

# Many-Facet Rasch Measurement 適用上での閾値増大順序性、 評定値統合、及び、データ一次元充足度考察事例

The Natural Ordering of the Thresholds' Estimates, the Recategorization Process, and  
Unidimensionality on Linacre's Many-Facet Rasch Measurement

井澤 廣行\*  
Hiroyuki Izawa

This paper describes the natural ordering of the thresholds' estimates, the recategorization process, and unidimensionality on the Many-Facet Rasch Measurement. It includes a reference to Linacre's (1989) derivation of the judges' common-step model from the specifications of the Rasch model, i.e., unidimensionality and invariance; and a recognition of the model's principle identical to Andrich's (1978) Rating Scale model. In the analysis of an actual dataset, it has been found that there is a great extent of incongruence between an almost perfect degree of natural ordering of the thresholds' estimates and a low degree of unidimensionality indication.

Key words : Linacre, Many-facet Rasch measurement, judges' common-step model, thresholds, unidimensionality.

## I. はじめに

本稿の目的は、Andrich<sup>1)</sup>(1978)考案の Rating Scale モデルに評定者を変数として加えた Linacre<sup>2)</sup>(1989)案出による Many-Facet Rasch Measurement の適用分析要領を記述することである。日本におけるその適用上での研究報告が秋山<sup>3)</sup>(2000)を含めて増大しているが、研究成果報告上でのその適用分析基本要領への認識欠如の節が窺われる。それが本稿執筆の動機である。遺憾ながら、筆者にその適用に具するデータが取得されていない。従って、Linacre<sup>2)</sup>(1989)著作の *Many-Facet Rasch Measurement* において図 A.5 (p. 11)に供与されている現実データをその目的に使用する。なお、それは、“Complete” judging plan for the essay data (Courtesy: Robert G. Cameron of the College Board)と題されており、Linacre<sup>2)</sup>(1989)によってはデータ一例としてのみの掲示で以って分析はなされていない。Dr. John M. Linacre による「データ公共性」に関する示唆(2008年4月7日)への筆者自身の判断に基づくそのデータ借用自由裁量(本稿末尾「謝辞にかえて」参照)で以って、そのデータ説明と分析結果を含めて本論を第三節と第四節に与える。次節では、Many-Facet Rasch

Measurement 理解促進への意図の下に、Linacre<sup>2)</sup>(1989)による評定者「共通段階モデル」導出を参照する。又、その導出参照過程の上で、筆者によって Linacre<sup>2)</sup>(1989)にその言明は見出されないけれども、評定者「共通段階モデル」は Andrich<sup>1)</sup>(1978)の Rating Scale モデルの付帯特性とされる潜在閾値母数順次増大性を内在共有しているとの理解一端を与える。

## II. Linacre (1989)による評定者「共通段階モデル」導出とそのモデル付帯特性としての潜在閾値母数順次増大性

Linacre<sup>2)</sup>(1989)の評定者「共通段階(common-step)モデル」(p. 1)は対数オッズ表現式として下に与えられるものである。C<sub>j</sub>を面接官 j の評定難度母数とすれば、面接試験において能力母数 B を持つ受験者 n による困難度母数 D を持つ項目 i への応答に対して、順序カテゴリーでの第 k 段階閾値母数を F<sub>k</sub> として、面接官 j の順序評定として評定値 (k - 1)ではなく評定値 k が選択される確率の対数オッズとして示されている。従って、各母数の測定単位は log-odds units としてのロジット(logits)である(Linacre<sup>2)</sup>, 1989, p. 53)。この評定者「共通段階モデル」が帰納的観点の上で母数推定値の安定性と再現性において最も普遍度の高いモデル(Linacre<sup>2)</sup>, 1989, p. 98)とされている。

$$\ln(P_{nijk} / P_{nijk-1}) = B_n - D_i - C_j - F_k$$

この評定者「共通段階モデル」確率式の上で、そのモデル期待確率比同一性と尺度基準化概念で以って、Rasch モデルに特有とされる「固有客観性」( 'specific objectivity,' in Rasch<sup>4)</sup>, 1977)に基づくものとして Linacre<sup>2)</sup>(1989)により以下の記述構成枠の下で導出されている(pp. 47-52)。

1) k > h として、二つの順序カテゴリー k と h から成る一つの項目に対する二人の受験者 O<sub>n</sub> と O<sub>m</sub> による想定上での無限に近い多数回試行における一評定者のカテゴリー評定選択上での受験者 O<sub>n</sub> と O<sub>m</sub> に関する能力比較は F<sub>kh</sub> / F<sub>hk</sub> で示される。F<sub>kh</sub> は O<sub>n</sub> が k 評価で O<sub>m</sub> が h 評価である生起度数であり、F<sub>hk</sub> は O<sub>n</sub> が h 評価で O<sub>m</sub> が k 評価である生起度数である。

2) 上記の想定多数回試行における生起度数比率 F<sub>kh</sub> / F<sub>hk</sub> を無限回試行として一般普遍化すれば、k > h として、二つの順序カテゴリー k と h から成る一つの項目に対する二人の受験者 O<sub>n</sub> と O<sub>m</sub> の反応に関する一評定者のカテゴリー評定選択上での受験者 O<sub>n</sub> と O<sub>m</sub> についての能力比較は潜在的に P<sub>kh</sub> / P<sub>hk</sub> で示される。P<sub>kh</sub> は評定者 J<sub>j</sub> によって O<sub>n</sub> に k 評価、及び、O<sub>m</sub> に h 評価が付与される確率であり、P<sub>hk</sub> は評定者 J<sub>j</sub> によって O<sub>n</sub> に h 評価、及び、O<sub>m</sub> に k 評価が付与される確率である。

3) 上記の 2)は二つの順序カテゴリー k と h (k > h)から成る一つの項目について一人の評定者 J<sub>j</sub> による二人の受験者 O<sub>n</sub> と O<sub>m</sub> に関する能力比較を示す確率式である。三つ以上の順序カテゴリーから成る一つの項目についての上記と同じ状況下での能力比較確率式においては、「固有客観性」充足のために「不変性」が隣接評定順序カテゴリー組み合わせの各一対間に設定される必要がある。

Linacre<sup>2)</sup>(1989)による上記説明構成枠における 1)は容易に理解される故に、それ以上の注釈は必

要とされない。従って、2)と3)に基づく Linacre<sup>2)</sup>(1989)の評定者「共通段階モデル」確率式導出(pp. 48-52)を更に詳しく以下に参照する。

2) 受験者  $O_n$  と  $O_m$  の潜在能力比較 =  $P_{kh} / P_{hk}$

$P_{kh}$  は評定者  $J_j$  によって  $O_n$  の項目  $A_i$  への応答に  $k$  評価、及び、 $O_m$  の項目  $A_i$  への応答に  $h$  評価が付与される確率であり、 $P_{hk}$  は評定者  $J_j$  によって  $O_n$  の項目  $A_i$  への応答に  $h$  評価、及び、 $O_m$  の項目  $A_i$  への応答に  $k$  評価が付与される確率である。「固有客観性」充足のためには評定者  $J_j$  による各受験者への各評価は独立していなければならない。独立事象の確率乗法定理により、 $P_{kh} = P_{nij} * P_{mijh}$ 、並びに、 $P_{hk} = P_{nijh} * P_{mijk}$  と表示される。従って、

$$\text{受験者 } O_n \text{ と } O_m \text{ の潜在能力比較} = P_{kh} / P_{hk} = (P_{nij} * P_{mijh}) / (P_{nijh} * P_{mijk})$$

「固有客観性」充足のためには、更に、上式が順序評定テスト構成概念に沿う如何なる項目  $A_i$  にも充たされなければならない。従って、

$$\begin{aligned} \text{受験者 } O_n \text{ と } O_m \text{ の潜在能力比較} &= P_{kh} / P_{hk} = (P_{nij} * P_{mijh}) / (P_{nijh} * P_{mijk}) \\ &= (P_{ni'jk} * P_{mi'jh}) / (P_{ni'jh} * P_{mi'jk}) \end{aligned}$$

上式を変換して、

$$P_{nij} / P_{nijh} = (P_{mijk} / P_{mijh}) * (P_{ni'jk} / P_{ni'jh}) * (P_{mi'jh} / P_{mi'jk})$$

ここで、「固有客観性」充足のためには、受験者  $O_n$  と潜在能力上で比較される受験者  $O_m$  については如何なる受験者の場合にも上式が当てはまる必要がある。受験者  $O_n$  と潜在能力上で比較されるその受験者  $O_m$  に能力尺度上で原点に位置する一人の基準受験者  $O_0$  を上式に挿入する。

$$P_{nij} / P_{nijh} = (P_{0ijk} / P_{0ijh}) * (P_{ni'jk} / P_{ni'jh}) * (P_{0i'jh} / P_{0i'jk})$$

「固有客観性」充足のためには、更に、項目  $A_i$  と順序評定テスト構成概念上で相対関係にある項目  $A_j$  に項目困難度尺度上で原点に位置する一つの基準項目  $A_0$  を上式に挿入する。

$$P_{nij} / P_{nijh} = (P_{0ijk} / P_{0ijh}) * (P_{n0jk} / P_{n0jh}) * (P_{00jh} / P_{00jk})$$

更に、上式における  $(P_{0ijk} / P_{0ijh})$  と  $(P_{n0jk} / P_{n0jh})$  のそれぞれが考察される。まず、後者の  $(P_{n0jk} / P_{n0jh})$  については、受験者  $O_n$  と  $O_m$  の潜在能力比較において、同一項目  $A_i$  における順序カテゴリー  $k$  と  $h$  ( $k > h$ ) に対する二人の評定者  $J_j$  と  $J_j$  による付与確率が「固有客観性」充足のため同一とされる。従って、

$$\begin{aligned} \text{受験者 } O_n \text{ と } O_m \text{ の潜在能力比較} &= P_{kh} / P_{hk} = (P_{nij} * P_{mijh}) / (P_{nijh} * P_{mijk}) \\ &= (P_{nij'k} * P_{mij'h}) / (P_{nij'h} * P_{mij'k}) \end{aligned}$$

上式を変換して、

$$P_{nij} / P_{nijh} = (P_{mijk} / P_{mijh}) * (P_{nij'k} / P_{nij'h}) * (P_{mij'h} / P_{mij'k})$$

「固有客観性」充足上での上記と同様の考え方で以って、 $O_m$  に替わる基準受験者  $O_0$ 、 $A_i$  に替わる基準項目  $A_0$ 、並びに、 $J_j$  に替わる基準評定者  $J_0$  を上式に挿入する。

$$P_{n0jk} / P_{n0jh} = (P_{00jk} / P_{00jh}) * (P_{n00k} / P_{n00h}) * (P_{000h} / P_{000k})$$

次に、 $(P_{0ijk} / P_{0ijh})$  が考察される。同一受験者  $O_n$  について二つの項目  $A_i$  と  $A_j$  への応答上で二人

の評定者  $J_j$  と  $J_j$  により独立に評定  $k$  と  $h$  ( $k > h$ ) が付与される確率で以って、その二人の評定者に関する潜在評定難度比較は次式により表される。

$$\begin{aligned} \text{評定者 } J_j \text{ と } J_j \text{ の潜在評定難度比較} &= (P_{nij} * P_{ni'jh}) / (P_{nijh} * P_{ni'jk}) \\ &= (P_{nij'k} * P_{ni'jh}) / (P_{nij'h} * P_{ni'jk}) \end{aligned}$$

上式を変換して、

$$P_{nij} / P_{nijh} = (P_{ni'jk} / P_{ni'jh}) * (P_{nij'k} / P_{nij'h}) * (P_{ni'jh} / P_{ni'jk})$$

再び、「固有客観性」充足上での上記と同様の考え方で以って、 $O_n$  に替わる基準受験者  $O_0$ 、 $A_i$  に替わる基準項目  $A_0$ 、並びに、 $J_j$  に替わる基準評定者  $J_0$  を上式に挿入する。

$$P_{0ijk} / P_{0ijh} = (P_{00jk} / P_{00jh}) * (P_{0i0k} / P_{0i0h}) * (P_{000h} / P_{000k})$$

上に求められた重要な三式は以下のものとして要約される。

1.  $P_{nij} / P_{nijh} = (P_{0ijk} / P_{0ijh}) * (P_{n0jk} / P_{n0jh}) * (P_{00jh} / P_{00jk})$
2.  $P_{0ijk} / P_{0ijh} = (P_{00jk} / P_{00jh}) * (P_{0i0k} / P_{0i0h}) * (P_{000h} / P_{000k})$
3.  $P_{n0jk} / P_{n0jh} = (P_{00jk} / P_{00jh}) * (P_{n00k} / P_{n00h}) * (P_{000h} / P_{000k})$

従って、上記第 1 式に 2 式と 3 式の挿入の上で整理すれば、

$$P_{nij} / P_{nijh} = (P_{n00k} / P_{n00h}) * (P_{0i0k} / P_{0i0h}) * (P_{00jk} / P_{00jh}) * (P_{000h} / P_{000k})^2$$

これが、二つの順序カテゴリー  $k$  と  $h$  ( $k > h$ ) から成る一つの項目  $A_i$  への一人の受験者  $O_n$  の応答に対して一人の評定者  $J_j$  により評定  $k$  あるいは  $h$  が付される確率について Linacre<sup>2)</sup>(1989)により「固有客観性」に基づいて導出されたオッズ表現式(p. 50)である。

3) Linacre<sup>2)</sup>(1989)により、更に、三つ以上の順序カテゴリーから成る一つの項目  $A_i$  について一人の評定者  $J_j$  による受験者  $O_n$  と  $O_m$  の潜在能力比較を示す確率式が考察される(pp. 51-52)。順序カテゴリーのその順序構造で以って、二人の受験者  $O_n$  と  $O_m$  の潜在能力比較上で「固有客観性」充足のためには、隣接評定順序カテゴリー組み合わせの各一对間に評定者  $J_j$  による順序カテゴリー評定選択上での確率において「不変性」の設定が必要とされる(Linacre<sup>2)</sup>, 1989, p. 51)。順序カテゴリー上で  $k = h + 1$ 、並びに、 $k' = h' + 1$  とすれば、「不変性」の設定が次式により示される。

$$\begin{aligned} \text{受験者 } O_n \text{ と } O_m \text{ の潜在能力比較} &= P_{kh} / P_{hk} = (P_{nij} * P_{mijh}) / (P_{nijh} * P_{mijk}) \\ &= (P_{nij'k} * P_{mijh'}) / (P_{nij'h} * P_{mijk'}) \end{aligned}$$

上式を変換して、

$$P_{nij} / P_{nijh} = (P_{mijk} / P_{mijh}) * (P_{nij'k} / P_{nij'h}) * (P_{mijh'} / P_{mijk'})$$

「固有客観性」充足上での上記 2)においてと同じ思考で以って、 $O_m$  に替わる基準受験者  $O_0$ 、 $A_i$  に替わる基準項目  $A_0$ 、並びに、 $J_j$  に替わる基準評定者  $J_0$  を上式に挿入する。

$$P_{n00k} / P_{n00h} = (P_{000k} / P_{000h}) * (P_{n00k'} / P_{n00h'}) * (P_{000h'} / P_{000k'})$$

右辺における項を並べ替えて、

$$P_{n00k} / P_{n00h} = (P_{n00k'} / P_{n00h'}) * (P_{000h'} / P_{000k'}) * (P_{000k} / P_{000h})$$

上式における  $(P_{n00k'} / P_{n00h'}) * (P_{000h'} / P_{000k'})$  は、評定者  $J_j$  による受験者  $O_n$  への順序カテゴリー評定

上での隣接カテゴリー一対に関する生起確率の「不変性」を示している。その二つの項が受験者  $O_n$  に関する順序評定「不変性」標示としての  $P_{n00}$  に置換されて、次式が与えられる。

$$P_{n00k} / P_{n00h} = P_{n00} * (P_{000k} / P_{000h})$$

項目  $A_i$  に関する  $(P_{0i0k} / P_{0i0h})$  と評定者  $J_j$  に関する  $(P_{00jk} / P_{00jh})$  についても、順序評定「不変性」の上でそれぞれ同様に次式が導かれる。

$$P_{0i0k} / P_{0i0h} = P_{0i0} * (P_{000k} / P_{000h}); \quad P_{00jk} / P_{00jh} = P_{00j} * (P_{000k} / P_{000h})$$

従って、Linacre<sup>2)</sup>(1989)による記述構成枠上での 2)(pp. 48-50)において末尾に記された二つの順序カテゴリー  $k$  と  $h (k > h)$  についての式が次の様に変換されて、三つ以上の順序カテゴリー  $(k = h + 1)$  に拡張充当されるものとなる。

$$\begin{aligned} P_{nijk} / P_{nijh} &= (P_{n00k} / P_{n00h}) * (P_{0i0k} / P_{0i0h}) * (P_{00jk} / P_{00jh}) * (P_{000h} / P_{000k})^2 \\ &= [P_{n00} * (P_{000k} / P_{000h})] * [P_{0i0} * (P_{000k} / P_{000h})] \\ &\quad * [P_{00j} * (P_{000k} / P_{000h})] * (P_{000h} / P_{000k})^2 \\ &= P_{n00} * P_{0i0} * P_{00j} * (P_{000k} / P_{000h}) \end{aligned}$$

ここで、上式右辺各項の対数について次の定義が与えられる。

受験者  $O_n$  の潜在能力  $B_n = \ln(P_{n00})$ ;

項目  $A_i$  の潜在困難度  $D_i = -\ln(P_{0i0})$ ;

評定者  $J_j$  の潜在評定難度  $C_j = -\ln(P_{00j})$ ;

順序カテゴリー  $h (= k - 1)$  から  $k$  への潜在段階困難度  $F_k = -\ln(P_{000k} / P_{000h})$

上の定義で以って、上式が次の様に対数変換される。

$$\ln(P_{nijk} / P_{nijh}) = \ln(P_{nijk} / P_{nijk-1}) = B_n - D_i - C_j - F_k$$

これが、Linacre<sup>2)</sup>(1989)によって案出された Many-Facet Rasch Measurement における評定者「共通段階モデル」の対数オッズ表現式(p. 52)である。

更に、上式に基づく  $P_{nijk}$  の Linacre<sup>2)</sup>(1989)による指数表現式導出(pp. 53-54)を以下に参照する。それは、Linacre<sup>2)</sup>(1989)の評定者「共通段階モデル」は Andrich<sup>1)</sup>(1978)によって考案された Rating Scale モデルにおける被験者位置母数と項目位置母数に加えて評定者位置母数挿入の上でのその拡張モデル化であることを以降に示すための序となるものである。

Linacre<sup>2)</sup>(1989)により与えられた評定者「共通段階モデル」の対数オッズ表現式(p. 52)について順序カテゴリー一評定値 0 から  $k$  に渡る隣接カテゴリー一対上での生起確率比較が次の様に示される。

$$\ln(P_{nij1} / P_{nij0}) = \ln(P_{nij1}) - \ln(P_{nij0}) = B_n - D_i - C_j - F_1$$

$$\ln(P_{nij2} / P_{nij1}) = \ln(P_{nij2}) - \ln(P_{nij1}) = B_n - D_i - C_j - F_2$$

$$\ln(P_{nij3} / P_{nij2}) = \ln(P_{nij3}) - \ln(P_{nij2}) = B_n - D_i - C_j - F_3$$

$$\begin{array}{ccc} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{array}$$

$$\ln(P_{nijk-2} / P_{nijk-3}) = \ln(P_{nijk-2}) - \ln(P_{nijk-3}) = B_n - D_i - C_j - F_{k-2}$$

$$\ln(P_{nijk-1} / P_{nijk-2}) = \ln(P_{nijk-1}) - \ln(P_{nijk-2}) = B_n - D_i - C_j - F_{k-1}$$

$$\ln(P_{nij} / P_{nijk-1}) = \ln(P_{nij}) - \ln(P_{nijk-1}) = B_n - D_i - C_j - F_k$$

従って、上式すべてを加算すれば、

$$\ln(P_{nij}) - \ln(P_{nij0}) = \ln(P_{nij} / P_{nij0}) = k(B_n - D_i - C_j) - \sum_{s=1}^k F_s$$

上式を指数表現式に変換して、

$$P_{nij} / P_{nij0} = \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s]$$

上式を  $P_{nij}$  について解けば、

$$P_{nij} = P_{nij0} * \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s]$$

ここで、上式の  $P_{nij}$  は受験者  $n$ 、項目  $i$ 、評定者  $j$  に関する第  $k(=1, 2, \dots, K)$  段階の生起確率であり、 $P_{nij0}$  が  $P_{nij}$  に含まれていないことに留意される。順序カテゴリー評定値  $k(=0, 1, 2, \dots, K)$  に関する生起確率の上で、 $\sum_{k=0}^K P_{nij} = 1$  である。従って、次式が導かれる。

$$\begin{aligned} \sum_{k=0}^K P_{nij} = 1 &= P_{nij0} + \sum_{k=1}^K \{P_{nij0} * \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s]\} \\ &= P_{nij0} * \{1 + \sum_{k=1}^K \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s]\} \end{aligned}$$

上式を  $P_{nij0}$  について解けば、

$$P_{nij0} = 1 / \{1 + \sum_{k=1}^K \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s]\}$$

上式を  $P_{nij} = P_{nij0} * \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s]$  に代入すれば、次式が導かれる。

$$P_{nij} = \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s] / \{1 + \sum_{k=1}^K \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s]\}$$

繰り返しになるが、上式の  $P_{nij}$  は第  $k(=1, 2, \dots, K)$  段階の生起確率であり、 $P_{nij0}$  が  $P_{nij}$  に含まれていない。従って、存在しない第 0 段階の  $F_0$  を 0 とみなして、 $P_{nij}$  を順序カテゴリー評定値  $k(=0, 1, 2, \dots, K)$  の生起確率とする式が以下の様に導出される。

上に示された様に、

$$P_{nij0} = 1 / \{1 + \sum_{k=1}^K \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s]\}$$

上式に  $1 = \exp(0)$  を代入すれば、

$$P_{nij0} = \exp(0) / \{\exp(0) + \sum_{k=1}^K \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s]\}$$

ここで、 $F_0 \equiv 0$  とみなされている故に、上式は、第  $k(=1, 2, \dots, K)$  段階の生起確率  $P_{nij} = \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s] / \{1 + \sum_{k=1}^K \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s]\}$  を参照して、次式に変換される。

$$P_{nij0} = \exp[0(B_n - D_i - C_j) - F_0] / \{ \exp[0(B_n - D_i - C_j) - F_0] + \sum_{k=1}^K \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s] \}$$

$$= \exp[0(B_n - D_i - C_j) - F_0] / \sum_{k=0}^K \exp[k(B_n - D_i - C_j) - \sum_{s=0}^k F_s]$$

$F_0 = 0$  として、上式を  $P_{nij} = P_{nij0} * \exp[k(B_n - D_i - C_j) - \sum_{s=1}^k F_s]$  に代入すれば、 $P_{nij}$  を順序カテゴリー評定値  $k(=0, 1, 2, \dots, K)$  の生起確率とする次式が与えられる。

$$P_{nij} = \exp[k(B_n - D_i - C_j) - \sum_{s=0}^k F_s] / \sum_{k=0}^K \exp[k(B_n - D_i - C_j) - \sum_{s=0}^k F_s]$$

以上が、Linacre<sup>2)</sup>(1989)による評定者「共通段階モデル」の指数表現式導出(pp. 53-54)過程である。なお、読者の理解を容易にするためと想像されるが、Linacre<sup>2)</sup>(1989)は、記号重複を避けて、上式を次式で以って示している(p. 54)。

$$P_{nij} = \exp[k(B_n - D_i - C_j) - \sum_{s=0}^k F_s] / \sum_{h=0}^K \exp[h(B_n - D_i - C_j) - \sum_{s=0}^h F_s]$$

導出過程を含めて上に参照された Linacre<sup>2)</sup>(1989)の評定者「共通段階モデル」は、Andrich<sup>1)</sup>(1978)によって考案された Rating Scale モデルにおける被験者位置母数と項目位置母数に加えて評定者位置母数挿入の上でのその拡張モデル化であることが次の様に示される。それは、Andrich<sup>1)</sup>(1978)考案による Rating Scale モデルの付帯特性としての潜在閾値母数順次増大性が Linacre<sup>2)</sup>(1989)案出の評定者「共通段階モデル」(p. 54)に内在共有されていることを明示するものである。

Andrich<sup>1)</sup>(1978)の Rating Scale モデルにおいては、 $\beta$ 、 $\delta$ 、 $\kappa$  をそれぞれ被験者位置母数、項目位置母数、項目群全体について同数  $k(=0, 1, 2, \dots, m)$  から成る順序カテゴリーの母数として、一人の被験者による評定値  $x(=0, 1, 2, \dots, m)$  の選択確率が次式により表わされる(p. 569)。

$$p\{X = x \mid \beta, \delta, \kappa\} = \exp[\kappa_x + x(\beta - \delta)] / \sum_{k=0}^m \exp[\kappa_k + \kappa(\beta - \delta)]$$

$$\text{なお、} \kappa_x = -\sum_{k=1}^x \tau_k; \kappa_0 = 0 \quad x = 0, 1, 2, \dots, m$$

上式において  $(\beta - \delta)$  に修正項を加えることは可能である(Andrich<sup>5)</sup>, 1997, p. 879)とされている故に、添え字が付された  $(\beta_n - \delta_i)$  に  $(-\lambda_j)$  を加える。

$$P_{xnij} = \exp[\kappa_x + x(\beta_n - \delta_i - \lambda_j)] / \sum_{k=0}^m \exp[\kappa_k + \kappa(\beta_n - \delta_i - \lambda_j)]$$

$$= \exp[-\sum_{k=0}^x \tau_k + x(\beta_n - \delta_i - \lambda_j)] / \sum_{k=0}^m \exp[-\sum_{k=0}^k \tau_k + \kappa(\beta_n - \delta_i - \lambda_j)]$$

$$\text{なお、} \kappa_x = -\sum_{k=1}^x \tau_k; \kappa_0 = 0 \quad x = 0, 1, 2, \dots, m$$

上式の記号を Linacre<sup>2)</sup>(1989, p. 54)が与える表示に変更する。

$$P_{nij} = \exp[-\sum_{s=0}^k F_s + k(B_n - D_i - C_j)] / \sum_{k=0}^K \exp[-\sum_{s=0}^k F_s + h(B_n - D_i - C_j)]$$

$$= \exp[k(B_n - D_i - C_j) - \sum_{s=0}^k F_s] / \sum_{h=0}^K \exp[h(B_n - D_i - C_j) - \sum_{s=0}^h F_s]$$

$$\text{なお、} F_0 \equiv 0 \quad k = 0, 1, 2, \dots, K$$

Andrich<sup>1)</sup>(1978)の Rating Scale モデルは、順序カテゴリー選択確率上での各閾値母数における同一弁別性をその潜在尺度の上で規定している。更に、その構成要素としてのガットマン完全順序尺度性に基づく閾値母数順次増大性  $\tau_1 < \tau_2 < \tau_3, \dots, < \tau_m$  をモデル付帯特性としている。Rating Scale モデル測定出力上でのデータ属性としての閾値母数推定値増大順序性がモデル付帯特性としての潜在閾値母数順次増大性により質問紙調査項目群カテゴリー順序段階尺度化妥当性の上で必然と

される(Andrich, 1978<sup>1)</sup>, 2004<sup>6)</sup>, and 2005<sup>7)</sup>)。この点に関する重要な留意事項は、Rating Scale モデル測定閾値母数推定上でのその潜在閾値順次増大性への準拠・違反はデータの属性である(Andrich<sup>7)</sup>, 2005, p. 49)との認識である。モデル付帯特性としての潜在閾値母数順次増大性に対する閾値母数推定結果としてのその違反データは、質問紙調査設定項目群全体に係わる一元的構成概念上でのカテゴリー順序段階尺度化妥当性に欠けることを示している(Andrich<sup>6)</sup>, 2004, p. 184)。Andrich<sup>7)</sup>(2005)によって Rating Scale モデル測定上での統計的モデル適合度指標は最優先確認事項としての閾値母数推定値増大順序性とは無関連である(p. 53)とも指摘されている。これは、閾値母数推定値増大順序性が Rating Scale モデル測定 Facets (Linacre<sup>8)</sup>, 1989–2001, Winsteps.com)出力上で各項目、各被験者、並びに、各順序カテゴリーに対してモデル適合度指標として付される Infit・Outfit 統計量からは分離独立しているとの示唆である。なお、Rating Scale モデル閾値母数推定値増大順序性は質問紙項目群カテゴリー順序段階尺度化妥当性検証への必要条件ではある(Andrich, 1978<sup>1)</sup>, 2004<sup>6)</sup>, and 2005<sup>7)</sup>)が、潜在特性次元性と局所独立性を併せた構成概念妥当性を保証する十分条件ではないということが井澤<sup>9)</sup>(2007)により Rating Scale モデル適用一事例上で査察されている(pp. 14–21)。

Andrich<sup>1)</sup>(1978)考案による Rating Scale モデルに内在する数理規定が、本節で参照詳述された Linacre<sup>2)</sup>(1989)案出による評定者「共通段階モデル」の原理である。故に、筆者によって Linacre<sup>2)</sup>(1989)にその言明は見出されないけれども、又、Linacre (2001<sup>10)</sup> and 2002<sup>11)</sup>)への参照で以って、Rating Scale モデル規定としての各閾値母数での隣接順序カテゴリー間同一弁別性、並びに、モデル付帯特性としての潜在閾値母数順次増大性が評定者「共通段階モデル」に内在共有されていると理解される。その認知に関する Andrich (1996<sup>12)</sup> and 2005<sup>7)</sup>)参照上での以下に示す証左一端で以って本節の締めとする。

Andrich<sup>1)</sup>(1978)による Rating Scale モデルの対数オッズ表現式は次のものである。

$$\ln(P_{nik} / P_{nik-1}) = \beta_n - \delta_i - \tau_k$$

従って、

$$\ln(P_{nik} / P_{nik-1}) - \ln(P_{nik+1} / P_{nik}) = (\beta_n - \delta_i - \tau_k) - (\beta_n - \delta_i - \tau_{k+1})$$

$$\rightarrow \ln[(P_{nik})^2 / (P_{nik-1} * P_{nik+1})] = \tau_{k+1} - \tau_k$$

$$\rightarrow (P_{nik})^2 / (P_{nik-1} * P_{nik+1}) = \exp(\tau_{k+1} - \tau_k)$$

Rating Scale モデルの付帯特性は $(\tau_{k+1} - \tau_k) > 0$  である故に、上式は次式に変換される。

$$(P_{nik})^2 / (P_{nik-1} * P_{nik+1}) > 1$$

これは、Rating Scale モデルの付帯特性として、 $\tau_k < \beta_n < \tau_{k+1}$  である場合には  $P_{nik}$  が他のどの順序カテゴリー生起確率よりも大きくなることを示している(Andrich<sup>12)</sup>, 1996, p. 21、参照)。それは、Andrich<sup>12)</sup>(1996)により各順序カテゴリー評定値分布の単一形態性(unimodality)と呼称され(p. 21–22)、Rating Scale モデルが構成要素とするガットマン完全順序尺度の充足性を示唆する(Andrich<sup>7)</sup>, 2005, p. 51)ものとなる。前記参照された様に、Linacre<sup>2)</sup>(1989)による評定者「共通段階





図1が、Linacre<sup>2)</sup>(1989)によって言及不在で供与されている図A.6 (p. 12)に対応する当該データの Facets 出力による評定者「共通段階モデル」測定マップである。カテゴリ順序評定データ構成要因のモデル規定分離測定が定める同一関係枠(‘the same frame of reference,’ in Linacre<sup>2)</sup>, 1989, p. 5)で以って、「母数結合一次元性」(Wright<sup>13)</sup>, 1991, p. 158)をロジット(対数オッズ)単位上で示したものである。評定者難度推定値平均と項目困難度推定値平均のそれぞれがロジット尺度値0に設定されており、尺度値が大きくなるにつれて受験者能力、評定者難度、項目困難度のそれぞれが増大する。なお、図1の最右列に付与されている尺度はロ

ジット値に対応する Rasch 測定確率期待値としての順序尺度評定値である(Linacre<sup>2)</sup>, 1989, p. 12、参照)。受験者32名中の30名が-1.0から1.0のロジット値能力範囲内に見出されて、0を中心として狭小に分布する評定者難度推定値と項目困難度推定値に対しての32名能力推定値分布の拡がり具合が目視される。

表2 当該データの評定者「共通段階モデル」測定による Facets 出力上での閾値母数推定値とその標準誤差

順序評定	1	2	3	4	5	6	7	8	9
閾値母数推定値	*	-2.51	-1.65	-0.90	0.04	0.22	0.88	1.17	2.75
標準誤差	*	0.27	0.13	0.08	0.07	0.07	0.09	0.13	0.38

表2に当該データの評定者「共通段階モデル」測定による Facets 出力上での閾値母数推定値とその標準誤差を与えており、閾値母数推定値増大順序性への完全準拠が観察される。Rating Scale モデル測定上での隣接閾値母数推定値間の差については、Linacre<sup>11)</sup>(2002)により、項目群について三つのカテゴリから成る二段階順序尺度構成データと五つのカテゴリから成る四段階順序尺度構成データにおけるその差のモデル期待最小値はそれぞれ1.4ロジット値、1.0ロジット値である(pp. 102-103)と指摘されている。その参照に基づく九つのカテゴリから成る八段階順序尺度構成データにおける隣接閾値母数推定値間の差のモデル期待最小値は0.6ロジット値前後と単純に類推される。当該データ隣接閾値母数推定値間でのその0.6ロジット値の差を満たさない隣接順序評定カテゴリ組み合わせは5と6の対と7と8の対の二組であるが、その差はそれぞれ0.18と0.29であり0.6と較べて大変に小さいという値でもない。従って、カテゴリ数の多さをも踏まえれば、一元的論文評定構成概念上でのかなり高い順序評定段階尺度化妥当性が当該データについて窺われる。又、閾値母数推定値の標準誤差は信頼性指標であり(Linacre<sup>2)</sup>, 1989, p. 11)、

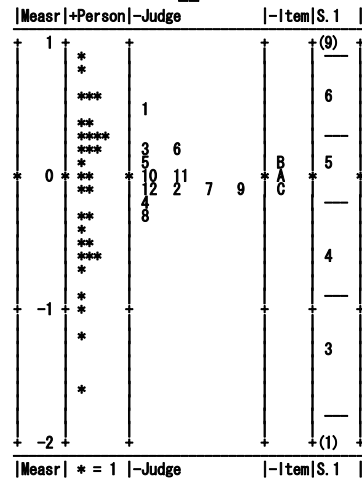


図1 Facets 出力による当該データの評定者「共通段階モデル」測定マップ

当該データ構成要因母集団での高い再現性が当該順序評定尺度構成に示唆されている。特に、順序評定値 4、5、6、7 のそれぞれに付された閾値母数推定値にかなり高いその再現性が認められる。

下表 3 は、当該データへの評定者「個別段階(judge-step)モデル」測定適用上での Facets 出力による評定者 12 名それぞれについての各閾値母数推定値とその増大順序性を示すものである。記号 ○ はモデル付帯特性としての潜在閾値母数順次増大性に準拠している評定者であり、× は閾値母数推定値が少なくとも一つの順序評定段階において潜在閾値母数順次増大性に反している評定者である。その×が付された評定者個別に値が一つ前の閾値母数推定値より小さくなっている箇所を薄い灰色で表示している。なお、参考までに、評定者「共通段階モデル」と評定者「個別段階モデル」それぞれによる測定出力上での各評定者難度位置母数推定値をも併せて表提示に含めている。

表 3 当該データの評定者「個別段階モデル」測定 Facets 出力上での閾値母数推定値とその増大順序性

評定者番号	1	2	3	4	5	6	7	8	9	10	11	12	平均値
評定者共通段階モデル位置母数推定値	0.45	-0.08	0.18	-0.16	0.10	0.17	-0.12	-0.30	-0.15	0.00	-0.04	-0.06	0.00
評定者個別段階モデル位置母数推定値	0.18	-0.15	0.20	0.09	0.07	0.03	-0.28	-0.40	-0.14	0.02	0.41	-0.01	0.00
評定者個別段階モデル閾値母数推定値	*												
閾値順次増大性	×	×	×	○	○	○	×	×	×	×	×	×	*
1	*	*	*	*	*	*	*	*	*	*	*	*	*
2	-2.65	-1.54	-2.49	*	-2.42	-1.66	-2.11	-2.34	-1.72	-1.21	*	-2.49	-2.06
3	-1.78	-0.82	-2.23	-1.98	-1.74	-1.16	-1.19	-0.28	-0.65	-1.83	-1.31	-1.67	-1.39
4	-0.82	-1.25	-0.20	-1.60	-0.48	-0.29	-0.22	-0.58	-1.15	-1.09	-0.89	-0.15	-0.73
5	0.87	0.69	0.39	-0.32	-0.02	0.04	0.02	0.31	0.66	0.27	-0.37	0.67	0.27
6	1.14	-0.03	1.04	0.38	0.72	0.89	0.71	0.45	-0.04	0.26	0.08	0.33	0.49
7	1.90	1.97	-0.07	1.41	0.98	1.03	1.47	1.05	0.94	1.41	0.54	0.80	1.12
8	1.33	0.98	1.31	2.11	2.96	1.16	1.30	1.38	1.95	2.19	1.56	0.45	1.56
9	*	*	2.24	*	*	*	*	*	*	*	0.39	2.07	1.57

評定者「個別段階モデル」は、Rasch 測定上で質問紙調査順序尺度データに適用される項目「個別段階モデル」に相当する Masters<sup>14)</sup>(1982)考案の Partial Credit モデルが Linacre<sup>2)</sup>(1989)により Many-Facet Rasch Measurement として評定者順序評定に拡張されたものである。Partial Credit モデルでも、Andrich<sup>1)</sup>(1978)の Rating Scale モデルと同様に、Rasch 項目分析モデル(Rasch<sup>15)</sup>, 1960)を特徴づける「固有客観性」が隣接カテゴリー間での段階における規定とされている(平越<sup>17)</sup>, 2009)。然しながら、Rasch 項目分析モデルと Rating Scale モデルの属性である項目弁別力同一性ないしは項目特性曲線平行性が Partial Credit モデルにおいては属性とされてはいない(Andrich<sup>16)</sup>, 1988; 井澤<sup>18)</sup>, 2008)。従って、項目群と受験者群の母集団に関する構成尺度普遍性ないしは母数推定値再現性に係わる推察妥当性の観点上で、質問紙調査順序尺度データへの適用に際して Rating Scale モデルが Partial Credit モデルに優る。同様の論理が評定者「共通段階モデル」と評定者「個別段階モデル」の関係にも正しく当てはまる。但し、Partial Credit モデルと評定者「個別段階モデル」による測定出力への参照有意義性は一元的に妥当性と信頼性の高い測定尺度を「作る」[Wright and

Stone<sup>19)</sup>(2004)の書名 *Making Measures* 参照]上での参照資料価値にあるとの筆者理解である。それが表 3 提示の意図である。

参考として、評定者「共通段階」と評定者「個別段階」の両モデルによる測定出力上での表 3 に付加された評定者難度位置母数推定値間でのピアソン相関係数とケンドール順位相関係数  $\tau_{\text{b}}$  の値はそれぞれ 0.590、0.515 である。井澤<sup>9)</sup>(2007)における表 1 (p. 16)での Rating Scale と Partial Credit の両モデル測定 Facets 出力による 15 項目位置母数推定値間でのそれぞれの相関係数の大変に高い値 0.988、0.981 と非常に異なる結果が上に示されている。その異なる結果を生じさせている正確な原因は筆者に不詳であるが、Many-Facet Rasch Measurement における母数推定変数増大に起因していると推測される。当該データに関する分析推察妥当性の観点からは、上記の評定者「共通段階モデル」優位性によりその測定出力推定値が参照されて然るべきである。なお、評定者「共通段階」と評定者「個別段階」の両モデル測定出力上での当該データ受験者群 32 名位置母数推定値についての相関係数は 1.00 であり、井澤<sup>9)</sup>(2007)での Rating Scale と Partial Credit の両モデル測定出力上での質問紙調査被験者群 692 名位置母数推定値に関するものと同じである。

前掲表 2 における評定者「共通段階モデル」測定出力上での閾値母数推定値増大順序性の観察により当該評定データ一元的構成概念上でのかなり高い順序評定段階尺度化妥当性が窺われたけれども、表 3 はその信憑性を覆すものである。評定者「個別段階モデル」付帯特性としての閾値母数順次増大性に準拠している評定者は僅かに三人であることが示されている。なお、評定者「個別段階モデル」測定上での閾値母数順次増大性については、以下に参照引用される Partial Credit モデル測定上での「項目」を「評定者」に置き換えた「評定者弁別力」に関するものと理解される。

- 1) 『閾値母数推定値の増大順序性が保持されていることを以ってして、その値の差が大きくなればなる程、その項目弁別力は順じて小さくなる』(Andrich<sup>16)</sup>, 1988, p. 370)。
- 2) 『閾値母数推定値間の差が小さくなればなる程、その項目弁別力は順じて大きくなる』(Andrich<sup>16)</sup>, 1988, p. 370)。
- 3) 『閾値母数推定値増大順序性に違反があれば、その項目弁別力は更に大きくなる』(Andrich<sup>16)</sup>, 1988, p. 370)。
- 4) 『閾値増大順序性への違反は、その項目が潜在特性に関する限られた範囲の上で過度に弁別的であることの示唆である』(Linacre<sup>20)</sup>, 1999, p. 675)。

前掲表 2 について閾値母数推定値間で 0.6 ロジット値の差を満たさない隣接順序評定カテゴリ一組み合わせは 5 と 6 の対と 7 と 8 の対の二組であると言及されたが、表 3 においてはその二組に加えて順序評定値 3 と 4 の尺度統合が示唆されている。更に、順序評定値 9 は僅かに三人の評定者のみによって使用されているに過ぎず、その三人の中の評定者番号 11 については順序評定値 9 に閾値順次増大性違反が観察される。従って、当該データ順序評定値上での 3 と 4 の統合、5 と 6 の統合、7 と 8 と 9 の統合に基づく順序評定値 1、2、3、4、5 から成る四段階尺度構成に妥当性が高いとの示唆である。更に、当該元データ上での評定値 1 の生起度数が 15 ときわめて微小

であり、評定者番号 4 と 11 により評定値 1 が付与されていないことから、3 と 4 の統合、5 と 6 の統合、7 と 8 と 9 の統合に加えて、評定値 1 と 2 の統合に基づく順序評定値 1、2、3、4 から成る三段階尺度構成妥当性も検討される必要がある。以下に、順序評定値統合上でのその四段階データと三段階データの閾値母数推定値増大順序性を観察する。その関連の上で、元データである Linacre<sup>2)</sup>(1989)提示の順序評定八段階データ(p. 11)、及び、上記の要領で作成された順序評定値統合上での四段階データと三段階データの潜在特性次元充足度を次節において比較考察する。

表 4 四段階データの評定者「共通段階モデル」測定 Facets 出力上での閾値母数推定値と標準誤差

順序評定	1	2	3	4	5
閾値母数推定値	*	-1.97	-1.70	1.03	2.65
標準誤差	*	0.27	0.13	0.07	0.09

表 5 三段階データの評定者「共通段階モデル」測定 Facets 出力上での閾値母数推定値と標準誤差

順序評定	1	2	3	4
閾値母数推定値	*	-2.30	0.33	1.97
標準誤差	*	0.13	0.07	0.09

表 6 四段階データの評定者「個別段階モデル」測定 Facets 出力上での閾値母数推定値とその増大順序性

	評定者番号	1	2	3	4	5	6	7	8	9	10	11	12	平均値
	評定者共通段階モデル位置母数推定値		0.75	0.01	0.20	-0.30	0.09	0.27	-0.24	-0.51	-0.19	0.03	-0.06	-0.06
評定者個別段階モデル位置母数推定値		0.44	0.02	-0.15	0.14	-0.09	0.22	-0.30	-0.48	-0.15	0.06	0.60	-0.30	0.00
	評定者個別段階モデル閾値母数推定値													*
	閾値順次増大性	○	×	×	○	○	○	○	○	○	×	○	○	*
評定者順序評定	1	*	*	*	*	*	*	*	*	*	*	*	*	*
	2	-2.69	-1.47	-1.97	*	-2.07	-1.63	-1.91	-2.06	-1.48	-1.02	*	-2.02	-1.83
	3	-2.48	-1.88	-2.01	-2.97	-1.89	-1.57	-1.51	-0.81	-1.46	-2.53	-1.85	-1.57	-1.88
	4	1.55	0.60	1.86	-0.01	0.94	0.77	0.67	0.68	0.75	0.63	0.10	1.69	0.85
	5	3.61	2.75	2.12	2.98	3.03	2.43	2.75	2.19	2.19	2.92	1.75	1.90	2.55

表 4 と 6 の対、並びに、表 5 と 7 の対が、それぞれ、Linacre<sup>2)</sup>(1989)提示の順序評定八段階データ(p. 11)から上に記された方法で新たに作成された四段階データと三段階データの評定者「共通段階」と評定者「個別段階」の両モデル測定による Facets 出力上での閾値母数推定値に関する情報を与えるものである。表 4 と 5 の提示法、及び、表 6 と 7 の提示法は、それぞれ、Linacre<sup>2)</sup>(1989)により与えられた順序評定八段階データ(p. 11)に関する前掲表 2、及び、表 3 と同じである。

前掲表 3、及び、表 6 と 7 のそれぞれにおける評定者に関するモデル付帯特性としての閾値母数順次増大性への準拠を示す記号○の数により、八段階元データ、及び、評定値統合四段階データと比較の上での新規作成三段階データの当該データに関する順序段階尺度化妥当性の上での秀逸性が一目瞭然である。表 4 と 5 に目を転じれば、表 4 においては順序評定値 2 から 3 への閾値母数推定値増大度として 1.0 に満たない 0.27 の値が観察されるが、表 5 における隣接閾値母数推定値間での差の値はそれぞれ 1.5 を超えて 2.63 と 1.64 である。前記に参照された様に、Linacre<sup>11)</sup>(2002)により、三つのカテゴリーから成る二段階順序尺度構成データと五つのカテゴ

リーから成る四段階順序尺度構成データについての Rating Scale モデル測定上での隣接閾値母数推定値間での差のモデル期待最小値はそれぞれ 1.4 ロジット値、1.0 ロジット値である(pp. 102-103)と指摘されている。従って、表 4 と 5 によっても、三段階データの四段階データに優る順序段階尺度化妥当性が顕示されている。又、三段階データのその秀逸な順序段階尺度化妥当性の示唆一端として、表 5 での評定者「共通段階モデル」測定上での各閾値母数推定値と表 7 での評定者「個別段階モデル」測定上での 12 名閾値母数推定値に関する各平均値との顕著な近似性が観察される。

表 7 三段階データの評定者「個別段階モデル」測定 Facets 出力上での閾値母数推定値とその増大順序性

	評定者番号	1	2	3	4	5	6	7	8	9	10	11	12	平均値
		評定者共通段階モデル位置母数推定値	0.84	-0.02	0.22	-0.30	0.10	0.24	-0.25	-0.53	-0.23	0.01	-0.04	-0.06
	評定者個別段階モデル位置母数推定値	0.81	0.04	-0.01	-0.45	0.08	0.29	-0.18	-0.32	-0.13	-0.02	0.04	-0.15	0.00
	評定者個別段階モデル閾値母数推定値													*
	閾値順次増大性	○	○	○	○	○	○	○	○	○	○	○	○	*
評定者順序評定	1	*	*	*	*	*	*	*	*	*	*	*	*	*
	2	-3.32	-2.21	-2.60	-3.05	-2.52	-1.96	-2.09	-1.47	-1.83	-2.60	-1.89	-2.19	-2.31
	3	0.63	0.02	1.18	0.02	0.22	0.15	0.01	-0.02	0.19	0.15	0.12	0.99	0.31
	4	2.69	2.19	1.43	3.02	2.31	1.81	2.09	1.49	1.64	2.45	1.77	1.20	2.01

参照として興味ある傾向一端を示すものが表 8 である。それは、前掲表 3、6、7 における各データへの評定者「共通段階」と評定者「個別段階」の両モデル測定による評定者 12 名位置母数推定値の分布指標値、並びに、各データ両モデル測定間でのその推定値の相関係数である。なお、各平均値は同一値 0 であるから表 8 には含めていない。八段階データ、四段階データ、三段階データの順に僅かにしても評定者位置母数推定値分布幅が大きくなっていることが観察される。これにより、両モデル測定上での評定者位置母数推定値ばらつき程度の大きさと評定者「個別段階モデル」測定上での閾値母数推定値増大順序性充足度の呼応傾向がこの一事例上で窺われる。但し、三つのデータすべてにおいて評定者位置母数推定値分布幅がかなりの程度に狭小であることに注目される。

各データ両モデル測定上での評定者 12 名位置母数推定値の相関係数については、三段階データについての顕著にかなり高い相関係数が他の二つのデータについてのその対応値と比較の上で顕示されている。それは、三段階データに関して上述された表 5 での評定者「共通段階モデル」測定上での各閾値母数推定値と表 7 での評定者「個別段階モデル」測定上での 12 名閾値母数推定値に関する各平均値の顕著な近似性に対応していると推測される。なお、八段階データに下回る四段階データについての低い評定者位置母数推定値相関係数には留意される。従って、両モデル測定上での評定者位置母数推定値相関係数と評定者「個別段階モデル」測定上での閾値母数推定値増大順序性充足度の関係は一概に高低整合するものではないと示されている。

表 8 八段階元データ、四段階データ、三段階データそれぞれの「共通段階モデル」と「個別段階モデル」測定 Facets 出力上での評定者 12 名位置母数推定値の分布指標値、並びに、各データ両モデル測定間での評定者位置母数推定値相関係数

評定者番号		最小値	最大値	標準偏差	ピアソン相関係数	ケンドール順位相関係数 $\tau_b$
8 段階(元)データ	評定者共通段階モデル位置母数推定値	-0.30	0.45	0.20	0.590	0.515
	評定者個別段階モデル位置母数推定値	-0.40	0.41	0.22		
4 段階データ	評定者共通段階モデル位置母数推定値	-0.51	0.75	0.32	0.574	0.388
	評定者個別段階モデル位置母数推定値	-0.48	0.60	0.31		
3 段階データ	評定者共通段階モデル位置母数推定値	-0.53	0.84	0.35	0.938	0.779
	評定者個別段階モデル位置母数推定値	-0.45	0.81	0.32		

本節で明確にされたことは、Linacre<sup>2)</sup>(1989)提示の順序評定八段階データ (p. 11)への Many-Facet Rasch Measurement 適用上での順序評定値統合三段階データのその元データと順序評定値統合四段階データに優る秀逸な順序段階尺度化妥当性である。但し、その三段階データの高い順序段階尺度化妥当性は評定者「共通段階モデル」測定上での閾値母数推定値増大順序性準拠、隣接閾値母数推定値間での差の値として 1.5 を超える評定者弁別力準許容性、並びに、評定者「個別段階モデル」測定上での評定者閾値母数推定値増大順序性の完全充足に基づく判定である。Rasch 評定尺度モデル上での評定者特性曲線想定によるその示唆は、評定者「共通段階モデル」において規定されている評定者弁別力同一性が、その規定を属性としない評定者「個別段階モデル」測定においてもかなり高い程度に満たされているということのみである(Andrich<sup>16)</sup>, 1988, 参照)。

Andrich<sup>7)</sup>(2005)によって Rating Scale モデル測定上での統計的モデル適合度指標は最優先確認事項としての閾値母数推定値増大順序性とは無関連である(p. 53)と指摘されている。これは、閾値母数推定値増大順序性が評定者「共通段階モデル」測定 Facets 出力上で各項目、各被験者、各評定者、並びに、各順序カテゴリーに対してモデル適合度指標として付される Infit・Outfit 統計量からは分離独立しているとの示唆である。それに関係する実証研究考察結果として、Rating Scale モデル測定適用上での閾値母数推定値増大順序性観点に基づく評定値統合順序尺度化最良データ探索精査における Infit・Outfit 統計量の無用性が<sup>5)</sup> Zhu, Updyke, and Lewandowski<sup>21)</sup>(1997)によって報告されている。その明示一事例上で、評定値統合順序尺度化最良データは必ずしも Infit・Outfit 統計量基準によるモデル適合度上での最良データではないと付言されている(Zhu, Updyke, and Lewandowski<sup>20)</sup>, 1997, p. 300)。Infit・Outfit 統計量基準のみならず位置母数推定値「不変性」基準によるデータ次元充足度と閾値母数推定値増大順序性充足度の連関性有無について Many-Facet

Rasch Measurement 適用上での一考察を次節に与える。

#### IV. Linacre(1989)提示の順序評定八段階データ(p. 11)、順序評定値統合上での四段階データ、及び、三段階データの潜在特性一次元充足度比較考察

Rasch 測定理論の真髄は「不変性」(invariance)である(Andrich<sup>7)</sup>, 2005, p. 53)。それは、モデル基底数理としての「母数分離定理」(Rasch<sup>15)</sup>, 1960, pp. 171-178)から概念化された「固有客観性」(Rasch<sup>4)</sup>, 1977, p. 66)の実体であり、Rasch<sup>22)</sup>(1961)の理念として表現された「測定のためのモデル」(pp. 331-332)の根拠となっている。それは、又、Wright<sup>13)</sup>(1991)によって定義されたモデル規定「母数結合一次元性」(p. 158)に由来するデータの普遍的尺度構成妥当性への本質条件としての潜在特性一次元性充足によって顕現される特性でもある。0・1 データと質問紙調査順序尺度データの一次元充足度確認法としての Rasch 測定母数推定値不変性成立程度の検証法は項目群と受験者群それぞれについての二分割データによってなされる。項目群折半各個別データ Rasch 測定上での同一受験者群能力推定値、及び、受験者群折半各個別データ Rasch 測定上での同一項目群困難度推定値の各一群についての相関係数出力観察が簡便な方法である(Henning<sup>23)</sup>, 1988)。各折半データ上での Rasch 測定母数推定値一対個別に推測統計学上の t 値算出で以って、絶対値 2 以上の値が付された推定値個数の全数に対する割合を観察するのが精密な方法である(Bond and Fox<sup>24)</sup>, 2001, pp. 56-57; McNamara<sup>25)</sup>, 1991; Smith, R. M. and Suh<sup>26)</sup>, 2003; Wright and Stone<sup>27)</sup>, 1979, pp. 112-117)。但し、後者二分割標準 t 値の顕著なデータ折半法依存性が 0・1 データ一事例上で示されており(井澤<sup>28)</sup>, 2006a, pp. 48-50)、その方法を Rasch モデル不適合項目群検出に利用することは筆者に疑念を抱かせる。母数推定値不変性成立程度検証法に具されるデータ折半法として識者により簡易最善最有効とされているものが項目群と受験者群それぞれに関する Rasch 測定標準化残差主成分分析上での第 1 主成分負荷量正負分別に基づく折半法である(Linacre<sup>29)</sup>, 1998; Smith<sup>30)</sup>, E. V., 2002; Tennant and Pallant<sup>31)</sup>, 2006; Wright and Stone<sup>19)</sup>, 2004, p. 21)。その折半法の他に優る有効性が 0・1 データ一事例上で井澤<sup>32)</sup>(2005a)によっても確認されている。

0・1 データ、並びに、質問紙調査順序尺度データとは異なり、Many-Facet Rasch Measurement における評定者数、受験者数、及び、項目数にはその尺度構成の上で標本としての大きな限界があり、一群それぞれについての母集団想定と無作為抽出性に難がある。これが、Many-Facet Rasch Measurement 上での潜在特性一次元充足度確認を無視させる理由の一つとも考えられる。従って、評定データ構成概念妥当性が一般的には Rasch モデル適合度指標としての Infit・Outfit 統計量のみによって確認される(Linacre<sup>2)</sup>, 1989, p. 11、参照)傾向が窺われる。然しながら、Rasch 測定理論の真髄としての「不変性」(Andrich<sup>7)</sup>, 2005, p. 53)は Many-Facet Rasch Measurement 上での評定者「共通段階モデル」においても当てはまるとの筆者理解である。故に、評定者群と受験者群の母集団代表性と無作為抽出性についての問題内在を承知の上で、Linacre<sup>2)</sup>(1989)提示の順序評定八段階データ(p. 11)、及び、前節で作成された順序評定値統合上での四段階データと三段階データの潜



在特性次元充足度を本節で比較考察する。又、Rating Scale モデル測定上での統計的モデル適合度指標は最優先確認事項としての閾値母数推定値増大順序性とは無連関であるとの Andrich<sup>7)</sup> (2005)による指摘(p. 53)を踏まえて、各データの評定者「共通段階モデル」測定 Facets 出力に基づく Infit・Outfit  $\chi^2$  平均平方統計量上でのモデル適合度比較観察をも含める。

受験者群、評定者群、及び、項目群から成る順序評定尺度データは三つの構成要因から成っており、前節掲示表 1 での評定値配置から察知される様に、その Rasch 測定標準化残差の主成分分析第 1 主成分負荷量正負に基づく各構成要因折半データの簡易な作成が不可能となっている。更に、本考察に使用している Linacre<sup>2)</sup>(1989)提示の順序評定八段階データ(p. 11)の項目数は三個であり、その項目群折半データの作成は無意味である。従って、折半データの対象は評定者群と受験者群に限定される。それぞれについての単純で機械的な方法による折半データに併せて、Rasch 測定標準化残差の使用に替わる素点上での主成分分析負荷量正負に基づく折半データによる母数推定値不変性成立程度の確認が必要と判断されるならば、その折半データの作成要領は次のものとなる。各評定者による各受験者への三つの項目評定値の合計点数を各評定者と各受験者の組み合わせの下での個別評定値とすれば、評定者群と受験者群のそれぞれについて素点上での主成分分析が可能となる。受験者群、評定者群、及び、項目群から成る順序尺度評定素点データへのこの種の主成分分析適用妥当性についての識者による見解を参照文献上で筆者は眼にしていない。受験者群と項目群の二要因から成る質問紙調査順序尺度データに関しては、Wright<sup>33)</sup>(1996)により、線形性を保持しない順序尺度データへの因子分析適用は不適切であり、第 1 因子抽出後のその非線形残差に基づく第 2 因子抽出は局所独立性欠如の性質を帯びている可能性が高い(p. 509)と指摘されている。一方、0・1 データへの次元項目応答理論分析適用に際しては、事前にその適用妥当性を視認する上での潜在特性次元充足度確認法として項目間のテトラコリック相関係数に基づく因子分析主因子法初期解による第 1 因子の寄与率とその突出程度が一般的に検討される(Ohtomo et al., 1987<sup>34)</sup> and 1989<sup>35)</sup>; 芝<sup>36)</sup>編、1991, p. 169; 渡辺・野口<sup>37)</sup>、1999, pp.23-24)。芝<sup>36)</sup>編(1991)では、第 1 固有値が全体の 50%前後 (p. 169)、Ohtomo<sup>34)</sup> et al. (1987)では 33.9% (p. 115)により因子数が 1 と判定されており、20%がその下限と見なされている(Reckase<sup>38)</sup>、1979, p. 228; 豊田<sup>39)</sup>編、2002, p. 30)。従って、当該データの評定者群と受験者群についての素点上での主成分分析負荷量正負に基づく折半データ作成以前に、因子分析主因子法初期解と同等である主成分分析上での評定者群と受験者群それぞれについての第 1 固有値の大きさは各一群に関する次元充足度の目安となる。更に、本考察に使用している Linacre<sup>2)</sup>(1989)提示の順序評定八段階データ(p. 11)の項目数は三個であるからその項目群折半データの作成は無意味であるが、各受験者に関する各項目上での評定者 12 名による評定値の合計点数を各項目と各受験者の組み合わせの下での個別評定値とすれば、項目群について素点上での主成分分析が可能となる。Linacre<sup>2)</sup>(1989)提示の順序評定八段階データ(p. 11)、及び、前節で作成された順序評定値統合上での四段階データと三段階データに関して、評定者群、受験者群、及び、項目群それぞれについての上記素点評定値新規作

成に基づく主成分分析上での第1固有値の大きさとその寄与率を与えたものが表9である。なお、クロンバックのアルファ係数をも参照表示として加えており、出力はすべてSPSS (Version 15.0, 2006)によるものである。

表9 八段階(元)データ、四段階データ、及び、三段階データそれぞれにおける評定者群、受験者群、項目群についての素点上での主成分分析第1固有値、その寄与率、及び、クロンバックの $\alpha$

データ類	対象群	第1固有値	第1固有値寄与率 (%)	クロンバックの $\alpha$
8段階(元)データ	評定者群(12名)	8.89	74.1	0.966
	受験者群(32名)	8.46	26.4	0.880
	項目群(3個)	1.97	65.6	0.736
4段階データ	評定者群(12名)	8.39	69.9	0.959
	受験者群(32名)	7.34	22.9	0.838
	項目群(3個)	1.97	65.5	0.727
3段階データ	評定者群(12名)	8.34	69.5	0.958
	受験者群(32名)	7.31	22.8	0.841
	項目群(3個)	2.02	67.3	0.755

表9における各データ内での評定者群、受験者群、及び、項目群に関する第1固有値寄与率への目視により、いずれのデータについても評定者群と項目群に関する素点上での第1主成分統合性が65%を超えており、両群それぞれ素点上での一次元充足度はかなり高いと判断される。一方、受験者群に関する第1主成分統合性は八段階データにおける26.4%を最大値としており、いずれのデータにおいても受験者群能力一次元充足度の低さが窺われて、各データ受験者群全体としての評定者「共通段階モデル」への低い適合度が予見される。表9に関して筆者に興味を持たれることは次の二点である。前掲表3、6、7に示された評定者「個別段階モデル」測定閾値増大順序性準拠についての八段階データと四段階データに圧倒的に上回る三段階データのその閾値増大順序完壁性が、表9での八段階データと四段階データに僅かに下回る三段階データの素点上での評定者群第1主成分統合性に反映されていないことである。又、項目群に関する素点上での第1主成分統合性については、三段階データが八段階データと四段階データに微小ながらも優っていることが観察される。なお、クロンバックのアルファ係数については、その性質としてデータ列数増大とデータ列間相関度増大につれて大きな値になることが知られている(Zhu, Updyke, and Lewandowski<sup>21)</sup>, 1997, 参照)。参照として表9に付加されたその値群からは、データ列数とその基準値0.80から判断して、いずれのデータについても評定者群、受験者群、及び、項目群における素点上での列間相関度に基づく内的一貫性の程度はかなり高い。然しながら、各データ受験者群についての低い素点主成分分析第1固有値寄与率との比較の上で、その内的一貫性指標値の過大評価性が顕著に認められる。

評定者群と受験者群それぞれについての単純で機械的な方法による折半データに併せて、Rasch測定標準化残差の使用に替わる素点上での主成分分析負荷量正負記号に基づく折半データによる母数推定値不変性成立程度の確認必要性を前提として、その折半データ作成に具されるものが表

10 と 11 である。それぞれの表に第 1 主成分負荷量をも参照として含めているが、その意味する第 1 主成分と各行要素とのピアソン相関係数と前掲表 9 での各データにおける評定者群と受験者群それぞれについての第 1 固有値寄与率との相対関係の顯示がその提示意図である。

表 10 八段階(元)データ、四段階データ、及び、三段階データそれぞれにおける評定者群素点主成分分析上での第 1 主成分と第 2 主成分の負荷量

8 段階データ評定者群 主成分分析			4 段階データ評定者群 主成分分析			3 段階データ評定者群 主成分分析		
評定者 番号	主成分負荷量		評定者 番号	主成分負荷量		評定者 番号	主成分負荷量	
	第 1 主成分	第 2 主成分		第 1 主成分	第 2 主成分		第 1 主成分	第 2 主成分
1	0.726	0.047	1	0.688	0.527	1	0.704	0.513
2	0.909	-0.082	2	0.908	-0.084	2	0.892	-0.101
3	0.860	0.166	3	0.810	0.106	3	0.805	0.129
4	0.851	-0.044	4	0.823	-0.118	4	0.823	-0.117
5	0.739	0.465	5	0.747	0.364	5	0.737	0.414
6	0.897	0.204	6	0.863	0.153	6	0.848	0.148
7	0.795	-0.463	7	0.802	-0.320	7	0.805	-0.283
8	0.872	-0.027	8	0.829	0.140	8	0.822	0.073
9	0.926	-0.215	9	0.880	-0.241	9	0.890	-0.269
10	0.933	-0.166	10	0.926	-0.138	10	0.924	-0.132
11	0.899	0.131	11	0.873	-0.039	11	0.879	0.003
12	0.891	0.040	12	0.857	-0.193	12	0.847	-0.216

表 11 八段階(元)データ、四段階データ、及び、三段階データそれぞれにおける受験者群素点主成分分析上での第 1 主成分と第 2 主成分の負荷量

8 段階データ受験者群 主成分分析			4 段階データ受験者群 主成分分析			3 段階データ受験者群 主成分分析		
受験者 番号	主成分負荷量		受験者 番号	主成分負荷量		受験者 番号	主成分負荷量	
	第 1 主成分	第 2 主成分		第 1 主成分	第 2 主成分		第 1 主成分	第 2 主成分
1	0.386	-0.472	1	0.123	-0.239	1	0.126	-0.182
2	-0.142	-0.286	2	-0.170	-0.306	2	-0.189	-0.380
3	0.430	0.430	3	0.516	0.252	3	0.509	0.227
4	0.504	0.692	4	0.371	0.575	4	0.359	0.517
5	0.331	-0.067	5	0.366	-0.011	5	0.355	0.044
6	0.514	0.215	6	0.620	0.099	6	0.610	0.067
7	0.771	0.426	7	0.672	0.630	7	0.678	0.604
8	0.610	0.349	8	0.534	0.473	8	0.522	0.507
9	0.806	-0.151	9	0.811	-0.149	9	0.802	-0.168
10	0.742	-0.567	10	0.780	-0.189	10	0.798	-0.198
11	0.041	0.094	11	-0.152	0.042	11	-0.161	-0.009
12	0.248	0.028	12	0.365	-0.036	12	0.356	0.001
13	0.509	-0.476	13	0.384	-0.664	13	0.365	-0.693
14	-0.201	0.693	14	0.054	0.683	14	0.065	0.726
15	0.185	0.807	15	0.104	0.799	15	0.139	0.817
16	0.484	-0.547	16	0.576	-0.614	16	0.556	-0.635
17	0.575	-0.119	17	0.538	-0.190	17	0.494	-0.085
18	0.640	0.054	18	0.443	-0.012	18	0.476	-0.043
19	0.133	-0.513	19	-0.238	-0.147	19	-0.213	-0.162
20	0.800	-0.117	20	0.708	-0.106	20	0.714	-0.066
21	0.426	0.022	21	0.315	-0.028	21	0.428	-0.029
22	0.613	-0.497	22	0.441	-0.672	22	0.460	-0.629
23	0.671	-0.459	23	0.542	-0.663	23	0.546	-0.680
24	0.022	-0.083	24	0.040	-0.104	24	-0.064	-0.240
25	0.661	0.288	25	0.460	0.345	25	0.463	0.324
26	0.564	-0.732	26	0.578	-0.578	26	0.595	-0.580
27	0.638	0.415	27	0.714	0.218	27	0.709	0.236
28	0.003	-0.129	28	-0.308	-0.025	28	-0.215	0.034
29	0.399	0.533	29	0.187	0.522	29	0.201	0.481
30	0.713	0.382	30	0.646	0.405	30	0.657	0.379
31	0.351	0.576	31	0.232	0.500	31	0.141	0.573
32	0.568	0.138	32	0.655	0.046	32	0.639	-0.004

前述した様に、受験者群、評定者群、及び、項目群から成る順序評定尺度データは三つの構成要因から成っており、その Rasch 測定標準化残差の主成分分析第 1 主成分負荷量正負に基づく各構成要因折半データの簡易な作成が不可能である。0・1 データ、並びに、質問紙調査順序尺度データにおいて母数推定値不変性成立程度を視る上で簡易最善最有効とされている項目群と受験者群それぞれに関する標準化残差主成分分析上での第 1 主成分負荷量正負分別に基づく折半法 (Linacre<sup>28</sup>, 1998; Smith<sup>29</sup>, E. V., 2002; Tennant and Pallant<sup>30</sup>, 2006; Wright and Stone<sup>18</sup>, 2004, p. 21) に関して、その第 1 主成分が素点データからの Rasch 次元抽出後の主たる Rasch 測定誤差複合物であるとみなされる (Linacre<sup>28</sup>, 1998; Wright<sup>32</sup>, 1996)。従って、項目群と受験者群それぞれに関する素点データ上での主成分分析における第 2 主成分負荷量が上記 Rasch 測定標準化残差主成分分析上での第 1 主成分負荷量にその両者異同程度無視の上でほぼ対応する。当該各データについての母数推定値不変性成立程度を視る上でのデータ折半法に表 10 と 11 が使用されるべき根拠がそこに在る。各データにおける評定者群と受験者群それぞれに関する素点主成分分析第 2 主成分負荷量正負に基づく各折半個別一対データの評定者「共通段階モデル」測定による母数推定値不変性成立程度を、別の単純なデータ折半法によるものと併せて、与えたものが表 12 と 13 である。

表 12 八段階(元)データ、四段階データ、及び、三段階データにおいて異なる四つの評定者群折半法上での受験者 32 名能力推定値不変性成立程度(評定者群 12 名一次元充足度)

データ類	評定者群 12 名折半法	有効 受験者数	ピアソン 相関係数 (p 値)	ケンドールの タウ b (p 値)
8 段階(元) データ	正の負荷量 6 名 対 負の負荷量 6 名	32	0.899 (p=0.000)	0.748 (p=0.000)
	評定難度上位 6 名 対 評定難度下位 6 名	32	0.911 (p=0.000)	0.754 (p=0.000)
	評定者番号前半 6 名 対 評定者番号後半 6 名	32	0.927 (p=0.000)	0.787 (p=0.000)
	奇数番号 6 名 対 偶数番号 6 名	32	0.944 (p=0.000)	0.846 (p=0.000)
4 段階 データ	正の負荷量 5 名 対 負の負荷量 7 名	32	0.866 (p=0.000)	0.751 (p=0.000)
	評定難度上位 6 名 対 評定難度下位 6 名	32	0.893 (p=0.000)	0.771 (p=0.000)
	評定者番号前半 6 名 対 評定者番号後半 6 名	32	0.889 (p=0.000)	0.740 (p=0.000)
	奇数番号 6 名 対 偶数番号 6 名	32	0.929 (p=0.000)	0.855 (p=0.000)
3 段階 データ	正の負荷量 6 名 対 負の負荷量 6 名	32	0.859 (p=0.000)	0.744 (p=0.000)
	評定難度上位 6 名 対 評定難度下位 6 名	32	0.897 (p=0.000)	0.771 (p=0.000)
	評定者番号前半 6 名 対 評定者番号後半 6 名	32	0.893 (p=0.000)	0.748 (p=0.000)
	奇数番号 6 名 対 偶数番号 6 名	32	0.928 (p=0.000)	0.856 (p=0.000)

表 13 八段階(元)データ、四段階データ、及び、三段階データにおいて異なる四つの受験者群折半法上での評定者 12 名評定難度推定値不変性成立程度(受験者群 32 名一次元充足度)

データ類	受験者群 32 名折半法	有効 受験者数	ピアソン 相関係数 (p 値)	ケンドールの タウ b (p 値)
8 段階(元) データ	正の負荷量 17 名 対 負の負荷量 15 名	12	0.290 (p=0.361)	0.215 (p=0.335)
	能力上位 16 名 対 能力下位 16 名	12	0.601 (p=0.039)	0.419 (p=0.062)
	受験者番号前半 16 名 対 受験者番号後半 16 名	12	0.829 (p=0.001)	0.554 (p=0.013)
	奇数番号 16 名 対 偶数番号 16 名	12	0.811 (p=0.001)	0.615 (p=0.006)
4 段階 データ	正の負荷量 14 名 対 負の負荷量 18 名	12	0.227 (p=0.478)	0.110 (p=0.627)
	能力上位 16 名 対 能力下位 16 名	12	0.556 (p=0.060)	0.448 (p=0.051)
	受験者番号前半 16 名 対 受験者番号後半 16 名	12	0.782 (p=0.003)	0.534 (p=0.016)
	奇数番号 16 名 対 偶数番号 16 名	12	0.781 (p=0.003)	0.520 (p=0.022)
3 段階 データ	正の負荷量 15 名 対 負の負荷量 17 名	12	0.292 (p=0.356)	0.156 (p=0.488)
	能力上位 16 名 対 能力下位 16 名	12	0.585 (p=0.046)	0.457 (p=0.044)
	受験者番号前半 16 名 対 受験者番号後半 16 名	12	0.755 (p=0.004)	0.531 (p=0.018)
	奇数番号 16 名 対 偶数番号 16 名	12	0.784 (p=0.003)	0.540 (p=0.018)

表 12 と 13 から次の三点が傾向として明確に認知される。まず、評定者群と受験者群の折半上でいずれについても、各データにおけるピアソン相関係数とケンドール順位相関係数タウ b それぞれの最小値が正の負荷量群対負の負荷量群によって与えられている。従って、その折半法が母数推定値不変性成立程度を視る精度の上での下限信頼性を示しており、その有用価値の高さが特に表 13 での受験者群折半において顕示されている。次に、前掲表 9 での評定者群と受験者群についての素点主成分分析第 1 固有値寄与率大小の相対関係が再現されており、次元充足度のかなり高い評定者群と次元充足度の非常に低い受験者群であると判定される。但し、その判定精度観点の上では、Rasch 測定による表 12 と 13 が優先参照されて、表 9 はそれを補完するものとみなされる。更に、評定者群と受験者群それぞれについての上記相関係数指標上での次元充足度については三つのデータ間にその顕著な優劣は観察されない。故に、母数推定値「不変性」基準によるデータ次元充足度と、前節で観察された三段階データの評定者「個別段階モデル」測定における閾値母数推定値増大順序性への準拠完璧に基づく秀逸な順序段階尺度化妥当性との比較の上で、両者の無連関性一端が示されている。これは、Andrich<sup>7)</sup>(2005)による Rating Scale モデル適用最優先確認事項としての閾値母数推定値増大順序性とモデル適合度指標との無連関性への指摘(p. 53)に準ずるものと理解される。

Rasch 測定 Facets 出力におけるモデル適合度指標とはその標準化残差に基づく Infit・Outfit 統計量であり、Outfit  $\chi^2$  平均平方統計量は重み付けのないモデル残差指標値であり、Infit  $\chi^2$  平均平方統計量ははずれ値の影響を軽減すべく重み付けされたモデル残差指標値である(Wright and Masters<sup>40)</sup>, 1982, pp. 99-101)。いずれの指標値においても、1.0 を基準にして 0 に近接するにつれて同一構成要因他要素(群)との局所従属性が増大し、1.0 より大きくなるにつれて同一構成要因他要素群との次元性から乖離すると一般的に解釈される(Smith<sup>41)</sup>, E. V., 2005, p. 151、参照)。その指標値 0.8 未満で示唆されるモデル過剰適合局所従属項目は他のある一つあるいは複数の項目との内容同一性(Shen<sup>42)</sup>, 1996, p. 485)による統計的情報力寄与欠如を間接的に示しており(Linacre<sup>43)</sup>, 1997, p. 546)、『[構成された]尺度の意味を乱すものではない』(Linacre<sup>44)</sup>, 2000, p. 755; Linacre and Wright<sup>45)</sup>, 1994, p. 361)とされている。又、Infit・Outfit 統計量は Rasch モデル規定母数次元性に離反する弁別力異同を察知する(McNamara<sup>46)</sup>, 1996, p. 175 and p. 181; Smith<sup>47)</sup>, R. M., 2004, p. 432)とも指摘されており、特に Outfit  $\chi^2$  平均平方統計量のその検知鋭敏性が一つの 0・1 データについて井澤<sup>48)</sup>(2005b)により観察されている。但し、Infit・Outfit  $\chi^2$  平均平方統計量指標値 1.2 を超えてモデル不適合と判定される項目群が必ずしも「特異項目」ないしは「偏向項目」ではない(Seol<sup>49)</sup>, 1999; Smith, R. M., 1996<sup>50)</sup> and 2004<sup>47)</sup>)との実証分析に基づく指摘には留意される。井澤<sup>32)</sup>(2005a)により、0・1 データ Rasch 測定標準化残差主成分分析第 1 主成分負荷量正負に基づくデータ折半上での母数推定値不変性成立程度と Infit・Outfit 統計量によって示唆されるデータのモデル適合度との間に大きな乖離が観察されたことは、上記に関係する。

Andrich<sup>7)</sup>(2005)の指摘から類推される Many-Facet Rasch Measurement 適用上での閾値母数推定値

順序性とモデル適合度指標との無連関性(p. 53)を確認するものが表 14 である。Facets による評定者「共通段階モデル」測定の上で各項目、各評定者、各受験者、及び、各順序カテゴリーに対して適合度指標として付される Infit・Outfit  $\chi^2$  平均平方統計量において、許容値範囲 0.8 ~ 1.2 (Linacre<sup>2</sup>, 1989, p. 10)を逸脱する各一群での個数を各データについて調べた結果の一覧表である。なお、各順序カテゴリーに関しては、Outfit  $\chi^2$  平均平方統計量のみが Facets により出力される。

表 14 各データ構成要因群別 Infit・Outfit  $\chi^2$  平均平方統計量はずれ値個数

データ類	項目群 (総数3)				評定者群 (総数 12)				受験者群 (総数 32)				順序カテゴリー	
	Infit		Outfit		Infit		Outfit		Infit		Outfit		Outfit	
	0.7 以下	1.3 以上	0.7 以下	1.3 以上	0.7 以下	1.3 以上	0.7 以下	1.3 以上	0.7 以下	1.3 以上	0.7 以下	1.3 以上	0.7 以下	1.3 以上
8段階(元)データ	0	0	0	0	2	1	2	1	8	6	8	6	0	0
4段階データ	0	0	0	0	1	2	2	2	6	4	7	4	0	0
3段階データ	0	0	0	0	2	2	2	1	5	4	5	4	0	0

Infit・Outfit  $\chi^2$  平均平方統計量の上で各データにおける項目群と順序評定カテゴリーはモデル適合とされ、0.7 以下と 1.3 以上の値それぞれによって示唆されるモデル過剰適合局所従属性とモデル一次元性乖離不適合性については、評定者群においてデータ間に差異はほとんど観察されない。受験者群のみにおいて、八段階データ、四段階データ、三段階データの順にモデル過剰適合局所従属性が減少しており、モデル不適合性に関しては八段階データに較べて四段階データと三段階データが同程度に減少している。然しながら、受験者群のモデル不適合性に関するデータ間での異同数は 2 名であり、 $2/32 \approx 0.06$  としての百分率差である。受験者群のみについてのこのモデル不適合率 6%の違いが、前掲表 3 と 7 での八段階データと三段階データの間に存在する評定者「個別段階モデル」測定上での閾値母数推定値増大順序性準拠・違反に関する大きな違いに相応しているとは考えられない。従って、Andrich<sup>7</sup>(2005)指摘による閾値母数推定値増大順序性とモデル適合度指標との無連関性(p. 53)が当該データの Many-Facet Rasch Measurement 適用上でも確認される。

## V. おわりに

Linacre<sup>2</sup>(1989)による評定者「共通段階モデル」導出(pp. 47-54)参照とその Rating Scale モデル(Andrich<sup>1</sup>, 1978)との原理共通性への認知を含めて、Many-Facet Rasch Measurement 適用上での推定閾値増大順序性、評定値統合、及びデータ一次元充足度についての考察事例を詳述した。Rating Scale モデル(Andrich<sup>1</sup>, 1978)の本質特有規定とされる母数分離性、各閾値母数での隣接順序カテゴリー間同一弁別性(Andrich<sup>7</sup>, 2005)、及び、「母数結合一次元性」(Wright<sup>13</sup>, 1991, p. 158)に併せて、Rating Scale モデル上でのガットマン完全順序尺度性に基づくモデル付帯特性としての潜在閾値順次増大性が Linacre<sup>2</sup>(1989)の評定者「共通段階モデル」に内在共有されている。その認知は両

モデルが示す数理同一性の点から筆者に疑念の余地はない。従って、順序評定尺度取得データ上でのカテゴリ順序段階尺度化妥当性と評定構成概念妥当性に関する高い充足度が測定尺度構成に必須のものであるとされる限りにおいて、閾値母数推定値増大順序性とデータ次元充足度の確認は評定者「共通段階モデル」測定の上でも必然である。これが、筆者初体験としての Many-Facet Rasch Measurement 適用上で強固に意識化されたことである。その適用一事例のみの上ではあるが、Andrich<sup>7)</sup>(2005)の指摘による閾値母数推定値増大順序性とモデル適合度指標との無連関性(p. 53)が観察されている。更に、それに関係すると思われる閾値母数推定値増大順序性とデータ次元充足度の間に存在する事象乖離性にその認識重要性の上で特記注目される。

### 謝辞にかえて

この論考は、John M. Linacre<sup>2)</sup>(1989)著 *Many-Facet Rasch Measurement* 所載図 A.5(p. 11)のデータ分析を本論としている。そのデータの所有者である Robert G. Cameron of the College Board の所在探知不首尾の上で、本稿でのその使用は井澤自身によるその「データ公共性」としての認可判断に基づくものである。その「データ公共性」への示唆を賜った(2008年4月7日)Dr. John M. Linacre に心よりお礼を申し上げる次第である。

### 【参考文献】

- 1) D. Andrich. 1978. A rating formulation for ordered response categories. *Psychometrika*, 43, 4, pp. 561–573.
- 2) J. M. Linacre. 1989. *Many-facet Rasch measurement*. Chicago: MESA Press.
- 3) 秋山朝康 2000. 「スピーキングテストの分析と評価 — 項目応答理論を使つての研究」STEP (The Society for Testing English Proficiency, Inc.) *Bulletin*, 12, pp. 67–78.
- 4) G. Rasch. 1977. On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. In Blegvad, M. (Ed.), *The Danish Yearbook of Philosophy*, (pp. 58–94). Copenhagen: Munksgaard.
- 5) D. Andrich. 1997. Rating scale analysis. In Keeves, J. P. (Ed.), *Educational research, methodology, and measurement: An international handbook (2nd. Edition)* (pp. 874–880). Oxford: Elsevier Science Ltd.
- 6) D. Andrich. 2004. 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* (pp. 167–200). Maple Grove, Minnesota: JAM Press.
- 7) D. Andrich. 2005. The Rasch model explained. In Alagumalai, S, Curtis, D. D., & Hungi, N (Eds.), *Applied Rasch measurement: A book of exemplars* (pp. 27–59). The Netherlands: Springer.
- 8) J. M. Linacre. 1989–2001. *A user's guide to FACETS: Rasch measurement computer program*. Chicago: Winsteps.com.
- 9) 井澤廣行 2007. 「Rasch 評定尺度モデル解題とそのモデル上での閾値順序性」『流通科学大学論集 — 人間・社会・自然編』第20巻、第1号、pp. 1–25.
- 10) J. M. Linacre. 2001. Rating Scales, disordering and inferential falsification. *Rasch Measurement Transactions*,

- 15, 1, pp. 795–797.
- 11) J. M. Linacre. 2002. Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3, 1, pp. 85–106.
  - 12) D. Andrich. 1996. Measurement criteria for choosing among models with graded responses. In Eye, A. V. & Clogg, C. C. (Eds.), *Categorical variables in developmental research: Methods of analysis* (pp. 3–35). San Diego, California: Academic Press, Inc.
  - 13) B. D. Wright. 1991. Scores, reliabilities and assumptions. *Rasch Measurement Transactions*, 5, 3 in Linacre, J. M. (Ed.), 1995, *Rasch Measurement Transactions, Part 1* (pp. 157–158). Chicago: MESA Press.
  - 14) G. N. Masters. 1982. A Rasch model for partial credit scoring. *Psychometrika*, 47, 2, pp. 149–174.
  - 15) G. Rasch. 1960. *Probabilistic models for some intelligence and attainment tests*. The Danish Institute for Educational Research. (Reprinted in 1980 by the University of Chicago Press with a Foreword and Afterword by B. D. Wright)
  - 16) D. Andrich. 1988. A general form of Rasch's extended logistic model for partial credit scoring. *Applied Measurement in Education*, 1, 4, pp. 363–378.
  - 17) 平越裕之 2009. 「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』第 21 巻、第 2 号(掲載予定)
  - 18) 井澤廣行 2008. 「Partial Credit モデルにおける項目弁別力と閾値母数推定値増大順序性違反についての探査一考察」 『流通科学大学論集 — 人間・社会・自然編』第 21 巻、第 1 号、pp. 11-35.
  - 19) B. D. Wright & M. H. Stone. 2004. *Making measures*. Chicago: The Phaneron Press.
  - 20) J. M. Linacre. 1999. Category disordering vs. step (threshold) disordering. *Rasch Measurement Transactions*, 13, 1, pp. 675–678.
  - 21) W. Zhu, W. F. Updyke, & C. Lewandowski. 1997. Post-hoc Rasch analysis of optimal categorization of an ordered-response scale. *Journal of Outcome Measurement*, 1, 4, pp. 286–304.
  - 22) G. Rasch. 1961. On general laws and the meaning of measurement in psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Theory of Probability, Vol. IV* (pp. 321–333). Berkeley: University of California Press.
  - 23) G. Henning. 1988. The influence of test and sample dimensionality on latent trait person ability and item difficulty calibrations. *Language Testing*, 5, 1, pp. 83–98.
  - 24) T. G. Bond & C. M. Fox. 2001. *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
  - 25) T. McNamara. 1991. Test dimensionality: IRT analysis of an ESP listening test. *Language Testing*, 8, 2, pp. 139–159.
  - 26) R. M. Smith & K. K. Suh. 2003. Rasch fit statistics as a test of the invariance of item parameter estimates. *Journal of Applied Measurement*, 4, 2, pp. 153–163.
  - 27) B. D. Wright & M. H. Stone. 1979. *Best test design*. Chicago: MESA Press.
  - 28) 井澤廣行 2006a. 「Rasch 項目分析モデル適合度指標についての一考察」 『流通科学大学論集 — 人間・社会・自然編』第 19 巻、第 2 号、pp. 39–52.
  - 29) J. M. Linacre. 1998. Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, 2, 3, pp. 266–283.



- 30) E. V. Smith, Jr. 2002. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3, 2, pp. 205–231.
- 31) A. Tennant & J. F. Pallant. 2006. Unidimensionality matters. *Rasch Measurement Transactions*, 20, 1, pp.1048–1051.
- 32) 井澤廣行 2005a. 「Rasch 測定標準化残差主成分分析に基づく母数推定値不変性成立程度確認の重要性と有効性」 『流通科学大学論集 — 人間・社会・自然編』第 18 巻、第 1 号、pp. 1–14.
- 33) B. D. Wright. 1996. Local dependency, correlations and principal components. *Rasch Measurement Transactions*, 10, 3, pp. 509–511.
- 34) K. Ohtomo, H. Asano, T. Hattori, & M. Yoshie. 1987. Item difficulty of English language tests for Japanese students: The Rasch model calibration. *JACET Bulletin*, 18, pp. 109–125.
- 35) K. Ohtomo, H. Asano, T. Hattori, & M. Yoshie. 1989. Test-free person measurement in tests of English for Japanese students. *JACET Bulletin*, 20, pp. 71–87.
- 36) 芝 祐順編著 1991. 『項目反応理論 — 基礎と応用』 東京大学出版会
- 37) 渡辺直・野口裕之編著 1999. 『組織心理測定論 — 項目反応理論のフロンティア』 東京：白桃書房
- 38) M. D. Reckase. 1979. Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 3, pp. 207–230.
- 39) 豊田秀樹編著 2002. 『項目反応理論[事例編]』 東京：朝倉書店
- 40) B. D. Wright & G. N. Masters. 1982. *Rating scale analysis*. Chicago: MESA Press.
- 41) E. V. Smith, Jr. 2005. Effect of item redundancy on Rasch item and person estimates. *Journal of Applied Measurement*, 6, 2, pp. 147–163.
- 42) L. Shen. 1996. Quantifying item dependency. *Rasch Measurement Transactions*, 10, 1, p. 485.
- 43) J. M. Linacre. 1997. Investigating judge local independence. *Rasch Measurement Transactions*, 11, 1, pp. 546–547.
- 44) J. M. Linacre. 2000. Redundant items, overfit and measure bias. *Rasch Measurement Transactions*, 14, 3, p. 755.
- 45) J. M. Linacre & B. D. Wright. 1994. Chi-square fit statistics. *Rasch Measurement Transactions*, 8, 3 in Linacre, J. M. (Ed.), 1996, *Rasch Measurement Transactions*, Part 2 (pp. 360–361). Chicago: MESA Press.
- 46) T. McNamara. 1996. *Measuring second language performance*. London: Longman.
- 47) R. M. Smith. 2004. Detecting item bias with the Rasch model. *Journal of Applied Measurement*, 5, 4, pp. 430–449.
- 48) 井澤廣行 2005b. 「Rasch モデル母数一次元性規定に離反する項目・受験者弁別力異同への Infit・Outfit 統計量検知鋭敏性」 『流通科学大学論集 — 人間・社会・自然編』第 18 巻、第 2 号、pp. 1–14.
- 49) H. Seol. 1999. Detecting differential item functioning with five standardized item-fit indices in the Rasch model. *Journal of Outcome Measurement*, 3, 3, pp. 233–247.
- 50) R. M. Smith. 1996. A comparison of the Rasch separate calibration and between-fit methods of detecting item bias. *Educational and Psychological Measurement*, 56, 1, pp. 403–418.