An Extension of the MIRID Model for Polytomous Responses

Yongsang Lee

(Associate Research Fellow, Korea Institute for Curriculum and Evaluation)

════════════════ ≪ ABSTRACT ≫ ════════════════

The purpose of this paper is to demonstrate how to extend the MIRID (Model with Internal Restrictions on Item Difficulty) model to polytomous situations. The original MIRID model is designed to explore the underlying relationship between the complex cognitive task and its basic components with binary response data. If polytomous response data is treated in a dichotomous way, however, then much information about examinees might be lost. This phenomenon is explained in both empirical and simulation studies. The interpretation of the models is illustrated in an example with data from a study of verbal aggression. In order to verify the accuracy and stability of the estimates for the Partial Credit MIRID (PC-MIRID), a simulation study was conducted under 16 conditions using SAS NLMIXED. The simulation study in this paper confirms that the PC-MIRID provides accurate and stable parameter estimates under appropriate conditions.

Key Words : the partial credit MIRID model, the partial credit model, the Rasch model

Ⅰ. Introduction

In an item explanatory model (De Boeck and Wilson, 2004) such as the linear logistic test model (LLTM; Fischer, 1973), we seek to explain the item parameters using properties of the items. For example, one of the item properties is the item format, and we can attempt to explain item difficulty

using item format in order to see how item format affects item difficulty because an item difficulty may differ depending on whether it is a multiple choice item or a open ended item. Within the LLTM framework, every item parameter reflects certain item properties to some degree (assumed known), so item properties are thus external variables. In cases where we believe that certain parameters of certain items are dependent on other item parameters, however, those (first) item properties would be predicted by other item parameters, and hence would be internal variables. An example of this is where a certain item asks the respondent to summarize his or her views, which have been asked about in previous items-it is reasonable to hypothesize that the "summary" item's difficulty will be in part, predictable from the parameters of the earlier items. At this point, the item predictors are the counterparts of the item properties in the LLTM. In an effort to capture these ideas in terms of a psychometric model, Butter, De Boeck and Verhelst (1998) introduced a "model with internal restrictions on item difficulty" (i.e., the MIRID). Subsequent work along this line has also concentrated on dichotomous responses (Bechger, Verhelst, and Verstralen, 2001; Smits and De Boeck, 2003; Smits, De Boeck and Verhelst, 2003; Maris and Bechger, 2004). Many measurement instruments in both cognitive and affective domains, however, commonly employ polytomous response item formats. One of the main reasons for this is that polytomous items can be more informative than binary responses and hence can form more reliable and useful tests than dichotomous items. If we reduce polytomous response data to dichotomous data, we might lose significant amounts of information about examinees. It is, therefore, worthwhile to extend the MIRID for dichotomous data to a model for polytomous data.

In psychometric research, measurement situations in which a complex cognitive task can be decomposed to basic components (or elements) are frequently discussed, and various item response theory (IRT) models have been developed and applied to this type of measurement context (Hoskens and De Boeck, 1995, 2001; Embretson, 1980, 1997, 2000; Fischer, 1973, 1977, 1983; Sternberg, 1985, 1997; Wainer and Kiely, 1987; Fischer and Formann, 1982; Whitely, 1980, Samejima, 1983). Since the IRT models that they use address not only a universal task but also the components of it, their models in general have been called componential IRT models. In the respect that the universal task is explained by its components, these kinds of models can be also referred as explanatory IRT models (De Boeck and Wilson, 2004).

In this paper, I propose the Partial Credit MIRID (PC-MIRID) and investigate the quality of estimation under various measurement situations with simulation studies. I begin by describing the original dichotomous MIRID model, and then show how it can be extended to a polytomous framework consistent with the Partial Credit Model (PCM)--that is the PC-MIRID. Coincidently, there

has been another group of people working on the PC-MIRID (Wang and Jin, 2009). In their paper, the simulation studies were conducted only with varying the test length and sample size, and the study focused on extending the MIRID model. In order to test the performance of the PC-MIRID, the current study, however, tested it with more comprehensive simulation conditions such as the correlation among components, the size of component weights and the number of categories.

II. Models

1. MIRID for Dichotomous data

The MIRID for dichotomous data was originally proposed by De Boeck (1991) and developed by Butter, De Boeck and Verhelst (1994, 1998) to explore the structure of a cognitive or affective construct (or task) in terms of its components. The emotion of guilt, for example, can be explained in terms of a number of components such as responsibility, norm violation, negative self-evaluation, worrying and restitution (Smits and De Boeck, 2003). The MIRID model is designed to handle situations in which one wants to know the underlying relationship between these components and guilt, but does not know the values of the components. Under the MIRID framework, guilt is treated as a composite concept while others (i.e., situations) are treated as component concepts. If the components' values are externally known, then the LLTM should be applied. If, however, we are considering situations where these values are unknown, the MIRID would be necessary.

To specify the MIRID model, two different types of items are required: component items and composite items. The composite item is an item that measures a concept that can be broken down into components. One composite item with a set of component items is called an item family. One may, for example, want to investigate quality of life. Quality of life can be explained by many different aspects (components) such as income, job satisfaction, and so on. In order to explore quality of life, one thus ask about the quality of life itself in his or her questionnaire, and then asks about each component such as income and job satisfaction. Since quality of life can differ depending on the place where people live, one can specify a place in each question. In this questionnaire, an item for the quality of life is a composite item, and items for income and job satisfaction are component items. The place which one specifies is a situation; and, consequently, each situation has one composite item and a set of component items. This composite item with a set of component items is called an item

family. [Figure 1] visualizes the terms in the MIRID model. In this Figure, there situations (Seoul, Hong Kong, and Tokyo) are specified, and thus there are three item families and each item family has one composite item with two component items.
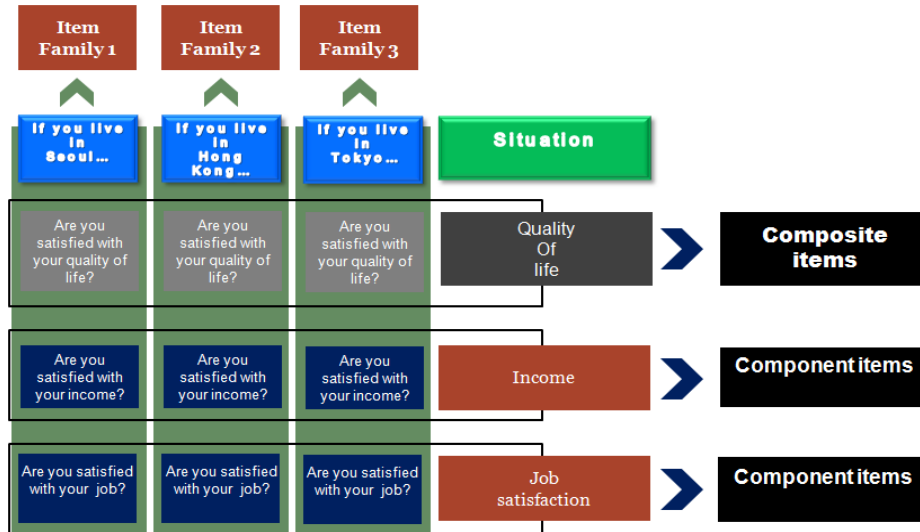
〔FIGURE 1〕 Measurement contexts and terms in the MIIRD model

In the MIRID model, we assume that the effect of the composite item is expressed as a linear function of the effect of the component items. Within this framework, an item matrix with a fixed effect part B can be formulated as the product of a latent item predictor matrix A and a componential weight matrix $\Psi^{'}$. [Figure 1] shows a latent predictor matrix A and [Figure 2] shows a componential weight matrix Ψ for one item family. [Figure 3] is the product of a latent item predictor matrix A and a componential weight matrix $\Psi^{'}$. $A\Psi^{'}_{sr}$ is the cell for a component item r in an item family s and $A\Psi^{'}_{s(r+1)}$ is the composite item in an item family s.

$$
\begin{array}{c}
\begin{array}{ccccc} \text{Component.1} & \cdots & \text{Component.r} & \text{Constant.} \end{array} \\
\begin{array}{c} \text{Item family 1} \\ \vdots \\ \text{Item family s} \end{array}
\begin{pmatrix} \beta_{11} & \cdots & \beta_{1r} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{s1} & \cdots & \beta_{sr} & 1 \end{pmatrix}
\end{array}
$$

〔FIGURE 2〕 Latent item predictor matrix A for dichotomous items.

$$
\begin{array}{c}
\begin{array}{cccc}
\text{Component. 1} & \text{...} & \text{Component. r} & \text{Intercept}
\end{array} \\
\begin{array}{l}
\text{Component item type. 1} \\
\vdots \\
\text{Component item type. r} \\
\text{Composite item}
\end{array}
\left(
\begin{array}{cccc}
1 & ... & 0 & 0 \\
\vdots & ... & \vdots & \vdots \\
0 & ... & 1 & 0 \\
\lambda_1 & ... & \lambda_r & \tau
\end{array}
\right)
\end{array}
$$

〔FIGURE 3〕 Componential weight matrix Ψ for one item family.

$$
\begin{array}{c}
\begin{array}{ccc}
\text{Comp. 1 ... Comp. r} & & \text{Composite item}
\end{array} \\
\begin{array}{l}
\text{Item family. 1} \\
\vdots \\
\text{Item family. S}
\end{array}
\left(
\begin{array}{cccc}
\beta_{11} & \cdots & \beta_{1r} & \sum_{r=1}^{R} \lambda_r \beta_{1r} + \tau \\
\vdots & \ddots & & \vdots \\
\beta_{s1} & \cdots & \beta_{sr} & \sum_{r=1}^{R} \lambda_r \beta_{sr} + \tau
\end{array}
\right)
\end{array}
$$

〔FIGURE 4〕 Item parameter matrix for dichotomous items

The MIRID model is given by the following equation:

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

for the component item $\beta_i = A\Psi'_{sr}$

for the composite item $\beta_i = A\Psi'_{s(r+1)} = \sum_{r=1}^{R} \lambda_r \beta_{sr} + \tau$

where "θ_p" indicates a person ability parameter for person p,

"β_i" indicates an item parameter for item i,

"Y_{pi}" indicates the response of person p, with ability θ_p to item i,

"r" indicates the component,

"s" indicates the item family,

"i" indicates the pairs (s, r), and

"p" indicates the subject.

Since the composite item parameter is restricted to a linear function of item parameters (β_{sr}), a weight for each component (λ_r) and an intercept (τ), the MIRID is in fact a restriction on the Rasch model as is the LLTM. Since item parameters are invariant as long as the ability scale is not changed, rescaling the linear function of the β_{sr} should lead to the same result as forming a linear function of the rescaled β_{sr}. As in the Rasch model, the translation of the item parameters in the MIRID is identified up to an additive constant, so this principle can be expressed as follows:

$$\sum_{r=1}^{R} \lambda_r \beta_{sr} + \tau + c = \sum_{r=1}^{R} \lambda_r (\beta_{sr} + c^{'}) + \tau^{*} \tag{2}$$

To satisfy this equation, τ^{*} should be the same as $\tau + c(1 - \sum_{r=1}^{R} \lambda_r)$. As the intercept is not invariant under the scale transformations, we can make parameters identifiable by choosing 'c' so that the sum of the component weights does not be equal to one ($\sum_{r=1}^{R} \lambda_r \neq 1$).

2. The Partial Credit Model

The Partial Credit Model (PCM) was originally developed by Masters (1982) as an extension of the Rasch model (Rasch, 1960). It is designed to analyze test items that have ordered polytomous categories of response (Embretson & Reise, 2000). This model allows flexible step difficulties which means that there are no assumptions about the relative difficulties of the steps within any item. In other words, the relative difficulties of the steps in each item might be different from item to item.

The Partial Credit Model is given by the following equation.

$$P(Y_{pi} = m | \theta_p) = \frac{\exp(\sum_{j=0}^{m} (\theta_p - \beta_{ij}))}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^{k} (\theta_p - \beta_{ij})} , \ \text{m} = 0, 1, \cdots, m_i \tag{3}$$

where　"Y_{pi}" represents the response of person p, with ability θ_p to item i,

　"β_{ij}" is an item level parameter for item i associated with score m,

$$\beta_{i0} = 0 \text{ , and } \sum_{j=0}^{0} (\theta_p - \beta_{ij}) = 0.$$

.

The numerator $\exp \sum_{j=0}^{m} (\theta_p - \beta_{ij})$ represents the difference between person ability and the item difficulty of the first m completed steps, and the denominator is the sum of all $m_i + 1$ possible numerators.

3. MIRID for Polytomous data: PC-MIRID

For illustrative purposes, let me take the case in which each item has three ordinal categories; thus, each item has two step parameters. Since SAS NLMIXED does not allow estimation of item and step parameters separately, let me use the term "level parameter (β_{ij}) to mean the sum of item parameter (δ_i) and step parameter (δ_{ij}). Therefore β_{ij} corresponds to the j-th level parameter of item i, and in the latent item predictor matrix of [Figure 5]., β_{sr1} and β_{sr2} indicate level one and level two parameters of the r-th component item in the s-th item family, respectively.

	Component.1 Level 1.	Component.1 Level 2.	...	Component. r Level 1.	Component. r Level 2.	Constant.
Item family 1	β_{111}	β_{112}	...	β_{1r1}	β_{1r2}	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Item family s	β_{s11}	β_{s12}	...	β_{sr1}	β_{sr2}	1

〔FIGURE 5〕 Latent item predictor matrix A for polytomous items.

As can be seen in [Figure 6], each row of the componential weight matrix specifies the level parameters of each component or composite item similar to the Rasch-MIRID.

	Component.1 Level 1.	Component.1 Level 2.	...	Component. r Level 1.	Component. r Level 2.	Intercept
comp. 1 level 1	1	0	...	0	0	0
comp. 1 level 2	0	1	...	0	0	0
⋮	⋮	⋮	...	⋮	⋮	⋮
comp. r level 1	0	0	...	1	0	0
comp. r level 2	0	0	...	0	1	0
composite level 1	λ_{11}	0	...	λ_{r1}	0	τ_1
composite level 2	0	λ_{21}	...	0	λ_{r2}	τ_2

〔FIGURE 6〕 Componential weight matrix Ψ for one item family.

To construct the formula for the PC-MIRID, one might need to assume a relationship between the level parameters of composite items and of component items. As can be seen in [Figure 7], the level parameter of the composite item is formulated as a linear function of the same level parameters of component items, weights, and intercept. The level-one parameter of the composite item in the first item family, for example, can be expressed as a linear combination of level-one parameters of component items (β_{1r1}) their weights (λ_{r1}) and an intercept (τ_1). According to research perspectives, however, different formulations might be preferred. One might assume, for instance, that level-one parameters of component items contribute to level-two parameters as well as level-one parameters of composite items. One might, furthermore, be interested in how item parameters (δ_i) and step parameters (δ_{ij}) of component items contribute to composite item parameters and step parameters, respectively. These issues need to be further investigated, but are beyond the scope of this paper.

	Component1 Level 1	Component.1 Level 2	...	Component. r Level 1	Component. r Level 2	Composite Level 1	Composite Level 2
Item family 1	β_{111}	β_{112}	...	β_{1r1}	β_{1r2}	$\sum_{r=1}^{R} \lambda_{r1}\beta_{1r1} + \tau_1$	$\sum_{r=1}^{R} \lambda_{r2}\beta_{1r2} + \tau_2$
⋮	⋮	⋮	...	⋮	⋮	⋮	⋮
item family s	β_{s11}	β_{s12}	...	β_{sr1}	β_{sr2}	$\sum_{r=1}^{R} \lambda_{r1}\beta_{sr1} + \tau_1$	$\sum_{r=1}^{R} \lambda_{r2}\beta_{sr2} + \tau_2$

〔FIGURE 7〕 Item parameter matrix for polytomous items

In the same way as the Rasch-MIRID, the linear component (linear predictor of the person's response) and the PC-MIRID can be expressed as follows:

$$P(Y_{pi} = m | \theta_p) = \frac{\exp \sum_{j=0}^{m} (\theta_p - \beta_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^{k} (\theta_p - \beta_{ij})} , \quad \text{m} = 0, 1, \cdots, m_i \quad (4)$$

for the component item $\beta_{ij} = A \Psi'_{sr(j)}$

for the composite item $\beta_{ij} = A \Psi'_{s(r+1)(j)} = \sum_{r=1}^{R} \lambda_{rj} \beta_{sr(j)} + \tau_j$

where "θ_p" indicates a person ability parameter for person p,

"β_{ij}" indicates an item parameter for item i associated with score m,

"Y_{pi}" indicates the response of person p, with ability θ_p to item i,

"r" indicates the component,

"s" indicates the item family,

"i" indicates the pairs (s, r), and

"p" indicates the subject.

Regarding the identifiability of the PC-MIRID, a similar condition to that of the Rasch-MIRID is needed to make the PC-MIRID identifiable. The PC-MIRID condition, however, is level parameter specific, while the Rasch-MIRID has an item-specific condition. The PC-MIRID condition, therefore, is that the sum of the weights for the j-th level parameters of the component items should not be equal to one (1) to identify the model. This condition can be expressed as follows:

$$\tau^*_{j} = \tau_j + c(1 - \sum_{r=1}^{R} \lambda_{rj}) , \text{ where } \sum_{r=1}^{R} \lambda_{rj} \neq 1 \quad (5)$$

Ⅲ. Estimation

For the parameter estimation of the Rasch-MIRID, a conditional maximum likelihood (CML) and a marginal maximum likelihood (MML) approach have been used (Smits and De Boeck, 2003; Smits, De Boeck, Verhelst, and Butter, 2001). Currently the MIRID CML and the PROC NLMIXED of SAS are available for each estimation method. The MIRID CML is a program for the estimation of the parameter values of the Rasch-MIRID that uses a conditional maximum likelihood (CML)

approach, whereas PROC NLMIXED is a procedure of SAS for non-linear mixed models that uses a marginal maximum likelihood (MML) approach. Smits and De Boeck (2003) found that these two programs provide somewhat different estimates for the person parameters while they estimate the item parameters equally well. They explained that this result might be from the different assumptions about the person distribution. Like the Rasch-MIRID, both a CML and a MML formulation can also be used for the PC-MIRID parameters estimation. The MIRID CML program, however, is only designed to deal only with binary data, thus the PROC NLMIXED was used to investigate the estimate of the PC-MIRID parameters as it can be adapted to the polytomous case. In PROC NLMIXED, an integration method and an optimization technique are specified with the Gauss-Hermite quadrature and the Quasi Newton-Raphson, respectively. In fact, the quasi Newton-Raphson optimization technique requires only the first derivatives to be calculated, so it is a time efficient method.

Ⅳ. Empirical study

1. Data: Verbal aggression

To illustrate the use of the new models, an example data set about verbal aggression (Vansteelandt, 2000) was used. This data includes 316 respondents and 24 items that were designed in terms of three factors: 'Behavior Mode', 'Situation type', and 'Behavior type' (see Appendix A). Each item has three categories scored as 0 (no), 1 (perhaps), and 2 (yes), which means each item has two step parameters.

In this application, 'Behavior Mode' was used to distinguish composite items from component items. In the data set on verbal aggression, a respondent's aggressive reaction to each situation is measured in two different ways which is called the 'Behavior Mode': (1) whether one wants to display the corresponding reaction in that situation (the action tendency), and (2) whether one actually would display the reaction (Smits, 2003; De Boeck and Wilson, 2004 see Appendix A). The first mode can be considered a component for the second mode, so twelve item families have been assigned, and each item family is composed of two items, which are thus components of the composite item; that is, the component item is designed to explore whether one wants to display the corresponding reaction in that situation (Action tendency), and the composite item is designed to

investigate whether one actually would display the reaction (Smits, 2003; De Boeck and Wilson, 2004). The descriptive statistics for each item can be seen in <Table 1>.

⟨TABLE 1⟩ Descriptive statistics on "verbal aggression" items.

Item type	Item	Mean	Standard deviation	Minimum Score	Maximum score
Component ('Want' Mode)	x1	1.123	0.828	0	2
	x2	0.930	0.852	0	2
	x3	0.712	0.778	0	2
	x4	1.222	0.774	0	2
	x5	0.959	0.840	0	2
	x6	0.734	0.816	0	2
	x7	0.810	0.766	0	2
	x8	0.462	0.654	0	2
	x9	0.282	0.534	0	2
	x10	0.978	0.774	0	2
	x11	0.589	0.744	0	2
	x12	0.424	0.684	0	2
Composite ('Do' Mode)	x13	1.082	0.808	0	2
	x14	0.832	0.817	0	2
	x15	0.468	0.710	0	2
	x16	1.003	0.834	0	2
	x17	0.684	0.781	0	2
	x18	0.326	0.616	0	2
	x19	0.576	0.693	0	2
	x20	0.294	0.557	0	2
	x21	0.104	0.345	0	2
	x22	0.883	0.786	0	2
	x23	0.566	0.725	0	2
	x24	0.225	0.513	0	2

2. Results

The models were compared in terms of item parameters and model fit; furthermore, by applying the PC-MIRID, whether one's actual display of aggressive reaction can be explained by one's desire to display that aggressive reaction was examined.

The correlation between item parameters estimated under the PCM and the PC-MIRID is very high (0.951 and 0.983) even though the composite item parameters are reconstructed based on parameter estimates and the weight of the component item. For the level-one parameters, the correlations between item parameters estimated under the PCM and the PC-MIRID is 0.983. For the level-two parameters, the correlations between item parameters estimated under the PCM and the PC-MIRID is 0.951.

〈TABLE 2〉 Correlation between item parameter estimates between the two models.

	PCM	PC−MIRID
Level-one		
PCM	1.000	
PC-MIRID	0.983	1.000
Level-two		
PCM	1.000	
PC-MIRID	0.951	1.000

Note that the correlations between the models for the level-two parameters are slightly lower than the correlations for the level-one parameters. This result can be explained by the number of people who responded to Category 3 in each item. For all items, the mean raw score is less than 1.5 which means that people selected 0 or 1 much more than they did 2. As a result, the standard errors of the level-two parameter estimates tend to be higher than for the level-one parameters. This higher standard error causes a discrepancy between the PCM and the PC-MIRID.

As can be seen in <Table 3>, the weight of action tendency is larger (1.315) than 1, which means that action tendency has a large effect on aggressive behavior (actual action). The negative constant values (τ_1, τ_2) indicate that people are inclined to actually do less than what they want to do. It appears also that the constant for the level-two parameters (τ_2) is smaller than the constant for the level-one parameters (τ_1) in the PC-MIRID. Since the constant can be interpreted as the person

threshold (Smits, 2003), these results show that (at least for this data) people tend to support aggressive behavior less than action tendency.

〈TABLE 3〉 Estimates and standard error
for the component weight and variance of the person parameter(PC-MIRID).

Parameter	Value	(Standard Error)
λ_1(weight of Action tendency)	1.315	(0.077)
τ_1(constant for level-one parameter)	-0.459	(0.075)
τ_2(constant for level-two parameter)	-0.679	(0.110)
σ_θ^2	0.8923	(0.082)

Ⅴ. Simulation study

1. Simulation models

Simulation studies were conducted to verify the accuracy and stability of the estimates for the PC-MIRID. For this purpose, The data under the PC-MIRID model were simulated with various conditions. The simulated data included two component items for each item family. Since the PC-MIRID is designed to deal with polytomous data, one can assume that each component has multiple weights (λ_{rj}) and each level parameter of the composite items has its own intercept (τ_j) depending on one's research hypothesis. In order to choose a simulation model from among these alternatives, these alternative models were applied to the verbal aggression data and then the results were compared. Since these alternative models provided identical results, for convenience, only one intercept was specified for the composite items and one weight for each component in the simulation models.

2. Simulation conditions

For the simulation conditions, the following design factors were varied: (a) number of subjects, (b) magnitude of weights, (c) degree of correlation between subtask difficulties and (d) number of categories of response. The conditions of the simulation study are summarized in <Table 4>. There

were 16 conditions (24), based on the trying out of two levels for each of these design factors. One hundred (100) data sets with five (5) item families from the PC-MIRID model for each of the 16 conditions (24: a, b, c, d) were simulated, and then analyzed using the PC-MIRID.

〈TABLE 4〉 The conditions of the simulation study.

# of subjects	Magnitude of weights	Correlation between component difficulties	# of categories
m=500 or 1000	λ_1=1.0, λ_2=1.0 λ_1=0.2, λ_2=0.8 Both with τ=0.5	0 or 0.7	3 or 5

3. Results

The parameter recovery was examined under simulation conditions with the PC-MIRID. For this purpose, the Pearson correlation coefficient between the true values of parameters and their estimates ($\gamma_{p\hat{p}}$) and the mean squared error (MSE) over the estimates were calculated as measures of estimator stability and accuracy. The average correlation was computed between parameters (λ_{rj}, τ, and β_{ij}) and their estimates over 100 replications. The MSE was obtained based on the discrepancy between the true and estimated parameter for each parameter (λ_{rj}, τ, and β_{ij}) as shown in the following:

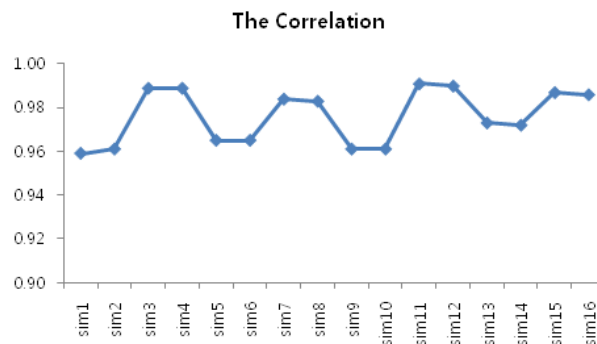$$MSE = \frac{1}{100} \sum_{i=1}^{100} (\hat{p_i} - p_i)^2$$

where p_i is the true parameter, and

$\hat{p_i}$ is an estimate of the true parameter.
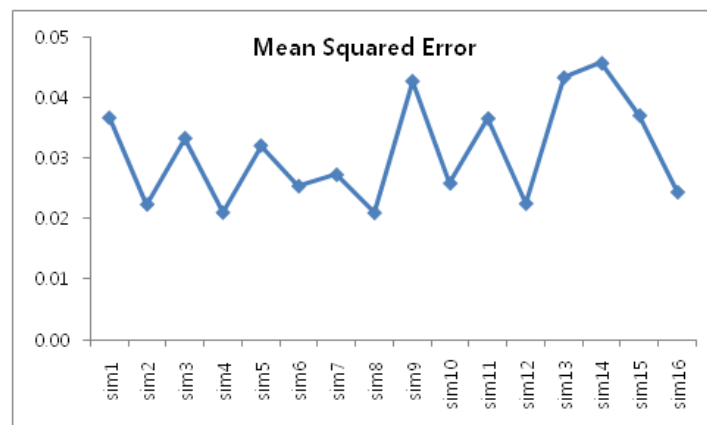
1) PC-MIRID

<Table 5> summarizes the simulation results with 16 combinations of conditions under the PC-MIRID model. This table shows the simulation conditions, the average correlation between parameters and their estimates ($\gamma_{p\hat{p}}$), and the MSE values. As can be seen in this Table, correlations between the parameters and their estimates are consistently greater than 0.95 across all simulation conditions even though composite item parameters were reconstructed using $\widehat{\lambda_{rj}}$ $\hat{\tau}$ and $\widehat{\beta_{ij}}$. A closer

look at these correlations, however, reveals an interesting trend. Equal weights for the components lead to a higher correlation than unequal weights. Regardless of the combinations with other conditions, this phenomenon is consistently observed in [Figure 10]. As can be seen in this figure, the simulations with unequal weights (sim3, sim4, sim7, sim8, sim11, sim12, sim15, and sim16) provide a higher correlation than others.

〔FIGURE 10〕 The average correlation between estimated and true parameters.

With regard to the MSE, the simulation results suggest that a bigger sample size and a smaller number of categories result in a smaller MSE value. As we expected, the sample size had a profound effect on the MSE value regardless of other conditions; furthermore, the results from the three-category data consistently showed smaller MSE values than those from the five-category data. It appears, however, that the correlation between component item parameters does not have any substantial effect on the MSE value. In order to visualize these trends, [Figure 11] is provided. By referring to both <Table 5> and [Figure 11], it can be seen that the main sources of MSE difference are sample size and number of categories.

〔FIGURE 11〕 The mean of mean squared error statistic for all simulation conditions.
    After looking at the correlations and the MSE values, several factors which affect the accuracy and stability of the estimator: component weights, sample size and number of categories were identified. The simulation results, however, indicate that the effects of these conditions are somewhat limited in the PC-MIRID model. Although a difference in simulation results can be observed as the conditions are changed, this difference might not be a serious problem. They also show that estimates for component items are not biased while estimates for composite items are somewhat. As we now know, however, the composite item parameters in the PC-MIRID are not estimated directly from the model but are reconstructed using the component item parameter ($\widehat{\beta_{ij}}$), its weight ($\widehat{\lambda_{rj}}$), and the intercept ($\hat{\tau}$) without allowing any residuals. Thus, this bias of composite item parameters was expected.

〈TABLE 5〉 The simulation results.

Simulation NO	# of categories	Correlation b/w components	# of subjects	λ_1/λ_2	Average $\gamma_{p\hat{p}}$	MSE		
						Mean	Min	Max
sim1	5	0	500	0.2/0.8	0.959	0.037	0.014	0.127
sim2	5	0	1000	0.2/0.8	0.961	0.022	0.007	0.109
sim3	5	0	500	1/1	0.989	0.033	0.011	0.080
sim4	5	0	1000	1/1	0.989	0.021	0.009	0.089
sim5	3	0	500	0.2/0.8	0.965	0.032	0.012	0.169
sim6	3	0	1000	0.2/0.8	0.965	0.025	0.006	0.173
sim7	3	0	500	1/1	0.984	0.027	0.013	0.122
sim8	3	0	1000	1/1	0.983	0.021	0.006	0.122
sim9	5	0.7	500	0.2/0.8	0.961	0.043	0.015	0.310
sim10	5	0.7	1000	0.2/0.8	0.961	0.026	0.007	0.216
sim11	5	0.7	500	1/1	0.991	0.037	0.014	0.189
sim12	5	0.7	1000	1/1	0.990	0.022	0.007	0.100
sim13	3	0.7	500	0.2/0.8	0.973	0.043	0.010	0.323
sim14	3	0.7	1000	0.2/0.8	0.972	0.046	0.005	0.530
sim15	3	0.7	500	1/1	0.987	0.037	0.010	0.193
sim16	3	0.7	1000	1/1	0.986	0.024	0.006	0.122

Ⅵ. Summary and Discussion

In this paper, how to extend the MIRID model to polytomous situations is demonstrated. The interpretation of the models is illustrated in an empirical context, and simulation studies to test the model under various conditions are conducted. The simulation results indicate that the PC-MIRID provides accurate and stable parameter estimates regardless of given conditions. Since the proposed model is a componential model, one of the contributions of this study is to provide insight into the underlying relationships between components and a composite task or concept using polytomous data.

The MIRID model has been heavily studied and applied to investigate the complex cognitive process (Bechger, Verhelst, and Verstralen, 2001; Smits and De Bock, 2003). For the componential approach, the Multicomponent Latent Trait Model (MLTM; Whitely, 1980) and, as a generalization of MLTM, the General Component Latent Trait Model (GLTM; Embretson, 1984) have been developed. These models are designed to measure specified cognitive components (van der Linden and Hambleton, 1997). Like MIRID models, GLTM addresses componential weights, but unlike MIRID models, they are expressed using item response probability and the product of the probabilities of individual component tasks.

Unlike in Rasch-MIRID models, one might be interested in the relationship among level parameters of the PC-MIRID models. One might assume, for example, that the level-one parameter of component items would contribute to both level-one and level-two parameters of composite items. In fact, depending upon one's formulations, one might have a different understanding of the relationship of subtasks (components) and the universal task at these different levels. The investigation of this issue allows one to be able to gain more detailed information about this underlying relationship; hence, this issue can be another research question for future study.

References

Bechger, T. M., Verhelst, N. D., & Verstralen, H. H. F. M. (2001). Identibility of nonlinear logistic test models. *Psychometrika, 66*, 357-372.

Butter, R. (1994). Item response models with internal restriction on item difficulty. Unpublished doctoral dissertation, 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*, 47-63.

De Boeck, P. (1991). Componential IRT models. Unpublished manuscript, University of Leuven, Belgium.

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. NY: Springer.

Embretson, S. E. (1984). A general multicomponent latent trait model for response process. *Pychometrika, 49*, 175-186.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. NJ : Lawrence Erlbaum Associates.

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.

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.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Maris, G., & Bechger, T. M. (2004). Equivalent MIRID models. *Psychometrika, 69*, 627-639.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.

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

Smits, D. J. M., De Boeck, P., & Verhelst, N. D. (2003). Estimation of the MIRID: A program and a SAS based approach. Behavior Research Methods. *Instruments, and Computers, 35,* 537-549.

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

Vansteelandt, K. (2000). Formal models for contextualized personality psychology. Unpublished doctoral dissertation, K. U. Leuven, Belgium.

Van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory.* NY: Springer.

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

Wang, W., & Jin. K. (2009). A generalized model with internal restrictions on item difficulty for polytomous items. *Educational and Psychological Measurement, 70*, 181-198.

ABSTRACT

The purpose of this paper is to demonstrate how to extend the MIRID (Model with Internal Restrictions on Item Difficulty) model to polytomous situations. The original MIRID model is designed to explore the underlying relationship between the complex cognitive task and its basic components with binary response data. If polytomous response data is treated in a dichotomous way, however, then much information about examinees might be lost. This phenomenon is explained in both empirical and simulation studies. The interpretation of the models is illustrated in an example with data from a study of verbal aggression. In order to verify the accuracy and stability of the estimates for the Partial Credit MIRID (PC-MIRID), a simulation study was conducted under 16 conditions using SAS NLMIXED. The simulation study in this paper confirms that the PC-MIRID provides accurate and stable parameter estimates under appropriate conditions.

Key Words : the partial credit MIRID model, the partial credit model, the Rasch model

요 약

이 용 상(한국교육과정평가원 부연구위원)

MIRID 모형은 인지적 구인과 그 구인을 이루고 있는 요인들 사이의 관계를 문항 모수 측면에서 탐색할 수 있도록 설계된 모형으로서, 본 연구의 목적은 이분 문항 반응 자료 분석만이 가능하였던 MIRID 모형(the Rasch MIRID)을 다분 문항 반응 자료 분석이 가능한 부분 점수 MIRID 모형(Partial Credit MIRID)으로 확장하는 것이다. 본 연구에서는 우선 경험적 자료 분석을 통해 결과 해석 및 MIRID 모형의 활용 방안에 대해서 논의 하였다. 경험적 자료의 분석을 위해 부분점수 모형(partial credit model)과 부분점수 MIRID 모형이 이용되었으며, 두 모형을 적용하여 분석 한 결과, 두 모형간 문항 모수 추정치가 매우 유사하게 나타나는 것으로 확인되었다. 부분 점수 모형과 비교하여 부분 점수 MIRID 모형은 문항 모수 간의 관계에 대한 정보까지도 제공하므로 모형의 활용도 측면에서 부분 점수 모형보다 유용한 모형으로 판단되어졌다. 다음으로 부분 점수 MIRID 모형의 모수 추정 양호도를 검증하기 위해 16가지 조건하에서 모의실험을 실시하였다. 모의 실험 결과 부분 점수 MIRID 모형의 모수 추정치가 양호한 것으로 확인되었다.

주제어 : 부분 점수 MIRID 모형, 라쉬 모형, 부분 점수 모형

Appendix A. Verbal aggression Items

Items	Situation type	Behavior type
Behavior Mode : Want		
A bus fails to stop for me. I would want to curse.	Other-to-blame	*Curse*
A bus fails to stop for me. I would want to scold.		*Scold*
A bus fails to stop for me. I would want to shout.		*Shout*
I miss a train because a clerk gave me faulty information. I would want to curse.		*Curse*
I miss a train because a clerk gave me faulty information. I would want to scold.		*Scold*
I miss a train because a clerk gave me faulty information. I would want to shout.		*Shout*
The grocery store closes just as I am about to enter. I would want to curse.	Self-to-blame	*Curse*
The grocery store closes just as I am about to enter. I would want to scold.		*Scold*
The grocery store closes just as I am about to enter. I would want to shout.		*Shout*
The operator disconnects me when I had used up my last 10 cents for a call. I would want to curse.		*Curse*
The operator disconnects me when I had used up my last 10 cents for a call. I would want to scold.		*Scold*
The operator disconnects me when I had used up my last 10 cents for a call. I would want to shout.		*Shout*
Behavior Mode : Do		
A bus fails to stop for me. I would curse.	Other-to-blame	*Curse*
A bus fails to stop for me. I would scold.		*Scold*
A bus fails to stop for me. I would shout.		*Shout*
I miss a train because a clerk gave me faulty information. I would curse.		*Curse*
I miss a train because a clerk gave me faulty information. I would scold.		*Scold*
I miss a train because a clerk gave me faulty information. I would shout.		*Shout*
The grocery store closes just as I am about to enter. I would curse.	Self-to-blame	*Curse*
The grocery store closes just as I am about to enter. I would scold.		*Scold*
The grocery store closes just as I am about to enter. I would shout.		*Shout*
The operator disconnects me when I had used up my last 10 cents for a call. I would curse.		*Curse*
The operator disconnects me when I had used up my last 10 cents for a call. I would scold.		*Scold*
The operator disconnects me when I had used up my last 10 cents for a call. I would shout.		*Shout*

Appendix B. SAS code for the PC-MIRID(Verbal aggression)

```
Proc NlMixed data=MIRID.VA method=gauss technique=quanew noad
Qpoints=15 update=dfp maxfunc=20000 maxiter=20000;
Parms a1_1-a1_12=1 a2_1-a2_12=1 b1=0   b0=0 b00=0 sd=0.5;

alpha1=a1_1*X1;
alpha2=a1_2*X2;
alpha3=a1_3*X3;
alpha4=a1_4*X4;
alpha5=a1_5*X5;
alpha6=a1_6*X6;
alpha7=a1_7*X7;
alpha8=a1_8*X8;
alpha9=a1_9*X9;
alpha10=a1_10*X10;
alpha11=a1_11*X11;
alpha12=a1_12*X12;

alpha13=a2_1*X1;
alpha14=a2_2*X2;
alpha15=a2_3*X3;
alpha16=a2_4*X4;
alpha17=a2_5*X5;
alpha18=a2_6*X6;
alpha19=a2_7*X7;
alpha20=a2_8*X8;
alpha21=a2_9*X9;
alpha22=a2_10*X10;
alpha23=a2_11*X11;
alpha24=a2_12*X12;

beta1=b1*a1_1*X1;
```

```
beta2=b1*a1_2*X2;
beta3=b1*a1_3*X3;
beta4=b1*a1_4*X4;
beta5=b1*a1_5*X5;
beta6=b1*a1_6*X6;
beta7=b1*a1_7*X7;
beta8=b1*a1_8*X8;
beta9=b1*a1_9*X9;
beta10=b1*a1_10*X10;
beta11=b1*a1_11*X11;
beta12=b1*a1_12*X12;

beta13=b1*a2_1*X1;
beta14=b1*a2_2*X2;
beta15=b1*a2_3*X3;
beta16=b1*a2_4*X4;
beta17=b1*a2_5*X5;
beta18=b1*a2_6*X6;
beta19=b1*a2_7*X7;
beta20=b1*a2_8*X8;
beta21=b1*a2_9*X9;
beta22=b1*a2_10*X10;
beta23=b1*a2_11*X11;
beta24=b1*a2_12*X12;

exp1=exp(-theta +(1-Co)*(alpha1 + alpha2+ alpha3
+ alpha4 + alpha5 + alpha6 + alpha7 + alpha8 + alpha9 + alpha10 + alpha11
+ alpha12)
+Co*(beta1+beta2+beta3+beta4+beta5+beta6+beta7+beta8+beta9+beta10+beta
11+beta12+b0));

exp2=exp(-2*theta
+(1-Co)*(alpha1+alpha2+alpha3+alpha4+alpha5+alpha6+alpha7+alpha8+alpha9+
alpha10+alpha11+alpha12
+alpha13+alpha14+alpha15+alpha16+alpha17+alpha18+alpha19+alpha20+alpha21
```

```
+alpha22+alpha23+alpha24)
+Co*(beta1+beta2+beta3+beta4+beta5+beta6+beta7+beta8+beta9+beta10+beta
11+beta12+
beta13+beta14+beta15+beta16+beta17+beta18+beta19+beta20+beta21+beta22+
beta23+beta24+b00));

denom=1+exp1+exp2;
if (y=0) then p=1/denom;
else if (y=1) then p=exp1/denom;
else if (y=2) then p=exp2/denom;
if (p>le-8) then ll=log(p);
else ll=-le100;
model y~general(ll);
Random Theta~Normal(0, sd**2) Subject=person;
ESTIMATE 'sd**2' sd**2;
run;
```