流通科学大学論集—人間・社会・自然編—第 22 巻第 2 号，87-97（2010）

多値 Rasch モデルにおけるカテゴリ閾値逆順性と項目間従属性の連関性についての一考察

Disordering of Polytomous Rasch Model Thresholds and Items' Local Dependency

平越裕之*，井澤廣行**

Hiroyuki Hirakoshi, Hiroyuki Izawa

This paper investigates degree of difference in the test information function values, data uni-dimensionality, and items' local independence between ordered and disordered thresholds of the Rating Scale model (RSM). Findings based on RSM-fit simulated data are as follows: 1) disordered thresholds don't necessarily indicate a large scale of the test information function values affected greatly by the distribution of difficulty estimates; 2) there is no big difference in the degree of data uni-dimensionality between ordered and disordered thresholds; and 3) there is no definite observation made about the relationship between disordered thresholds and local dependency. Relevant key notions are provided.

Key words: Polytomous Rasch model, disordered thresholds, test information function, local independence.

Ⅰ. はじめに

多値順序カテゴリカルデータを扱う確率モデルである評定尺度モデル(rating scale model, by Andrich[1], 1978, 以降 RSM)と部分採点モデル(partial credit model, by Masters[2], 1982, 以降 PCM)は，共に Rasch 項目分析モデル(Rasch[3], 1960)に準じていることから，一般に Rasch モデル族(Rasch model family)と呼ばれている。また，非 Rasch モデルである項目反応理論(item response theory, 以降 IRT)体系では，段階反応モデル(graded response model, by Samejima[4], 1969, 以降 GRM) が多値カテゴリカルデータを扱う代表的なモデルとして位置づけられている。

多値順序カテゴリカルデータ Rasch モデル(polytomous Rasch model, 以降 PRM)と非 Rasch モデルの最大の相違は，モデルに適合した条件で行われるテストにおいて，項目や被験者によらず，測定に客観性(objectivity)が備わっているか否かである。ここでいう客観性とは，モデルパラメータの分離によって PRM の二つの隣接カテゴリにおける上位カテゴリ生起比率の上で獲得されている客観性である。つまり，多値項目中の隣接カテゴリ，すなわち，段階こそが Rasch モデル族としての客観性

* 流通科学大学情報学部、〒 651-2188　神戸市西区学園西町 3-1
** 流通科学大学サービス産業学部、〒 651-2188　神戸市西区学園西町 3-1

(2009 年 9 月 11 日受理)

を保っているということである(平越[5]，2009，参照)。

PRMにおけるこの客観性への信認により，非Raschモデルよりもその母数推定値不変性への高い期待で以って，特に，一回のテストにおいて全ての項目に渡ってカテゴリ閾値が同一とされるRSMのその閾値正順性へのデータ確認が容易であるとも考えられる。そのため，中小規模のテストあるいはアンケートのRSMへの適用分析出力によるカテゴリ閾値順序性とその閾値間差の値大小で以って，データ測定尺度構成の妥当性可否が論じられてもいる。しかしながら，PRM適用の際には，必然的に2値データを扱う以上に複雑なそのモデル特徴を理解する必要がある。PRMへのデータ適合度ないしはデータ一次元充足度の判定をカテゴリ閾値推定値の正順性とその閾値間差の値大小への参照のみに頼ることは取得データの特徴理解を妨げる可能性も存在する。

本稿では，RSMのカテゴリ閾値を概観し，RSM適用時におけるカテゴリ閾値の検討についてその順序性と項目独立性からの考察を行う。その要となる視点は，Andrich[6] (1998)による『閾値推定値逆転性はデータに問題ありとの証左』(p. 648)との言明が，その隣接カテゴリ逆転閾値推定値間において観察される差の大小程度が示唆するパラメータ推定の上での精度の違いにかかわらず，モデル不適合を暗示する一律的なデータ欠陥性として理解されるべきものであるか否かである。

Ⅱ．評定尺度モデル（RSM）と部分採点モデル（PCM）

本章ではRSM及びPCMが一次元能力測定テストに適用されることを想定する。つまり，受験者が項目を受験し，結果として項目のカテゴリを得るということを想定する。モデル上，受験者は能力パラメータ，項目は困難度パラメータ，カテゴリは閾値パラメータを持っているが，いずれも潜在特性である故に取得データから推定されることとなる。能力パラメータは値が大きいほど能力が高く，困難度パラメータは値が大きいほど項目は困難であり，能力と困難度は同次元の指標であるとみなされる。また，カテゴリについては項目得点と対応させ，低い数値が低い得点となる。なお，テストではなくアンケートを想定する際には，困難度・能力・カテゴリを適宜対応する概念標示により把握する。カテゴリ閾値パラメータは，次に示す二つのモデル式によって解釈を与える。

RSMでは，受験者iが項目jに対してカテゴリkの得点となる確率を次のように表す(Andrichi[1]，1978，参照)。

$$P_i(x=k\mid\delta_j)=\frac{\exp\{\kappa_k+k(\beta_i-\delta_j)\}}{\sum_{l=0}^{m}\exp\{\kappa_l+l(\beta_i-\delta_j)\}} \tag{1}$$

β_i：受験者iの能力値

δ_j：項目jの困難度

τ_m：カテゴリmの閾値（全項目共通），$\sum\tau_m=0$

$$\kappa_m=-\sum_{k=1}^{m}\tau_k,\kappa_0\equiv 0$$

一方，PCM では，受験者 i が項目 j に対してカテゴリ k の得点となる確率を次のように表す (Masters[2], 1982, 参照)。

$$P_i(x=k\mid\delta_j)=\frac{\exp\{\kappa_{j,k}+k(\beta_i-\delta_j)\}}{\sum_{l=0}^{m}\exp\{\kappa_{j,l}+l(\beta_i-\delta_j)\}} \tag{2}$$

β_i : 受験者 i の能力値

δ_j : 項目 j の困難度

$\tau_{j,m}$: 項目 j のカテゴリ m の閾値, $\sum_m \tau_{j,m}=0$

$$\kappa_{j,m}=-\sum_{k=1}^{m}\tau_{j,k},\ \kappa_{j,0}\equiv 0$$

両モデルはそれぞれの導出時[1) 2)]においては異なるアプローチがとられているが，どちらも Rasch が示唆したモデル式[3)]に基づいており，結果として導出されたモデルは形式的にはほとんど同じである。RSM と PCM の相違は，カテゴリ閾値 τ が 1 テスト内での全ての項目について同一であるか個別項目ごとに異なるかとの一点のみである。このため，RSM においては全ての項目においてカテゴリ数が同一である必要があるが，PCM ではその制限はない。なお，上式以外によってもモデル定義は可能であるが，本質的な差異は存在しない。

Ⅲ. 評定尺度モデル(RSM)におけるカテゴリ閾値逆順性と項目間従属性

1. カテゴリ閾値の順序性

RSM においては, 2 値 Rasch モデル(Rasch[3)], 1960)には存在しないカテゴリ閾値がパラメータとして存在している。2 値 Rasch モデルでは，テストと項目によらず正答確率（正答を 1，誤答を 0 とした時の期待得点）曲線の形状は同一であり，能力軸の上で平行移動すれば重なる。つまり，正答確率曲線形状は潜在的な特性ではなく，そのモデル規定である。しかし，RSM では項目に対する期待得点曲線を決定するカテゴリ閾値は潜在特性であり，テストごとに異なることが許容されている。RSM におけるこの要因がカテゴリ閾値に対する様々な解釈と評価を与えることとなっている(Andrich[7)], 2004, p. 191; Masters[8)], 1988, p. 287, 参照)。RSM を含む Rasch モデル族においては，そもそも項目間での確率的な独立性が前提であり(Hambleton, 野口訳，池田他編訳[9)], 1992, p. 218)，データでの項目間従属性はモデル適用逸脱を意味する。現実テストにおいて全ての項目を確率的に完全な独立とすることは不可能であろうが，許容される程度に項目間独立性を保つことは必須である。

平越[5)](2009)では，RSM におけるカテゴリ閾値の順序性に規定はないと結論付けている。しかしながら，多くの RSM 識者・適用者はカテゴリ閾値の正順性こそが RSM 適用時に第一に確認すべき事項として重要視している(Andrich[1)], 1978; Luo[10)], 2005; Smith et al.[11)], 2003; Tennant[12)], 2004; Zhu et al.[13)], 1997)。カテゴリ閾値正順性の重要性は，概ね次の二点において主張されている。

・カテゴリの推移を「通過」と考えれば，閾値逆順の拠り所がない(Andrich[14], 2005, p. 51)。

・カテゴリー閾値が逆順であると項目弁別力が大きくなり，項目間の独立性が低くなる(Andrich[15], 1998, pp. 370-371 and 2006[16] ; Linacre[17], 1999a)。

「カテゴリの推移を『通過』と考えれば，閾値逆順の拠り所がない」という前者の指摘については，モデル化された時点でカテゴリ閾値は「通過」という概念のみで以って説明できるものではない付帯属性を有している(平越[5]，2009，参照)。カテゴリ閾値が逆順である場合においても，項目の期待得点は（能力β－困難度δ）の増加に伴い単調増加する。図 3-1 は，3 カテゴリから成る RSM 項目においてカテゴリ閾値$\tau_1(=-\tau_2)$の変化に伴う期待得点を示しており，いずれのタウ値においても（能力β－困難度δ）の増大に伴って期待得点が単調増加していることは明らかである。RSM 上でのモデル化の際に順序カテゴリの「推移・通過」という概念を利用しながらも，カテゴリ閾値にこの属性が付与されている。従って，順序カテゴリの「推移・通過」という概念は RSM の本質を標示するものではないと考えられる。項目内の複数カテゴリにおけるカテゴリ閾値の順次「推移・通過」との視点ではなく，一項目全体としての上に述べた期待得点，及び，各カテゴリの選択確率(平越[5]，2009，参照)を評価すべきである。

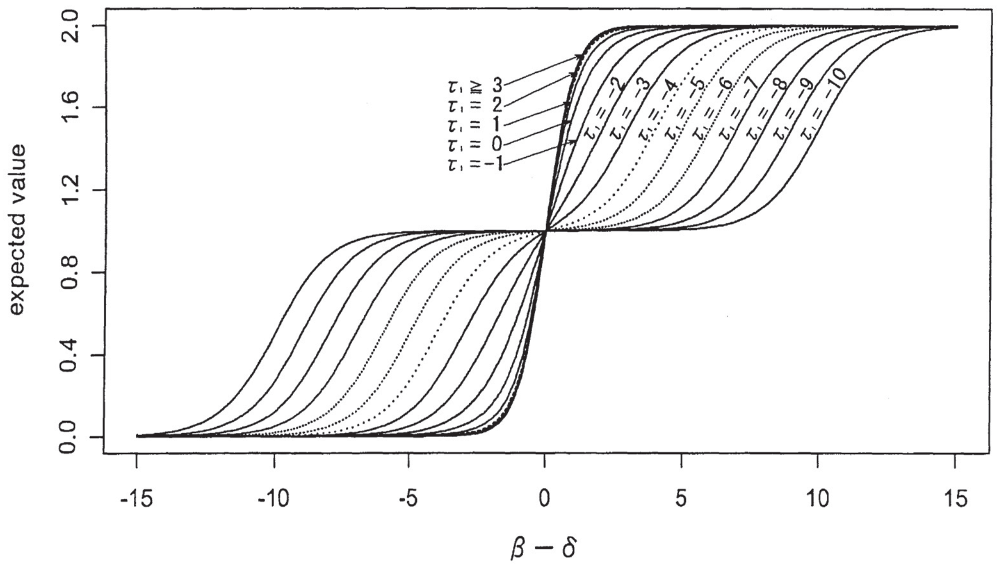

図 3-1　3 カテゴリ項目（満点２）の期待得点

次に，「カテゴリ閾値推定値が逆順であると項目弁別力が大きくなり，項目間の独立性が低くなる」という主張は，「同一概念を構成する一次元項目群は相似程度において余りにも過度に相似してはならない」(Linacre[18], 1999b, p. 695，参照)との確率モデルの上で仮定されている項目間独立性に関係する。図 3-1 でもわかるように，3 カテゴリ項目での閾値が逆順である場合，つまり$\tau_1>0$である場合

には，2つのカテゴリ閾値の平均値が（能力β－困難度δ）になる点において期待得点曲線の傾きは大きくなり，（能力β－困難度δ）＝0の点で大きな弁別力を持つことになる。このように大きな弁別力を持つような場合には，「項目間での各受験者得点の相関係数が高くなり，従ってテスト全体での弁別力が高すぎて項目間の独立性が低くなり，モデルの前提を外す方向となる」との見解である。それがカテゴリ閾値正順重要性の根拠とされているように思われる。

以降，カテゴリ閾値が逆順であっても，つまり各項目の最大弁別力が大きくとも，テスト全体の弁別力は必ずしも大きくならない場合もあることを示す。さらに，項目間の相関係数が高くなるように恣意的に作成したデータを用いて，項目間相関係数が全体的に高い場合においても，必ずしもカテゴリ閾値の逆順方向への牽引力にはならないことを示す。

2．カテゴリ閾値，項目困難度分布，及び，テスト情報関数

テスト全体の弁別力を示す指標としてテスト情報関数が認知されており(豊田[19]，2005, pp. 65-67)，それは各項目の情報関数の和である。項目情報関数とは項目の期待得点の微分値であり，それは弁別力の指標である。つまり，テスト全体の弁別力指標としてテスト情報関数が有効である。本節では，シミュレーションによりカテゴリ閾値，困難度推定値分布幅，及び，素点データの主成分分析第一主成分寄与率の関係一例を示し，これらの間に存在する関係の一端を明示する。

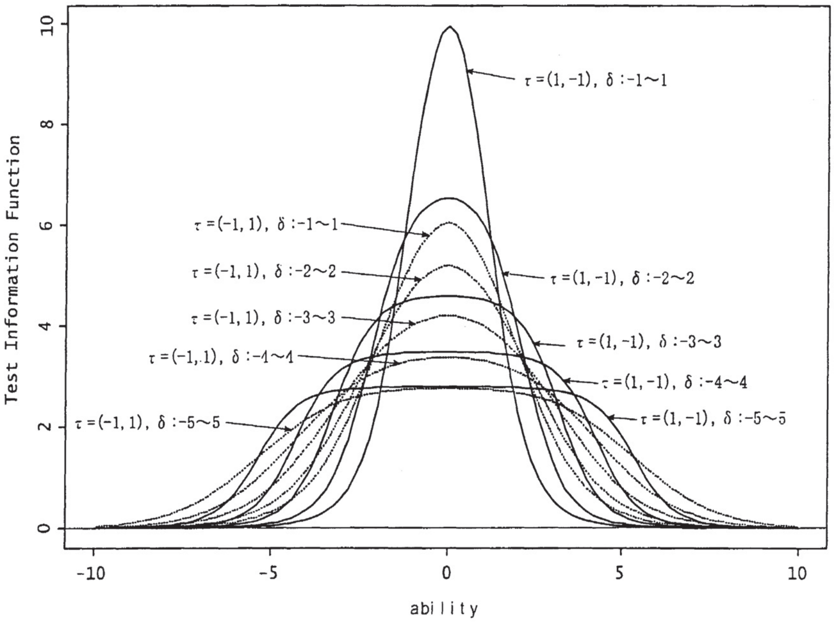

図3-2　テスト情報関数

まず，カテゴリ数を3，カテゴリ閾値を$(\tau_1, \tau_2) = (1,-1), (-1,1)$とする2通りの設定で以って，15項

目のテストを想定する。また，15 項目の困難度は，その最小値と最大値を(最小値,最大値) $=(-1,1), (-2,2), (-3,3), (-4,4), (-5,5)$ として，その間隔を 14 等分して 15 項目の困難度が均等に配置されたテストの情報関数を求める。その設定されたカテゴリ閾値とテスト情報関数の関係を図 3-2 に与える。タウ値が逆順である $(\tau_1, \tau_2)=(1,-1)$ と設定されたカテゴリ閾値の場合，項目困難度の分布幅が狭くなれば，テスト情報関数は能力 0 近辺で大きく尖り，それ以外の部分では低くなっていることがわかる。これは能力 0 近辺にテストの持つ弁別力が集中しているからである。しかし，項目困難度分布幅のみを大きくしていった場合，中央の尖りは緩やかになっていき，能力 0 近辺の弁別力はその周囲へと移動している。カテゴリ閾値を $(\tau_1, \tau_2)=(-1,1)$ とする正順である場合は破線で示している。同条件で比較すれば，タウ値の逆順よりも正順において能力 0 近辺の尖りがなだらかであり，広い能力値範囲にテスト弁別力が分布していることがわかる。図 3-2 より，カテゴリ閾値の逆順はその正順と比べて能力 0 近辺のテスト弁別力が大きいことは明らかである。しかしながら，項目困難度の分布幅を加味して比較した場合，困難度の分布幅が大きくなるにつれてカテゴリ閾値の正順と逆順の間におけるテスト情報関数分布の違いが小さくなる。これにより，テスト全体の弁別力を表すテスト情報関数は，カテゴリ閾値の順序性に加えて，項目困難度分布幅に大きく影響されることがわかる。つまり，カテゴリ閾値の順序性は，単独でテスト全体の弁別力指標とはならないということである。

次に，標準正規分布 N(0,1)に従う能力値を持つ受験者 100 名を想定して，図 3-2 での 10 通りのグラフと同一条件であるテスト受験に関するシミュレーションデータを各 100 通り作成する。合計 1000 通りの各条件素点データの項目群主成分分析による第一主成分寄与率（以降第一寄与率）に関する統計値を表 3-1 に与える。

表　3-1　第一主成分寄与率の変化(受験者は N(0,1)から 100 人)

(τ_1, τ_2)	困難度分布幅	平均値	標準偏差	100平均の正規性の Kolmogorov-Smirnov検定	
				D値	p値
(1, −1)	−1〜1	0.3345	0.0180	0.0829	0.4983
	−2〜2	0.2819	0.0153	0.0657	0.7819
	−3〜3	0.2366	0.0151	0.0508	0.9586
	−4〜4	0.2063	0.0170	0.0612	0.8482
	−5〜5	0.1819	0.0169	0.0801	0.5430
(−1, 1)	−1〜1	0.2774	0.0173	0.0649	0.7941
	−2〜2	0.2609	0.0178	0.0687	0.7330
	−3〜3	0.2353	0.0181	0.0780	0.5764
	−4〜4	0.2107	0.0169	0.0716	0.6846
	−5〜5	0.1876	0.0153	0.0518	0.9513

表 3-1 より，困難度分布幅が狭い場合には，カテゴリ閾値の正順よりも逆順において第一寄与率は高くなっていることがわかる。しかし，困難度分布幅が-3～3 より広い場合には，逆順，正順の間に第一寄与率の違いは認められない。これは，図 3-2 での-3～3 より広い項目困難度分布幅に関するカテゴリ閾値の逆順と正順でのテスト情報関数分布の間に差異がほとんどないことに符合している。なお，各条件での第一寄与率に対して，正規分布を対象とする Kolmogorov-Smirnov 検定の上で，非常に高い p 値が付与されている。また，累積頻度グラフと正規分布の累積分布関数を描いた図への目視による検討においても正規分布性が十分に認められたことを付け加えておく。

以上により，テスト全体の弁別力は，単純にカテゴリ閾値の順序性により単独で影響を受ける訳ではなく，テストにおける各項目の困難度がどのように配置されているかという困難度分布幅に大きく影響を受けることが明らかである。従って，カテゴリ閾値の順序性を単独のテスト弁別力指標として単純に議論せず，テスト情報関数に困難度推定値分布をも併せて参照することが重要であることがわかる。

3．項目独立性とカテゴリ閾値推定値

RSM 適用のパラメータ推定結果として得られるカテゴリ閾値が逆順である場合には項目間の独立性が低いとの論理が，Andrich (1998[15], pp. 370-371 and 2006[16])による PCM (Masters[2], 1982)適用に関するその視点から導かれる。確率モデルとしての理論的な展開は困難である故に，本節では，シミュレーションデータに基づいて項目間の確率的独立性とカテゴリ閾値順序性の関係一端を検証する。

まず，標準正規分布 N(0,1)に従う能力値を持つ受験者 300 名，3 カテゴリの閾値を$(\tau_1, \tau_2) = (-1,1)$とする 15 項目の困難度分布(困難度最小値,困難度最大値)$= (-1,1), (-2,2), (-3,3)$で以って，項目間が確率的に完全独立とされるシミュレーションデータを 10 通りずつ作成する。次に，各シミュレーションデータにおける 1～50 人目の 50 行を 6 回繰り返して 300 名として，恣意的に項目間の相関を強めたデータを作成する。項目間の従属性を強める場合には，項目データを何度か繰り返し，相関を強める恣意的データ作成方法もある。しかし、その方法では複数の項目について全く同じ得点パターンが出現することにより，多値を扱う PRM 適用においては特にデータの不自然さが際立ってしまう。そのため，ここでは受験者を繰り返して結果的に項目間の従属性を高めることとした。この方法によれば、項目間に同じ得点パターンが出現しないことにより項目間の検討に関してより自然なデータで以って検討出来る。また，この受験者繰り返しによる方法においてもほとんど全ての素データで項目間の相関は強くなる。なお，補完のため，項目を繰り返したいくつかのデータによっても同様のシミュレーションを試みたが，結果として上記両方法による大きな相違点は特に認められない。

各データの RSM 適用出力の上でのカテゴリ閾値推定値と各データの項目群主成分分析第一寄与率を表 3-2 に与える。3 通りの困難度分布幅いずれの場合における各一対データに付与されている第一寄与率についてオリジナルデータよりも繰り返しデータが高く，後者での項目間相関が強くなっていることがうかがえる。一方，オリジナルデータと繰り返しデータのカテゴリ閾値推定値に関しては，

その順序性に項目間相関高低との連関法則はみいだせない。つまり，当該シミュレーションデータ群においては，1) 項目間の独立性が低くなることとカテゴリ閾値が逆順になることに明確な連関性が存在しない，あるいは，2) 項目間の独立性が低くなるとカテゴリ閾値が逆順になるということが仮に論理的に正しくとも，その連関性が顕示されていないと判断される。これにより，一回の現実テストへの RSM 適用上でのカテゴリ閾値の逆順と項目間独立性欠如程度の正確な連関性を推察することは不可能に近いと思われる。

表 3-2 項目相関とカテゴリ閾値

困難度分布	オリジナルデータ		繰り返しデータ(50人×6)		オリジナルと繰り返しでの値の変化	
	τ_1, τ_2推定値	第一寄与率	τ_1, τ_2推定値	第一寄与率	τ_2推定値(正値が正順方向)	第一寄与率
-1～1	−1.06, 1.06	0.304	−1.30, 1.30	0.378	0.24	0.074
	−1.08, 1.08	0.293	−1.01, 1.01	0.308	−0.07	0.015
	−1.06, 1.06	0.302	−1.06, 1.06	0.359	0.00	0.056
	−0.95, 0.95	0.290	−0.99, 0.99	0.335	0.04	0.045
	−1.08, 1.08	0.298	−1.04, 1.04	0.331	−0.04	0.034
	−1.12, 1.12	0.303	−1.22, 1.22	0.340	0.10	0.037
	−1.04, 1.04	0.298	−0.98, 0.98	0.338	−0.06	0.040
	−1.09, 1.09	0.294	−1.22, 1.22	0.378	0.13	0.084
	−1.12, 1.12	0.310	−1.14, 1.14	0.343	0.02	0.033
	−1.12, 1.12	0.289	−1.26, 1.26	0.313	0.14	0.024
-2～2	−1.09, 1.09	0.267	−1.11, 1.11	0.317	0.02	0.049
	−1.13, 1.13	0.279	−0.94, 0.94	0.303	−0.19	0.024
	−1.12, 1.12	0.294	−1.15, 1.15	0.352	0.03	0.058
	−1.12, 1.12	0.277	−1.31, 1.31	0.280	0.19	0.004
	−1.16, 1.16	0.280	−1.28, 1.28	0.304	0.12	0.024
	−1.07, 1.07	0.280	−1.13, 1.13	0.321	0.06	0.041
	−1.09, 1.09	0.274	−0.97, 0.97	0.277	−0.12	0.003
	−1.13, 1.13	0.275	−1.19, 1.19	0.317	0.06	0.041
	−1.12, 1.12	0.277	−1.32, 1.32	0.294	0.20	0.017
	−1.15, 1.15	0.267	−1.02, 1.02	0.282	−0.13	0.015
-3～3	−1.24, 1.24	0.263	−1.32, 1.32	0.294	0.08	0.030
	−1.14, 1.14	0.242	−1.24, 1.24	0.275	0.10	0.033
	−1.05, 1.05	0.219	−1.06, 1.06	0.232	0.01	0.013
	−1.14, 1.14	0.252	−1.20, 1.20	0.335	0.06	0.083
	−1.15, 1.15	0.247	−0.99, 0.99	0.282	−0.16	0.035
	−1.09, 1.09	0.252	−1.30, 1.30	0.311	0.21	0.059
	−1.15, 1.15	0.261	−1.23, 1.23	0.292	0.08	0.030
	−1.06, 1.06	0.235	−1.11, 1.11	0.237	0.05	0.002
	−1.10, 1.10	0.240	−1.08, 1.08	0.257	−0.02	0.018
	−1.26, 1.26	0.262	−1.32, 1.32	0.317	0.06	0.055

なお，平越[5](2009)での論説の通り，カテゴリ閾値は本質的に項目群の設計によって変化するものであり，カテゴリ閾値のみで以ってテスト全体の検証を行うことは困難である。従って，カテゴリ閾値のみの評価にとらわれずに，項目群設計の検討，項目群困難度推定値の分布，受験者群能力推定値の分布を含めて総合的に検証することが必要とされる。

Ⅳ. おわりに

本稿において，第一章末尾に引用した『閾値推定値逆転性はデータに問題ありとの証左』との Andrich[6] (1998, p. 648)による言明における『問題あり』のその所在への考察を与えた。RSM 適用に

関して本考察に使用したカテゴリ閾値の逆順 $(\tau_1, \tau_2) = (1,-1)$ と正順 $(\tau_1, \tau_2) = (-1,1)$ の間に，以下の点においてほぼ同様な検証結果であると要約される。なお，下記 1)は，平越[5](2009) に基づく概念論である。

1) 各カテゴリー閾値「通過」を下敷きとするその推定値順序性への個別的主観判断よりも，各項目全体としての期待得点に併せて各カテゴリの選択確率の全体客観評価を優先させる。
2) カテゴリー閾値推定値の逆順は必ずしもテスト全体の弁別力の大きさを示唆するものではなく，テスト弁別力すなわちテスト情報関数は項目群困難度推定値分布幅に大きく影響される。
3) カテゴリ閾値推定値の正順と逆順の場合において、項目群素点得点によるデーター次元充足度に大きな違いは認められない。
4) カテゴリ閾値推定値逆順と項目間独立性侵害との直接明示的な連関関係は顕現されがたい。

従って，隣接カテゴリ逆転閾値推定値間において観察される差の大小程度が示唆するパラメータ推定の上での精度の違いにかかわらず，カテゴリ閾値推定値の逆転をモデル不適合と暗示する一律的なデータ欠陥性として捉えるべきでなないと結論される。PRM を含む Rasch モデル族は，一度のみの試行で得られる離散値データからパラメータを推定する確率モデルそのものの実質的限界性，例えば，パラメータの推定値域や推定方法による推定精度・推定値信頼性等の限界性が内在するモデルである。従って，現実に実施されるテストにおける項目群と受験者群のモデル不適合性は，モデル限界性と切り離して考えなければならない。一口に「データ欠陥性」とまとめずに，モデルに内在する本質的な限界性とデータのモデル適合度という二つの概念を峻別理解することが重要であろう。

RSM について，Andrich[14](2005)は『カテゴリ閾値推定値の逆順データが一般的な統計的モデル適合度検定の上で必ずしも不適合と示されることはない』(p. 53)と述べている。この指摘に付加されるべき言葉は，『しかし，その隣接閾値推定値間に逆順での大きな差があるデータは，その正順での大きな差があるデータと同じく，パラメータ推定精度の悪い可能性が大である』ということである。一般的に，RSM 適合シミュレーションデータにおける隣接閾値推定値間に逆順での大きな差があるデータにおいては中位カテゴリが両端カテゴリに比べてその生起度数が極端に小さなものとなる。正にこのパラメータ推定精度の点においてのみ，『閾値推定値逆転性はデータに問題ありとの証左である』(Andrich[6], 1998, p. 648)と理解される。然しながら，本考察でその一端が明示された様に，閾値推定値逆順は本質としてデータの PRM への不適合の指標ではないということであり，又，それは一次元充足度と項目間独立性侵害程度の指標でもないということである。

最後に，付記として，本稿でのカテゴリ閾値順序性，テスト情報関数，項目群困難度推定値分布，及び，項目間従属性の間における相互連関性有無程度への検証に使用したカテゴリ閾値は $(\tau_1, \tau_2) = (1,-1)$ と $(\tau_1, \tau_2) = (-1,1)$ の二種類のみであることに言及しておく。閾値逆順が項目間従属性の指標となることへの反例であるから，他の値でも構わない。その二種類にカテゴリ閾値を限定した理由は，検証の負担軽減に併せて，現実 3 カテゴリデータへの RSM 適用の上でのその出現可能性が

大きいとの想定，及び，上記のパラメータ推定精度における良質性による。各一対閾値の幅を広めるにつれて，パラメータ推定精度が悪くなり，従って，閾値逆順と項目間従属性の連関性についての実質的な特徴を見逃す恐れがある。その配慮の上での上記閾値設定である。

文献

1) D. Andrich: "A rating formulation for ordered response categories", *Psychometrika*, 43, 4(1978) 561-573.
2) G. N. Masters: "A Rasch model for partial credit scoring", *Psychometrika*, 47, 2(1982) 149-174.
3) Rasch, G: *Probabilistic models for some intelligence and attainment tests*, The Danish Institute for Educational Research(1960). (Reprinted in 1980 by the university of Chicago Press with a Foreword and Afterword by Wright, B. D.　G.ラッシュ，内田良男(完訳): 心理テストの確率モデル，名古屋大学出版会(1985))
4) Samajima, F: "Estimation of latent ability using a response pattern of graded scores", *Psychometrika, Monograph Supplement No. 17* (1969).
5) 平越裕之:「多値 Rasch モデルのカテゴリ閾値に関する一考察」,『流通科学大学論集 — 経済・経営情報編』17，No.2(2009)　87-104.
6) D. Andrich: "Thresholds, steps and rating scale conceptualization", *Rasch Measurement Transactions*, 12, 3(1998)　648-649.
7) D. Andrich: "Understanding resistance to the data-model relationship in Rasch's paradigm", A reflection for the next generation, In Smith, Jr., E. V. & Smith, R. M. (Eds.), *Introduction to Rasch Measurement* (Maple Grove, Minnesota, JAM Press, 2004)　167-200.
8) G. N. Masters: "The analysis of partial credit scoring", *Applied Measurement in Education*, 1, 4(1988) 279-298.
9) R. K. Hambleton, 野口裕之訳，池田央・藤田恵璽・柳井晴夫・繁桝算男編訳，ロバート L. リン編:『教育測定学原著第３版』上巻（発行：C.S.L.学習評価研究所，発売：みくに出版, 1992).
10) G. Luo: "The relationship between the rating scale and partial credit models and the implication of disordered thresholds of the Rasch models for polytomous responses", *Journal of Applied Measurement*, 6, 4(2005) 443-455.
11) E. V. Smith Jr., M. B. Wakely, R. E. L. de Kruif, & C. W. Swartz: "Optimizing rating scales for self-efficacy (and other) research", *Educational and Psychological Measurement*, 6, 3(2003)　369-391.
12) A. Tennant: "Disordered thresholds:　An example from the functional independence measure", *Rasch Measurement Transactions*, 17, 4(2004)　945-948.
13) W. Zhu, W. F. Updyke, & C. Lewandowski: "Post-hoc Rasch analysis of optimal categorization of an

ordered-response scale", *Journal of Outcome Measurement*, 1, 4(1997) 286-304.

14) D. Andrich: The Rasch model explained. In Alagumalai, S, Curtis, D. D., & Hungi, N (Eds.), *Applied Rasch measurement: A book of exemplars* (The Netherlands, Springer, 2005) 27-59.

15) D. Andrich: "A general form of Rasch's extended logistic model for partial credit scoring", *Applied Measurement in Education*, 1, 4(1988) 363-378.

16) D. Andrich: "Item discrimination and Rasch-Andrich thresholds revisited", Rasch Measurement Transactions, 20, 2(2006) 1055-1057.

17) J. M. Linacre: "Category disordering vs. step (threshold) disordering", Rasch Measurement Transactions, 13, 1(1999) 675-678.

18) J. M. Linacre: Comment on *the New Rules of Measurement* (edited by Embretson, S. E. & Hershberger, S. L. 1999. Mahway, NJ: Lawrence Erlbaum), *Rasch Measurement Transactions*, 13, 2(1999) 692-695.

19) 豊田秀樹編著: 『項目反応理論[理論編]』(朝倉書店, 2005).