

# 客観式能力テストへの段階反応モデル適用についての検討

○仁田光彦 舛田博之  
株式会社 リクルート HR・斡旋カンパニー 新卒事業本部  
インフローソリューションディビジョン  
ソリューション開発部 測定技術研究所

## 1. 背景と目的

項目反応理論(Item Response Theory, 以下 IRT と記す)を用いたテスト採点においては、2 値項目反応モデルが用いられる場合が多く、日本では2 パラメータロジスティックモデル(以下 2PLM と記す)が用いられるのが一般的である。2PLM は、モデルとして扱いやすい一方で、正答と誤答の 2 値のデータしか利用できないという限界がある。そのため、この 2PLM を発展させたモデルとして、多値項目反応モデルが存在する。多値項目反応モデルでは、正答と誤答の 2 値のデータだけでなく、受検者がどのように間違えたのかという誤答の情報を含んだ反応を扱っているため、2PLM と比較して、多くの情報量が得られることが知られている (Bock, 1972; 平井, 1993)。

多値項目反応モデルにはいくつかの種類が存在するが、多肢択一式の客観式能力テストの採点では、一般的に名義反応モデル (Nominal Response Model, 以下 NRM と記す) が用いられる場合が多い。NRM では、選択肢ごとに選択確率を定義することで、能力レベルごとに選びやすい誤答を知ることができ、能力を測定する上で、より多くの情報を得ることができる(豊田, 2005)。

一方で、NRM はテスト作成者側にとっては、いくつかのデメリットが存在する可能性がある。例えば、テスト作成者は、テスト項目を作成する際、意図を持って正答位置を設定するが、NRM においては、事前に正答位置の情報を加味することが不可能である。また、多値項目反応モデルでは、選択数が少ないとパラメータの推定値が不安定になりやすい特徴があり、さらに NRM ではパラメータ数も多く、モデル自体も複雑であるためその特徴がより顕著に現れやすい。そのため、推定のためにはある程度大量のデータが必要とされるが、実際のテスト開発時には限られたデータしかない場合も多く、NRM への適用が困難なことも多い。

多値項目反応モデルのその他の例としては、段階反応モデル(Graded Response Model, 以下 GRM と記す)が存在する。GRM は、一般的にリッカートタイプの質問項目といった、選択肢(もしくはカテゴリ)の順序が決まっている場合のテスト採点に利用される(豊田, 2005)。GRM は 2PLM と非常に親和性が高いモデルであり、テスト作成者側にとっては、扱いやすいモデルである一方で、能力テストで用いる場合には、事前に選択肢の順序を決めておく必要がある。平井(1993)によれば、能力テストの採点においては、能力推定精度の向上という点では、GRM と NRM に大きな差はなく、よりモデルがシンプルでパラメータ推定の安定した GRM を利用する方が望ましいとされている。

本研究では、多肢択一式の客観式能力テストを研究に用いて、事前の項目分析の情報と作成者側の意図を反映した正答位置を考慮に入れ、選択肢を序列づけすることで、段階反応モデルへの適用を検討する。

上記テストを用いて、2PLM と多値項目反応モデルの結果を比較することで、多

肢択一式の客観式能力テストにおける多値項目反応モデルの活用価値を再確認する。さらに、NRM と GRM の結果を比較することで、多肢択一式の客観式能力テスト採点において、GRM を実運用場面で適用するための 1 つの方法を示したい。

## 2. 方法

### 2-1. 使用テスト

6 肢もしくは 10 肢択一式の客観式テスト、構造的把握力検査(リクルート)を用いた。テストは、33 項目、3 形式で構成されており、1 尺度である。

### 2-3. 対象データ

対象データは、2010 年 2 月～3 月に WEB 上でデータ収集を行った際のデータを使用した。6 肢択一式の形式のみ 15 項目を選び、さらに、欠損値のあるデータを取り除いた 320 名分のデータを使用した。

### 2-3. 選択肢の序列づけ

段階反応モデルを適用するにあたって、事前に選択肢の序列づけを行った。選択肢の序列づけは、各項目の選択データを名義反応として多重対応分析を行い、第 1 成分への成分スコアを項目分析の情報として用いた。第 1 成分はテストの測定対象となる能力に対応した特性を表していると考えられるため、各選択肢の成分スコアは、一次元性を最大化する最適重みづけ係数となる。本研究では、各項目の成分スコアの最大値が、正答位置と同じであったため、この成分スコアを参考にして、誤答選択肢の序列づけを行った。

### 2-4. 分析方法

各モデルでの推定は R 上で動作する MCMC パッケージである Brugs を用いて、2PLM, NRM, GRM のパラメタ推定を行った。

各モデル間の比較をするため、各受検者の  $\theta$  推定値の散布図を確認し、さらに、2PLM と GRM のテスト情報量を比較した。また、測定精度を確認するための指標として、MCMC での  $\theta$  推定値の標準偏差を標準誤差とみなして、受検者全体での平均値を算出し、各モデル間で比較した。

加えて  $\theta$  推定値について NRM と GRM の相関係数を算出し、GRM の実運用場面での適用可能性について確認した。

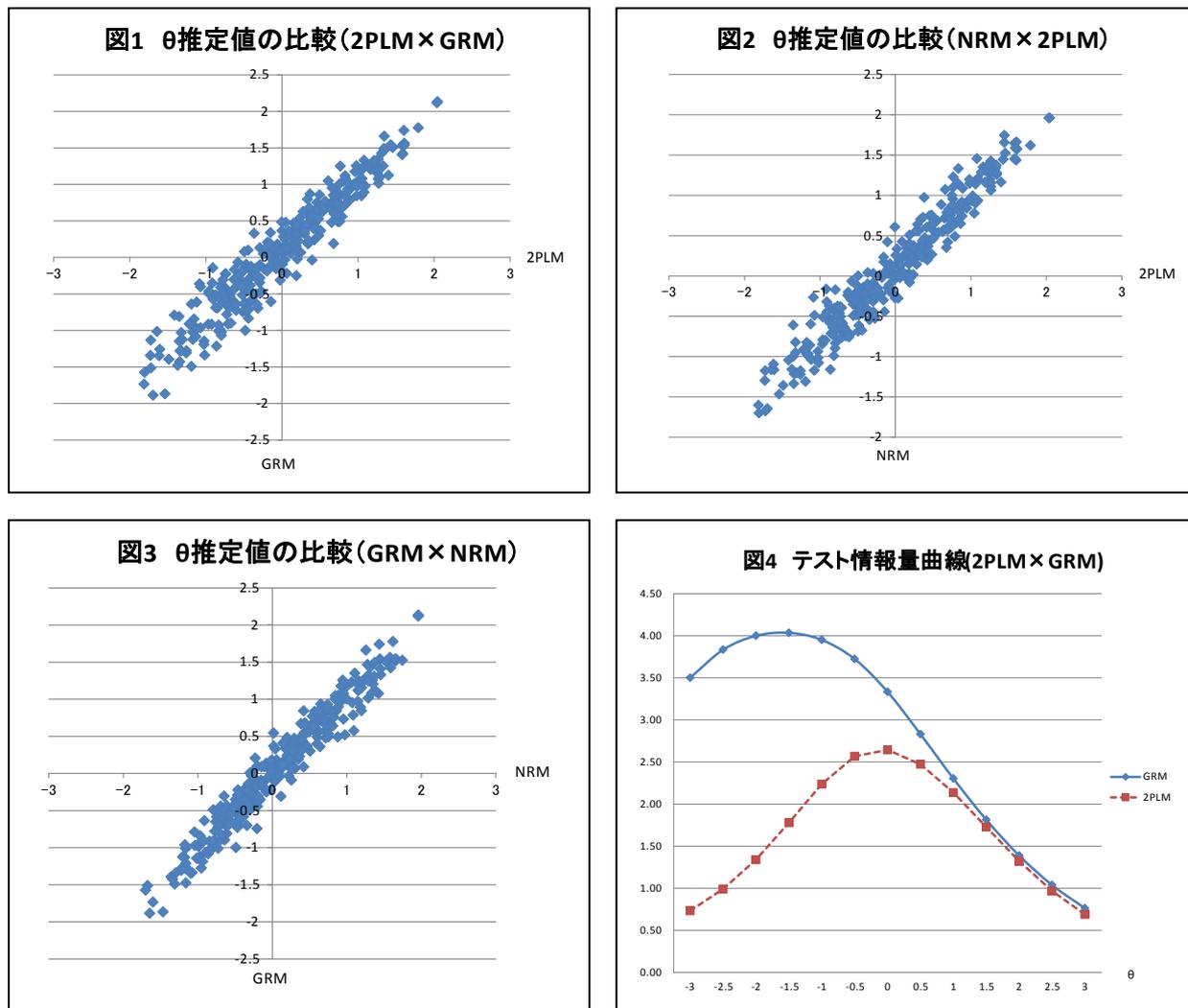
## 3. 結果

各受検者の  $\theta$  推定値の散布図を図 1, 2, 3 に示す。図 3 の GRM と NRM の散布図と比較して、図 1, 2 の 2PLM と多値項目反応モデル(GRM, NRM)の散布図では、 $\theta$  推定値のばらつきが、特に  $\theta$  推定値が低い部分で大きくなっていることが確認できた。

また、図 4 は 2PLM と GRM のテスト情報量曲線を示している。2PLM と比較して GRM では、 $\theta$  推定値の低い部分で、情報量が著しく上昇していた。

さらに、表 1 に各モデルの  $\theta$  推定値の標準誤差の平均を示す。2PLM では、0.564、多値項目反応モデルでは、GRM で 0.508、NRM で 0.495 となっており、多値項目反応モデルで測定精度が高くなっていることを示している。

多値項目反応モデル間の比較である，GRM と NRM の  $\theta$  推定値の相関係数は 0.976 であった。



**表1 標準誤差の平均**

	2PLM	GRM	NRM
標準偏差の平均	0.564	0.508	0.495

#### 4. 考察

図 1, 2 の散布図を確認すると，2PLM と多値項目反応モデルの間では  $\theta$  の低い層でばらつきがみられた。表 1 より各モデル間の標準誤差は，多値項目反応モデル (GRM=0.508, NRM=0.495) が 2PLM(0.564) より小さく，測定精度が高いことが確認できた。この 2PLM と多値項目反応モデル間の標準誤差の差は，散布図で見られている低い層でのばらつきの状況に表れていると考えられる。

また，2PLM と GRM の比較では，図 4 が示す通り，GRM を利用することにより，2PLM と比較して，特に  $\theta$  の低い層で多くの情報量を得られているが，これは，これまでの研究と同様の結果である (Bock, 1972; 平井, 1993)。2PLM では正答以外

の選択肢は全て誤答として扱われるが、GRM では受検者がどの誤答選択肢を選んだかの誤答パターンまで利用することで、情報量が増加しているためである。このように、GRM では $\theta$ の低い層で多くの情報量が得られることを考えると、正答率の低いテスト、つまり難しいテストの採点においてGRM は非常に有効である。

また、GRM と NRM の間の関係については、両者の標準誤差は表 1 より、GRM=0.508, NRM=0.495 となっており、測定精度は同程度の水準であることが確認できた。また、図 3 の散布図が示す通り 2PLM との関係と比べ、 $\theta$  推定値に大きな差はないと考えられる。さらに、GRM, NRM 間の $\theta$  推定値の相関は 0.976 であり、平井(1993)が示しているとおおり、両者の $\theta$  推定値の違いは実務上の観点からみれば、ほぼ差異はないと考えられる。このことから、多肢択一式の客観式能力テストにおいても、項目分析によって事前に選択肢を序列づけすることで、GRM を実践的に利用できることが確認できた。

これらの結果は、多肢択一式の客観式能力テスト採点において、多値項目反応モデルの中でも、よりモデルがシンプルな GRM 活用の可能性について示唆している。

また、選択肢数が多い項目が含まれる場合、項目分析での結果を用いて同程度の成分スコアの選択肢を統合して、同じ選択肢数にすることで、GRM を適用することができる。その際に、選択数が少ない選択肢を統合することでより安定した推定結果が期待できる。

## 5. まとめ

本研究では、多肢択一式の能力テストについて、多重対応分析による項目分析と正答位置を考慮に入れて、事前に選択肢を序列づけしておくことで、GRM が実践的に利用できることを確認した。その結果、GRM を用いた場合、2PLM よりも情報量が上昇し、NRM と同等の情報量が得られることが確認された。このことは、多値項目反応モデルの中でも、よりモデルのシンプルな GRM を多肢択一式の客観式能力テストに活用するための1つの方法について示唆している。

今後は、項目分析による選択肢順と正答位置の情報が合致しない場合に、正答位置の情報を選択肢の序列づけに反映した場合、どのような影響を与えるかなど、実践への展開にあたり、適用する上での条件と限界について検討を重ねていきたい。

## 参考文献

Bock, R. D.(1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.

平井洋子(1993) 多肢選択式テストにおける測定の精緻化の試み(東京大学教育学部紀要 第33巻 1993)

豊田秀樹(2005) 項目反応理論[理論編]-テストの数理- (第2章, p58-67, p78-87)