The Graded Neural Test Model: A Neural Test Model for Ordered Polytomous Data

SHOJIMA Kojiro

September 2007

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

The graded neural test model: A neural test model for ordered polytomous data

Kojiro Shojima

Abstract

Neural test theory (NTT; Shojima, 2007) was developed for the purpose of analyzing test data. It is a remodeled one-dimensional self-organizing map (SOM) which was originally a statistical learning technique to cluster samples into several classes. The model presented here is an extension of the NTT model for dichotomous data, and is applicable to polytomously ordered data. We call this the graded neural test (GNT) model. An item category reference profile is obtained for each item category by the GNT model, and it provides an effective way to examine the behavior of the selection ratio of each item category. In addition, the boundary category reference profile (BCRP) expresses the selection ratio of the item category or higher, and a method of imposing a monotonically increasing constraint on the BCRPs is also proposed. Furthermore, we show an example of the analysis of earth science test data.

Key words: neural test theory, neural network, self-organizing map, nonparametric model, polytomous response.

段階ニューラルテストモデル: 多値の順序データのためのニューラルテストモデル

荘島宏二郎

要約

本研究では，多値の順序データに対応するニューラルテスト理論 (neural test theory, NTT; Shojima, 2007) を提案した．NTT は，テストデータを分析するためのモデルであり，1 次元の自己組織化マップを改良した統計的学習モデルである．本研究で提案されるモデルは，2 値データのための NTT モデルの拡張であった．各項目のカテゴリに対する反応確率を見るには，項目カテゴリ参照プロファイル (item category reference profile, ICRP) が有効であった．また，各カテゴリ以上の反応確率を見るには，境界カテゴリ参照プロファイル (boundary category reference profile, BCRP) が便利であるが，BCRP に単調増加制約を課す方法も提案した．また，地学テストのデータの分析例を示した．

キーワード: ニューラルテスト理論，ニューラルネットーワーク，自己組織化マップ，ノンパラメトリックモデル，多値モデル．

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

1 Introduction

A test given to a person is a tool for measuring knowledge or ability, but its resolution is not particularly sensitive because it cannot distinguish between two examinees with nearly equal abilities. In contrast, mechanical means of measurement, such as a weight scale, can detect very slight differences between two objects. Apparently, the reliability of human testing is significantly lower than that of measurement by machine. Therefore, the continuous scale assumed in, for example, item response theory (IRT; e.g., Lord, 1980; Hambleton & Swaminathan, 1985) is not necessary. Instead, preparing a rank-ordered scale for clustering examinees into several grades is satisfactory given the lack of sensitivity and resolution in human testing.

Neural test theory (NTT; Shojima, 2007) is a remodeled one-dimensional self-organizing map (SOM; Kohonen, 1995), which is a neural network model for clustering samples on a two-dimensional lattice, and it is widely used in marketing research. NTT is like a nonparametric and nonlinear principal regression analysis, as is SOM (e.g., Shojima, 2007; Ritter, Martinetz, & Schulten, 1992; Kohonen, 1995; Mulier, & Cherkassky, 1995). The latent scale assumed in NTT is rank-ordered, where neurons connected like a chain stand for the latent ranks, and examinees are located on the latent rank scale according to their ability levels. The item reference profile (IRP; Shojima, 2007) is useful for examining the behavior of each item's correct answer rate at each latent rank.

The NTT model proposed by Shojima (2007) is applicable to binary data (true/false items). That is, the Shojima (2007) model can be called a dichotomous neural test model. However, tests sometimes contains testlets, which are usually composed of a few or several small questions, multiple-choice multiple-answer items which are not single-answer items, and Likert-type items in which people are asked to select the closest remark among, for example, a five-point agree-disagree scale. Although polytomous data can be nominal and rank-ordered, in this study we propose a polytomous neural test model for rank-ordered polytomous data; we call this the graded neural test (GNT) model. The GNT model is a natural extension of the dichotomous NTT model, so it is reduced to the dichotomous model when all items are binary.

2 Method

Let us assume that the sample size is N, and the response data of the examinees is

$$\boldsymbol{X} = \{x_{ij} | x_{ij} \in \{0, \cdots, C_j - 1\}\} \quad (N \times n), \tag{1}$$

where x_{ij} is the response of sample i to item j. Let us also suppose that the number of categories of item j is C_j, and the examinees with the upper latent rank select the larger category. From $\boldsymbol{X}$, we can then obtain

$$\boldsymbol{U} = \{\boldsymbol{u}_i\} \quad (i = 1, \cdots, N), \tag{2}$$

where

$$\boldsymbol{u}_i = [\boldsymbol{u}'_{i1} \cdots \boldsymbol{u}'_{in}]' \quad \Big(\Big(\sum_{j=1}^{n} C_j - n\Big) \times 1\Big), \tag{3}$$

$$\boldsymbol{u}_{ij} = \{u_{ijk} | u_{ijk} \in \{0, 1\}\} \quad ((C_j - 1) \times 1), \tag{4}$$

and

$$u_{ijk} = \begin{cases} 1, \text{ If } x_{ij} \geq k \ (k = 1, \cdots, C_j - 1) \\ 0, \text{ otherwise} \end{cases} . \tag{5}$$

Let us further assume that the number of latent ranks is Q, the nodes which represent the latent ranks are R_1, $\cdots$, R_Q, and Node R_q $(q = 1, \cdots, Q)$ has a reference vector as follows:

$$\boldsymbol{v}_q = [\boldsymbol{v}'_{q1} \cdots \boldsymbol{v}'_{qn}]' \quad \Big(\Big(\sum_{j=1}^{n} C_j - n\Big) \times 1\Big). \tag{6}$$

The latent scale assumed in the graded neural test (GNT) model can be graphically expressed as shown in Figure 1. The big black circles are the neurons or nodes which represent the latent ranks, and the number of latent ranks in the figure is seven. Small dark circles on each node stand for the reference vector. The number of items in the figure is five. Each reference vector has $\sum_j (C_j - 1)$ elements. This is because, for example, Item 1 has three categories, so the number of elements corresponding to Item 1 in the reference vector is two. Similarly, there is only one element for Item 2 in each reference vector because Item 2 is binary. Therefore, in the figure, the total number of elements in each reference vector is 11 because the sum of categories of the five items is 16.

The essence of NTT modeling is the process of the statistical learning (e.g., Hastie, Tibshirani, & Friedman, 2001) of the reference matrix $\boldsymbol{V} = \{\boldsymbol{v}_q\}$ $(Q \times (\sum_j C_j - n))$. Following

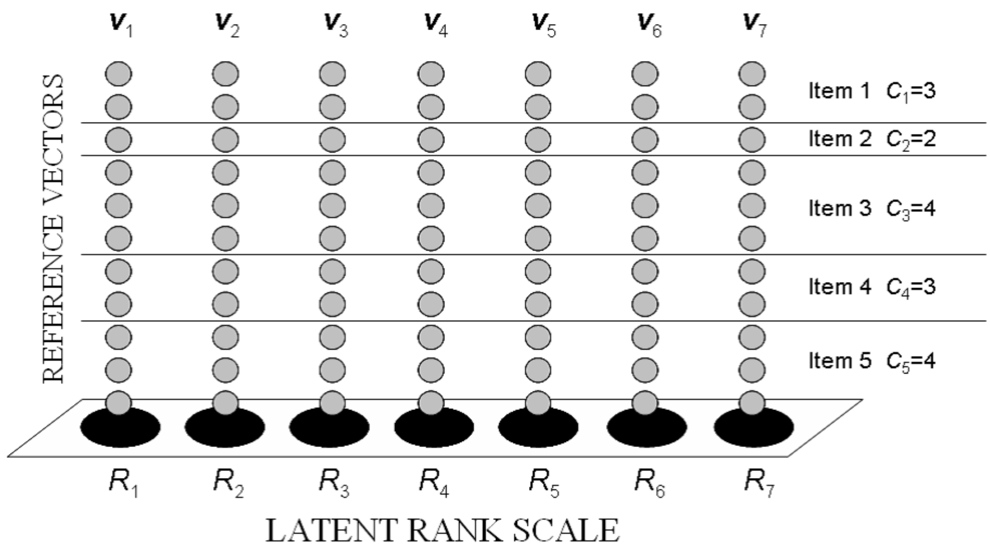


Figure 1: Latent Scale of Polytomous NTT Model

the statistical learning procedure of the dichotomous NTT model, the learning procedure of the GNT model becomes as follows:

For (t=1; $t \leq T$; $t = t + 1$) (L1)

– $\boldsymbol{U}^{(t)} \Leftarrow$Randomly sort the row vectors of $\boldsymbol{U}$. (L2)

For (h=1; $h \leq N$; $h = h + 1$) (L3)

— Select the winner for $\boldsymbol{u}_h^{(t)}$ by d. (L4)

— Obtain $\boldsymbol{V}_h^{(t)}$ by updating $\boldsymbol{V}_{h-1}^{(t)}$. (L5)

— $\boldsymbol{V}^{(t+1)} \Leftarrow \boldsymbol{V}_N^{(t)}$. (L6)

The above procedure can be outlined as follows. Lines (L4) and (L5) make up a routine which is repeatedly applied when the value of h is from 1 to N, and then t adds 1 after (L6) is executed. This repetition continues until t approaches T. As for details, randomly sorting the row vectors of $\boldsymbol{U}$ in Line (L2) is needed to cancel the systematic effect of the data input order. As to the input data in Line (L4),

$$\boldsymbol{u}_h^{(t)} = [\boldsymbol{u}_{h1}^{(t)\prime} \cdots \boldsymbol{u}_{hn}^{(t)\prime}]' \quad \left(\left(\sum_{j=1}^{n} C_j - n\right) \times 1\right), \tag{7}$$

which is the h-th row vector of $\boldsymbol{U}^{(t)}$, the node with the closest reference vector to $\boldsymbol{u}_h^{(t)}$ is selected as the “winner” in terms of the discrepancy function d. As in dichotomous NTT and conventional SOM applications, the square of the Euclidian distance is recommended as the descrepancy function. That is,

$$R_w: \ w = \arg\min_{q \in Q} ||\boldsymbol{v}_q^{(t)} - \boldsymbol{u}_h^{(t)}||^2. \tag{L4a}$$

However, (L4a) is strongly affected by items with large numbers of categories. Therefore,

$$R_w: \ w = \arg\min_{q \in Q} \sum_{j=1}^{n} \frac{||\boldsymbol{v}_{qj}^{(t)} - \boldsymbol{u}_{hj}^{(t)}||^2}{C_j - 1} \tag{L4b}$$

is desirable as the discrepancy function for the polytomous NTT model. Equation (L4b) becomes identical to the discrepancy function of the dichotomous NTT model described by Shojima (2007) when all items are binary.

Next, to update the reference vectors in Line (L5), the reference vectors of the nodes that are close to the winner should be more updated to make them numerically closer to the input data. That is,

$$\begin{aligned} &\text{For } (q{=}1;\, q \le Q;\, q = q+1) \\ &— \boldsymbol{v}_{qh}^{(t)} = \boldsymbol{v}_{qh-1}^{(t)} + h_{qw}(t)(\boldsymbol{u}_h^{(t)} - \boldsymbol{v}_{qh-1}^{(t)}), \end{aligned} \tag{L5a}$$

where

$$h_{qw}(t|\alpha_t, \sigma_t^2) = \alpha_t \exp\left\{-\frac{(R_q - R_w)^2}{2\sigma_t^2}\right\}, \tag{8}$$

$$\alpha_t = \frac{T - t + 1}{T}\alpha_1, \tag{9}$$

and

$$\sigma_t = \frac{(T-t)\sigma_1 + (t-1)\sigma_0}{T-1}. \tag{10}$$

Equation (8) is called the “tension” that regulates the updated size of the reference vector of each node which becomes larger as the node is geographically located closer to the winner. Equation (9) is the factor that controls the amount of tension, and it becomes linearly smaller from α_0 to α_0/T as t approaches T. In addition, (10) is the function that determines the propagation range of the statistical learning around the winner, and it becomes smaller from σ_1 to σ_0 as t increases.

Alternatively, we can stop computing the statistical learning when a certain criterion is satisfied. For example,

$$C^{(t)} = \sum_{h=1}^{N} \sum_{j=1}^{n} \frac{||\boldsymbol{u}_{hj}^{(t)} - \boldsymbol{v}_{whj}^{(t)}||^2}{C_j - 1} \tag{11}$$

is a valid candidate, where $\boldsymbol{v}_{wh}^{(t)} = [\boldsymbol{v}_{wh1}^{(t)}{}' \cdots \boldsymbol{v}_{whn}^{(t)}{}']'$ is the winner node for the input data $\boldsymbol{u}_h^{(t)}$. Stopping rules can then be determined as follows:

$$C^{(t)} < C_1, \tag{12}$$

or

$$|C^{(t+1)} - C^{(t)}| < C_2, \tag{13}$$

where C_1 and C_2 are constants as decided by the analyst.

3 Analysis

3.1 Example 1

The analysis of results from an earth science test with the proposed GNT model is examined in this section. The sample size of the test was 1,424, the number of items was 9, and the number of categories for all the items was 4 $(= 0, 1, 2, 3)$. Figure 2 shows the score distribution of the test.

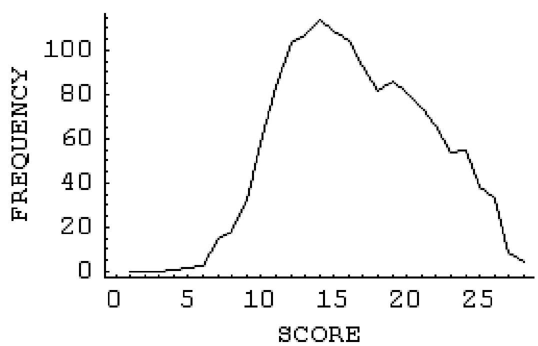


Figure 2: Score Distribution

With the number of latent ranks equal to 10, the parameters prerequisite for analyzing data were set to be $(T, \alpha_1, \sigma_1, \sigma_0) = (500, Q^{-1}, Q, 1.0)$. The $(\sum_{l=1}^{j-1} l + k)$-th column vector (row vector in Figure 1) of the finall reference matrix $\boldsymbol{V}^{(T)}$ is called the boundary category reference profile (BCRP) of item j's category k. That is,

$$\boldsymbol{\rho}_{jk} = [v_{1jk}^{(T)} \cdots v_{Qjk}^{(T)}]' \quad (k = 0, \cdots, C_j;\ j = 1, \cdots, n). \tag{14}$$

Figure 3(a) shows BCRP plots of the nine items. Each BCRP expresses the selection ratio of the item category or higher. For example, the BCRP of category 2 is the ratio of selecting

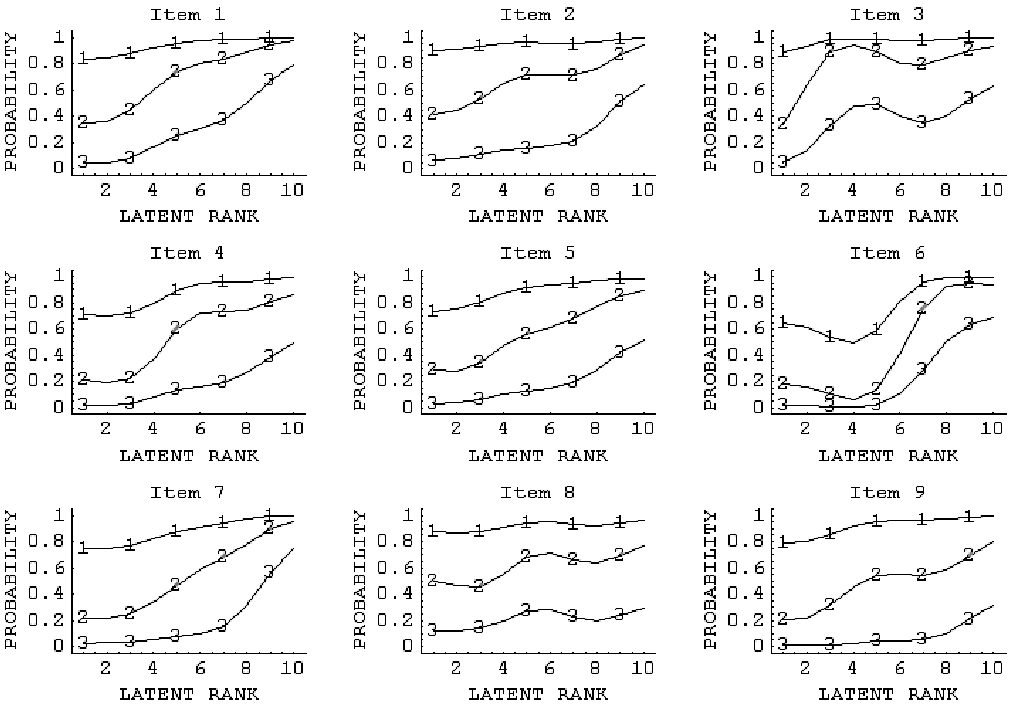


Figure 3: (a) Boundary Category Reference Profiles

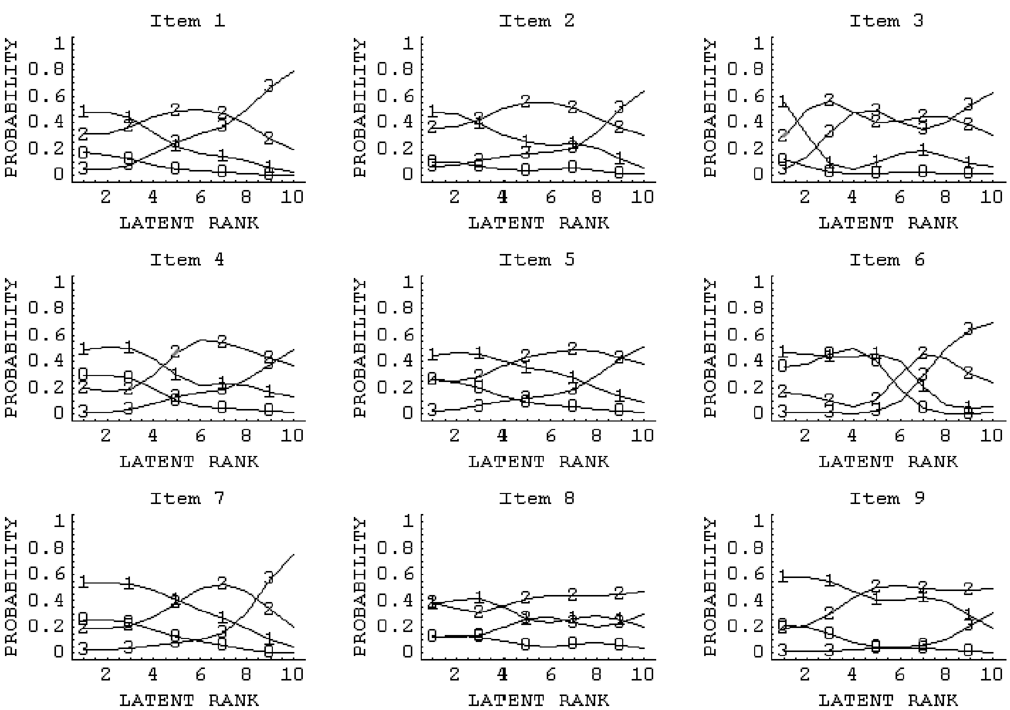
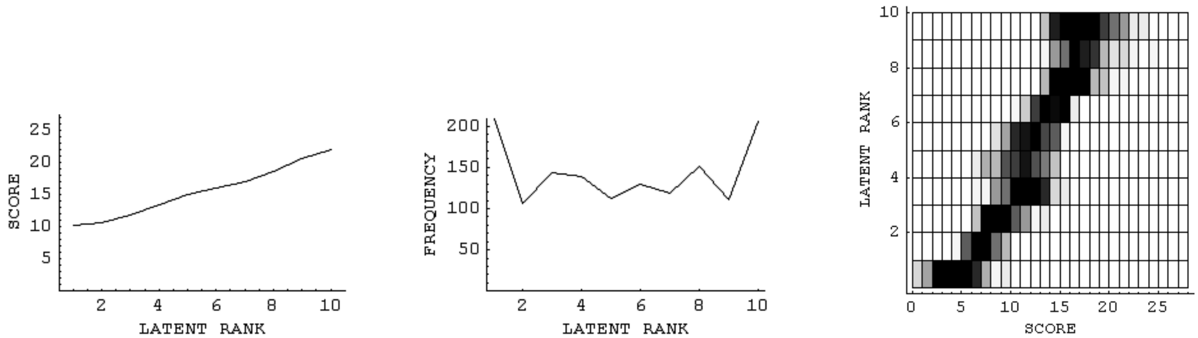


Figure 3: (b) Item Category Reference Profiles



(a) Test Reference Profile (b) Latent Rank Distribution (c) Scatter Plot

Figure 4: TRP, LRD, and Scatter Plot of Scores and Ranks

categories 2, 3, and 4 at each latent rank. Apparently, the BCRP of an upper category never goes across the BCRPs of its lower categories through latent ranks.

For the BCRP of item j's category k, $\boldsymbol{\rho}_{jk}$, we can obtain the following item category reference profile (ICRP). That is,

$$\boldsymbol{\pi}_{jk} = \boldsymbol{\rho}_{jk} - \boldsymbol{\rho}_{jk+1} \quad (k = 0, \cdots, C_j - 1), \tag{15}$$

where

$$\boldsymbol{\rho}_{j0} = \mathbf{1}_Q, \tag{16}$$

and

$$\boldsymbol{\rho}_{jC_j} = \mathbf{0}_Q. \tag{17}$$

Each ICRP tells us the transition of the selection ratio of the relevant item category through the latent ranks. Generally speaking, the upper categories are more frequently chosen by examinees with higher latent ranks.

Figure 4(a) shows the test reference profile (TRP; Shojima, 2007) which is the weighted sum of the ICRPs. That is,

$$\boldsymbol{\tau} = \sum_{j=1}^{n} \sum_{k=1}^{C_j - 1} k \boldsymbol{\pi}_{jk}. \tag{18}$$

The TRP expresses the expected score of the examinees at each latent rank. For instance, the number of correct answers for the examinees at R_6 is slightly over 15. Although not every BCRP is monotonically increasing, the obtained TRP shape is almost exclusively monotonically increasing as observed in Shojima (2007) as well as many SOM applications.

The NTT algorithm is an improvement upon the SOM algorithm, and the mapping executed by the SOM algorithm resembles nonlinear and nonparametric principal component analysis (Ritter, Martinetz, & Schulten, 1992; Kohonen, 1995; Mulier, & Cherkassky, 1995). NTT also has such a feature, which is strong evidence that the latent scale assumed in NTT is rank-ordered. Although some BCRPs are not monotonically increasing, the latent scale is rank-ordered because the obtained TRP is monotonically increasing. Nodes that are simply chained are not arranged in rank-order from the beginning.

The latent rank of each examinee can be calculated by (L4b), and Figure 4(b) shows the latent rank distribution of all examinees. Every test has its own target ability, and the latent ranks of the examinees outside the target ability of the earth science test were estimated to be at both ends of the latent scale. In addition, Figure 4(c) is a scatter plot of the scores and the latent ranks. The darker the area, the higher the density was. We found that the latent ranks of examinees with the same scores were not always the same.

3.2 Example 2

The number of ranks in the latent scale depends on practical circumstances. That is, the number of latent ranks is up to the analyst or the test administrator. As Q becomes larger, the rank-ordered scale becomes more like a continuous one. However, Q may not be over 20 in practical usage because testing reliability is not high in general. In this section, analysis results with $Q = 5$ are shown. The other parameters are set to the same values as in the previous section. Figures 5(a) and 5(b) shows BCRPs and ICRPs, and Figures 6(a), 6(b), and 6(c) show the TRP, the latent rank distribution (LRD), and the scatter plot of the scores and the latent ranks, respectively.

The basic shapes of BCRPs and ICRPs between Figures 3 and 5 are similar, but Figure 5(a) shows that BCRPs tend to become simpler and monotonically increase when Q is smaller, even if the BCRPs do not monotonically increase under $Q = 10$ (e.g., Items 3, 6 and 8 in Figure 3). This tendency was also observed by Shojima (2007). Monotonic BCRPs are useful for test administrators, but it is difficult to precisely grade examinees if Q is too small.

3.3 Example 3

In this section, we look at a method of imposing a monotonically increasing constraint on the BCRPs. As seen in Sections 3.1 and 3.2, BCRPs tend to be nonmonotonic when the number of latent ranks is large. Although nonmonotonic BCRPs represent the real state of

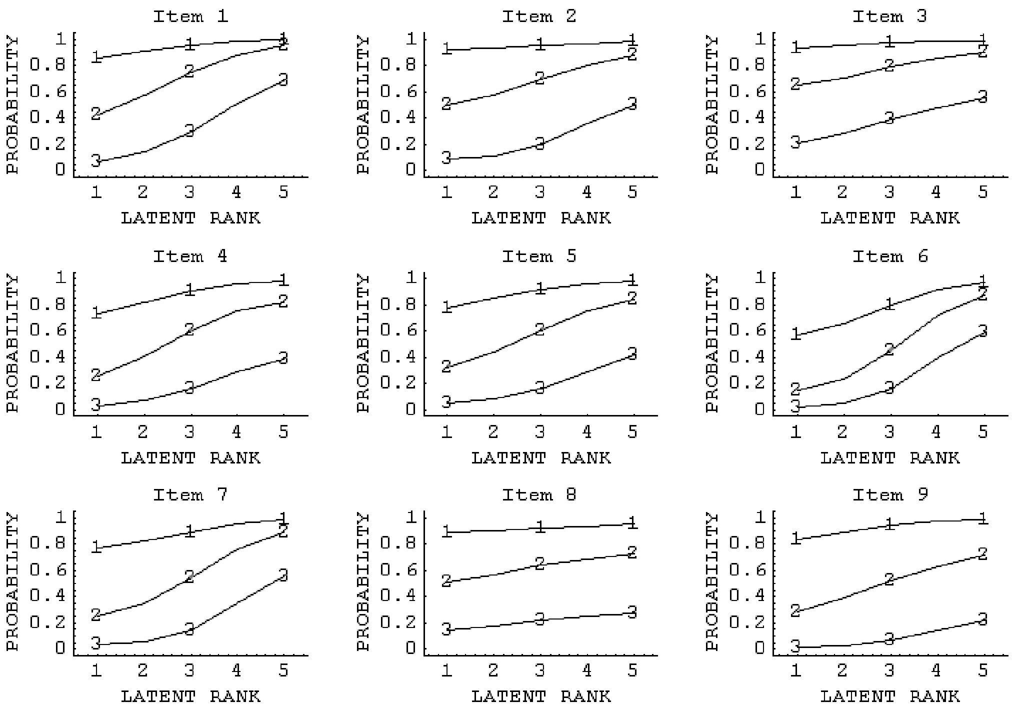


Figure 5: (a) Boundary Category Reference Profiles (Q=5)

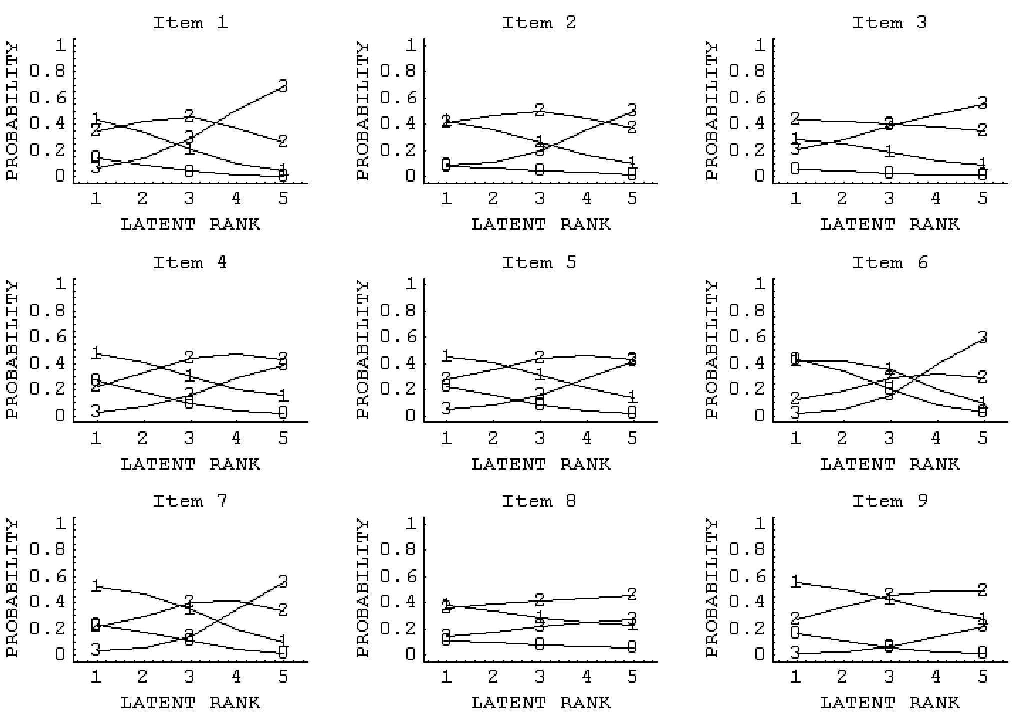
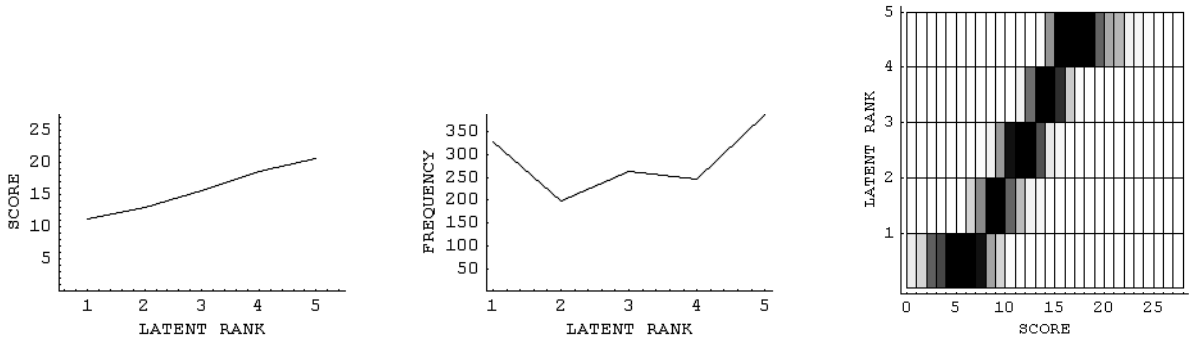


Figure 5: (b) Item Category Reference Profiles (Q=5)



(a) TRP (Q=5) (b) LRD (Q=5) (c) Scatter Plot (Q=5)

Figure 6: TRP, LRD, and Scatter Plot of Scores and Ranks ($Q = 5$)

the data, some test administrators prefer monotonically increasing BCRPs even when Q is not small.

For example, we can simply add the following steps after Line (L6) to make BCRPs monotonic. That is,

$$\begin{aligned} &\text{For } (j{=}1;\ j \le n;\ j = j+1) \\ &\quad \text{For } (k{=}1;\ k \le C_j - 1;\ k = k+1) \\ &\quad\quad \text{For } (q{=}1;\ q \le Q-1;\ q = q+1) \\ &\quad\quad \text{— If } v^{(t+1)}_{q+1,jk} \le v^{(t+1)}_{qjk}, \text{ then } v^{(t+1)}_{q+1,jk} = v^{(t+1)}_{qjk}, \end{aligned} \tag{L5b}$$

or

$$\begin{aligned} &\text{For } (j{=}1;\ j \le n;\ j = j+1) \\ &\text{— Sort}(\boldsymbol{v}^{(t)}_{jk}). \end{aligned} \tag{L5c}$$

With the same parameter setting as in Sections 3.1 and 3.2, we analyzed the earth science test data under $Q = 10$ and constraint (L5c). Figures 7(a) and 7(b) show the BCRP and ICRP results, and the TRP, the latent rank distribution, and a scatter plot of the scores and the latent ranks are shown in Figure 8. As Figure 7(a) shows, each obtained BCRP was monotonically increasing as the latent rank became higher. In rare instances, a BCRP for the upper category could step over the BCRPs of lower categories depending on the circumstances.

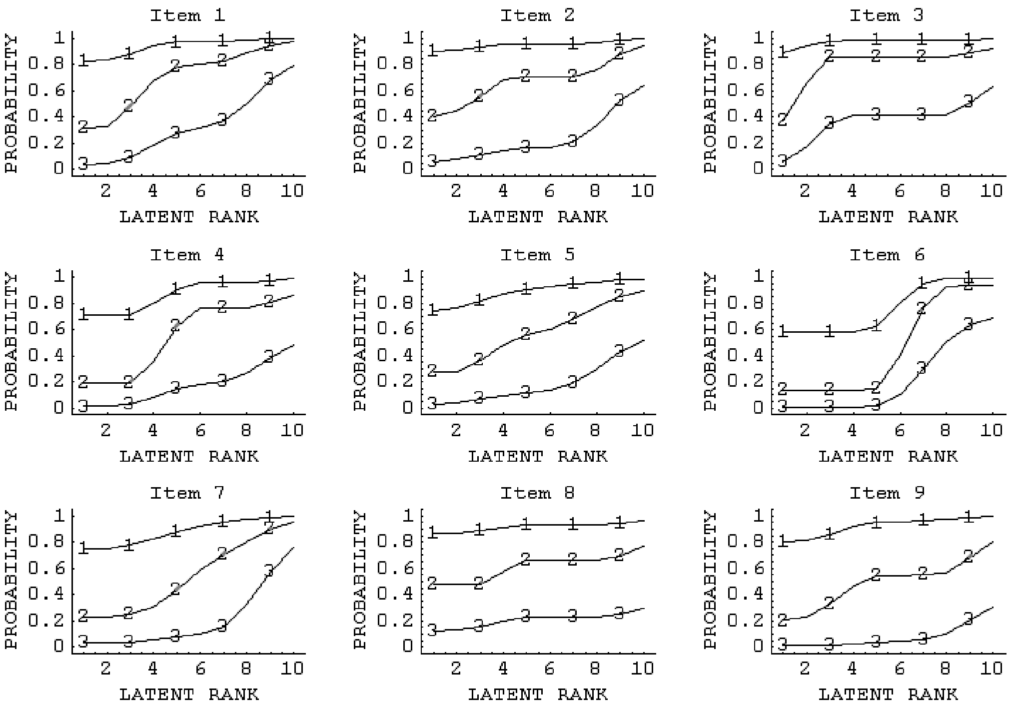


Figure 7: (a) Boundary Category Reference Profiles (Mono. Inc.)

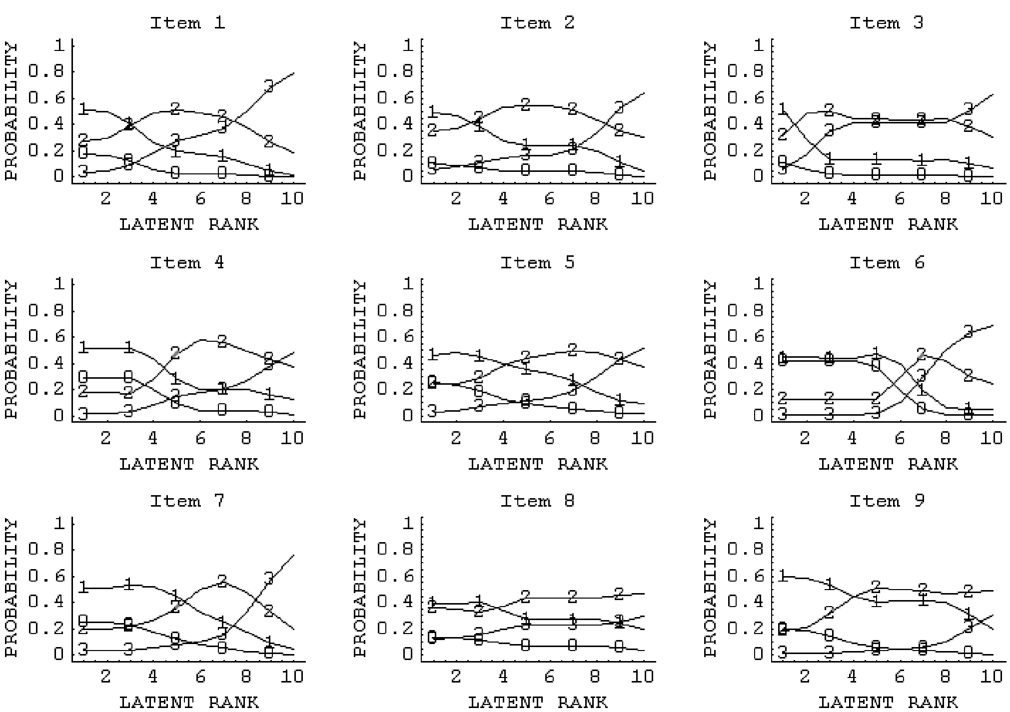
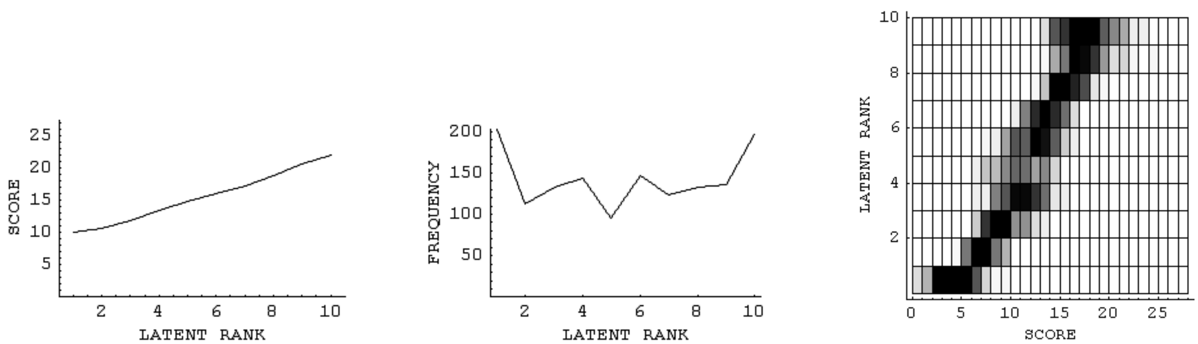


Figure 7: (b) Item Category Reference Profiles (Mono. Inc.)



(a) TRP (Mono. Inc.) (b) LRD (Mono. Inc.) (c) Scatter Plot (Mono. Inc.)

Figure 8: TRP, LRD, and Scatter Plot of Scores and Ranks (Mono. Ince.)

4 Discussion

The proposed graded neural test (GNT) model is a neural test model applicable to polytomous rank-ordered data. The GNT model is reduced to the dichotomous NTT model when all items are binary. The boundary category reference profile (BCRP) is useful for reviewing the selection ratio of a certain item category or higher at each latent rank, while the item category reference profile (ICRP) shows the selection ratio of the item category itself through latent ranks.

In addition, analysts or test administrators can determine the number of latent ranks. BCRPs tend to be simpler and monotonically increasing when the number of latent ranks is small. Furthermore, the monotonically increasing constraint can be imposed on the process of statistical learning to make BCRPs monotonic. Such monotonic BCRPs will be useful in practical situations, although nonmonotonic BCRPs can capture certain types of the real features of a phenomenon.

As future work, a technique is needed to determine the optimal number of latent ranks when Q is not theoretically or practically determined in advance. In addition, a polytomous neural test model which can be applied to nominal-polytomous data is required.

References

Amari, S. (1980) Topographic organization of nerve fields. *Bulletin of Mathematical Biology*, **42**, 339-364.

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.

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

Mulier, F. & Cherkassky, V. (1995) Self-organization as an iterative kernel smoothing process. *Neural Computation*, **7**, 1165-1177.

Ritter, H., Martinetz, T. & Schulten, K. (1992) *Neural computation and self-organizing maps: An introduction*. Addison-Wesley.

Ritter, H. & Schulten, K. (1986) On the stationary state of Kohonen's self-organizing sensory mapping. *Biological Cybernetics*, **54**, 99-106.

Shojima, K. (2007) Neural test theory. *DNC Research Note*, 07-02.