Title

Random Item MIRID Modeling and Its Application

Journal
Applied Psychological Measurement, 41(2)

ISSN
0146-6216

Authors
Lee, Yongsang
Wilson, Mark

Publication Date
2017-03-01

DOI
10.1177/0146621616675835

Peer reviewed

Article

Random Item MIRID Modeling and Its Application

Applied Psychological Measurement
2017, Vol. 41(2) 97–114

DOI: 10.1177/0146621616675835
journals.sagepub.com/home/apm

Yongsang Lee[1] and Mark Wilson[2]

Abstract

The Model With Internal Restrictions on Item Difficulty (MIRID; Butter, 1994) has been useful for investigating cognitive behavior in terms of the processes that lead to that behavior. The main objective of the MIRID model is to enable one to test how component processes influence the complex cognitive behavior in terms of the item parameters. The original MIRID model is, indeed, a fairly restricted model for a number of reasons. One of these restrictions is that the model treats items as fixed and does not fit measurement contexts where the concept of the random items is needed. In this article, random item approaches to the MIRID model are proposed, and both simulation and empirical studies to test and illustrate the random item MIRID models are conducted. The simulation and empirical studies show that the random item MIRID models provide more accurate estimates when substantial random errors exist, and thus these models may be more beneficial.

Keywords

MIRID model, random item, Rasch model, componential item response model

Introduction

Although educators and psychologists have recognized that a person's outcome on the achievement test, or attitude survey, is made up of several component processes, they have struggled to model this type of complex thinking. Often, they have ended up with modeling the product or the outcome only. However, in fact, they may also be interested in knowing about the processes that contribute to it. To meet the rising demand for models dealing with the relationship between an outcome and its component processes, various item response models have been introduced. Such models include the linear logistic test model (LLTM; Fischer, 1973, 1983), the multi-component latent trait model (MLTM; Whitely, 1980), the general component latent trait model (GLTM; Embretson, 1984), and the Model With Internal Restrictions on Item Difficulty (MIRID; Butter, 1994; Butter, De Boeck, &Verhelst, 1998).

Among these models, the MIRID model has been found to be useful for investigating a cognitive behavior in terms of its underlying cognitive processes. For instance, to test the underlying process in the structure of feeling guilty, Smits and De Boeck (2003) provided situations to

[1]Korea Institute for Curriculum and Evaluation, Seoul, Korea
[2]University of California, Berkeley, CA, USA

Corresponding Author:
Yongsang Lee, Korea Institute for Curriculum and Evaluation, Jeongdong Bldg., 21-15, Jeongdong-gil, Jung-gu, Seoul 100-784, Korea.
Email: yong21c@kice.re.kr

their participants with two types of questions about feeling guilty. The first type of questions were about the processes (components) that may lead to feeling guilty, and the second type of question was about feeling guilty itself. As the first type of questions concerned only one component process (e.g., norm violation, worrying, etc.), they were named as component items. On the contrary, the second type of question was designed to cover all the component processes (e.g., feeling guilty), and this type of question (e.g., summary item or universal task item) was named as a composite item (Butter et al., 1998).

The original MIRID model is, indeed, a fairly restricted model for a number of reasons. First, the original MIRID model (i.e., the Rasch-MIRID) was designed for only binary responses. Its application has been somewhat limited because many measurement instruments in both the cognitive and affective domains commonly use polytomous response item formats. Recent psychometric contributions (Lee & Wilson, 2009; Wang & Jin, 2009), however, enable us to deal with polytomous responses within the MIRID framework. Second, the original MIRID model does not allow for any individual differences when interacting with items. For some people, for example, guilt feelings may depend mainly on whether they feel that they have violated a moral, ethical, religious, or personal code in the situation (norm violation); whereas, for other people, a norm violation may be a less important factor in generating feelings of guilt. There may thus be individual differences in the process structure of feeling guilty, and it may be reasonable to assume that the effects of components vary from person to person. By introducing random weight effects for the item part of the MIRID model (e.g., the RW-MIRID; Smits, 2003; the random weight partial credit MIRID model [RW-PC-MIRID]; Lee & Wilson, 2009), it is now possible to investigate individual differences in the effects of the component processes. Because these extended MIRID models do not specify any random item effects, however, they still have limitations as follows:

a. They do not include an error term in the linear function for expressing the composite item parameter; thus, the underlying assumption of this model is that the composite item parameter is fully explained by the component item parameters. This assumption implies that the explanation of composite items about feeling guilty is perfect, which is unlikely to be true.
b. As items are fixed, this model cannot quantify the uncertainty of an item parameter (i.e., variance); however, in the context of item banks or automated item generation, the reliability of the item parameter is of concern, and the quantification of this uncertainty is needed to provide reliability information for an item.
c. The fixed-item MIRID model also might provide a biased parameter estimate of the component items' effect on the composite item parameter when substantial random item variance (e.g., random errors for the composite item) exists. Because the main purpose of the MIRID model is to show how component item parameters contribute to the composite item parameter, the quality of this component weight parameter is a main concern.

The need for random item effects may seem to be more theoretical and technical, but, indeed, there are substantive reasons. It may be more reasonable to treat items as random effects rather than fixed effects in certain situations, and, thus, the model may need to reflect these random item effects. In the application of the MIRID model, items may be random in two ways. First, items and the situations that the items address can be seen as a random sample from an item or situation universe. In fact, this is not a new concept and it has been addressed in Generalizability theory (G-theory; Brennan, 2001) and Generalizability in the Item Response Model (GIRM; Briggs & Wilson, 2007), though the formulation is weaker in both of these as there is no specific prediction from component items to composite items. Second, a composite item parameter may not be fully explained by component item parameters and there will be random errors. Although theories might identify the major underlying component processes, there may still be minor processes that the model does not reflect, and, perhaps, there may be more factors that affect the composite item parameter in addition to the component item parameters (e.g., variation of the scenarios that the items address). It is reasonable, therefore, to consider item effects as random under these considerations. To propose the MIRID model with random item effects, this article first illustrates the Rasch-MIRID model (e.g., the original MIRID model), and then addresses three types of MIRID models with random item effects.

The MIRID Model for Dichotomous Data: The Rasch-MIRID Model

The MIRID model for dichotomous data was described by Butter (1994) following original contributions by Paul De Boeck (cited in Butter (1994)) to explore the structure of a cognitive or affective construct (or task) in terms of its components. The phenomenon of feeling guilty, for example, can be explained in terms of a number of components such as responsibility, norm violation, negative self-evaluation, worrying, and tendency to rectify (Smits & De Boeck, 2003). The MIRID model is designed to handle situations in which one wants to know the underlying relationship between these components and feeling guilty, but does not know the values of its components (i.e., component item difficulty values). Under the MIRID framework, feeling guilty is treated as a composite concept whereas the other concepts (e.g., norm violation, worrying, and tendency to rectify) are treated as component concepts. To formulate the MIRID model, two different types of items are required (i.e., component items and composite items).

The composite item is an item that measures a concept that is composed of components. The composite item, with its relevant set of component items, is named as an ''item family'' in which it is assumed that the composite item effect is expressed as a linear function of the effect of the component items. The number of item families is decided based on how many situations are given to test the concept. To test guilt, for example, Smits and De Boeck (2003), in a study of Guilt, gave people 10 situations along with a set of items. Among these 10 situations, Table 1 displays only two situations along with items for illustrative purposes. As can be seen in this table, their instrument has three component items (i.e., norm violation, worrying, and tendency to rectify) and one composite item (i.e., guilt) for each situation. Each item family is thus composed of these four items, and because 10 situations were given to the people, 10 item families are specified.

The MIRID model for dichotomous responses is given by the following equation:

$$P\left(Y_{pi}=1|\theta_p\right) = \frac{\exp\left(\theta_p - \beta_i\right)}{1+\exp\left(\theta_p - \beta_i\right)}, \tag{1}$$

where θ_p indicates a person ability parameter for person p, and $\theta_p \sim N(0, \sigma_\theta^2)$, β_i indicates an item parameter for item i, which will vary in its functional form depending on whether the item is a component item or a composite item, and Y_{pi} indicates the response of person p, with ability θ_p to item i.

For the component items, $\beta_i = \beta_{sr}$, where β_{sr} indicates the rth component item in the sth item family. For the composite items, $\beta = \beta_{s(R+1)}$, where $\beta_{s(R+1)}$ indicates the composite item in the sth item family; it would be expressed as $\sum_{r=1}^{R} \gamma_r \beta_{sr} + \tau$. To interpret this expression, refer to the Guilt example described above (and in Table 1). In the Guilt example, there are three component items (norm violation, worrying, and tendency to rectify) and 10 item families responding to 10 situations (see the appendix). The composite item difficulty—which is about

Table 1. The Example of Guilt Items.

Family	Item type	Items	Situations
Item Family 1	Component 1 (Norm violation)	Do you feel like having violated a moral, an ethical, a religious, and/or a personal code?	You have been dating for some time a person you are not really in love with. When you break up, you find out that he or she was in love with you (and was taking the relationship very seriously). The break up hurts him or her considerably. (Break up)
	Component 2 (Worrying)	Do you worry about what you did or failed to do?	
	Component 3 (Tendency to rectify)	Do you want to do something to rectify what you did or failed to do?	
	Composite (Guilt)	Do you feel guilty about what you did or failed to do?	
Item Family 2	Component 1 (Norm violation)	Do you feel like having violated a moral, an ethical, a religious, and/or a personal code?	You have been a member of a brass band for some years now. As a result, you learned to play the trumpet for free. Now that you are skilled enough, you leave the band because you do not like the members of the band any more. (Trumpet)
	Component 2 (Worrying)	Do you worry about what you did or failed to do?	
	Component 3 (Tendency to rectify)	Do you want to do something to rectify what you did or failed to do?	
	Composite (Guilt)	Do you feel guilty about what you did or failed to do?	

feeling guilty—will, therefore, be expressed as a linear combination using three component item difficulties and the intercept in the MIRID model. As the composite item parameter is restricted to being a linear function of component item parameters (β_{sr}), a weight for each component (γ_r), and an intercept (τ) without an error term, the MIRID model is in fact a restriction on the simpler Rasch model just as is the LLTM (Fischer, 1973, 1983). In general, therefore, the MIRID model may not fit as well as a Rasch model (just as is the case for the LLTM).

The Random Item MIRID (RI-MIRID) Models

The Three Types of RI-MIRID Models

Although the concept of random item parameters is a relatively new concept, its usability has been discussed in a number of research papers (Brennan, 2001; Briggs & Wilson, 2007; De Boeck, 2008; Gonzalez, De Boeck, & Tuerlinckx, 2008; Janssen, Schepers, & Peres, 2004; Mislevy, 1988; Smits & De Boeck, 2004). Brennan (2001) and Briggs and Wilson (2007) assumed that items are a random sample from an item universe, and, accordingly, they treated items as random. In research aimed at the explanation of item parameter estimates using item predictors (item properties, item group, etc.), Janssen et al. (2004) divided items into multiple groups based on item properties and hence identified a multilevel model for the items. For example, when each group represents a specific content area, items within a group can be seen as a random sample from the item universe for that content, and, accordingly, they allow within-group variation in their model. As well, Mislevy (1988) and De Boeck (2008) shared essentially the same idea that items cannot be fully explained by item predictors, and the error term should be incorporated in the model. Finally, Gonzalez et al. (2008) saw the data as the product of interactions of respondents, items, and situations (i.e., three-mode data; Kroonenberg, 2005), and incorporated the random variation from both individual and situational differences into the model (the double structure structural equation model; 2sSEM).

The 2sSEM (Gonzalez et al., 2008) was designed initially in a context of investigating emotions: They modeled a causal relationship among people's emotional traits (De Boeck & Smits, 2006; Gonzalez et al., 2008) by considering both individual differences and situational differences in the structure of emotion. The 2sSEM and MIRID models could both be applied to investigate the structure of emotion with three-mode data, but their approaches are quite different, and the scientific questions that they embody are also different. The 2sSEM identifies the structure of an emotion in terms of people's emotional latent variables whereas the MIRID approach attempts to explain it in terms of item difficulties.

The MIRID model specifies two types of items (component and composite items), and, accordingly, two random effects (the random component items and the random composite intercept) can be formulated in the RI-MIRID model. Depending on which random effect is incorporated into the model, three different types of the RI-MIRID model can be formulated: (a) the fixed component-random composite intercept MIRID (the FR-MIRID), (b) the random component-fixed composite intercept MIRID (the RF-MIRID), and (c) the random component-random composite intercept MIRID (the RR-MIRID), which specifies the random component items and the random composite intercept. It is worth to particularly note that when the component items are random, the composite items become random as well because the composite items are expressed as a linear function of the component items. Thus, we must distinguish between the ''random composite'' models and the ''random composite intercept'' models. This article uses the term ''FF-MIRID'' (the fixed component and fixed composite MIRID) for the original MIRID model to differentiate it from the three types of RI-MIRID models. Table 2 summarizes these four models showing which random effect is incorporated into each model.

Table 2. Three Types of the Random Item MIRID Models and the Fixed-Item MIRID Model.

		Random effect for the component items	
		No	Yes
Random intercept for the composite item	No	1. FF-MIRID	3. RF-MIRID
	Yes	2. FR-MIRID	4. RR-MIRID

Note. MIRID = Model With Internal Restrictions on Item Difficulty; FF-MIRID = fixed component and fixed composite MIRID; RF-MIRID = random component-fixed composite intercept MIRID; FR-MIRID = fixed component-random composite intercept MIRID; RR-MIRID = random component-random composite intercept MIRID.

For all three RI-MIRID models, the basic model formulation is the same as for the FF-MIRID (i.e., the original MIRID) model:

$$p\left(Y_{pi}=1|\theta_p,\beta_i'\right)=\frac{\exp\left(\theta_p-\beta_i'\right)}{1+\exp\left(\theta_p-\beta_i'\right)},\quad \theta_p\sim N\left(0,\sigma_\theta^2\right). \tag{2}$$

For the component item, β_i' is β_{sr}, and for the composite item, β_i' is a linear combination of the component item parameters (β_{sr}), component weights (γ_r), and the intercept (τ). As the RI-MIRID models incorporate random effects for items, the expression of the item parameters will be different from the FF-MIRID model, and this expression will depend on which effects are considered random in the model. The FF-MIRID model expresses the composite item parameter as a linear function of the component item parameters (β_{sr}), their weight (γ_r), and an intercept (τ) without any random error.

The first type of random item MIRID model, the FR-MIRID, relaxes one underlying assumption of the FF-MIRID model.

FR-MIRID:

$$\begin{aligned}\beta_{s0} &= \sum_{r=1}^{R}\gamma_r\beta_{sr}+\tau+\zeta_s, \qquad \zeta_s\sim \mathrm{N}\left(0,\sigma_\zeta^2\right)\\ \beta_{sr} &= \beta_{sr}.\end{aligned} \tag{3}$$

Note that, in this equation (as for the two that follow), β_{s0} indicates a composite item parameter, β_{sr} indicates a component item parameter for the rth component in the sth item family, and ζ_s indicates the random effect for the composite item. By adding an error term to the linear function, the FR-MIRID model relaxes the underlying assumption that the components must exactly predict the composite. As this model treats component items as fixed effects and the composite intercept as a random effect by incorporating the random error term into the composite item, the model is called the fixed component-random composite intercept MIRID model (i.e., the FR-MIRID). This model is useful when there is a considerable random error for the composite item parameter. Because the main goal of the MIRID model is to quantify the effect of the component item parameters on the composite item parameter, the component weight is the main parameter of interest. If there is a considerable random error, and if the MIRID model does not take this considerable random error into account, the model may estimate the component weight parameter inaccurately. The FR-MIRID model, however, specifies this random error in the model; and it is expected to provide more accurate component weight parameter estimates, which is a key advantage of the FR-MIRID model.

The second type of random item MIRID model is the RF-MIRID. This model addresses measurement situations where component item effects are random because of randomly sampled scenarios without considering any random composite intercept. This measurement situation implies that the scenarios (situations or passages) that students face can be seen as random subsets of the universe of possible scenarios and that they affect students' responses to an item, and therefore the item parameter.

RF-MIRID:

$$\beta_{s0} = \sum_{r=1}^{R} \gamma_r \beta_{sr} + \tau, \qquad (4)$$
$$\beta_{sr} = \beta_r + \varepsilon_{sr}, \qquad \varepsilon_{sr} \sim N\left(0, \sigma^2_{\beta(r)}\right).$$

In the case of the guilt data, for example, students are given 10 different scenarios to test each component (e.g., *norm violation*). Students are asked whether they feel like they violated a moral, an ethical, a religious, or a personal code given 10 different scenarios to test the norm violation feeling. Their feelings, in fact, are dependent on the scenario (e.g., *break up, pen pal*). They might feel as if they violated a moral code much more in a scenario where they break up than in a scenario where they do not respond to their pen pal's letter. Their response to a norm violation item might consequently vary from scenario to scenario, and so would the norm violation item parameter. As one can come up with many different scenarios to test this component, it might be reasonable to assume that these scenarios are just a subset of a universe of scenarios, and are sampled from that universe at random. At this point, the norm violation item would have a random effect as a result of the randomness of the scenarios. The RF-MIRID model is designed to identify this random effect. In the equation, ε_{sr} indicates this random effect for the component item.

The third type of random item MIRID model is the RR-MIRID. This model addresses both random effects in the previous two models: the randomness among scenarios and the random error of the composite item.

RR-MIRID:

$$\beta_{s0} = \sum_{r=1}^{R} \gamma_r \beta_{sr} + \tau + \zeta_s, \qquad \zeta_s \sim \mathrm{N}\left(0, \sigma^2_{\zeta}\right) \qquad (5)$$
$$\beta_{sr} = \beta_r + \varepsilon_{sr}, \qquad \varepsilon_{sr} \sim N\left(0, \sigma^2_{\beta(r)}\right).$$

As the RR-MIRID model considers these two random effects, it is a model that covers both issues described in the FR-MIRID and the RF-MIRID models. In terms of the random item effect, the RR-MIRID model is the most complex model, and its complexity pays off when both of these random effects are considerable. If the model does not consider any random effects in either the component or composite item parameters, the estimator of the component weight might not provide an accurate estimate. This is investigated under various conditions through the comprehensive simulation studies.

Estimation Method: Markov Chain Monte Carlo (MCMC) Algorithm

Because the RI-MIRID models consider person and item parameters (θ_p and β_i) as random effects, they are referred to as crossed random effect models. For estimation in these models, the MCMC estimation is often applied. The MCMC method (Albert, 1992; Gelfand & Smith, 1990; Geman & Geman, 1984; Tanner, 1996) is frequently used in these kinds of complex models, and many approaches to implement MCMC have been introduced (Bernardo & Smith, 1994; Carlin & Louis, 1996; Gelfand & Smith, 1990; Gelman, Carlin, Stern, & Rubin, 1995;

Geman & Geman, 1984), and, accordingly, various computer programs have been developed (Albert, 1992; Patz & Junker, 1999; Spiegelhalter, Thomas, Best, & Lunn, 2003). Among these programs, WinBUGS (Bayesian inference Using Gibbs Sampling: Spiegelhalter et al., 2003) is the most popular and accessible, and thus, this study adopts this software to implement both the simulations and the empirical studies. As WinBUGS adopts the Bayesian approach for parameter estimation, a prior must be specified for all parameters.

In the MCMC methods, unknown parameters are drawn from their posterior distribution via the Gibbs sampler. For Gibbs sampling, suppose that θ is a person location parameter, β is the mean of the component item parameters, σ^2_β is the variance of the component item parameters, γ is a component weight, and σ^2_ζ is the variance of random errors for the composite item parameters. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_r)'$, $\boldsymbol{\sigma}^2_{\boldsymbol{\beta}} = (\sigma^2_{\beta(1),} \sigma^2_{\beta(2),} \ldots, \sigma^2_{\beta(r)})'$, and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_r)'$; then $\boldsymbol{\omega} = (\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\sigma}^2_{\boldsymbol{\beta}}, \sigma^2_\zeta)$ would be a vector of parameters governing the response of person p to item i, which is represented by a random variable y_{pi} where p = 1,2, . . . , P and i = 1,2, . . . ,I, and $\mathbf{Y}$ is a P $\times$ I response matrix. As a result, the full posterior distribution of the parameter given in the data is the conditional posterior for parameter ω_i given the other parameters $\omega_{(-i)}$, $p(\omega_i|\omega_{(-i)}, \mathbf{Y})$ where i = 1,2, . . . k. Therefore, sampling a parameter in the MCMC algorithm is conditional on the other parameters and the data, so the simulated sample of parameters represents a sample from the marginalized posterior (Kim, 2001). In fact, over a Markov Chain, $\omega^{(t)}$ is updated to $\omega^{(t+1)}$, and the MCMC parameter estimates are obtained by averaging these values for ω.

Simulation Studies

Simulation Study Design

To help understand which MIRID model is the best choice for given conditions, a set of simulation studies was conducted. Previous research showed that the number of item families in the MIRID model affects the quality of the parameter estimate (Lee & Wilson, 2009), and that, as the length of the test and the sample size increase, the model fit improves in general (Kang & Cohen, 2007; Li, Cohen, Kim, & Cho, 2009). This indicates that the parameter quality and model fit might always be clear with a large sample and a long test. In fact, having a large sample size and a lengthy test comes with a cost. As large samples are quite challenging to obtain, it is important to examine the behavior of the models with relatively smaller samples and shorter tests. Furthermore, depending on the relative size of the random item effects, one MIRID model might perform better than others, and thus it is also important to explore model performance depending on the relative size of the random effect for the items (e.g., the random error of the intercept for the composite items and the random variation of the component items).

The conditions used in the data simulation include (a) the number of item families (10 item families and 30 item families), (b) the item standard deviation within each component (component item standard deviation, 0.5 and 1), (c) the standard deviation of random error of the intercept for the composite item (0, 0.5, and 1), and (d) the number of examinees (300 and 600). To make the data more realistic, these simulation studies used the parameter estimates from the empirical study with guilt data (Smits & De Boeck, 2003) as parameters for the item mean of each component (−0.07, −0.5, and −0.6) and component weights (0.025, 0.5, and 0.55). The intercept is set to be 0 for data generation. A total of 24 (2 $\times$ 2 $\times$ 3 $\times$ 2) conditions were simulated by varying the four simulation design factors. Note that Conditions (b) and (c) above determine which of the three RI-MIRID models are the data generators. First, as there is no condition in which both the item standard deviation within each component is 0 and the standard deviation of random error of the intercept for the composite item is 0, there is no condition under which FF-MIRID is the generating model. The simulations are not studied under which FF-MIRID is the generator, as this has been well studied in the past (Butter, 1994; Butter et al., 1998). Second, as there is no condition under which the item standard deviation within each component is 0, there is no condition under which FR-MIRID is the data generator. The simulations have been designed in this way because the main point that is intended to show in the simulation studies is that, when there are substantial random errors for the composite item parameter (which the authors believe is the common case in reality), the models that do not specify any random errors for the composite item parameter provide inaccurate component parameter estimates (De Boeck, 2008). Thus, all of the generating models are either RF or RR.

After the data were generated, the four different MIRID models (i.e., the FF-MIRID, the FR-MIRID, the RF-MIRID, and the RR-MIRID) were applied to the simulated data to estimate the parameters. For the MCMC estimation, priors are needed for each model. Priors for the random component item effects were $N(0, \sigma_\beta^2)$ and σ_β^2 distributed as an inverse gamma (0.001, 0.001). For the random composite intercept, the prior was $N(0, 1)$.

For analyses using the four different MIRID models under the 24 conditions, WinBUGS Version 1.4.3 was run with three chains for 3,000 iterations each, and the burn-in period was 1,000. For the iterative procedure of the simulation studies, R2WinBUGS (Sturtz, Ligges, & Gelman, 2005) was also applied. For the convergence check, the $\hat{R}$ index was used. The authors used the criterion that the $\hat{R}$ values should be below 1.1 to indicate that the estimation had converged (Gelman & Rubin, 1992). Depending on the simulation conditions and the model, the time taken to analyze each data set in WinBUGS varied from approximately 20 min to 1 hr.

After the simulation studies, parameter recovery was evaluated to ensure the quality of the estimators across the conditions based on the root mean square error (RMSE), and these RMSE values were compared across different conditions and models. The RMSE values are based on the discrepancy between true parameters and estimated parameters and indicate estimator stability and accuracy.

Simulation Results

First, the simulation results show the quality of the component weight estimation is affected by the random effect for the composite items (ζ_s). On one hand, when the true model does not specify the random intercept for the composite items, the models which do not consider this effect (e.g., FF-MIRID and RF-MIRID models) estimate the component weight more accurately than the other models do. On the other hand, if the true model assumes that there is a random effect for the composite intercept (ζ_s), the FR- and RR-MIRID models perform better than the FF- and RF-MIRID models (i.e., the models that do not consider the random intercept for the composite items (ζ_s)). The simulation results also show that the quality of the estimation is substantially affected by the simulation conditions that the current study varied, and that it depends also on the model applied.

The quality of the component weight ($\hat{\gamma}_r$) estimation is significantly different across models. Table 3 shows the average RMSE values for three component weights across simulations by the data generator, and box plots visualize these RMSE values. The box plots in Figure 1 show the distributions of the average RMSE values by simulation conditions when the RF-MIRID model is the data generator. The x axis indicates the simulation conditions, and the y axis indicates the RMSE values in the box plot. The first line of box plots shows the simulation results by changing the number of the item family, the second line is by the number of examinees, and the third line is by component item standard deviation. As can be seen in the box plots in Figure 1, the FF-MIRID and RF-MIRID models do better than the FR-MIRID and the RR-MIRID models.

Table 3. Average RMSE Values for the Component Weights Under Each Model.

True model	Number	Number of item family	Number of examinee	σ_β	σ_ζ	FF	FR	RF	RR
RF-MIRID	1	10	300	0.5	0	0.104	0.176	0.116	0.203
	2	10	300	1	0	0.073	0.106	0.075	0.108
	3	10	600	0.5	0	0.093	0.202	0.097	0.243
	4	10	600	1	0	0.102	0.445	0.205	0.377
	5	30	300	0.5	0	0.102	0.161	0.098	0.156
	6	30	300	1	0	0.059	0.067	0.059	0.073
	7	30	600	0.5	0	0.171	0.198	0.182	0.201
	8	30	600	1	0	0.080	0.114	0.081	0.112
					M	0.098	0.184	0.114	0.184
RR-MIRID	9	10	300	0.5	0.5	0.612	0.434	0.706	0.471
	10	10	300	0.5	1	1.890	0.279	1.657	0.276
	11	10	300	1	0.5	0.177	0.138	0.197	0.172
	12	10	300	1	1	0.575	0.249	0.598	0.252
	13	10	600	0.5	0.5	0.985	0.375	0.623	0.354
	14	10	600	0.5	1	2.952	1.219	1.772	1.171
	15	10	600	1	0.5	0.173	0.148	0.177	0.161
	16	10	600	1	1	4.492	0.433	1.543	0.452
	17	30	300	0.5	0.5	0.966	0.219	0.516	0.221
	18	30	300	0.5	1	3.413	0.241	2.695	0.237
	19	30	300	1	0.5	0.095	0.096	0.102	0.088
	20	30	300	1	1	0.623	0.071	0.422	0.066
	21	30	600	0.5	0.5	0.527	0.204	0.578	0.222
	22	30	600	0.5	1	1.516	0.412	1.637	0.418
	23	30	600	1	0.5	0.259	0.137	0.266	0.149
	24	30	600	1	1	0.609	0.321	0.617	0.328
					M	1.2415	0.311	0.882	0.315

Note. RMSE = root mean square error; FF = fixed component and fixed composite; FR = fixed component-random composite intercept; RF = random component-fixed composite intercept; RR = random component-random composite intercept; RF-MIRID = random component-fixed composite intercept MIRID; RR-MIRID = random component-random composite intercept MIRID; MIRID = Model With Internal Restrictions on Item Difficulty.

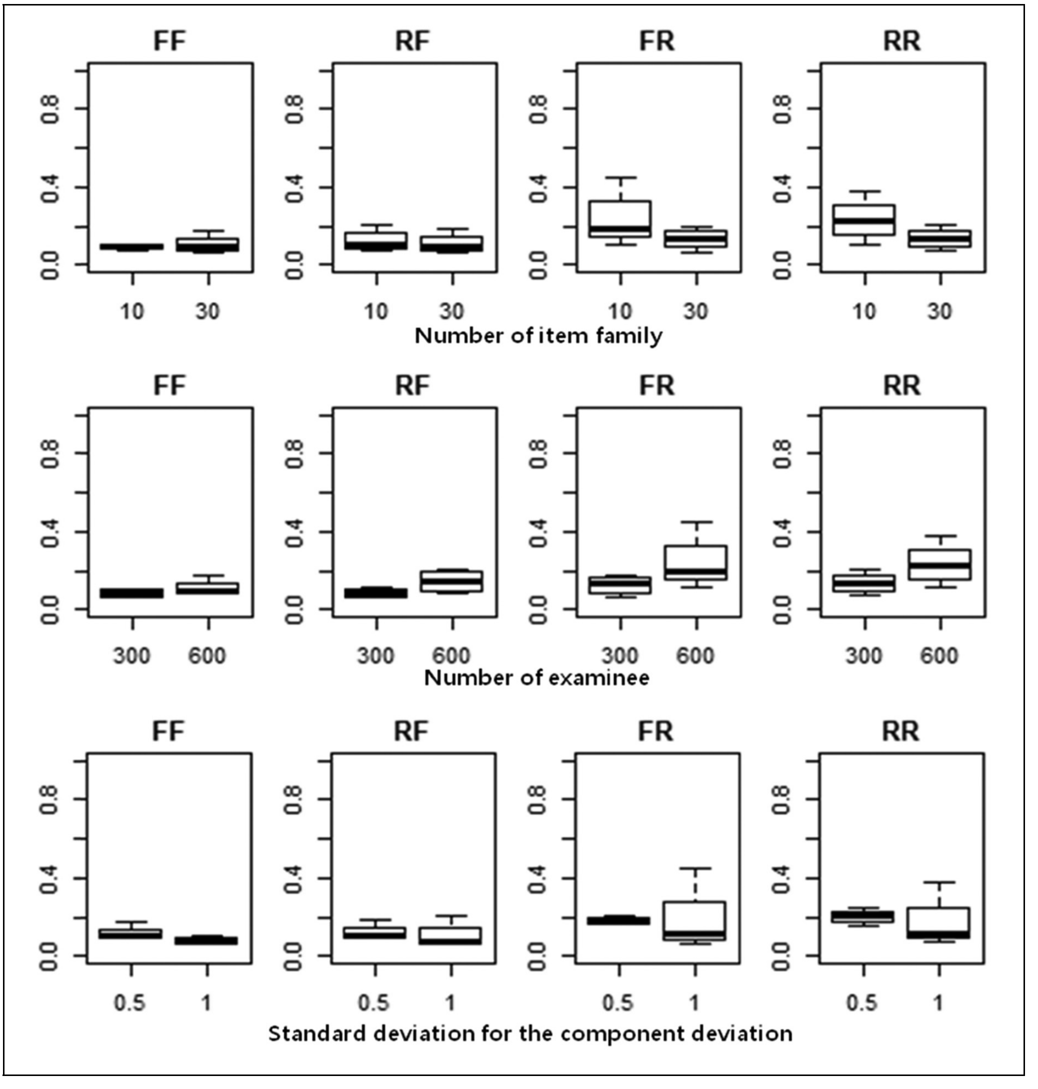

Figure 1. Average RMSE value distribution for the component weight by simulation conditions when the RF-MIRID is generator.

Note. RMSE = root mean square error; RF-MIRID = random component-fixed composite intercept MIRID; FF = fixed component and fixed composite; FR = fixed component-random composite intercept; RF = random component-fixed composite intercept; RR = random component-random composite intercept; MIRID = Model With Internal Restrictions on Item Difficulty.

As the true model is the RF-MIRID model in this figure, estimating error term for the composite intercept creates variability which leads to poor component weight estimates ($\hat{\gamma}_r$) of the FR-MIRID and RR-MIRID models. In addition, this figure clearly shows that the FF-MIRID performs better than the RF-MIRID model, and this result can be understood that the FF-MIRID model is more flexible than the RF-MIRID model because it allows the fixed component item parameters.

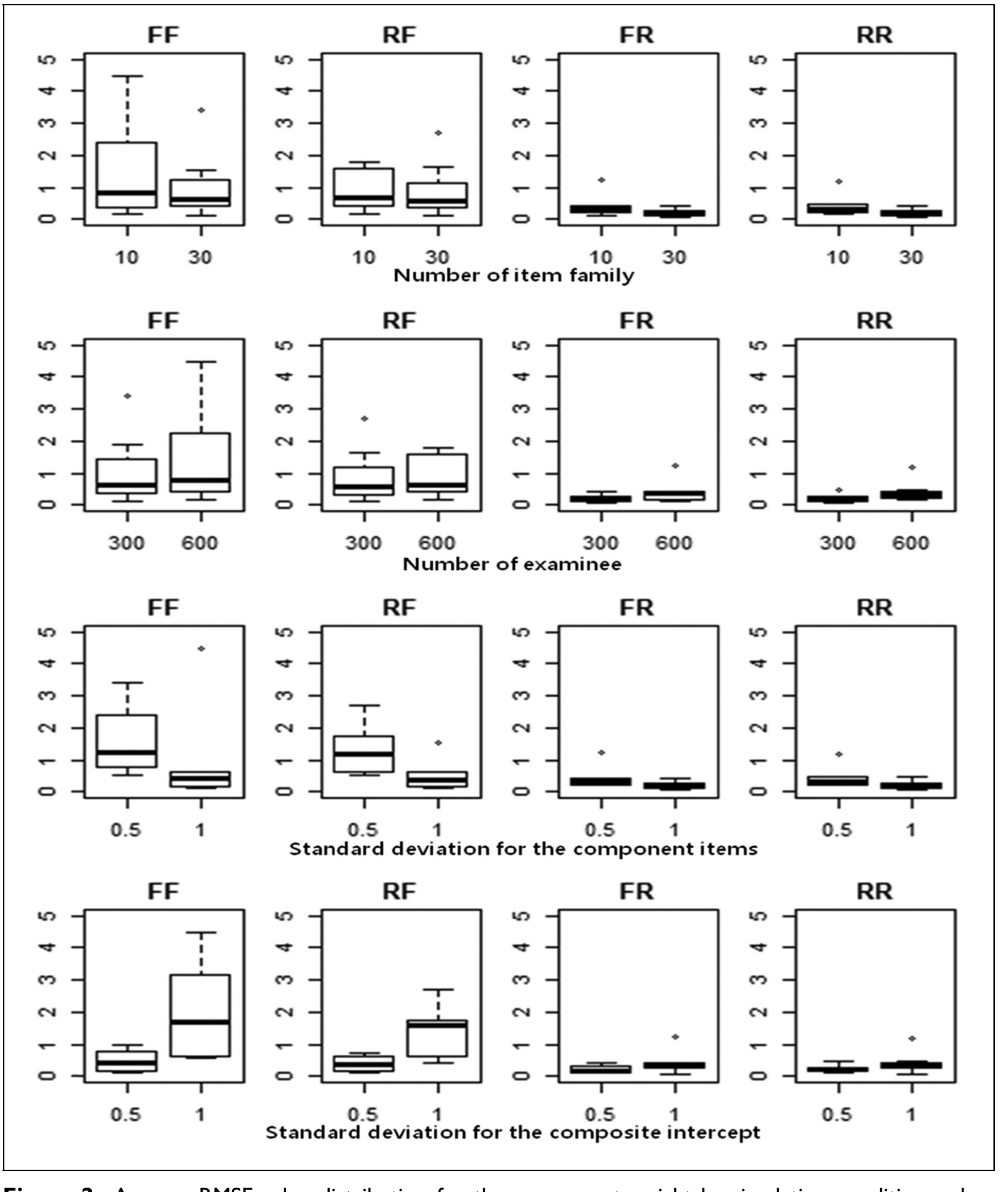

Figure 2. Average RMSE value distribution for the component weight by simulation conditions when the RR-MIRID is generator.

Note. RMSE = root mean square error; RR-MIRID = random component-random composite intercept MIRID; FF = fixed component and fixed composite; RF = random component-fixed composite intercept; FR = fixed component-random composite intercept; RR = random component-random composite intercept; MIRID = Model With Internal Restrictions on Item Difficulty.

Figure 2 also shows the distribution of the average RMSE values when the RR-MIRID is the data generator. As can be seen in the box plots, the FR-MIRID and the RR-MIRID models do much better than the FF-MIRID and RF-MIRID models. This result was, in fact, expected

Table 4. Means and Standard Deviations for Guilt Data.

Situation	Observation	Component 1 (Norm violation)		Component 2 (Worrying)		Component 3 (Tendency to rectify)		Composite (Guilt)	
Break up	268	0.49	(0.50)	0.77	(0.42)	0.61	(0.49)	0.67	(0.47)
Trumpet	268	0.09	(0.28)	0.15	(0.36)	0.14	(0.35)	0.14	(0.35)
Shoes	268	0.35	(0.48)	0.52	(0.50)	0.42	(0.49)	0.49	(0.50)
Movie	268	0.51	(0.50)	0.57	(0.50)	0.57	(0.50)	0.63	(0.48)
Discussion	268	0.77	(0.42)	0.84	(0.37)	0.90	(0.31)	0.86	(0.35)
Secret	268	0.91	(0.28)	0.87	(0.34)	0.84	(0.37)	0.92	(0.27)
Youth movement	268	0.75	(0.43)	0.84	(0.36)	0.67	(0.47)	0.83	(0.37)
Pen pal	268	0.48	(0.50)	0.45	(0.50)	0.52	(0.50)	0.49	(0.50)
Jacket	268	0.51	(0.50)	0.83	(0.37)	0.94	(0.24)	0.75	(0.43)
Homework	268	0.26	(0.44)	0.23	(0.42)	0.30	(0.46)	0.23	(0.42)

Note. Values in parenthesis indicate standard deviation.

because the FR-MIRID and the RR-MIRID models take the random errors in the intercept into account. As, in reality, it is quite rare that any dependent variable is explained without random errors (De Boeck, 2008), the FR-MIRID and the RR-MIRID models are more realistic as well as beneficial in the sense just described.

In this simulation study, it appears that the number of item families and examinees, and component item deviations (σ_β) affect the RMSE values, indicating the quality of estimator. The advantage of the FR- and RR-MIRID over other MIRID models, however, remains regardless of the simulation conditions when the intercept random errors exist.

Illustration of the RI-MIRID Models Using the Guilt Data

Data and Method

Smits and De Boeck (2003) investigated the components of guilt. They introduced the new approach using the MIRID model to test this componential structure of guilt, and showed advantages of the MIRID model for examining the componential structure of emotions. In their study, to test this structure, they collected 10 situations including ''break up'' and ''trumpet'' situations (see the appendix), and also derived five components of guilt from a literature review (*responsibility, norm violation, negative self-evaluation, worrying*, and *tendency to rectify*). After preliminary studies, they selected three components (e.g., *norm violation, worrying*, and *tendency to rectify*) to investigate guilt and its relationship with these three components (see the appendix for the items) with the MIRID model. They sampled 270 students and then gave them four questions to which they were to respond for 10 situations asking them to circle one of four choices (0 to 3). The scale represents as follows: ''0'' indicated ''no,'' ''1'' indicated ''not likely,'' ''2'' indicated ''likely,'' and ''3'' indicated ''yes.'' Two hundred sixty-eight students between 17 and 19 years old (138 females and 130 males) responded to their questionnaire.

In this example, the data are dichotomized for simplicity of illustration, so 0 and 1 are recoded as 0, and 2 and 3 are recoded as 1. Table 4 shows the mean and standard deviation for each item after dichotomization.

As shown in this table, the mean of responses for each item varies considerably across situations, but within each situation, the means of the three components and the composite items are fairly close to each other. These descriptive statistics imply that the random effect for

Table 5. Analysis Results Using Three RI-MIRID Models.

	FF	FR	RF	RR
γ_1	0.50	0.55	0.50	0.55
γ_2	0.55	0.59	0.54	0.40
γ_3	0.03	−0.05	0.04	0.13
τ	0.20	0.20	0.20	0.24
σ^2_θ	1.12	1.14	1.13	1.14
β_1			−0.07	−0.06
β_2			−0.59	−0.60
β_3			−0.51	−0.55
$\sigma^2_{\beta(1)}$			3.02	2.98
$\sigma^2_{\beta(2)}$			2.29	2.88
$\sigma^2_{\beta(3)}$			3.36	3.24
σ^2_ζ		0.020		0.019
Deviance	10,549	10,035	10,036	10,035
AIC	10,619	10,107	10,050	10,051

Note. RI-MIRID = The Random Item MIRID; FF = fixed component and fixed composite; FR = fixed component-random composite intercept; RF = random component-fixed composite intercept; RR = random component-random composite intercept; AIC = Akaike information criterion; MIRID = Model With Internal Restrictions on Item Difficulty.

component items might be considerable but the random error for the composite item is not indicating the component items might substantially explain the composite item.

In this example, three different RI-MIRID models are applied to the guilt data, and the consistency of the parameter estimates across models is examined. Second, the FF-MIRID model is compared with the RI-MIRID models in terms of model fit.

Results

Table 4 indicates that the results from three RI-MIRID models are consistent with previous research using the FF-MIRID model which is the original MIRID model (Smits & De Boeck, 2004). Again, depending on whether the model specifies the random effect in the component items and the composite intercept, the component weights are somewhat different from one another. The results indicate that the component item variances are fairly large, indicating that the component item parameters are substantially different across situations while the random error for the composite intercept is relatively small. The RF-MIRID model, thus, might be the best choice for this data. Akaike information criterion (AIC) values in Table 5 also indicate that these component item variances play a role in the model fit, and support that the RF-MIRID model is the best model for this data.

Discussion and Conclusion

This article presented a motivation for random item modeling for the MIRID model and proposed three types of RI-MIRID models. To clarify the distinctions among the random item approaches (e.g., the RF-MIRID, the FR-MIRID, and the RR-MIRID) and the fixed-item approach (e.g., the FF-MIRID), a set of simulation studies were conducted. For illustrative purposes, an empirical study with guilt data was presented as well.

As addressed in this article, the component weights are the most interesting parameters in the MIRID model, and thus the quality of the estimation of that parameter should be considered most crucial when using the MIRID model. Based on simulation studies, however, it is concluded that the FF-MIRID and RF-MIRID models do not estimate the component weights well when a substantial random error (ζ) for the composite intercept exists which might be the common case in reality. This might have been anticipated, as one can expect that the quality of component weight estimation is very sensitive to the size of the composite intercept deviation.

The empirical study displays the application of the RI-MIRID models to the data of feeling guilty and demonstrates the interpretation of the parameter estimates. The results indicate that the parameter estimates from the RI-MIRID models are consistent with previous research using the FF-MIRID model, and fit better compared with the FF-MIRID, although the RI-MIRID models are fairly restricted models. For the RF-MIRID and RR-MIRID results, the component item variances ($\sigma^2_{\beta(1)}$, $\sigma^2_{\beta(2)}$, $\sigma^2_{\beta(3)}$) were found to be substantial whereas the random error (ζ) for the composite intercept is relatively small indicating the RF-MIRID model can be the most appropriate for this data. As the component and composite items are identified with a unique situation (stimulus), they may be dependent. This local item dependency is a typical issue that one should consider, and it is well known that the violation of local independence can lead to inaccurate parameter estimates of the item response models. The advantage of the RI-MIRID model is that it allows one to detect this local item dependency. When the effect size of the component item parameter on the composite item parameter is substantial, that might indicate local item dependency across situations (e.g., stimulus, scenarios, or item families). The other advantage of the RI-MIRID model is that this model allows to expand the usual set of possible tactics for dealing with local item dependence (which includes multidimensionality, item bundling, differential item functioning [DIF], etc.) by introducing an item family into the model. As the RI-MIRID models take account of the local item dependency by specifying the item family in the model formulation, they may fit the data better than other Rasch models.

Further Steps to Develop MIRID Models

The models discussed in this article are all unidimensional, but different cognitive components do not always fit a unidimensional model (Butter, 1994); for example, *norm violation, worrying*, and *tendency to rectify* in the *guilt feeling* data may represent three different dimensions, and multidimensionality should be incorporated into the MIRID model. To deal with this multidimensional situation, Butter (1994) proposed the multidimensional MIRID model. His approach is, however, the consecutive approach to multidimensionality (i.e., treating each dimension independently) and does not consider random item effects. Therefore, a further study should be carried out to overcome the limitations of the consecutive approach within the random item MIRID model framework.

In the RF- and the RR-MIRID models of the study, a common variance is specified for all components. If there is, however, any strong theory or empirical evidence to support a unique variance for each component, this unique variance for each component should be formulated in the RF- and RR-MIRID models. The RF- and RR-MIRID models with unique component variances are, in fact, more flexible compared with the models in the current study, and thus fit data better reflecting measurement situations more accurately. These advantages may, however, come with challenges in estimation, thus need to be examined with various conditions.

Appendix

Guilt Items

Situation 1. You have been dating for some time a person you are not really in love with. When you break up, you find out that he or she was in love with you (and was taking the relationship very seriously). The break up hurts him or her considerably. (Break up)

Situation 2. You have been a member of a brass band for some years now. As a result, you learned to play trumpet for free. Now that you are skilled enough, you leave the band because you do not like the members of the band any more. (Trumpet)

Situation 3. During the holidays, you are working as a salesperson in a clothing and shoe store. One day, a mother with four children enters the store. One of the kids wants Samson-shoes (Samson is a popular doll figuring in a Belgian TV-series for children). The mother leaves the child with you while she goes on to look for clothes for the other children. The child tries on different types and sizes of shoes, but after a while the child gets tired of fitting the shoes and refuses to continue. She picks a pair she has not tried on before and you sell this pair to the mother afterward. The next day, the mother wants to return the shoes because they do not fit. Your boss takes back the shoes and reimburses the mother. The shoes have been worn, however, and they are dirty. Because of this, they cannot be sold anymore. Your boss says that it does not matter, and that everyone is capable of mistaking the size of shoes. (Shoes)

Situation 4. A not so close friend asks you if you want to join him or her to go to the movies. You tell him or her that you do not feel like it, and want to spend a quiet evening at home. That evening you do go out with a closer friend. (Movie)

Situation 5. During a discussion, you make a stinging remark toward one of your friends. You notice that it hurts him or her, but you pretend not to see it. (Discussion)

Situation 6. A friend tells you something in confidence, and adds that he or she would not like you to spread it around. Later, you do tell it to someone else. (Secret)

Situation 7. You are a member of a youth movement. One day the group leaders hang a rope between two trees, so you can glide from one tree to another. Jokingly, some other members make the stop of the pulley unclear. You see them doing it, but you do not help them. The following member, who wants to glide to the other tree, did not see that the stop was made unclear. You do not warn him or her. Halfway he falls from the rope, and he passes out. (Youth movement)

Situation 8. You have a pen pal. You get bored of writing with him or her, and suddenly, you stop corresponding with him or her. After 1½ year, he or she writes you again, and again, but you do not respond. (Pen pal)

Situation 9. You borrowed a jacket from a friend to wear when you go out. At the party, you leave the jacket on a chair. When you are about to leave, you notice the jacket has disappeared. In all probability, it has been stolen. (Jacket)

Situation 10. One evening, you do not feel like doing your homework. The following day, you copy the assignment of a friend who clearly has gone though a lot of trouble finishing it. You get a good grade for your assignment, the same grade as your friend. (Homework)

Items

1. Do you feel like having violated a moral, an ethic, a religious, and/or a personal code? (Norm Violation)
2. Do you worry about what you did or failed to do? (Worry)
3. Do you want to do something to rectify what you did or failed to do? (Tendency to rectify)
4. Do you feel guilty about what you did or failed to do? (Guilt)

Acknowledgments

The authors thank Dr. Smits and Dr. De Boeck for making the data available for this study. They also thank two anonymous reviewers for their insightful comments on the earlier version of this article.

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.

References

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics*, *17*, 251-269.
Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory*. Chichester, UK: Wiley.
Brennan, R. (2001). *Generalizability theory*. New York, NY: Springer.
Briggs, D. C., & Wilson, M. (2007). Generalizability in item response modeling. *Journal of Educational Measurement*, *44*, 131-155.
Butter, R. (1994). *Item response models with internal restriction on item difficulty* (Unpublished doctoral thesis). K.U. Leuven, Belgium.
Butter, R., De Boeck, P., & Verhelst, N. D. (1998). An item response model with internal restrictions on item difficulty. *Psychometrika*, *63*, 1-17.
Carlin, B. P., & Louis, T. A. (1996). *Bayes and empirical Bayes methods for data analysis*. London, England: Chapman and Hall.
De Boeck, P. (1991). *Componential IRT models*. Unpublished manuscript, University of Leuven, Belgium.
De Boeck, P. (2008). Random item IRT models. *Psychometrika*, *73*, 533-559.
De Boeck, P., & Smits, D. (2006). A double-structure structural equation model for the study of emotions and their components. In Q. Jing, M.R. Rosenzweig, G. d'Ydewalle, H. Zhang, H. Chen, & K. Zhang (Eds.), *Progress in psychological science around the world: Vol. 1. Neural, cognitive and developmental issues* (pp. 349–365). Hove, United Kingdom: Psychology Press.
Embretson, S. E. (1984). A general multicomponent latent trait model for response process. *Pychometrika*, *49*, 175-186.
Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, *37*, 359-374.
Fischer, G. H. (1983). Logistic latent trait models with linear constraints. *Psychometirka*, *48*, 3-26.
Gelfand, A. E., & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398-409.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London, England: Chapman and Hall.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457-511.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721-741.

Gonzalez, J., De Boeck, P., & Tuerlinckx, F. (2008). A double-structure structural equation model for three-mode data. *Psychological Methods*, *13*, 337-353.

Janssen, R., Schepers, J., & Peres, D. (2004). Models with item and item group predictors. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 189-212). New York, NY: Springer.

Kang, T., & Cohen, A. S. (2007). IRT model selection methods for dichotomous items. *Applied Psychological Measurement*, *31*, 331-358.

Kim, S. (2001). An evaluation of a Markov Chain Monte Carlo method for the Rasch model. *Applied Psychological Measurement*, *25*, 163-176.

Kroonenberg, P. (2005, April). Three-mode component and scaling methods. In B. Everitt & D. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 4, pp. 2032-2044). New York, NY: John Wiley.

Lee, Y., & Wilson, M. (2009). *An extension of the MIRID model for polytomous responses and random effects*. Paper presented at the annual meeting of American Educational Research Association, San Diego, CA.

Li, F., Cohen, A. S., Kim, S., & Cho, S. (2009). Model selection methods for mixture dichotomous IRT models. *Applied Psychological Measurement*, *33*, 353-373.

Mislevy, R. J. (1988). Exploiting auxiliary information about items in the estimation of Rasch item difficulty parameters. *Applied Psychological Measurement*, *12*, 725-737.

Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov Chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, *24*, 146-178.

Smits, D. J. M. (2003). *Item response model for self-report data on emotional responses* (Unpublished doctoral thesis). K.U. Leuven, Belgium.

Smits, D. J. M., & De Boeck, P. (2003). A componential IRT model for guilt. *Multivariate Behavioral Research*, *38*, 161-188.

Smits, D. J. M., & De Boeck, P. (2004). Latent item predictors with fixed effects. In P. De Boeck & M. Wilson (Eds.), *Explanatory item resposne models: A generalized linear and nonlinear approach* (pp. 267-287). New York, NY: Springer.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). WinBUGS (Version 1.4) [Computer program]. Cambridge, UK: MRC Biostatistics Unit, Institute of Public Health.

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, *12*, 1-16.

Tanner, M. A. (1996). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions* (2nd ed.). New York, NY: Springer.

Wang, W., & Jin, Y. (2009). Multilevel, two-parameter, and random-weights generalizations of the model with internal restrictions on item difficulty. *Applied Psychological Measurement*, *34*, 46-65.

Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, *45*, 479-494.