and the longer the test, the better the parameter estimation would be.

Due to space constraints, the above results showed only the overall consequences of ignoring MNAR. Actually, different items had different consequences. The authors took 1,000 persons and 10 items as an example. Figure 4a to 4c shows the bias in the δ -parameters across the three correlations under three conditions: (a) the UMNonM-C with missingness following the Rasch model, (b) the UMNonM-C with missingness following the unfolding model, and (c) the UMNonMD-C with both missingness and DK following the Rasch model. When missingness followed the Rasch model, as shown in Figure 4a and 4c, it appeared that the higher the correlation, the more serious the negative bias in the δ -parameters. This was because the higher the latent propensity, the higher the latent attribute and the fewer the missingness would be. Thus, persons with a lower latent attitude tended to yield more missingness than persons with a higher latent attitude. Treating such an MNAR as MAR in the UMNonM-C or UMNonMD-C would underestimate the δ -parameters. When missingness followed the unfolding model, as shown in Figure 4b, the negative δ -parameters tended to be overestimated, whereas the positive δ -parameters tended to be underestimated, especially when the correlation was as high as .8. This was because of the proximity characteristic of the unfolding process where persons located further away from item locations would yield more missingness, making the estimated scale of the δ -parameter shrunken toward the mean of zero.

Figure 2. MAB for each kind of parameters (vertical axis) against the correlation (horizontal axis) between the latent attitude and latent missing and DK variables when the UMNNonMD (upper two panels) and UMNNonMD-C (lower two panels) are fit, where the missingness and DK variables follow the Rasch model.

Note. p is the number of persons and i is the number of items (e.g., $p500-i10$ means 500 persons and 10 items). MAB = mean absolute bias; DK = don't know; UMNNonMD = unfolding model for nonignorable missingness and DK; UMNNonMD-C = unfolding model for nonignorable missingness and DK-constrained.

An Empirical Example

The empirical data came from the Survey Research Data Archive in Taiwan, which aimed to develop a religious belief scale about health for Taiwanese adults (Liu, 2010). Data were collected from a random sample of 565 Taiwanese adults. There were 20 five-option items (strongly agree, agree, don't know, disagree, and strongly disagree). A sample item was "Church or Buddha temple is a place that helps us build social relationships." A respondent

Figure 3. MRMSE for each kind of parameters (vertical axis) against the correlation (horizontal axis) between the latent attitude and latent missing and DK variables when the UMNNonMD (upper two panels) and UMNNonMD-C (lower two panels) are fit, where the missingness and DK variables follow the Rasch model.

Note. p is the number of persons and i is the number of items (e.g., p500-i10 means 500 persons and 10 items). MRMSE = mean root mean square error; DK = don't know; UMNNonMD = unfolding model for nonignorable missingness and DK; UMNNonMD-C = unfolding model for nonignorable missingness and DK-constrained.

could disagree with this item because he/she considered that Church or Buddha temple could not help build social relationship or that Church or Buddha temple could not only help build social relationship but also serve many other important functions. The authors conducted principal components analysis with parallel analysis procedure on the item responses using the free-ware FACTOR (Lorenzo-Seva & Ferrando, 2006) and obtained an additional spurious dimension besides a substantive dimension and a correlation of $-.54$ between dimensions. Unfolding models appeared appropriate (Tay & Drasgow, 2012).

Figure 4. Bias (vertical axis) in the δ -parameters across the three correlations (horizontal axis) when there were 1,000 people and 10 items: (a) the UMNonM-C with missingness following the Rasch model, (b) the UMNonM-C with missingness following the unfolding model, and (c) the UMNonMD-C with missingness and DK following the Rasch model.
Note. DK = don't know; UMNonM-C = unfolding model for nonignorable missingness–constrained; UMNonMD-C = unfolding model for nonignorable missingness and DK–constrained.

The missingness rates for the 20 items varied between 0.5% and 3% ($M = 1.4\%$), which were so low that treating them as ignorable in this study would do little harm (Schafer, 1999). However, the rates of DK were not low, ranging from 6% to 27% ($M = 19\%$). If DK responses were mistakenly treated as MAR, biased parameter estimates would be yielded, as demonstrated in the above simulation study. Three models were thus fitted: (a) the unfolding model with nonignorable DK (UMNonD) where DK was treated as measuring a latent variable (knowledge evaluation), and it might be correlated with the latent attitude that the test intended to measure; (b) the constrained unfolding model (UMNonD-C) where DK was treated as

ignorable and constrained to be uncorrelated with the latent attitude; and (c) the GUM where DK was treated as a middle category along the 5-point scale.

The burn-in period was 20,000, followed by drawing 30,000 samples in three parallel MCMC chains. The prior distributions for parameters were the same as those in the simulation study. The Gelman–Rubin’s diagnostic statistic was used to check the MCMC convergence (Gelman & Rubin, 1992). The point estimates for the potential scale reduction factors among the 20 items were all very close to 1.0, which were smaller than the criterion of 1.1, indicating convergence of the MCMC samples. In addition, the Cramér–von Mises statistic suggested stationary distributions of parameters. In terms of overall data fit, the authors adopted the posterior predictive model checking with the outfit measure as a discrepancy measure (Masters & Wright, 1997). The posterior predictive p value (PPP) near zero (or unity) indicates that the model is underpredicting (or overpredicting) the quantity of interest (Levy, Mislevy, & Sinharay, 2009). The PPP values for the UMNonD, UMNonD-C, and GUM were .39, .49, and .02, respectively. Therefore, both the UMNonD and UMNonD-C had a good overall fit, whereas the GUM did not, suggesting treating DK as a middle category was not appropriate. For model comparison between the UMNonD and UMNonD-C, the deviance information criterion (DIC) was 26,773 for the former and 26,829 for the latter. Using a difference in the DIC of more than 10 as a criterion (Lunn, Jackson, Best, Thomas, & Spiegelhalter, 2012), one found the UMNonD had a better fit than UMNonD-C, suggesting DK was nonignorable.

Item 3 (Church or Buddha temple is a place that helps us build social relationships) and Item 4 (When I feel sad, religion can ease my sadness) were taken as an example to demonstrate the underlying process. The item characteristic curves of these two items and their mean observed scores and mean expected scores against $\hat{\theta} - \hat{\delta}$ are shown in Online Appendix B, based on the parameter estimates of UMNonD. The mean observed and expected scores were very close to each other, suggesting a good fit, and they both exhibited an inverted U-shape, suggesting an unfolding process.

In the UMNonD, the correlation between θ and $\theta^{(DK)}$ was .58, suggesting a moderate correlation. An empirical test reliability coefficient can be computed as a criterion for comparison between competing models (Glas & Pimentel, 2008). The test reliability of θ was .81 in the UMNonD, .79 in the UMNonD-C, and .88 for the GUM. Taking the UMNonD as a gold standard (as it was the model of choice), one found that the test reliability was underestimated by the UMNonD-C (which was because the correlation between the two latent variables was not considered) and overestimated by the GUM (which was because DK was mistakenly treated as measuring the same latent attitude as that by the four substantive options).

With respect to the δ -parameter estimates, the difference in the estimates obtained from the UMNonD and UMNonD-C was between -0.03 and 0.08 , suggesting a minor impact on the δ -parameter estimates when DK was ignored. The minor impact was consistent to that of the above simulation study when the correlation was moderate. With respect to the person θ estimates, the UMNonD and UMNonD-C yielded very similar results; most differences were below 0.2, and only six were larger than 0.5. The large differences were because the item responses either were very extreme (e.g., strongly disagree with most items) or had large missing rates (e.g., more than 75% of items were missing), which made person θ estimates unreliable.

Conclusion and Discussion

Missingness and DK are common in surveys, and they may not be ignorable in the likelihood for parameter estimation. A two-by-two array was proposed to account for observed data and missing data in this study, in which both observed data and missing data can follow a dominance or an unfolding model. In addition to the two-by-two array, a hierarchy of missingness,

DK, and substantive responses and their corresponding IRT models were proposed. The new models are multidimensional in that missingness and DK are assumed to measure two distinct latent variables, in addition to the latent attitude. The correlation between the latent attitude and each of the additional latent variables describes the degree of nonignorability. The hierarchy in missingness and DK is an example that enables users to develop their customized models. The choice between dominance and unfolding models for the mechanism of missingness or DK, and the hierarchy of missingness, DK, and substantive responses, should reflect the mental process involved (Glas & Pimentel, 2008). Posttest interview on respondents may provide some suggestions. In case there is no evidence or strong believe on the choice or the hierarchy, exploratory analyses such as correspondence analysis (Polak et al., 2009) and the Bayesian model averaging approach (Lunn et al., 2012) might be helpful. More studies are warranted.

A series of simulations were conducted to evaluate the parameter recovery of the new models and the consequences on parameter estimation when nonignorable missingness or DK was ignored. As expected, the parameter recovery was satisfactory when nonignorable data were correctly treated and became very poor if they were ignored. The empirical example of religious attitude illustrates the implication and applications of the new models. Because the GUM had a poor fit, treating DK as a middle category along a 5-point scale was not appropriate. The UMNonD had a better fit than the UMNonD-C, suggesting DK was nonignorable. The test reliability was underestimated when DK was mistakenly treated as ignorable but overestimated when DK was mistakenly treated as a substantive response.

To facilitate the use of these new models, the authors suggest the following steps for data analysis. First, determine whether the observed item response follows an unfolding or dominance process. One may adopt those non-IRT methods suggested by Tay and Drasgow (2012) or fit a dominance (cumulative) IRT model and an unfolding IRT model to the data and compare their fit (Tay, Ali, Drasgow, & Williams, 2011). Second, calculate the missing rate and DK rate for each item. As a rule of thumb, a rate below 5% is inconsequential (Schafer, 1999); a rate above 10% deserves investigation into the effect of missingness or DK (Dong & Peng, 2013). Third, determine whether the missing data follow an unfolding process using the means shown in the first step. Fourth, DK should be considered as measuring an extra dimension if it is located on the side of a scale and analyzed accordingly (e.g., UMNonMD). If DK is located in the middle, it might measure an extra dimension or the same latent trait as the other categories. The two approaches (e.g., UMNonMD and UMNonM) should be used and compared.

Like other simulation studies of complicated IRT models, the simulation design in this study was restrictive with a medium number of replications (70) and a few independent variables. A more comprehensive design and more replications are welcome in future studies. The IRT models in this study focus on the GUM and Rasch model. Other IRT models are possible. To account for the variation in latent traits, one can add covariates to the models, which calls for explanatory IRT. For example, the latent propensity can be regressed on a set of relevant covariates such as income or socioeconomic status. In this study, the authors followed common practices to treat the correlation between the latent attitude and the latent propensity as linear. When appropriate, the correlation can be nonlinear.

In the UMNonM with missingness following an unfolding model, it was found that ignoring the nonignorable missingness would lead to a shrunken scale of the item locations, which was mainly due to the symmetric property of unfolding models. In case the symmetric missingness is not supported, asymmetric unfolding models should be adopted to accommodate asymmetric missingness. Besides DK, other options such as "No Opinion," "Refusal," and "Hard to Say" are sometimes provided in surveys. When the rates of these nonresponses are significant, the proposed hierarchy and their corresponding IRT models in this study should be extended to accommodate these nonresponses. Discussion on the properties of these nonresponses has been

documented in the literature (Faulkenberry & Mason, 1978; Shoemaker, Eichholz, & Skewes, 2002; Tourangeau & Yan, 2007). Finally, the latent attitude that a test intends to measure can be multidimensional, which calls for multidimensional unfolding models. They can be integrated into the hierarchical structure to accommodate nonresponses.

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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The second author was sponsored by the General Research Fund (No. 842709), Research Grants Council, Hong Kong.

Supplemental Material

The online appendix is available at <http://apm.sagepub.com/supplemental>

References

- Baka, A., Figgou, L., & Triga, V. (2012). "Neither agree, nor disagree": A critical analysis of the middle answer category in voting advice applications. *International Journal of Electronic Governance*, 5, 244-263.
- Coombs, C. H., & Coombs, L. C. (1976). "Don't know" item ambiguity or respondent uncertainty? *Public Opinion Quarterly*, 40, 497-514.
- De Boeck, P., & Partchev, I. (2012). IRTrees: Tree-based item response models of the GLMM family. *Journal of Statistical Software*, 48(1), 1-28.
- Dillman, D. A., Eltinge, J. L., Groves, R. M., & Little, R. J. A. (2002). Survey nonresponse in design, data collection, and analysis. In R. M. Groves, D. Dillman, J. L. Eltinge, & R. J. Little (Eds.), *Survey nonresponse* (pp. 3-26). New York, NY: John Wiley.
- Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, 2(1), 1-17.
- Faulkenberry, G. D., & Mason, R. (1978). Characteristics of nonopinion and no opinion response groups. *Public Opinion Quarterly*, 42, 533-543.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-472.
- Glas, C. A. W., & Pimentel, J. L. (2008). Modeling nonignorable missing data in speeded tests. *Educational and Psychological Measurement*, 68, 907-922.
- Hanisch, K. A. (1992). The Job Descriptive Index revisited: Questions about the question mark. *Journal of Applied Psychology*, 77, 377-382.
- Heidelberger, P., & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31, 1109-1144.
- Holman, R., & Glas, C. A. W. (2005). Modelling non-ignorable missing-data mechanisms with item response theory models. *British Journal of Mathematical and Statistical Psychology*, 58, 1-17.
- Jeon, M., & De Boeck, P. (2015). A generalized item response tree model for psychological assessments. *Behavior Research Methods*. Advance online publication. doi:10.3758/s13428-015-0631-y
- Korobko, O. B., Glas, C. A., Bosker, R. J., & Luyten, J. W. (2008). Comparing the difficulty of examination subjects with item response theory. *Journal of Educational Measurement*, 45, 139-157.
- Krosnick, J. A. (2002). The causes of no-opinion responses to attitude measures in surveys: They are rarely what they appear to be. In R. M. Groves, D. Dillman, J. L. Eltinge, & R. J. Little (Eds.), *Survey nonresponse* (pp. 87-100). New York, NY: John Wiley.

- Levy, R., Mislevy, R. J., & Sinharay, S. (2009). Posterior predictive model checking for multidimensionality in item response theory. *Applied Psychological Measurement, 33*, 519-537.
- Lillard, L., Smith, J. P., & Welch, F. (1986). What do we really know about wages? The importance of nonreporting and Census imputation. *Journal of Political Economy, 94*, 489-506.
- Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data* (Vol. 539). New York, NY: John Wiley.
- Liu, Y.-R. (2010). *The effect of religion on the elderly's health in Taiwan* (A.S. Institute of Sociology: Survey Research Data Archive). Retrieved from <https://srda.sinica.edu.tw/search/gensciitem/1249>
- Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods, 38*(1), 88-91.
- Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: CRC Press.
- Luo, G. (2001). A class of probabilistic unfolding models for polytomous responses. *Journal of Mathematical Psychology, 45*, 224-248.
- Masters, G. N., & Wright, B. D. (1997). The partial credit model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (p. 113). New York, NY: Springer.
- Nowlis, S. M., Kahn, B. E., & Dhar, R. (2002). Coping with ambivalence: The effect of removing a neutral option on consumer attitude and preference judgments. *Journal of Consumer Research, 29*, 319-334.
- Plummer, M. (2003, March). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*. Paper presented at the Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). Vienna, Austria.
- Polak, M., De Rooij, M., & Heiser, W. J. (2012). A model-free diagnostic for single-peakedness of item responses using ordered conditional means. *Multivariate Behavioral Research, 47*, 743-770.
- Polak, M., Heiser, W. J., & De Rooij, M. (2009). Two types of single-peaked data: Correspondence analysis as an alternative to principal component analysis. *Computational Statistics & Data Analysis, 53*, 3117-3128.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment test*. Copenhagen: Danmarks Paedagogiske Institute.
- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24*, 3-32.
- Roberts, J. S., Fang, H.-r., Cui, W., & Wang, Y. (2006). GGUM2004: A windows-based program to estimate parameters in the generalized graded unfolding model. *Applied Psychological Measurement, 30*, 64-65.
- Roberts, J. S., & Laughlin, J. E. (1996). A unidimensional item response model for unfolding responses from a graded disagree-agree response scale. *Applied Psychological Measurement, 20*, 231-255.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581-592.
- Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8*(1), 3-15.
- Shoemaker, P. J., Eichholz, M., & Skewes, E. A. (2002). Item nonresponse: Distinguishing between don't know and refuse. *International Journal of Public Opinion Research, 14*, 193-201.
- Sturgis, P., Roberts, C., & Smith, P. (2014). Middle alternatives revisited: How the neither/nor response acts as a way of saying "I Don't Know." *Sociological Methods & Research, 43*, 15-38.
- Tay, L., Ali, U. S., Drasgow, F., & Williams, B. (2011). Fitting IRT models to dichotomous and polytomous data: Assessing the relative model-data fit of ideal point and dominance models. *Applied Psychological Measurement, 35*, 280-295.
- Tay, L., & Drasgow, F. (2012). Theoretical, statistical, and substantive issues in the assessment of construct dimensionality: Accounting for the item response process. *Organizational Research Methods, 15*, 363-384.
- Tourangeau, R. (1984). Cognitive science and survey methods. In M. S. T. Jabine, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey design: Building a bridge between disciplines* (pp. 73-100). Washington, DC: National Academies Press.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*, 859-883.
- Wang, W.-C., Liu, C.-W., & Wu, S.-L. (2013). The random-threshold generalized unfolding model and its application of computerized adaptive testing. *Applied Psychological Measurement, 37*, 179-200.