The Batch-Type Neural Test Model:
A Latent Rank Model with The Mechanism of Generative Topographic Mapping

SHOJIMA Kojiro

June 2008

Department of Test Analysis and Evaluation, Research Division, The National Center for University Entrance Examinations

The batch-type neural test model: A latent rank model with the mechanism of generative topographic mapping

Kojiro Shojima

Abstract

This study investigated a batch-type learning version of the neural test model, in which the mechanism of a generative topographic mapping (GTM) and a linear filter smoothing were used for estimating the IRPs. The IRP estimates produced by the batch-type version of the NTT model were invariant in each calculation, and the computation time required for identifying the NTT model with the GTM mechanism was much shorter than that of the NTT model with the SOM mechanism. Furthermore, the optimal linear filter was estimated by minimizing an information criterion reflecting the smoothness of the IRPs in the degrees of freedom of the model, although the way of selecting the information criterion to be minimized emerges as a future issue.

Key words: neural test theory, latent rank theory, item reference profile, EM algorithm, smoothing, minimum information estimation, generative topographic mapping.

バッチ型ニューラルテストモデル: 生成トポグラフィックマッピングのメカニズムを利用した潜在ランク理論

荘島宏二郎

要約

本研究では，バッチ型学習のニューラルテスト理論モデルを提案した．そこでは，項目参照プロファイル (IRP) を推定する際に，生成トポグラフィックマッピングのメカニズムと線形フィルタによる平滑化を用いた．バッチ型 NTT モデルは，毎回の計算結果が変わらず，また，計算時間が大幅に短縮されるメリットがある．また，IRP を推定する際に，モデルの滑らかさをモデルの自由度に反映させた情報量基準を最小化することによって最適な線形フィルタを推定する方法を論じた．目的関数となる情報量基準をどのように選択するか課題が残った．

キーワード: ニューラルテスト理論, 潜在ランク理論, 項目参照プロファイル, EM アルゴリズム，平滑化，最小情報量推定，生成トポグラフィックマッピング.

Department of Test Analysis and Evaluation, Research Division, The National Center for University Entrance Examinations

1 Introduction

The neural test theory (NTT; Shojima, 2008a, 2008b) is a theory for test standardization that uses the mechanism of a self-organizing map (SOM; e.g., Kohonen, 1990). Although the assumed latent scale is continuous in the classical test theory and item response theory (e.g., Lord, 1980; Hambleton & Swaminathan, 1985), the NTT assumes that the latent scale is ordinal. This is because the resolution of tests generally is not high enough to measure human abilities on a continuous scale, where the resolution is the power to discriminate the difference between two or more things, and the most that tests can do is to grade examinees into several ranks (Shojima, 2008a, 2008b).

In the ordinal scale of the NTT, each examinee is estimated to belong to one of the latent ranks in $R_1, \cdots, R_Q$, where Q is the number of latent ranks. Although the size of Q can be determined by referring to some goodness-of-fit indices, it is basically up to the test administrator or data analyst. In addition, an item reference profile (IRP) that represents the transition of the correct answer ratio through the latent rank scale for each item is obtained in the NTT analysis.

The IRPs in the NTT corresponds to the item parameters in the IRT because the estimation of the IRPs is related to test standardization, especially, test scaling. Test standardization is a compositive concept that includes test scaling and equating. Test scaling is the procedure to define a scale or a space for an ability that the test measures and to clarify the statistical features of the test items in the space. In the NTT, the space is the latent ordinal scale, and an item can be said to be scaled when the IRP of the item is estimated, because the IRP represents the statistical feature of each item in the latent space. In addition, test equating is the procedure to prepare a common scale for comparing different test scales.

In addition, the IRP is important for computing the test reference profile (TRP; Shojima, 2008a) and rank membership profile (RMP; Shojima, 2008b) which are inevitable for tests administered using the NTT. The TRP is the weighted sum of the IRPs of all the items, and it is useful to review the expected scores of the examinees belonging to respective latent ranks. In addition, the RMP is the posterior distribution of the latent rank to which each examinee belongs.

However, the IRP estimate differs in each calculation. Differences between estimates can be obtained even when the estimation process is rerun under the same settings. This is because the NTT uses the SOM mechanism. The SOM is categorized as an unsupervised statistical learning model, which is one of the neural network models. In the estimation process of a statistical learning model, the model learns each data after the data has been input into the model. This is why, the learning process includes a step to randomize the input order to cancel out the sequence effect. This randomization step is effective to make the model learn the input data in an unbiased manner, which causes the slight differences between the IRP estimates in every calculation even for the same settings. The statistical features of the items are not invariant over time, and the differences can be reduced when the convergence criterion is set more rigorously. However, some data analysts and test administrators dislike the characteristics derived from the SOM mechanism.

The purpose of this study is to investigate a batch-type version of the NTT model to obtain invariant IRP estimates. This model uses the mechanism of the generative topographic mapping (GTM; Bishop, Svensen, & Williams, 1998). The GTM was originally developed as a batch-type SOM, and it uses EM (expectation maximization) algorithm (Dempster, Laird, & Rubin, 1977) in the statistical learning process. The procedure of the GTM is very similar to the estimation method of the latent class models (LCMs; e.g., Titterington, Smith, & Makov, 1985; McLachlan, & Peel, 2000; Croon, 2002). However, the IRPs become nonsmooth unlike the IRPs estimated by the SOM mechanism as a result of simply applying the GTM mechanism or the estimation method of the LCMs. In this paper, we propose an estimation framework for obtainining smooth and invariant IRPs by retrofitting the EM algorithm with an elastic mechanism.

2 Method

2.1 Statistical Learning Framework

Let us assume that the sample size is N, the number of items is n, and the response data of the N examinees for the n items is $\boldsymbol{U} = \{u_{ij}\}$ $(N \times n)$, where u_{ij} is a dichotomous variable coded 1 when the response of examinee i to item j is correct and 0 otherwise. In addition, let us also assume that $\boldsymbol{Z} = \{z_{ij}\}$ $(N \times n)$ is the missing indicator matrix, where z_{ij} is also a dichotomous variable coded 1 if u_{ij} is observed and 0 if the response is missing. Nonresponses are generally treated as observed $(z = 1)$ and incorrect $(u = 1)$.

Let us also assume that the number of latent ranks in the latent ordinal scale is Q and the q-th latent rank is denoted R_q $(q = 1, \cdots, Q)$, where the ability level of R_q is supposed to be higher than that of R_{q-1}. In addition, $\boldsymbol{F} = \{f_{iq}\}$ $(N \times Q)$ is the rank membership indicator (RMI) matrix, where f_{iq} is a dichotomous variable coded 1 if examinee i belongs to latent rank R_q and 0 otherwise.

Let us further assume that $\boldsymbol{V} = \{v_{qj}\}$ $(Q \times n)$ is the rank reference matrix (RRM;

Shojima, 2008b), where v_{qj} is the rank reference element (RRE) standing for the correct answer ratio of the examinees who belong to latent rank R_q for item j. That is,

$$\Pr(U_j = 1|F_q = 1) = v_{jq}, \tag{1}$$

where U_j is a random response variable for item j and F_q is also a random variable of the rank membership indicator of latent rank R_q. The j-th row vector of the RRM, $\boldsymbol{v}_j = \{v_{jq}\}$ $(Q \times 1)$, is the item reference profile (IRP) of item j, and the q-th column vector of the RRM, $\boldsymbol{v}_q = \{v_{jq}\}$ $(n \times 1)$, is the rank reference vector of latent rank R_q.

With reference to the procedure shown by Shojima (2008a, 2008b), the estimation framework for the batch-type NTT model is given by

$$\text{Obtain } \boldsymbol{Z} \text{ from } \boldsymbol{U}. \tag{2}$$

$$\text{Define } \boldsymbol{V}^{(0)}. \tag{3}$$

$$\text{For } (t{=}1;\ t \leq T;\ t = t+1) \tag{4}$$

$$\text{— Obtain } \boldsymbol{F}^{(t)} \text{ by using } \boldsymbol{U} \text{ and } \boldsymbol{V}^{(t-1)}. \tag{5}$$

$$\text{— Obtain } \boldsymbol{E}^{(t)} \text{ by using } \boldsymbol{F}^{(t)}. \tag{6}$$

$$\text{— Obtain } \boldsymbol{V}^{(t)} \text{ by using } \boldsymbol{E}^{(t)}. \tag{7}$$

Line (4) indicates that Lines (5)–(7) are repeatedly executed T times. In addition, $\boldsymbol{F}^{(t)}$ is the RMI at the t-th period in the statistical learning process. Similarly, $\boldsymbol{E}^{(t)}$ is the elastic RMI at the t-th period, which is obtained by operating $\boldsymbol{F}^{(t)}$, as explained later in Section 2.3. Furthermore, $\boldsymbol{V}^{(t)}$ is the RRM at the t-th period; Shojima (2008b) recommended an initial value $\boldsymbol{V}^{(0)}$ of

$$v_{jq}^{(0)} = q/(Q+1) \quad (q = 1, \cdots, Q;\ \forall j \in n). \tag{8}$$

2.2 Expected Log-Likelihood

The likelihood that $\boldsymbol{U}$ is observed provided that the RRM $\boldsymbol{V}$ and RMI $\boldsymbol{F}$ are given is

$$p(\boldsymbol{U}|\boldsymbol{V}, \boldsymbol{F}) = \prod_{i=1}^{N}\prod_{q=1}^{Q} p(\boldsymbol{u}_i|\boldsymbol{v}_q)^{f_{iq}} = \prod_{i=1}^{N}\prod_{q=1}^{Q}\prod_{j=1}^{n}\{v_{jq}^{u_{ij}}(1-v_{jq})^{1-u_{ij}}\}^{z_{ij}f_{iq}}, \tag{9}$$

where $\boldsymbol{u}_i = \{u_{ij}\}$ $(n \times 1)$ is the response vector of examinee i. The unknown variables in the above equation are $\boldsymbol{V}$ and $\boldsymbol{F}$. The RMI $\boldsymbol{F}$ is called nuisance parameters because the number of elements in $\boldsymbol{F}$ varies according to the sample size N. On the other hand, $\boldsymbol{V}$ is structural parameters because the number of elements in $\boldsymbol{V}$ is invariant with respect to N.

When the structural parameter(s) and nuisance parameter(s) are simultaneously included in the likelihood, the log-likelihood is generally marginalized with respect to the nuisance parameter(s). In such cases, the EM algorithm (Dempster, Laird, & Rubin, 1977) is a useful method.

The EM algorithm is a numerical solution method for estimating the structural parameters by repeatedly applying E-steps and M-steps. In the E-steps, the expected log-likelihood is obtained by integrating the log-likelihood with respect to the nuisance parameters over the posterior distribution of the nuisance parameters. The E-step in the t-th period is given by

$$\ln p(\boldsymbol{V}|\boldsymbol{U}) = E_{\boldsymbol{F}|\boldsymbol{U},\boldsymbol{V}^{(t-1)}}[\ln p(\boldsymbol{V}|\boldsymbol{U},\boldsymbol{F})], \tag{10}$$

where $\boldsymbol{V}^{(t-1)}$ is the estimate of the RRM obtained in the M-step in the $t-1$-th period. The right-hand side of the above equation is decomposed as

$$\begin{aligned} &E_{\boldsymbol{F}|\boldsymbol{U},\boldsymbol{V}^{(t-1)}}[\ln p(\boldsymbol{V}|\boldsymbol{U},\boldsymbol{F})] \\ =&E_{\boldsymbol{F}|\boldsymbol{U},\boldsymbol{V}^{(t-1)}}[\ln p(\boldsymbol{U}|\boldsymbol{V},\boldsymbol{F})] + E_{\boldsymbol{F}|\boldsymbol{U},\boldsymbol{V}^{(t-1)}}[\ln p(\boldsymbol{V}|\boldsymbol{F})] - E_{\boldsymbol{F}|\boldsymbol{U},\boldsymbol{V}^{(t-1)}}[\ln p(\boldsymbol{U}|\boldsymbol{F})] \end{aligned} \tag{11}$$

by using the Bayesian theorem. The first term of the above equation is

$$\begin{aligned} E_{\boldsymbol{F}|\boldsymbol{U},\boldsymbol{V}^{(t-1)}}[\ln p(\boldsymbol{U}|\boldsymbol{V},\boldsymbol{F})] &= \sum_{i=1}^{N}\sum_{q=1}^{Q} f_{iq}^{(t)} \ln p(\boldsymbol{u}_i|\boldsymbol{v}_q) \\ &= \sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{j=1}^{n} f_{iq}^{(t)} z_{ij}\big\{u_{ij}\ln v_{jq} + (1-u_{ij})\ln(1-v_{jq})\big\}, \end{aligned} \tag{12}$$

where $f_{iq}^{(t)}$ is the posterior probability that examinee i belongs to latent rank R_q at the t-th EM cycle given the RRM at the $t-1$-th period, $\boldsymbol{V}^{(t-1)}$, and the response vector of examinee i, $\boldsymbol{u}_i$. That is,

$$p(f_{iq}|\boldsymbol{u}_i,\boldsymbol{V}^{(t-1)}) = f_{iq}^{(t)} = \frac{p(\boldsymbol{u}_i|\boldsymbol{v}_q^{(t-1)})\pi_{iq}}{\sum_{q'=1}^{Q} p(\boldsymbol{u}_i|\boldsymbol{v}_{q'}^{(t-1)})\pi_{iq'}} \quad (i=1,\cdots,N;\ q=1,\cdots,Q), \tag{13}$$

where π_{iq} is the prior probability of f_{iq}. The above equation corresponds to Line (5) in the estimation framework shown in Section 2.2. In addition, the i-th row vector in $\boldsymbol{F}^{(t)}$, $\boldsymbol{f}_i^{(t)}$, is the posterior distribution of examinee i belonging to the respective latent ranks, and it is the rank membership profile (RMP; Shojima, 2008b) of examinee i.

Next, the second term of Equation (11) becomes

$$E_{\boldsymbol{F}|\boldsymbol{U},\boldsymbol{V}^{(t-1)}}[\ln p(\boldsymbol{V}|\boldsymbol{F})] = \ln p(\boldsymbol{V}), \tag{14}$$

and this term can be regarded as the prior distribution of the structural parameters. Furthermore, the third term of Equation (11) is reduced to a constant. Consequently, the expected log-likelihood of Equation (10) is given by

$$\ln p(\boldsymbol{V}|\boldsymbol{U}) = \sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{j=1}^{n} f_{iq}^{(t)} z_{ij}\big\{u_{ij}\ln v_{jq} + (1-u_{ij})\ln(1-v_{jq})\big\} + \ln p(\boldsymbol{V}) + const. \quad (15)$$

2.3 Scale Elasticity and Weakly Ordinal Alignment Condition

In the M-steps of the usual EM algorithm, the expected log-likelihood (Equation 15) is optimized with respect to the structural parameters $\boldsymbol{V}$. This procedure is also almost identical to a simple application of the GTM mechanism or the estimation method of the LCMs to the NTT. Accordingly, the ordinality of the latent scale in the NTT is not still satisfied because a mechanism that makes the latent scale ordinal is not built into the GTM mechanism or the estimation method for the LCMs themselves.

Shojima (2008b) defined weakly and strongly ordinal alignment conditions for expressing the degree of ordinality of the latent scale. The weak condition is satisfied when the test reference profile (TRP) is monotonically increasing but every IRP is not necessarily monotonic, where the TRP is the weighted sum of the IRPs of the test items and expresses the expected scores of the examinees belonging to the respective latent ranks. Meanwhile, the strong condition requires all the IRPs to be monotonic, so the TRP inevitably increases monotonically. The monotonicity of the TRP is the only evidence that can show the ordinality of the NTT scale. Therefore, at least the weak condition must be satisfied to attain the ordinality on the NTT scale.

In addition, the IRPs cannot become smooth by a direct application of the GTM mechanism or the estimation method for the LCMs to the estimation method for the NTT model, while the IRPs do become smooth in the NTT model with the SOM mechanism. This is because the SOM mechanism itself has a property that makes the IRPs smooth. In fact, the weak condition is easily attained in the NTT model with the SOM mechanism because the smoothness of each IRP has the effect of making the IRP monotonic, and the monotonicity of each IRP leads to the monotonicity of the TRP, which is the requirement of the weak condition.

Furthermore, it is important to make the IRPs smooth from the viewpoint of prediction because the IRP of each item can be said to be a nonlinear and nonparametric regression of the correct answer ratio of the item on the latent rank and is required to predict the correct answer ratio of the examinees belonging to each latent rank. It is well known that the predictive capability of a regression model for future data becomes precise when the regression line of the model is smooth (e.g., Green & Silverman, 1994; Ramsay & Silverman, 1997; Eubank, 1999; Hastie, Tibshirani, & Friedman, 2001).

In the NTT model with the SOM mechanism, the weakly ordinal alignment condition is highly likely to be satisfied by making the IRPs smooth. That is, it is necessary to improve the elasticity of the scale to satisfy the weak condition in the NTT model with the GTM mechanism. Therefore, it is of great significance to obtain the elastic rank membership indicator $\boldsymbol{E}^{(t)}$ from the rank membership indicator (RMI) $\boldsymbol{F}^{(t)}$ because the elasticity of $\boldsymbol{E}^{(t)}$ has a direct effect on the monotoniciy of each row vector of $\boldsymbol{V}^{(t)}$.

Although there are many ways to obtain the elastic RMI $\boldsymbol{E}^{(t)}$, one useful way is as follows:

$$\boldsymbol{E}^{(t)} = \boldsymbol{F}^{(t)}\boldsymbol{G}. \tag{16}$$

That is, the elastic RMI is given by weighting the RMI with a smoothing matrix $\boldsymbol{G}$, where

$$\boldsymbol{G} = [\boldsymbol{g}_1 \cdots \boldsymbol{g}_q \cdots \boldsymbol{g}_Q] \quad (Q \times Q) \tag{17}$$

and each column vector in $\boldsymbol{G}$ is standardized to make the sum of the vector one ($\boldsymbol{g}'_q\mathbf{1}_Q = 1$), where $\mathbf{1}_Q$ is a vector with size Q in which all the elements are one. Accordingly, each element of $\boldsymbol{E}^{(t)}$, $e^{(t)}_{iq}$, becomes a constant such that $\boldsymbol{f}^{(t)}_i$ (the i-th row vector of $\boldsymbol{F}^{(t)}$) is smoothed by $\boldsymbol{g}_q$ (the q-th column vector of $\boldsymbol{G}$).

In some cases, it is effective to set the smoothing matrix in advance. For example, for $Q = 5$, if $\boldsymbol{G}$ is given by

$$\boldsymbol{G} = \begin{bmatrix} 1/2 & 1/3 & 0 & 0 & 0 \\ 1/2 & 1/3 & 1/3 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/2 \\ 0 & 0 & 0 & 1/3 & 1/2 \end{bmatrix}, \tag{18}$$

then $e^{(t)}_{iq}$ becomes the moving average of $f^{(t)}_{iq-1}$, $f^{(t)}_{iq}$, and $f^{(t)}_{iq+1}$. The elements in the first and fifth columns are adjusted to make the sums of the columns equal to one. The weights like $\{1/3,\ 1/3,\ 1/3\}$ are called a linear filter. As for the linear filter $\boldsymbol{\phi}_b$ being desirable for the batch-type NTT model, the number of elements in the filter is odd, and the elements are symmetric about the central element and the more peripheral elements in the filter become smaller. In addition, it is satisfactory that the number of elements of the linear filter, b, is three or five by considering that the realistic range of the number of latent ranks is $3 \le Q \le 20$. Linear filters with three and five elements, $\boldsymbol{\phi}_3$ and $\boldsymbol{\phi}_5$, that satisfy such conditions are

$$\boldsymbol{\phi}_3 = \left\{\frac{1-\phi}{2}, \phi, \frac{1-\phi}{2}\right\} \quad \left(0 \leq \frac{1-\phi}{2} \leq \phi \leq 1\right), \tag{19}$$

and

$$\boldsymbol{\phi}_5 = \left\{\frac{1-\phi_1-2\phi_2}{2}, \phi_2, \phi_1, \phi_2, \frac{1-\phi_1-2\phi_2}{2}\right\} \quad \left(0 \leq \frac{1-\phi_1-2\phi_2}{2} \leq \phi_2 \leq \phi_1 \leq 1\right). \tag{20}$$

Furthermore, it is desirable to use a different linear filter according to the number of latent ranks. The parameter ϕ in Equation (19) that we recommend is

$$\phi = \begin{cases} 1.05 - 0.05Q & (1 \leq Q \leq 5) \\ 1.00 - 0.04Q & (5 \leq Q \leq 10) \\ 0.80 - 0.02Q & (10 \leq Q \leq 20) \end{cases}. \tag{21}$$

An extreme example of the linear filter is $\{\cdots, 0, 1, 0, \cdots\}$. In this case, the smoothing matrix becomes an identity matrix with size Q ($\boldsymbol{G} = \boldsymbol{I}_Q$), and the elastic RMI becomes identical to the RMI ($\boldsymbol{E}^{(t)} = \boldsymbol{F}^{(t)}$). In addition, when using another extreme linear filter, that is $Q^{-1}\mathbf{1}_b$ ($b \leq Q$), all the elements of the smoothing matrix then become $1/Q$, which subsequently causes all the elements of the elastic RMI to also becomes $1/Q$ ($\boldsymbol{E}^{(t)} = Q^{-1}\boldsymbol{F}^{(t)}\mathbf{1}_Q\mathbf{1}'_Q = Q^{-1}\mathbf{1}_N\mathbf{1}'_Q$).

Using the elastic RMI weighted by the smoothing matrix, we can reconstruct the expected log-likelihood obtained in the t-th E-step (Equation 15) as follows:

$$\ln p(\boldsymbol{V}|\boldsymbol{U}) = \sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{j=1}^{n} e_{iq}^{(t)} z_{ij} \left\{u_{ij} \ln v_{jq} + (1-u_{ij}) \ln(1-v_{jq})\right\} + \ln p(\boldsymbol{V}) + const. \tag{22}$$

Then, in the M-steps, the expected log-likelihood is optimized with respect to the structural parameters. When the prior probability of each RRE is a constant, the first derivative of the expected log-likelihood with respect to the structural parameters is given by

$$\frac{\partial \ln p(\boldsymbol{V}|\boldsymbol{U})}{\partial v_{jq}} = \sum_{i=1}^{N} e_{iq}^{(t)} z_{ij} \left\{\frac{u_{ij}}{v_{jq}} - \frac{1-u_{ij}}{1-v_{jq}}\right\}. \tag{23}$$

Equation (22) can be maximized with respect to each RRE. Therefore, the estimate of v_{jq} in the t-th M-step can be obtained by solving the above equation set to 0. That is, the RRE estimate at the t-th EM cycle is obtained as

$$v_{jq}^{(t)} = \frac{\sum_{i=1}^{N} z_{ij} u_{ij} e_{iq}^{(t)}}{\sum_{i=1}^{N} z_{ij} e_{iq}^{(t)}} = \frac{\sum_{i=1}^{N} z_{ij} u_{ij} \boldsymbol{g}'_q \boldsymbol{f}_i^{(t)}}{\sum_{i=1}^{N} z_{ij} \boldsymbol{g}'_q \boldsymbol{f}_i^{(t)}}. \tag{24}$$

2.4 Effective Degrees of Freedom and Minimum Information Estimation of Smoothing Matrix

The IRP shapes are greatly influenced by changing $\boldsymbol{G}$. Therefore, the idea that the optimal $\boldsymbol{G}$ is estimated from the data is worth considering. To begin with, it is necessary to evaluate the size of the effect of $\boldsymbol{G}$ controlling the elasticity of the model by referring to an information criterion.

We must start by considering the degrees of freedom of the model. The degrees of freedom of a statistical model are generally given by

$$df = p_1 - p_2, \tag{25}$$

where p_1 and p_2 are the numbers of parameters of the saturated and present models, respectively. The saturated model is the model that fits the data much better than the present model does and is used as the basis for comparison. When the saturated model for the batch-type NTT model is defined as the model in which the number of latent ranks is the number of response patterns, such a model can completely fit the data. However, such a model is a lofty ideal for the comparison basis of the present model. Therefore, the benchmark model (Shojima, 2008b) is used as the saturated model. This is a batch-type NTT model where the number of latent ranks is n (the number of items) and satisfies the strongly ordinal alignment condition. In addition, the smoothing matrix of the benchmark model is the identity matrix with size n. This means that the benchmark model is not smoothed. When the batch-type model is not smoothed, the RREs of the IRP are estimated independently. Accordingly, the number of parameters for each item of the benchmark model can be considered as $p_1 = n$.

Next, as for the number of parameters of the present model p_2, it was defined as the number of latent ranks Q per item in the NTT model with the SOM mechanism (Shojima, 2008b), which was equal to the number of elements (or the number of structural parameters) in the IRP of each item. Accordingly, the number of parameters for the whole model is $n \times Q$. However, the IRP elements are not estimated independently; instead, the size of the IRP estimates affect each other. Therefore, strictly speaking, the number of parameters should be smaller or equal to Q, although it is almost impossible to quantify the degree of interdependency among the elements in each IRP in the complicated estimation process of the NTT model with the SOM mechanism.

In the batch-type NTT model proposed in this study, however, the degree of interdependency is expressed in the smoothing matrix $\boldsymbol{G}$. Hastie, Tibshirani, & Friedman (2001) proposed the effective degrees of freedom (EDF), which was defined as the trace of the smoothing matrix. Here, the trace of the smoothing matrix $\boldsymbol{G}$ can be considered to be the number of parameters of the present model of the batch-type NTT model, and it is given by

$$p_2 = \text{tr } \boldsymbol{G}. \tag{26}$$

Let us consider the validity of regarding the trace of the smoothing matrix as the number of parameters of the present model by looking at two extreme examples. When the smoothing matrix is an identity matrix ($\boldsymbol{G} = \boldsymbol{I}_Q$), the IRP elements are estimated independently. In this case, the number of parameters is logically considered to be the number of latent ranks Q, and the trace of the smoothing matrix is then congruent with Q ($p_2 = \text{tr } \boldsymbol{I}_Q = Q$). In addition, when all the elements in the smoothing matrix are $1/Q$ ($\boldsymbol{G} = \mathbf{1}\mathbf{1}'/Q$), all the elements in the IRP of each item become equal to the item's correct answer ratio. In this case, the number of parameters can be considered to be 1, and the trace of the smoothing matrix also becomes 1 ($p_2 = \text{tr } (\mathbf{1}\mathbf{1}'/Q) = 1$). As can be seen above, a certain type of validity is satisfied by considering the trace of the smoothing matrix as the number of parameters of the present model. Therefore, the EDF of item j of the batch-type NTT model is given by

$$edf_j = n - \text{tr } \boldsymbol{G}. \tag{27}$$

The χ^2 statistic of the batch-type NTT model can then be computed as

$$C_j = 2\{\ln p(\hat{\boldsymbol{v}}_{Bj}|\boldsymbol{u}_j) - \ln p(\boldsymbol{v}_j^{(t)}|\boldsymbol{u}_j)\}, \tag{28}$$

where $\hat{\boldsymbol{v}}_{Bj}$ ($n \times 1$) is the IRP estimate of the benchmark model for item j. In addition, some information criteria for each item can be evaluated from the above statistic. For example, the Akaike information criterion (AIC; Akaike, 1987), consistent AIC (CAIC; Bozdogan, 1987), and Bayes information criterion (BIC; Schwarz, 1978) are calculated as follows:

$$AIC_j = C_j - 2edf_j, \tag{29}$$

$$CAIC_j = C_j - edf_j(\ln N + 1), \tag{30}$$

$$BIC_j = C_j - edf_j(\ln N). \tag{31}$$

In addition, the information criteria for the whole model are evaluated by using $C = \sum_j C_j$ and $edf = \sum_j edf_j = n(n - \text{tr } \boldsymbol{G})$.

The minimum information estimation (Akaike, 1974) is a method of estimating the model parameters by minimizing an information criterion. In the case of the batch-type NTT model, it is possible to explore an optimal smoothing matrix by setting an information criterion as the objective function. Let the following step be inserted after Line (5):

$$\text{— Obtain } \boldsymbol{G}^{(t)} \text{ by using } \boldsymbol{F}^{(t)} \text{ and } \boldsymbol{V}^{(t-1)}. \tag{32}$$

Given the RRM estimate at the $t-1$-th period ($\boldsymbol{V}^{(t-1)}$) and the RMI estimate at the t-th period ($\boldsymbol{F}^{(t)}$), the information criterion can be written as

$$IC(\boldsymbol{G}|\boldsymbol{V}^{(t-1)}, \boldsymbol{F}^{(t)}) = -2\sum_{i=1}^{N}\sum_{q=1}^{Q}\sum_{j=1}^{n}\boldsymbol{g}_q'\boldsymbol{f}_i^{(t)} z_{ij}\{u_{ij}\ln v_{jq}^{(t-1)} + (1-u_{ij})\ln(1-v_{jq}^{(t-1)})\}$$
$$+ kn \text{ tr } \boldsymbol{G} + const. \quad (33)$$

The terms related to the log-likelihood and the number of parameters of the benchmark model become disjoint during the optimization of the above equation with respect to $\boldsymbol{G}$. The above equation becomes the AIC when the constant k is equal to 2. In addition, the function also becomes the CAIC and BIC when $k = \ln N + 1$ and $k = \ln N$, respectively. The AIC tends to select more conservative models than the CAIC and BIC do. In other words, the AIC tends to be more supportive of a model with a bigger tr $\boldsymbol{G}$. Accordingly, the model selected by the AIC is more likely to be a less smooth model than hte one selected by the CAIC or BIC.

The optimization of Equation (33) is identical to minimization of the function with linear constraints. For example, the substantial number of parameters in the smoothing matrix is only one (ϕ) from Equation (19) when the number of elements in the linear filter is three, and this optimization must be executed under the linear constraints $0 \le (1-\phi)/2 \le \phi \le 1$. In addition, from Equation (20), the number of parameters in the smoothing matrix becomes two when the number of elements in the linear filter is five, and the linear constraints in this case are $0 \le (1-\phi_1-2\phi_2)/2 \le \phi_2 \le \phi_1 \le 1$.

The smoothing matrix that minimizes Equation (33) under the linear constraints is the estimate of the smoothing matrix at the t-th period, $\boldsymbol{G}^{(t)}$. Consequently, the RRM estimate at the t-th period is then obtained by

$$v_{jq}^{(t)} = \frac{\sum_{i=1}^{N} z_{ij}u_{ij}e_{iq}^{(t)}}{\sum_{i=1}^{N} z_{ij}e_{iq}^{(t)}} = \frac{\sum_{i=1}^{N} z_{ij}u_{ij}\boldsymbol{g}_q^{(t)\prime}\boldsymbol{f}_i^{(t)}}{\sum_{i=1}^{N} z_{ij}\boldsymbol{g}_q^{(t)\prime}\boldsymbol{f}_i^{(t)}}. \quad (34)$$

2.5 Stopping Rule and Strongly Ordinal Alignment Condition

The IRPs of the batch-type NTT model are estimated by executing Sections 2.2–2.4 repeatedly. Still, the result does not always satisfy the weakly ordinal alignment condition, although the possibility that the weak condition is satisfied is greatly increased by smoothing the IRPs. The NTT scale possesses no evidence that the scale is ordinal unless at least the weak condition is satisfied. Accordingly, one must judge whether the TRP is monotonic. If it is nonmonotonic, the data must be reanalyzed with a different smoothing matrix. The

TRP tends to strengthen the monotonicity when the central element in the linear filter is set smaller.

Every IRP does not always increase monotonically even if the weak condition is satisfied, and some test administrators or analysts might think that it is necessary for the IRPs of all the items to be monotonic in practical test administration because they might think that the correct answer ratios of all the items must be logically monotonic even if the correct ratios of somje items are not actually monotonic. When the IRPs of all the items are monotonic, the TRP is consequently monotonic. In this case, the strongly ordinal alignment condition is satisfied. To impose the monotonic increase constraint (MIC; Shojima, 2008a, 2008b) on the IRPs of all the items, it is necessary to insert a step after Line (7). For example, the step is given by

$$\text{For } (j{=}1;\ j \leq n;\ j = j+1) \tag{35}$$

$$\text{— Sort } \boldsymbol{v}_j^{(t)}. \tag{36}$$

The above step is very simple but effective to make the IRPs at the t-th period monotonic.

It is primarily necessary to estimate the latent rank and rank membership profile of each examinee after obtaining the IRPs. The maximum likelihood method or the Bayesian method shown in Shojima (2008b) is useful for estimating them.

3 Analysis

This section shows some examples of analyzing a geography test by the proposed method. The number of items in the test was 35 and the sample size was 5000. This test was an achievement test, so all missing data was treated as being incorrect. The number-right score distribution is shown in Figure 1 and the simple statistics of this test are listed in Table 1. From the figure and the skewness and kurtosis in the table, the number-right scores can be regarded as being almost normally distributed. In addition, the alpha coefficient (Cronbach, 1951) was about 0.7, which is not very low as the reliability for binary test data.

3.1 Example 1: Result with $Q = 10$ and Fixed Smoothing Matrix

Here, the result analyzed with 10 latent ranks is shown. The smoothing matrix was fixed, and the number of elements in the linear filter was three. That is, the linear filter used in this example was $\boldsymbol{\phi}_3 = \{0.2, 0.6, 0.2\}$ from Equation (21). In addition, the Bayesian estimation method was used, and the prior distribution of latent rank was a trapezoidal distribution

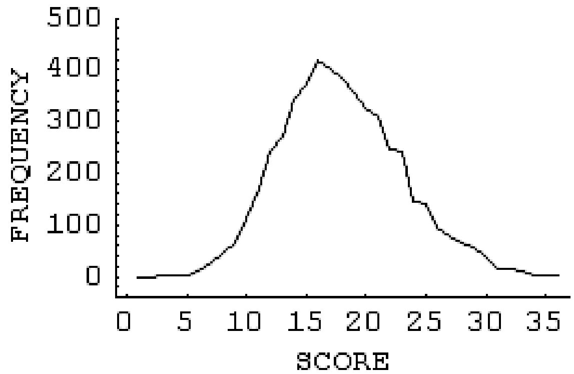


Figure 1: Number-Right Score Distribution

Table 1: Marginal Statistics of Number-Right Scores

Statistic	Value
N	5000
n	35
Median	17
Max	35
Min	2
Range	33
Mean	16.911
SD	4.976
Skew.	0.313
Kurt.	−0.074
Alpha	0.704

for which the prior probabilities at both ends of the latent rank scale were 0.095 (Shojima, 2008b). Table 2 shows the RRM estimated by the procedure described in Section 2. The number of EM cycles required for convergence was eight.

Figure 2 shows the IRPs of 35 items, and each IRP plots the corresponding row vector of the RRM in Table 2. The IRPs generally increase as the latent rank becomes higher. This reflects the correct answer ratio of higher latent rankers for each item generally being higher, although the IRPs of items 5, 7, 11, and 12 do not monotonically increase. In addition, the IRPs are basically smooth, which reflects the effect of using the smoothing matrix. Each IRP represents the characteristics of the corresponding item. For instance, it is clear that items 2, 7, and 20 were very difficult because the correct answer ratios of the items by the examinees belonging to even higher latent ranks were low. On the other hand, items 3 and

Table 2: Reference Matrix Estimate ($Q = 10$, Fixed Smoothing Matrix)

Item	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}
1	0.248	0.260	0.277	0.297	0.317	0.335	0.353	0.385	0.464	0.558
2	0.256	0.259	0.264	0.269	0.276	0.286	0.295	0.301	0.310	0.324
3	0.568	0.621	0.689	0.739	0.773	0.800	0.820	0.840	0.872	0.907
4	0.213	0.213	0.218	0.234	0.263	0.306	0.359	0.432	0.544	0.647
5	0.221	0.221	0.226	0.235	0.243	0.247	0.255	0.292	0.396	0.517
6	0.733	0.771	0.824	0.867	0.892	0.905	0.910	0.915	0.924	0.937
7	0.333	0.333	0.331	0.334	0.347	0.367	0.384	0.397	0.426	0.471
8	0.227	0.236	0.247	0.261	0.285	0.326	0.389	0.483	0.613	0.719
9	0.436	0.485	0.555	0.613	0.657	0.695	0.731	0.764	0.801	0.835
10	0.260	0.261	0.261	0.266	0.289	0.335	0.407	0.510	0.656	0.781
11	0.530	0.563	0.602	0.621	0.620	0.613	0.615	0.646	0.717	0.782
12	0.381	0.407	0.442	0.471	0.504	0.554	0.624	0.709	0.806	0.878
13	0.225	0.242	0.263	0.277	0.288	0.311	0.359	0.440	0.557	0.650
14	0.172	0.193	0.227	0.263	0.297	0.332	0.371	0.422	0.494	0.557
15	0.378	0.422	0.483	0.527	0.547	0.554	0.570	0.615	0.697	0.764
16	0.235	0.254	0.279	0.298	0.313	0.330	0.358	0.410	0.504	0.597
17	0.179	0.210	0.262	0.316	0.357	0.381	0.391	0.403	0.446	0.504
18	0.553	0.582	0.626	0.674	0.726	0.783	0.841	0.895	0.939	0.965
19	0.311	0.336	0.376	0.418	0.463	0.514	0.575	0.658	0.769	0.855
20	0.198	0.204	0.212	0.220	0.232	0.251	0.280	0.323	0.384	0.432
21	0.299	0.334	0.391	0.443	0.481	0.510	0.539	0.573	0.613	0.644
22	0.343	0.371	0.416	0.463	0.507	0.547	0.582	0.620	0.679	0.736
23	0.318	0.333	0.361	0.402	0.455	0.513	0.564	0.603	0.643	0.684
24	0.214	0.272	0.362	0.452	0.527	0.585	0.621	0.646	0.696	0.760
25	0.364	0.422	0.512	0.597	0.663	0.712	0.746	0.770	0.791	0.810
26	0.177	0.195	0.226	0.267	0.323	0.390	0.458	0.526	0.614	0.701
27	0.403	0.459	0.545	0.628	0.693	0.739	0.766	0.786	0.817	0.852
28	0.222	0.256	0.314	0.376	0.432	0.480	0.526	0.594	0.710	0.814
29	0.372	0.393	0.421	0.447	0.478	0.522	0.581	0.658	0.762	0.849
30	0.528	0.598	0.695	0.772	0.819	0.846	0.867	0.890	0.916	0.934
31	0.311	0.342	0.384	0.415	0.432	0.443	0.466	0.519	0.621	0.719
32	0.180	0.180	0.185	0.198	0.219	0.242	0.260	0.283	0.337	0.403
33	0.173	0.190	0.217	0.246	0.268	0.282	0.293	0.321	0.391	0.466
34	0.387	0.407	0.434	0.457	0.479	0.503	0.527	0.549	0.580	0.614
35	0.458	0.502	0.559	0.597	0.619	0.638	0.665	0.706	0.766	0.824

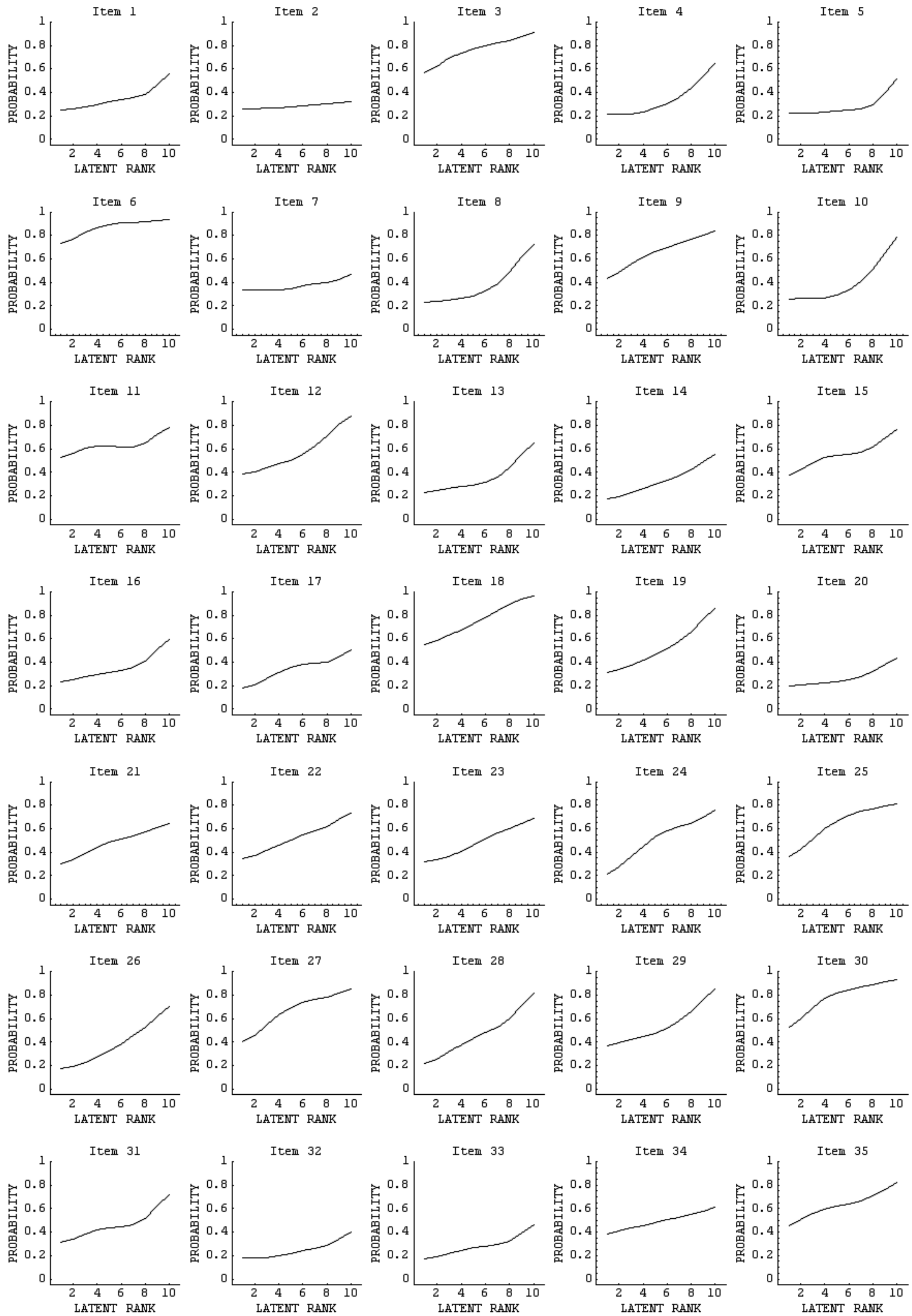


Figure 2: Item Reference Profiles ($Q = 10$, Fixed Smoothing Matrix)

6 were found to be easy because the correct ratios of these items even for lower latent rankers were high. In addition, items 8 and 10 were difficult for mid-level latent rankers and easy for high latent rankers, so these items can be said to have high resolution for discriminating higher latent rankers from mid-level latent rankers. Similarly, items 25 and 27 can discriminate lower rankers from mid-rankers. In addition, item 28 has high discriminancy over the entire latent rank scale. Furthermore, some items had a plateau around the mid-level latent ranks like items 11, 15, and 17.

Model-fit indices for each item can be calculated by the method shown in Shojima (2008b). Table 3 lists the χ^2 statistic, normed fit index (NFI; Bentler & Bonett, 1980), relative fit index (RFI; Bollen, 1986), incremental fit index (IFI; Bollen, 1989), Tucker-Lewis index (TLI; Bollen, 1989), comparative fit index (CFI; Bentler, 1990), root mean square error of approximation (RMSEA; Browne & Cudeck, 1993), AIC, CAIC, and BIC. These indices in the table are calculated from the expected log-likelihood, although they can also be computed by using the log-likelihood. As for the indices based on the expected log-likelihood, the IFI, TLI, and CFI are inclined to be 1.0 and the RMSEA tends to be 0.0. This is because the χ^2 statistic tends to be smaller than the EDF. The EDF of each item was $df_j = 35 - (0.6 \times 8 + 0.6/(0.6 + 0.2) \times 2) = 28.7$. In addition, Table 4 lists the model-fit indices for the whole test. The EDF was 1004.5 ($= 28.7 \times 35$). From these indices, we found that the model-fit was generally satisfactory.

The rank membership profiles (RMPs; Shojima, 2008b) of examinees 1–15 out of the 5000 samples are shown in Figure 3. The RMP represents the probabilities of the examinees belonging to the respective latent ranks. For example, the probability that examinee 1 belongs to latent rank R_2 was 0.304 and this probability was the highest in his/her RMP, so examinee 1 most likely belonged to latent rank R_2, although the probabilities that he/she belonged to R_1 and R_3 were also high: 0.281 and 0.227, respectively. In addition, the latent rank estimates of examinees 7 and 10 were the same, R_{10}, although the probability of examinee 7 belonging to R_{10} was 0.777 and that of examinee 10 was 0.601. Therefore, examinee 7 was more likely to have the ability of the R_{10} ranker than examinee 10 having it. In fact, examinee 7 correctly answered 28 items out of the total of 35 items, while the number-right score of examinee 10 was 24. Furthermore, the RMP of examinee 12 was found to be bimodal. This examinee could respond correctly to relatively difficult items 1, 16, and 20, but responded incorrectly to comparatively easy items 3, 9, and 27. Therefore, the RMP of the examinee was obtained as bimodal because eigther possibility—the ability of the examinee being high or low—was undetermined. In this way, the RMP can give us

Table 3: Item Fit Indices ($Q = 10$, Fixed Smoothing Matrix)

Item	$\chi^2_{28.7}$	NFI	RFI	IFI	TLI	CFI	RMSEA	AIC	CAIC	BIC
1	53.8	0.787	0.748	0.888	0.864	0.885	0.013	−3.6	−219.4	−190.7
2	3.2	0.791	0.752	1.000	1.000	1.000	0.000	−54.2	−269.9	−241.2
3	27.6	0.920	0.905	1.000	1.000	1.000	0.000	−29.8	−245.6	−216.9
4	28.3	0.944	0.934	1.000	1.000	1.000	0.000	−29.1	−244.8	−216.1
5	44.2	0.830	0.799	0.933	0.919	0.931	0.010	−13.2	−228.9	−200.2
6	21.1	0.900	0.881	1.000	1.000	1.000	0.000	−36.3	−252.0	−223.3
7	8.5	0.846	0.817	1.000	1.000	1.000	0.000	−48.9	−264.7	−236.0
8	43.6	0.931	0.919	0.976	0.971	0.975	0.010	−13.8	−229.6	−200.9
9	44.2	0.896	0.877	0.961	0.953	0.960	0.010	−13.2	−229.0	−200.3
10	67.1	0.912	0.896	0.948	0.938	0.947	0.016	9.7	−206.1	−177.4
11	28.1	0.807	0.772	1.000	1.000	1.000	0.000	−29.3	−245.1	−216.4
12	58.8	0.911	0.895	0.952	0.943	0.952	0.014	1.4	−214.3	−185.6
13	39.1	0.913	0.897	0.975	0.970	0.975	0.009	−18.3	−234.0	−205.3
14	32.5	0.915	0.899	0.989	0.987	0.989	0.005	−24.9	−240.6	−211.9
15	47.4	0.852	0.825	0.936	0.923	0.935	0.011	−10.0	−225.7	−197.0
16	27.4	0.908	0.891	1.000	1.000	1.000	0.000	−30.0	−245.8	−217.1
17	36.7	0.865	0.841	0.967	0.960	0.967	0.007	−20.7	−236.5	−207.8
18	35.6	0.945	0.935	0.989	0.987	0.989	0.007	−21.8	−237.5	−208.8
19	52.1	0.929	0.915	0.967	0.960	0.966	0.013	−5.3	−221.0	−192.3
20	39.2	0.791	0.753	0.934	0.919	0.932	0.009	−18.2	−234.0	−205.3
21	7.5	0.972	0.966	1.000	1.000	1.000	0.000	−49.9	−265.6	−236.9
22	21.7	0.938	0.927	1.000	1.000	1.000	0.000	−35.7	−251.5	−222.8
23	27.7	0.924	0.909	1.000	1.000	1.000	0.000	−29.7	−245.4	−216.7
24	113.8	0.855	0.828	0.887	0.866	0.887	0.024	56.4	−159.4	−130.7
25	60.1	0.897	0.878	0.943	0.932	0.943	0.015	2.7	−213.0	−184.3
26	68.9	0.907	0.889	0.943	0.932	0.943	0.017	11.5	−204.2	−175.5
27	41.2	0.926	0.913	0.976	0.972	0.976	0.009	−16.2	−232.0	−203.3
28	80.4	0.902	0.884	0.935	0.922	0.934	0.019	23.0	−192.8	−164.1
29	77.1	0.872	0.849	0.916	0.899	0.915	0.018	19.7	−196.1	−167.4
30	59.8	0.898	0.879	0.944	0.933	0.944	0.015	2.4	−213.3	−184.6
31	63.9	0.829	0.797	0.898	0.877	0.896	0.016	6.5	−209.3	−180.6
32	31.1	0.814	0.780	0.983	0.979	0.982	0.004	−26.3	−242.1	−213.4
33	18.8	0.910	0.893	1.000	1.000	1.000	0.000	−38.6	−254.4	−225.7
34	21.8	0.828	0.797	1.000	1.000	1.000	0.000	−35.6	−251.4	−222.7
35	49.8	0.844	0.816	0.928	0.913	0.926	0.012	−7.6	−223.3	−194.6

detailed information about each examinee such as an examinee could go up to the next rank if he/she studied a little harder and another examinee might go down a notch unless he/she strived to improve. That is, the RMP can be used as educational diagnostic information

Table 4: Test Fit Indices ($Q = 10$, Fixed Smoothing Matrix)

Index	Value
$\chi^2_{1004.5}$	1841.95
NFI	0.897
RFI	0.878
IFI	0.964
TLI	0.957
CFI	0.964
RMSEA	0.010
AIC	−527.05
CAIC	−8078.07
BIC	−7073.57

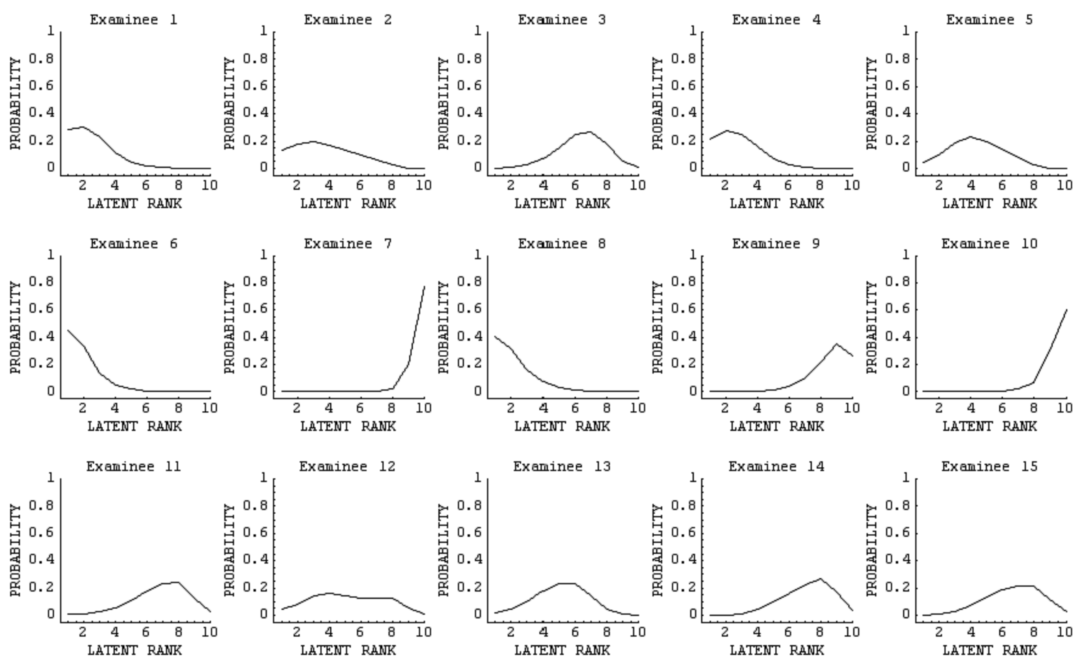
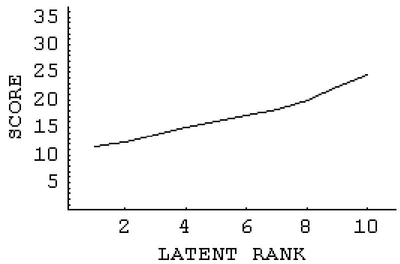


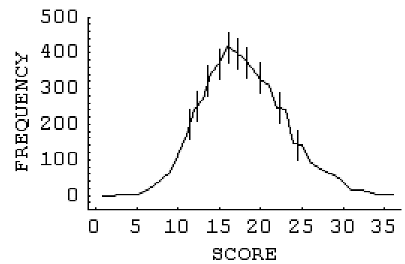
Figure 3: Rank Membership Profiles of Examinees 1–15 ($Q = 10$, Fixed Smoothing Matrix)

about each examinee's ability level.

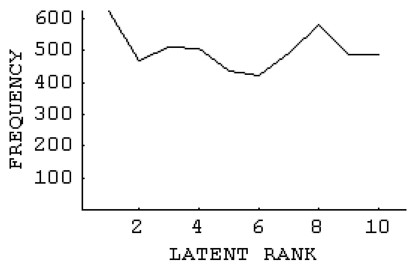
Some additional information obtained by the NTT analysis is shown in Figure 4. Figure 4(a) is the test reference profile (TRP; Shojima, 2008a, 2008b), which is the weighted sum of the IRPs and represents the expected test scores of the examinees belonging to the respective latent ranks. For example, the expected score of the examinees belonging to R_6 was found to be around 17 for this geography test. In the analysis of Example 1, the weakly ordinal



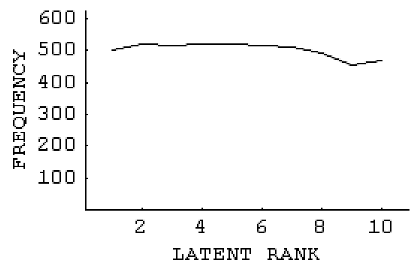
(a) Test Reference Profile (TRP)



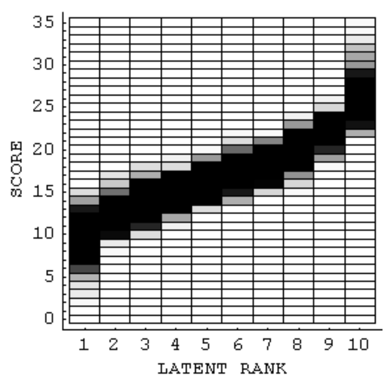
(b) TRP on Score Distribution



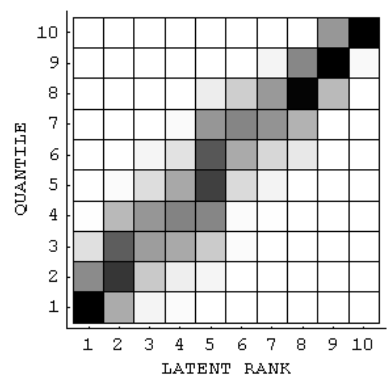
(c) Latent Rank Distribution (LRD)



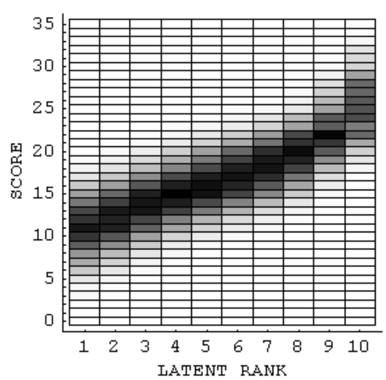
(d) Rank Membership Distribution (RMD)



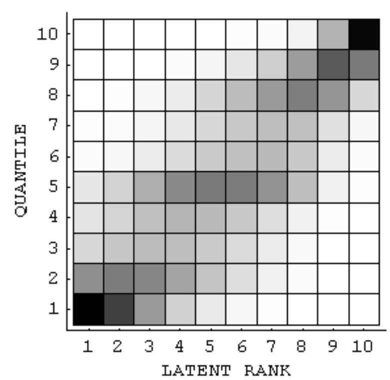
(e) Rank-Score Scatter Plot



(f) Rank-Quantile Scatter Plot



(g) Membership-Score Scatter Plot



(h) Membership-Quantile Scatter Plot

Figure 4: TRP, LRD, RMD, and Scatter Plots ($Q = 10$, Fixed Smoothing Matrix)

alignment condition was satisfied because the TRP monotonically increased, although not every IRP was monotonic. Therefore, the latent scale obtaied by the analysis was verified to be ordinal in terms of the TRP. If the TRP is not monotonic, the problem can be solved by imposing the MIC on some IRPs or using a smoothing matrix with larger diagonal elements. In addition, Figure 4(b) shows the TRP ticked on the number-right score distribution.

Figure 4(c) is the latent rank distribution (LRD; Shojima, 2008a, 2008b), which represents the frequencies of the latent rank estimates of the examinees. Furthermore, Figure 4(d) is the rank membership distribution (RMD; Shojima, 2008b), which is the simple sum of the RMPs of all the examinees. The LRD expresses the distribution of the latent rank estimates of the sample, while the RMD expresses that of the population.

Figure 4(e) is a scatter plot of the latent rank estimates and the number right scores of the examinees, where the darker area represents a higher frequency. In addition, we found from the scatter plot that the latent ranks of the examinees with the same number-right scores were not always identical. However, the Spearman's rank correlation coefficient between the ranks and the scores was 0.952, so the abilities measured by the latent rank scale and the number-right score scale were not totally different. The largeness of the coefficient shows a certain kind of validity for the latent rank scale. In addition, Figure 4(f) shows the scatter plot of the latent rank estimates and the decile scores (10 percentile scores), and the rank correlation coefficient between them was 0.948. Furthermore, Figures 4(g) and 4(h) are the RMDs stratified by the number-right scores and the decile scores, respectively, and they represent the characteristics of the scatter plot about the population, while Figures 4(e) and 4(f) show those of the sample.

3.2 Example 2: Result with $Q = 10$ and Free Smoothing Matrix

As explained in Section 2.4, the IRPs in this example were estimated with estimating the optimal smoothing matrix. The objective function was selected to be the BIC, so the constant used in Equation (33) was $k = \ln N$. The number of elements in the linear filter was three, and the linear filter estimate was then obtained to be $\hat{\boldsymbol{\phi}}_3 = \{0.039, 0.921, 0.039\}$. The number of EM cycles until convergence was eight.

Here, we selectively focus on some important results compared with Example 1. First, the IRP estimates of the 35 items are shown in Figure 5. It is evident from the linear filter estimate that these IRPs are not smoother than those in Figure 2. Each RRE estimate in the IRP kept higher independency because the EDF was smaller. That is, no adjacent RREs of the IRP in Example 2 were closer than those in Example 1.

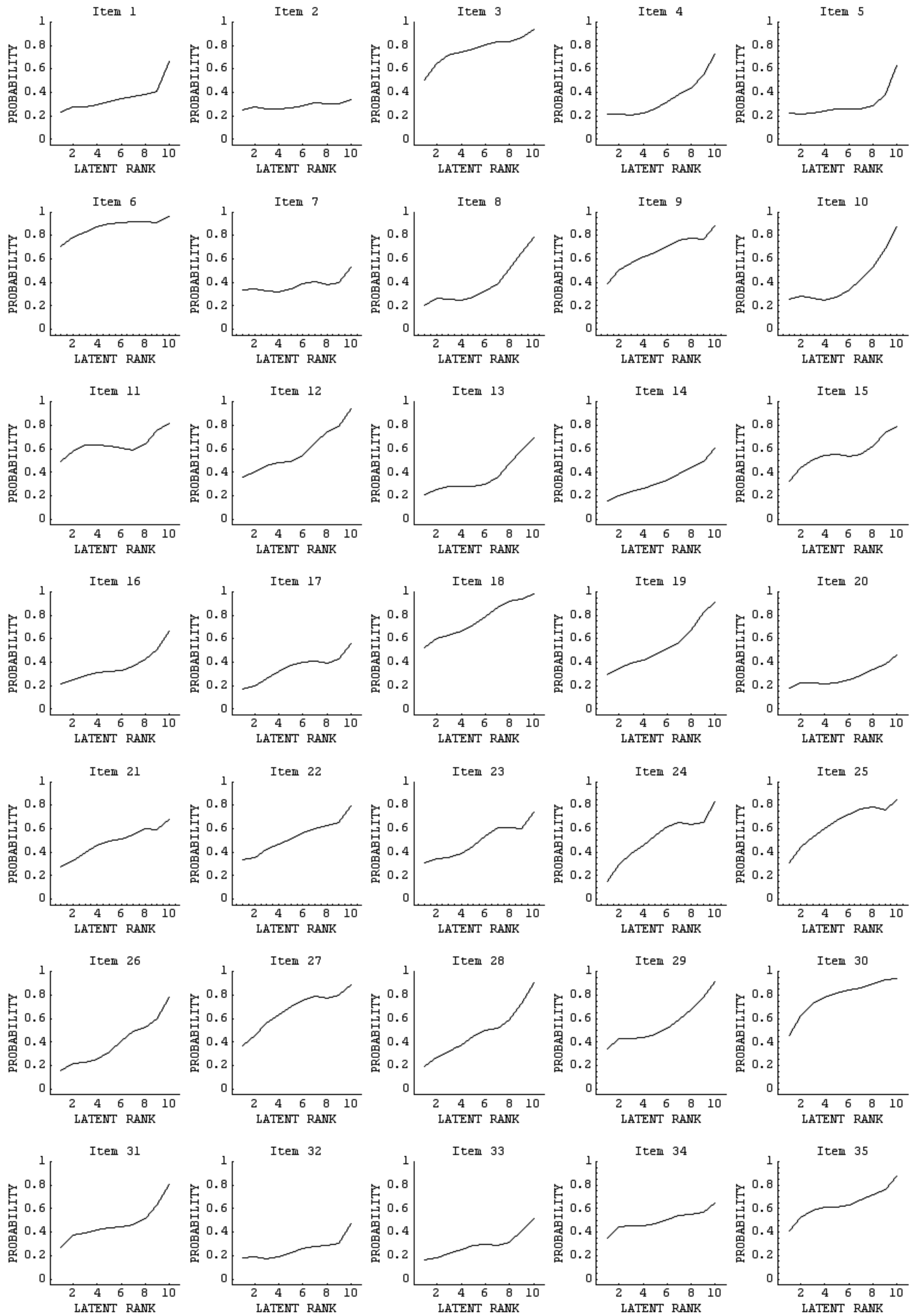


Figure 5: Item Reference Profile ($Q = 10$, Free Smoothing Matrix)

The model-fit indices for the whole test are listed in Table 5. The EDF of the model was 899.86 (= 25.71×35), where 25.71 (= $35 - (0.921 \times 8 + 0.921/(0.921 + 0.039) \times 2)$ is the EDF of each item. From these indices, the model-fit of Example 2 with the estimated smoothing matrix is clearly better than that of Example 1 with the fixed smoothing matrix, although it is likely that the model of Example 2 overfits the data, as seen from its IRPs. In addition, the linear filter estimate was obtained to be $\{1/3, 1/3, 1/3\}$ when the CAIC was used as the objective function. Accordingly, one important future task is to make an adequate constant k for Equation (33) defined within the range $\ln N < k < \ln N + 1$.

Table 5: Test Fit Indices ($Q = 10$, Free Smoothing Matrix)

Index	Value
$\chi^2_{899.86}$	559.88
NFI	0.961
RFI	0.949
IFI	1.000
TLI	1.000
CFI	1.000
RMSEA	0.000
AIC	−1239.84
CAIC	−8004.26
BIC	−7104.40

3.3 Example 3: Result with $Q = 5$ and Fixed Smoothing Matrix

The result with $Q = 5$ is shown in this subsection. The number of latent ranks is up to test administrator or teacher. If the test administrator wants to grade the examinees or the students into Excellent, Very Good, Good, Below Average, and Needs Improvement, he/she should analyze the data under the model with $Q = 5$.

The applied linear filter was $\{0.1, 0.8, 0.1\}$ from Equation (21). The number of EM cycles required for convergence was seven. The estimated IRPs are shown in Figure 6. The IRPs are inclined to monotonically increase as the number of latent ranks is smaller, although the shape of each IRP is basically similar to the IRPs of the model with $Q = 10$ in Example 1 (Figure 2). In addition, the TRP, LRD, RMD, and the scatter plots are shown in Figure 7. The LRD (Figure 7(d)) was almost flat, so the equiprobability scale can be said to be generated. In addition, Figure 7(f) is the scatter plot of the latent rank estimates and the 20 percentile scores.

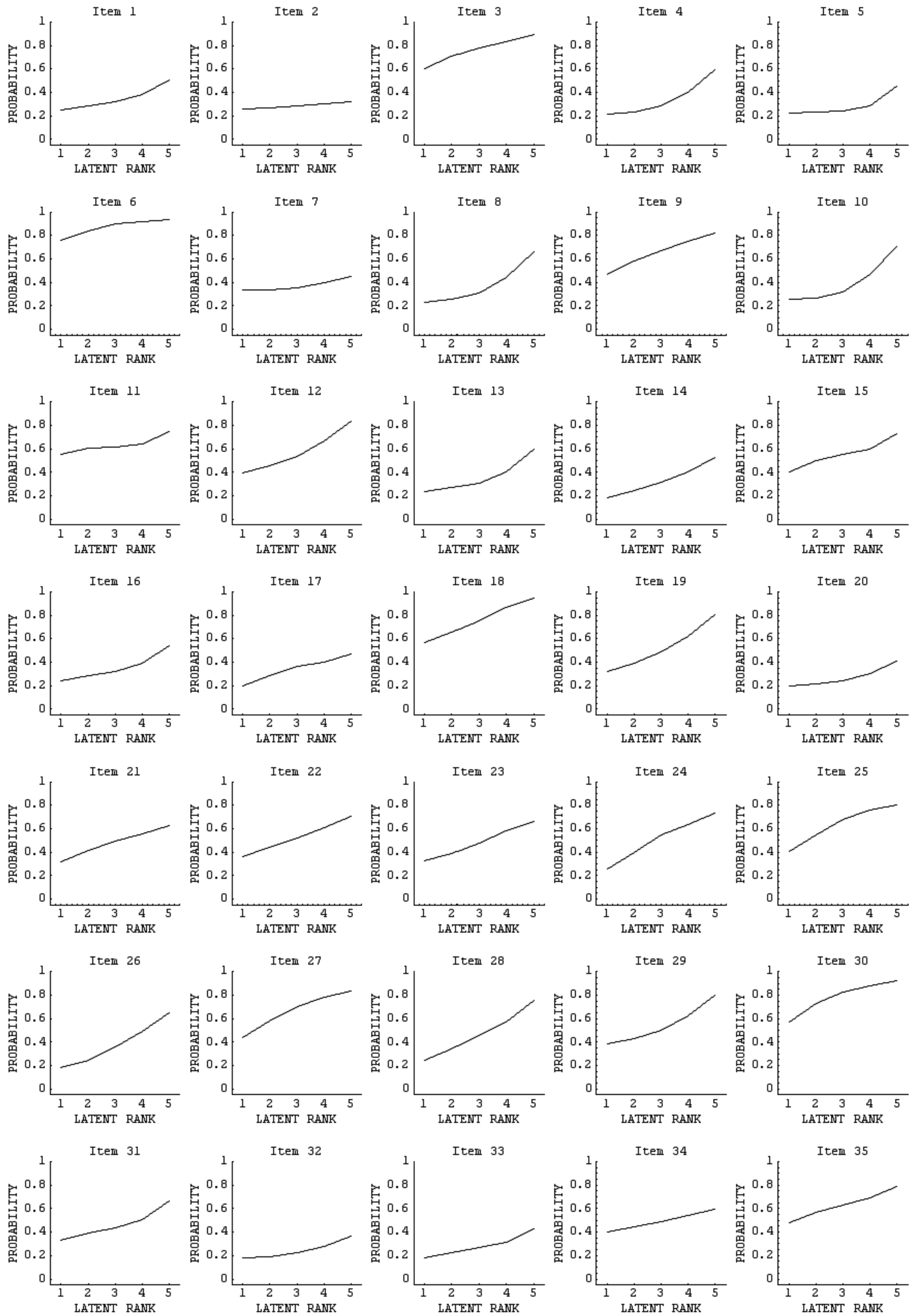
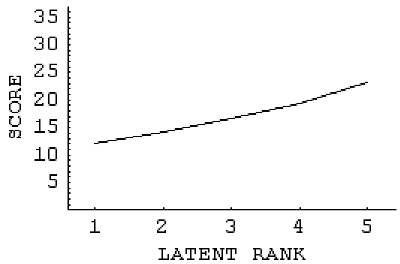
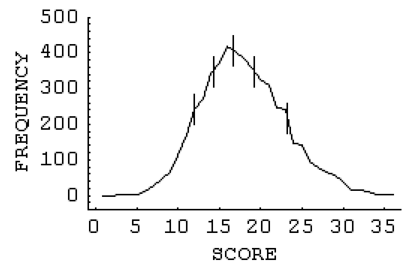


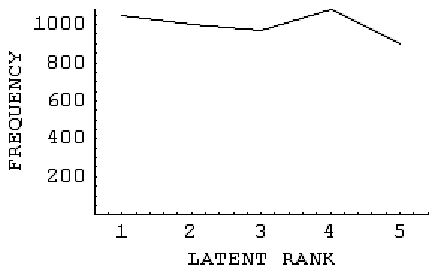
Figure 6: Item Reference Profiles ($Q = 5$, Fixed Smoothing Matrix)



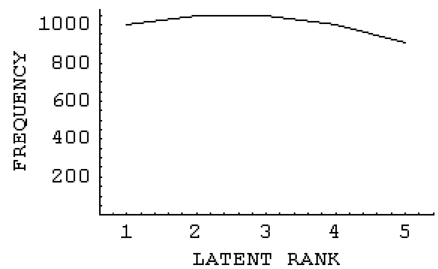
(a) Test Reference Profile (TRP)



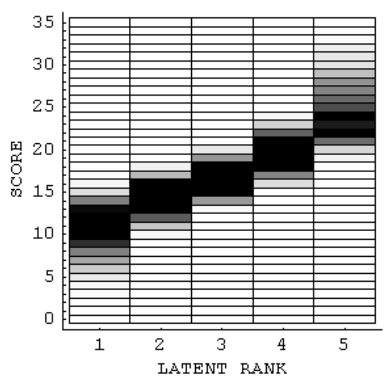
(b) TRP on Score Distribution



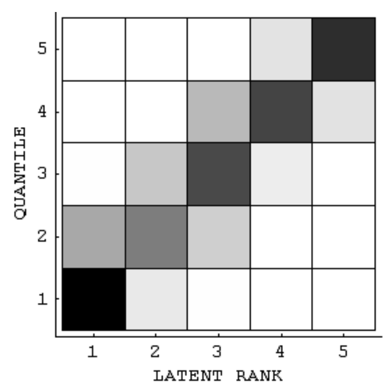
(c) Latent Rank Distribution (LRD)



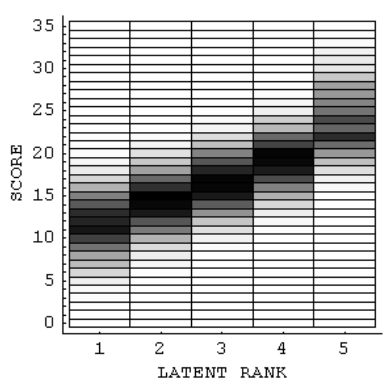
(d) Rank Membership Distribution (RMD)



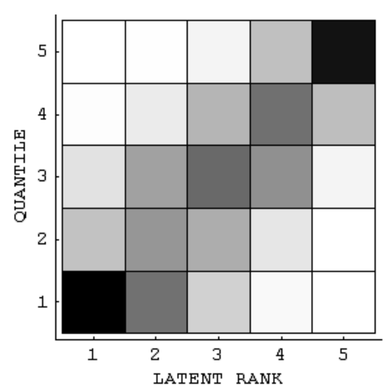
(e) Rank-Score Scatter Plot



(f) Rank-Quantile Scatter Plot



(g) Membership-Score Scatter Plot



(h) Membership-Quantile Scatter Plot

Figure 7: TRP, LRD, RMD, and Scatter Plots ($Q = 10$, Fixed Smoothing Matrix)

The model-fit indices of the whole test are given in Table 6. They are generally satisfactory. These indices serve as a useful reference in determining the number of latent ranks. Comparing the indices of Example 1 with those of Example 3, it is clear from all the listed indices that the model of Example 1 fits the data better than that of Example 3. Accordingly, the model with $Q = 10$ is better than the model with $Q = 5$. However, the number of latent ranks should be determined not only by the statistical indices, but also by the usage of the test. For example, test data should be analyzed by the model with $Q = 3$ even if many model-fit indices of the model with $Q = 3$ are unsatisfactory, when the test is a placement test for the purpose of roughly grading enrollees into three classes.

Table 6: Test Fit Indices ($Q = 5$, Fixed Smoothing Matrix)

Index	Value
$\chi^2_{1078.78}$	1978.76
NFI	0.863
RFI	0.848
IFI	0.932
TLI	0.925
CFI	0.932
RMSEA	0.013
AIC	−178.80
CAIC	−8288.18
BIC	−7209.40

3.4 Example 4: Result with $Q = 5$ and Free Smoothing Matrix

The last example is the result analyzed by the model with $Q = 5$ where the elements in the linear filter were treated as free parameters. The number of elements in the linear filter was three and the BIC was used as the objective function. The linear filter estimate was $\{0.016, 0.968, 0.016\}$. The number of EM cycles until convergence was nine.

As shown in Figure 8, the IRPs of the 35 items were hardly smooth. This is because the central element in the linear filter estimate was almost 1.0. In this case, the smoothing matrix became almost equal to an identity matrix, which leads that the elastic RMI becomes almost identical to the RMI ($\boldsymbol{E} \fallingdotseq \boldsymbol{F}$). Accordingly, the RREs of each IRP were obtained almost independently. The smoothness of the IRP is attained by each RRE referring to the neighboring RREs. Therefore, the IRPs are not obtained to be smooth when the smoothing

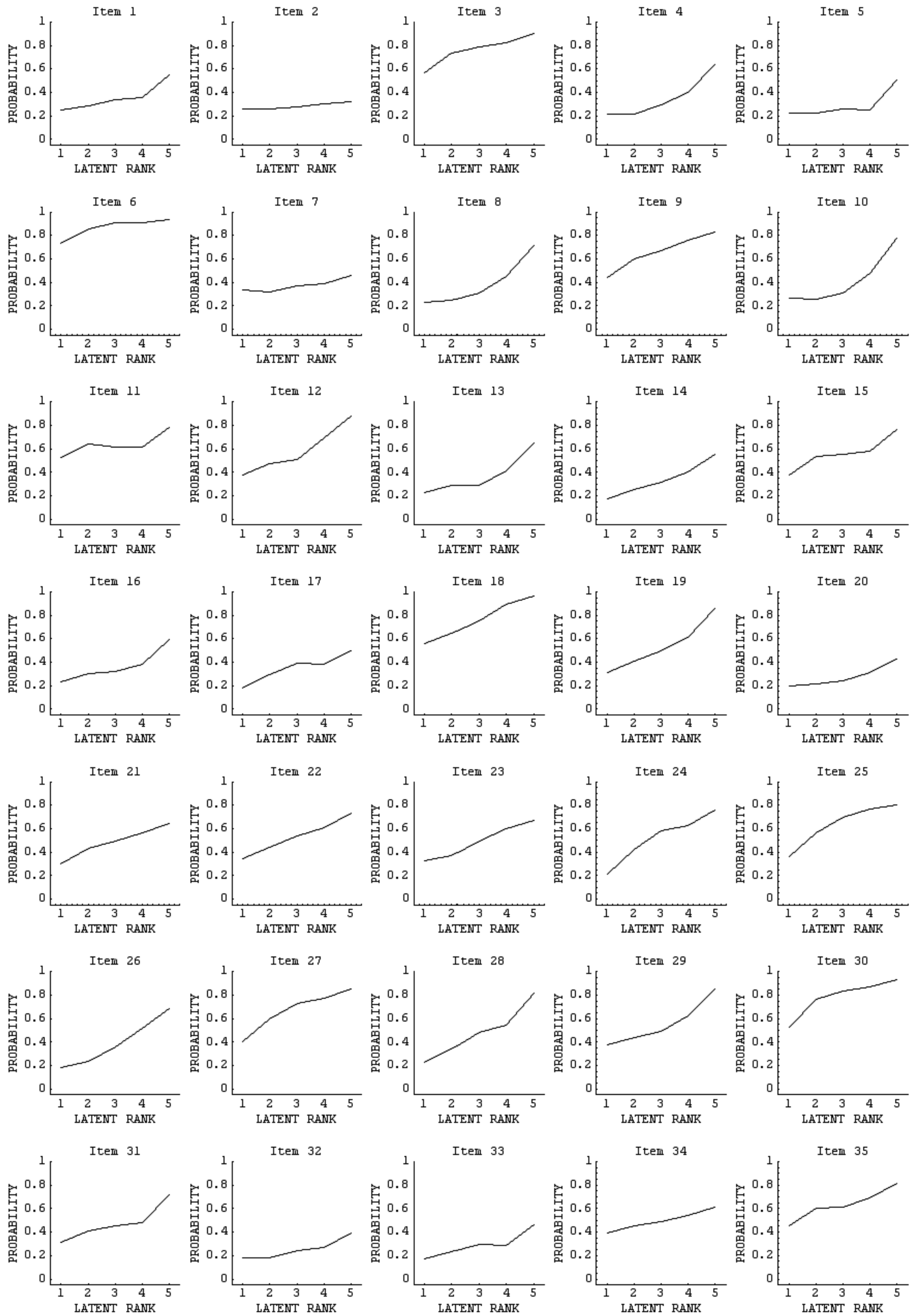


Figure 8: Item Reference Profile ($Q = 5$, Free Smoothing Matrix)

matrix is almost an identity matrix because the adjacent RREs do not become close to one another.

The goodness-of-fit indices of the model for the whole test are listed in Table 7. They were basically satisfactory. However, the IRPs of this model were not very smooth, as seen from Figure 8. Therefore, it might be better to select the models of Examples 1 and 3, that is the models with a fixed linear filter, from a comprehensive standpoint. This comfirms the statement made in Example 2 that a more appropriate constant k needs to be set.

Table 7: Test Fit Indices ($Q = 5$, Free Smoothing Matrix)

Index	Value
$\chi^2_{1054.49}$	1109.88
NFI	0.923
RFI	0.913
IFI	0.996
TLI	0.995
CFI	0.996
RMSEA	0.003
AIC	−999.10
CAIC	−8925.87
BIC	−7871.39

4 Discussion

A batch-type learning version of the NTT model was proposed in this study, where the mechanism of the GTM was applied to the statistical learning process and the smoothing method was also incorporated into the process. The nature of the NTT model with the SOM mechanism (Shojima, 2008a, 2008b) is such that the estimation result is slightly different in every calculation even if the parameter setting is identical. The batch-type NTT model is useful for someone who dislikes this nature. In addition, an advantage of the batch-type model is that the computation time required for identifying the model is much shorted than that for the NTT model with the SOM mechanism.

Furthermore, another advantage of the batch-type NTT model is that the degrees of freedom can be computed to evaluate the smoothness of the model. In the NTT model with the SOM mechanism, the number of degrees of freedom per item was $df_j = n - Q$. Although the number of degrees of freedom should be larger as the model becomes smoother, it was always $n - Q$ in the NTT model with the SOM mechanism because it was difficult to evaluate the degree of smoothness generated by the estimation process in the SOM mechanism. Therefore, Shojima (2008b) simply defined the difference between the number of parameters of the benchmark model and that of the present model as the degrees of freedom of the present model. However, in the batch-type NTT model, the model-fit could be evaluated taking into account the smoothness and flexibility of the model by defining the trace of the smoothing matrix as the effective degrees of freedom.

Furthermore, two estimation methods were shown when the smoothing matrix was fixed or free. A method for estimating the smoothing matrix by minimizing an information criterion was proposed. In four examples, the solutions with the fixed smoothing matrix were preferable to those with the free smoothing matrix. One of the future tasks must be how to select the information criterion to be optimized.

References

Akaike, H. (1987) Factor analysis and AIC. *Psychometrika*, **52**, 317-332.

Bentler, P. M. (1990) Comparative fit indexes in structural models. *Psychological Bulletin*, **107**, 238-246.

Bentler, P. M. & Bonett, D. G. (1980) Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, **88**, 588-606.

Bishop, C. M., Svensen, M., & Williams, C. K. I. (1998) GTM: The generative topographic mapping. *Neural Computation*, **10**, 215-234.

Bollen, K. A. (1986) Sample size and Bentler and Bonnet's nonnormed fit index. *Psychometrika*, **51**, 375-377.

Bollen, K. A. (1989) A new incremental fit index for general structural equation models. *Sociological Methods & Research*, **17**, 303-316.

Bozdogan, H. (1987) Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, **52**, 345-370.

Browne, M. W. & Cudeck, R. (1993) Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.) *Testing structural equation models*. Sage. (pp. 136-162).

Cronbach, L. J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika*, **16**, 297-334.

Croon, M. (2002) Ordering the classes. In J. A. Hagenaars & A. L. McCutcheon (Eds.) *Applied latent class analysis*. Cambridge University Press. (pp. 137-162).

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.

Eubank, R. L. (1999) *Nonparametric regression and spline smoothing*. Marcel Dekker Inc.

Green, P. J. & Silverman, B. W. (1994) *Nonparametric regression and generalized linear models*. Chapman & Hall, London.

Hambleton, R. K. & Swaminathan, H. (1985) *Item response theory*. Kluwer-Nijhoff.

Hastie, T., Tibshirani, R. & Friedman, J. (2001) *The elements of statistical learning:* . Springer-Verlag.

Kohonen, T. (1995) *Self-organizing maps*. Springer.

Loehlin, J. C. (2003) *Latent variable models: An introduction to factor, path, and structural equation analysis*. Lawrence Erlbaum.

Lord, F. M. (1980) *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates.

McLachlan, G. J. & Peel, D. (2000) *Finite mixture models*. Wiley.

Ramsay, J. O. & Silverman, B. W. (1997) *Functional data analysis*. Springer-Verlag.

Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461-464.

Shojima, K. (2008a) Neural test theory. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.) *New trends in psychometrics*. Universal Academy Press, Inc. (in press).

Shojima, K. (2008b) Neural test theory: A latent rank theory for analyzing test data. *DNC Research Note*, 08-01.

Titterington, D., Smith, A., & Makov, U. (1985) *Statistical analysis of finate mixture distributions*. John Wiley & Sons.