Behaviormetrika, Vol. 20, No. 2, 161-169
1993

A NEW ITEM RESPONSE MODEL WITH PARAMETERS REFLECTING STATE OF KNOWLEDGE

Kazuo Shigemasu* and Maomi Ueno*

The authors developed a new latent trait model for polytomous data. A unique feature of this model is that it explains the psychological process through which a subject reaches correct or wrong responses. This model is an integration of two parts. One part discriminates states of knowledge, namely, it classifies them into three categories, *i.e.*, a) complete knowledge, b) partial knowledge, and c) complete ignorance. Another feature of the model is that we have two different kinds of manifest data, *i.e.* confidence data and correct/wrong responses. The subject is required to choose whether he is confident of his answer or not. In this paper, we make use of confidence data as auxiliary information in making effective estimation of parameters. Then the new model was applied to mathematics test data of high school students.

1. Problem

The purpose of this paper is to develop a new item response model which explains the psychological process through which a subject reaches correct or wrong responses.

As is often pointed out, the high speed of recent computers has made it possible for psychometricians to deal with more complex statistical models. This increase of freedom of the model so that it could be used to explain cognitive aspects of item responses. If the model is appropriate and the parameters are not so lengthy as to decrease stability of parameter estimation, the model should be effective in reporting the characteristics of both test items and subjects. That is, we want to make the number of parameters as small as possible, yet large enough to fit the model to psychological reality.

Recently there have been quite a number of psychometric articles which make use of complex statistical models to reflect cognitive process involved in problem solving. Interested readers should refer to chapter 8 of "Educational Measurement, Third Edition" (Linn, 1989) and "Test Theory for a New Generation" (Frederiksen, Mislevy, and Bejar, 1993). These two books provide good information for this kind of psychometric literature. We do not intend to survey this area, but we reivew several reserch efforts which have similar objectives as this paper. Tatsuoka (1985) has studied performance on mathematics items in terms of the application of correct and incorrect rules, locating response vectors in a two dimensional space, where the first dimension is an index of lack of fit from the model. Fischer

Key Words and Phrases ; latent trait model, item response theory, cognitive psychology, latent class model, Bayes theory, confidence data
* Faculty of Engineering, Tokyo Kogyo Daigaku, 2-12-1, O-okayama, Meguro-ku, Tokyo, 152 Japan

(1973) proposed the linear logistic latent trait model (LLTM) which replaces the item difficulty parameters by the function of the complexity factors. Whitely (1980) developed the MultiComponent Latent Trait Model (MLTM) which models all cognitive components required for reaching correct answer. Embretson (1984) presented a more general model (which is called GLTM) that includes both features of the LLTM and MLTM. Embretson (1985) offers a model for alternative strategies in situations where subtask results can be observed in addition to the overall correctness or incorrectness of an item. Mislevy (1987) proposed a family of multiple-strategy IRT models that apply when each subject belongs to one of a number of exhaustive and mutually-exclusive classes that correspond to item-solving strategies, and within each class, a standard IRT model applies. Yamamoto (1993) offered HYBRID model which combines IRT and Latent Class Model to express different classes of the subjects' competency.

In this paper, the authors propose a different item response model depending on different states of knowledge. A unique feature of this model is that this model explains a psychological process through which a subject reaches a certain response. Another feature of the model is that we have two different kinds of manifest data, which are confidence data as well as correct/wrong responses. That is, the subject is required to choose whether he is confident in his answer or not. Evel (1968) proposed a method to eliminate the effect of guessing by using confidence data. In this paper, we make use of confidence data as auxiliary information in making effective estimation of parameters.

The following section gives a general description of the model.

2. Model

The model first distinguishes the subject's states of knowledge into three categories.

A subject may have perfect knowledge to solve a particular item. Or, he may only have partial knowledge, in which case, his knowledge may be ambiguous, and he may need to depend on some arbitrary heuristic rule to reach the answer. Or, he may not know at all how to solve the problem. These three categories are referred to as a) complete knowledge, b) partial knowledge, and c) complete ignorance. Next, the proposed model describes the probability of his being correct in his answer for each of three different states of knowledge.

If the subject has complete knowledge or knows the way to infer logically the answer from his knowledge, he could reach the correct answer without hesitation. When he has partial knowledge, whether his answer to the paticular item is correct or not is an uncertain event, which we can predict only statistically using a certain model as a function of the subject's true competency. In this paper, we employ a traditional two-parameter logistc model for this purpose.

When the subject is completely ignorant, the probability of his constructing a correct answer is zero, or in the case of multiple-choice test, this probability may be one over the number of choice categories of the item.

Now, let us express our model mathematically. We use $P(\ )$ as probability measure and $p(\ )$ as probability density function or probability mass function, and the other notations are listed below ;

i : the i-th subject,

j : the j-th item,

ξ_{ij} : latent dummy variable indicating the state of knowledge of the i-th subject on the j-th item, *i.e.*,

$$\xi_{ij}=\begin{cases} 1 & \text{complete knowledge} \\ 2 & \text{partial knowledge} \\ 3 & \text{complete ignorance,} \end{cases}$$

x_{ij} : a manifest dummy variable indicating whether the subject gets the correct answer ($x_{ij}=1$) or not ($x_{ij}=0$),

y_{ij} : a manifest variable indicating whether the i-th subject is confident of his answer for the j-th item ($y_{ij}=1$), or not ($y_{ij}=0$),

θ_i : a latent continuous variable of true competency of the i-th subject.

Thus, the probability of correct response is expressed as a function of θ_i by a law of total probability as follows ;

$$P(x_{ij}=1 \mid \theta_i)=\sum_{k=1}^{3} P(x_{ij}=1 \mid \xi_{ij}=k,\ \theta_i)P(\xi_{ij}=k \mid \theta_i). \tag{1}$$

As was discussed above, we assume

$$P(x_{ij}=1 \mid \xi_{ij}=1)=1, \tag{2}$$

and

$$P(x_{ij}=1 \mid \xi_{ij}=3)=0. \tag{3}$$

For the state of partial knowledge, as was mentioned above, we assume the usual two-parameter logistic model to express the relationship between the probability of correct response and the competency. That is,

$$P(x_{ij}=1 \mid \xi_{ij}=2)=\frac{1}{1+\exp(-1.7a_j(\theta_i-b_j))}. \tag{4}$$

For multiple-choice tiems, the equation (3) could be modified as follows ;

$$P(x_{ij}=1 \mid \xi_{ij}=3)=1/m, \tag{5}$$

where m is the number of alternatives,
and

$$P(x_{ij}=1 \mid \xi_{ij}=2)=c_j+\frac{1-c_j}{1+\exp(-1.7a_j(\theta_i-b_j))}, \tag{6}$$

where c_j is a guessing parameter. The state of knowledge relevant to the paticular

j-th item may well be dependent on θ_i, and we assume it can be represented by the Masters' model (Masters, 1982, 1985);

$$P(\xi_{ij}=1 \mid \theta_i)=(\exp((\theta_i-d_{j1})+(\theta_i-d_{j2})))/\psi_{ij}, \quad (7)$$
$$P(\xi_{ij}=2 \mid \theta_i)=(\exp((\theta_i-d_{j1}))/\psi_{ij}, \quad (8)$$
$$P(\xi_{ij}=3 \mid \theta_i)=1/\psi_{ij}, \quad (9)$$

where, $\psi_{ij}=1+\exp(\theta_i-d_{j1})+\exp((\theta_i-d_{j1})+(\theta_i-d_{j2}))$.

The auxiliary information y_{ij} can be used to identify the latent class ξ_{ij} through the use of the Bayes Theorem. That is, after we obtain the data y_{ij}, the probability of each latent class show a change as follows;

$$P(\xi_{ij}=k \mid y_{ij}, \theta_i) \propto P(y_{ij} \mid \xi_{ij}=k, \theta_i)P(\xi_{ij}=k \mid \theta_i). \quad (10)$$

We assume that

$$P(y_{ij}=1 \mid \xi_{ij}=1, \theta_i)=1(\text{for } {}^{\forall}\theta_i), \quad (11)$$
$$P(y_{ij}=1 \mid \xi_{ij}=3, \theta_i)=0(\text{for } {}^{\forall}\theta_i). \quad (12)$$

For the case of $\xi_{ij}=2$, so that the number of parameters should not increase, we assume the probability of the i-th subject's responding "confident" can be expressed as

$$P(y_{ij}=1 \mid \xi_{ij}=2, \theta_i)=\frac{P(\xi_{ij}=1 \mid \theta_i)}{P(\xi_{ij}=1 \mid \theta_i)+P(\xi_{ij}=3 \mid \theta_i)}. \quad (13)$$

When we observe $y_{ij}=1$, then the possibility of $\xi_{ij}=3$ vanishes, and the possibility of $\xi_{ij}=2$ remains only partially depending on (13).

Finally, the data generating model of x_{ij}'s with y_{ij}'s and all the parameters given is as follows;

$$P(x_{ij} \mid y_{ij}, a_j, b_j, \theta_i, \xi_{ij})=\sum_{k=1}^{3} P(x_{ij} \mid \xi_{ij}=k, \theta_i)P(\xi_{ij}=k \mid y_{ij}, \theta_i). \quad (14)$$

The auxiliary information y_{ij}'s are used here to identity the latent class for each subject.

3. Estimation

When the data x_{ij}'s and y_{ij}'s are given, the likelihood for the parmameters is given as

$$Q(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{y})=\prod_i \prod_j \pi_{ij}^{x_{ij}}(1-\pi_{ij})^{1-x_{ij}}, \quad (15)$$

where

$$\boldsymbol{a}=(a_1, a_2, \cdots, a_p)^t,$$
$$\boldsymbol{b}=(b_1, b_2, \cdots, b_p)^t,$$
$$\boldsymbol{d}_1=(d_{11}, d_{12}, \cdots, d_{1p})^t,$$
$$\boldsymbol{d}_2=(d_{21}, d_{22}, \cdots, d_{2p})^t,$$
$$\boldsymbol{\theta}=(\theta_1, \theta_2, \cdots, \theta_n)^t,$$

and

$$\pi_{ij}=P(x_{ij}\mid y_{ij},\ \theta_i)$$

We obtain the maximum likelihood estimate by numerically maximizing the log likelihood. The first derivatives of the log likelihood and the Hessian matrix are given in the Appendix 1. The numerical optimization algorithm is as follows;

Step 1: The initial estimates for $\boldsymbol{\theta}$, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ are given by the heuristic method as explained below.
Step 2: With $\boldsymbol{\theta}$, $\boldsymbol{a}$, and $\boldsymbol{b}$, optimize $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$.
Step 3: With $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ given, optimize $\boldsymbol{a}$ and $\boldsymbol{b}$.
Step 4: With $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ given, optimize θ_i.
Step 5: Repeat Step (2) through Step (4) till the estimates converge.

The initial estimates in step (1) are given as follows;

$$\hat{a}_j=\frac{\rho_j}{(1-\rho_j^2)^{1/2}}, \tag{16}$$

$$\hat{b}_j=\frac{1}{\alpha_j}\log\frac{1-\alpha_j}{\alpha_j}, \tag{17}$$

$$\hat{d}_{1j}=\log\frac{1-\alpha_{1j}}{\alpha_{1j}}, \tag{18}$$

$$\hat{d}_{2j}=\log\frac{1-\alpha_{2j}}{\alpha_{2j}}, \tag{19}$$

$$\hat{\theta}_i=\log\frac{\alpha_i}{(1-\alpha_i)}, \tag{20}$$

where, ρ_{1j}, indicates the four-fold point correlation coefficient between the test score $\sum_{j+1}^{p}x_{ij}$ and the learner's response x_{ij}. And $\alpha_j=\sum_{i=1}^{n}x_{ij}/n$, $\alpha_{1j}=\sum_{i=1}^{n}z_{1ij}/n$, $\alpha_{2ij}=\sum_{i=1}^{n}z_{2ij}z_{2ij}/n$, and $\alpha_i=\sum_{j=1}^{p}x_{ij}/p$. Here, $z_{1ij}=1$ if both $x_{ij}=1$ and $y_{ij}=1$ and otherwise $Z_{1ij}=0$. Also, $z_{2ij}=1$ if $x_{ij}=1$ or $y_{ij}=1$, and otherwise $z_{2ij}=0$.

The criterion for convergence is to stop the iteration steps when the maximum difference between the current estimate and that of the preceeding iteration is less than or equal to .001.

As is well known in mathematical statistics, the asymptotical variance and covariance matrix for maximum likelihood-estimation is given by the Fisher's information matrix. The Fisher's information matrix can be approximated by the Hessian matrix evaluated at the converged values.

4. Application

To demonstrate the feasibility and applicability of the model and estimation procedure, we applied the model to the following real data. We administered a 12 mathematics test items to 114 high school students. The items are given in Appendix 2.

Table 1
Estimated Parameters

Item	π_x	π_y	d_1	d_2	a	b	$\overline{P_1}$	$\overline{P_2}$	$\overline{P_3}$
1	.801	.448	−1.477 (0.010)	0.786 (0.008)	0.624 (0.049)	−0.436 (0.029)	0.316	0.437	0.136
2	.267	.155	−0.491 (0.012)	1.092 (0.004)	0.351 (0.032)	0.028 (0.036)	0.146	0.378	0.481
2	.819	.655	−1.457 (0.025)	0.666 (0.010)	0.699 (0.031)	−0.321 (0.014)	0.577	0.241	0.246
4	.646	.551	−0.656 (0.004)	0.743 (0.004)	0.869 (0.099)	−0.324 (0.016)	0.274	0.577	0.246
5	.897	.793	−0.497 (0.002)	−0.298 (0.014)	0.726 (0.026)	−0.421 (0.018)	0.611	0.126	0.244
6	.637	.620	−1.362 (0.016)	0.702 (0.021)	0.862 (0.026)	0.024 (0.022)	0.571	0.282	0.211
7	.664	.552	−1.329 (0.027)	0.766 (0.006)	0.829 (0.014)	−0.333 (0.015)	0.368	0.378	0.253
8	.828	.707	−1.476 (0.011)	0.223 (0.009)	0.654 (0.024)	−0.224 (0.004)	0.516	0.263	0.228
9	.741	.621	−1.328 (0.014)	0.611 (0.003)	0.738 (0.013)	−0.063 (0.014)	0.481	0.381	0.245
10	.362	.508	−1.268 (0.004)	0.776 (0.004)	0.826 (0.041)	0.211 (0.011)	0.298	0.524	0.311
11	.241	.293	−0.529 (0.001)	0.982 (0.008)	0.786 (0.031)	0.291 (0.039)	0.113	0.363	0.575
12	.371	.293	−.0741 (0.003)	0.926 (0.004)	0.729 (0.061)	0.424 (0.012)	0.241	0.342	0.358

Fig. 1 contrasts the usual item characteristic curve (ICC) and $P(x_{ij}=1 \mid \xi_{ij}=2)$ (*i.e.*, item characteristic curve for the latent class of partial knowledge). The ICC for the latent class is steeper than that for a whole population. The difference tends to be larger when the size of the second latent class is smaller. Fig. 1 shows $P(\xi_{ij}=k \mid \theta)$, *i.e.*, the probability of each latent class conditional on a particular θ. Table 1 shows the data (observed percentages of correct answer and confident response) and the estimate of the parameters with their approximate standard errors in parenthesis.

Table 1 also shows the data for subjects, $\hat{\theta}_i$ and the average of the probabilities of belonging to each latent class, which is given by

$$\bar{P}(\xi_{ij}=k)=\frac{1}{p}\sum_{j=1}^{p} P(\xi_{ij}=k \mid y_{ij},\ \hat{\theta}_i). \tag{21}$$

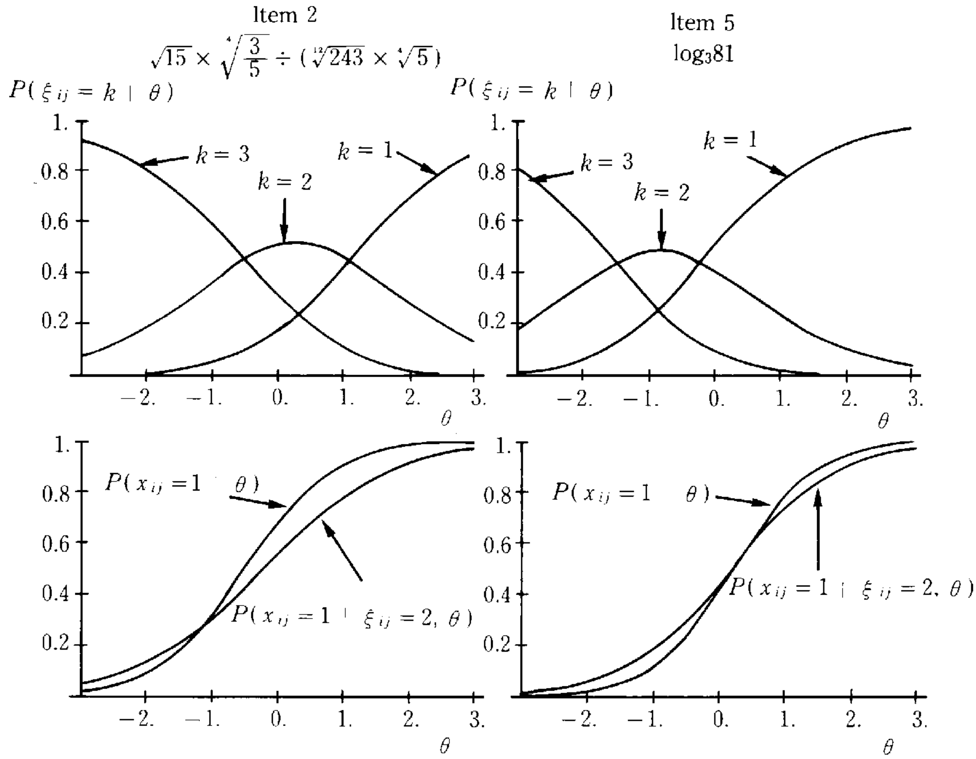

Fig. 1 Estimated item response curves

With these estimates, we can investigate the item characteristics in more detail. For example, we can learn in which items the subjects tend to have more confidence. Or, we can learn the performance of those who have only partial knowledge.

Also, with additional information provided by confidence data, we can expect more accurate estimation of the subjects' competence.

Appendix 1

The first derivatives and Hessian Matrix

To make the description compact, let f, π, μ, ϕ, and ψ be difined as follows;

$$f=P(x_{1ij}\,|\,y_{ij},\ a_j,\ b_j,\ \theta_i,\ \xi_{ij}),$$
$$\pi=\frac{1}{1+\exp(-1.7a_j(\theta_i-b_j))},$$
$$\mu=1+\exp(\theta_i-d_{2j})+\exp((\theta_i-d_{1j})+2(\theta_i-d_{2j})),$$
$$\phi=1+\exp(\theta_i-d_{1j})+\exp((\theta_i-d_{1j})+(\theta_i-d_{2j})),$$
$$\psi=1+\exp((\theta_i-d_{1j})+(\theta_i-d_{2j})).$$

The first derivatives are given below.

$$\frac{\partial f}{\partial d_{1j}}=\frac{\psi(\phi-1)}{\phi^2}-\frac{(1-\psi)}{\phi}+\frac{\pi\psi\exp(\theta_i-d_{2j})}{\mu^2}-\frac{\pi\psi}{\mu},$$
$$\frac{\partial f}{\partial d_2}=\frac{\psi(1-\psi)}{\phi^2}-\frac{(1-\psi)}{\phi}-\frac{-\pi\psi\exp(\theta_i-d_{2j})-2\exp(3\theta_i-d_{1j}-2d_{2j})}{\mu^2}-\frac{\pi\psi}{\mu},$$
$$\frac{\partial f}{\partial a}=\frac{1.7\pi^2(\theta_i-b_j)\exp(2\theta_i-d_{1j}-d_{2j}-1.7a_j(\theta_i-b_j))}{\mu},$$
$$\frac{\partial f}{\partial b}=\frac{-1.7\pi^2 a_j\exp(2\theta_i-d_{1j}-d_{2j}-1.7a_j(\theta_i-b_j))}{\mu}.$$

Hessian matrices are shown below.

$$\frac{\partial^2 f}{\partial^2 d_{1j}}=\frac{2\psi(1-\phi)}{\phi^3}+\frac{(3\psi-2)(1-\phi)}{\phi^2}+\frac{(\psi-1)}{\phi}+\frac{2\pi\psi^3\exp(2(\theta_i-d_{2j}))}{\mu^3}-\frac{3\pi\psi^2\exp(\theta_i-d_{2j})}{\mu^2}+\frac{\pi\psi}{\mu},$$
$$\frac{\partial^2 f}{\partial^2 d_{2j}}=\frac{2\psi\exp(-2d_{1j}-2d_{2j}+d\theta_i)}{\phi^3}-\frac{2\exp(4\theta_i-2d_{1j}-2d_{2j})}{\phi^2}-\frac{(1-\psi)\psi)}{\phi^2}+\frac{(1-\psi)}{\phi}+\frac{2\pi\psi\exp(2(\theta_i-d_{2j}))}{\mu^3}-\frac{3\pi\psi^2\exp(\theta_i-d_{2j})}{\mu^2}+\frac{\pi\psi}{\mu},$$
$$\frac{\partial^2 f}{\partial^2 a}=\frac{-1.7\pi^2\exp(2\theta_i-d_{1j}-d_{2j}-1.7a_j(\theta_i-b_j))}{\mu}+\frac{2.98a_j\pi(\theta_i-b_j)^2\exp(2\theta_i-d_{1j}-d_{2j}-1.7a_j(\theta_i-b_j))}{\mu},$$
$$\frac{\partial^2 f}{\partial^2 b}=\frac{-2.98a^2\pi^2\exp(2\theta_i-d_{1j}-d_{2j}-3.4a_j(\theta_i-b_j))}{\mu}.$$

Appendix 2

Items in mathematics test

Item Number	Content
1	$16^{\frac{2}{3}}\times 16^{\frac{1}{2}}\div 16^{\frac{1}{2}}$
2	$\sqrt{15}\times\sqrt[4]{\frac{3}{5}}\div(\sqrt[2]{243}\times\sqrt[4]{5})$
3	$\{(25^{\frac{5}{3}})^{\frac{3}{2}}\}^{\frac{1}{5}}$
4	$\sqrt{\sqrt[3]{27}\times\sqrt[3]{81}\times\sqrt[3]{243}}$
5	$\log_3 81$
6	$\log_2 63+\log_2 \frac{5}{7}-\log_2 45$
7	$\log_2 3\times\log_3 5\times\log_5 8$
8	Solve $2^{2x+1}=32$.
9	Solve $\log_4 (x-1)=2$.
10	Solve $\log_3 x<2$.
11	When $3^x+3^{-x}=6$, calculate $3^{2x}+3^{-2x}$.
12	What figure number is 6^{50}? Where, $\log_{10} 2=0.3010$, $\log_{10} 3=0.4771$.

REFERENCES

Ebel, R.E. (1968). Valid confidence testing. *Journal of Educational Measurement*, **5**, 4, 353-354.

Embretson, S.E. (1985). Multicomponent latent trait models for test design. In S.E. Embretson (Ed.), *Test design : Developments in Psychology and Psychometrics*. Orlando, Fl : Academic Press.

Fischer, G. (1973). Linear logistic test model as instrument in educational research. *Acta Psychologica*, **37**, 359-374.

Linn, R.L. (ed.) (1989). *Educational Measurement*. Macmillan Publishing Company.

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

Masters, G.N. (1985). A comparison of latent trait and latent class analysis of likert-type data. *Psychometrika*, **50**, 69-82.

Mislevy, R.J. (1987). Modeling item responses when different subjects employ different solution strategies. (*Reserch Report RR-87-47-ORN*). Princeton, NJ : Educational Testing Service.

Frederiksen, N., Mislevy, R.J. and Bejar, I.I. (Eds.) (1993). *Test Theory for A New Generation of Tests*. Hillsdale, NJ : Lawrence Erlbaum Associates, Publishers.

Tatsuoka, K.K. (1983). Rule space : An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, **20**, 345-354.

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

Yamamoto, K. (1993). *Application of a HYBRID model to a test of cognitive skill representation*. In N. Frederiksen, R.J. Mislevy and I.I. Bejar (Eds.). *Test Theory for a New Generation of Tests*. Hillsdale, NJ : Lawrence Erlbaum Associates, Publishers.

(Received January, 1993. Revised september, 1993)