

Rasch 測定 χ^2 平均平方統計量、項目反応理論二母数モデル 項目弁別力、及び、データ一次元性の連関性一考察

Relationships between the Rasch Infit/Outfit Mean-Squares,
the IRT Item Slopes, and Data Unidimensionality

平越 裕之*、井澤 廣行**
Hiroyuki Hirakoshi, Hiroyuki Izawa

This paper investigates the relationships between the Rasch Infit/Outfit mean-square statistics, the item slopes of the Item Response Theory (IRT) 2-Parameter Logistic Model, and data unidimensionality. It is confirmed that the expected values of the Infit/Outfit mean-squares have the negative ogive relationship with the slope parameters. The observation of some simulated data shows that 1) the positive slopes and data unidimensionality are essentially different kinds of substance and that 2) the Infit/Outfit mean-squares are sensitive to varying degrees of slopes for highly unidimensional data. Two other important concepts, i.e., response independence and dependence, are also touched upon.

Key words: Rasch item analysis model, Infit/Outfit mean-squares, IRT item slopes, Unidimensionality.

I. はじめに

Rasch 測定汎用コンピュータプログラム FACETS¹⁾の上で、Rasch 項目分析モデル(Rasch²⁾, 1960) 適合度指標値として Outfit・Infit χ^2 平均平方統計量(以降、それぞれ Outfit 統計量・Infit 統計量)が各項目と各受験者に関して出力される。この Outfit・Infit 統計量は、Wright and Panchapakesan³⁾(1969)における各項目についての χ^2 算出式(p. 45)に由来しており、その自由度適正配慮により現在では以下の算出式として与えられている(Wright and Masters⁴⁾, 1982, pp. 98-101)。なお、受験者数 M に替えて項目総数 N を挿入して、総和の方法を $\sum_{i=1}^M$ から $\sum_{j=1}^N$ に変換すれば各受験者についてのモデル適合度指標値の算出式となる。以下の記号使用において、 θ_i は受験者 i の能力母数、 β_j は項目 j の困難度母数、 u_{ij} は受験者 i の項目 j への応答 0 又は 1 である。

$$\text{Rasch 測定期待値 (期待正答確率): } P_{ij} = \frac{\exp(\theta_i - \beta_j)}{\exp(\theta_i - \beta_j) + 1} \quad (1)$$

* 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

$$\text{残差：} \quad y_{ij} = u_{ij} - P_{ij} \quad (2)$$

$$\text{標準化残差：} \quad z_{ij} = \frac{u_{ij} - P_{ij}}{\sqrt{P_{ij}(1 - P_{ij})}} \quad (3)$$

$$\text{Outfit } \chi^2 \text{ 平均平方統計量：} \quad O_j = \frac{\sum_{i=1}^M z_{ij}^2}{M} \quad (4)$$

$$\text{Infit } \chi^2 \text{ 平均平方統計量：} \quad I_j = \frac{\sum_{i=1}^M y_{ij}^2}{\sum_{i=1}^M (P_{ij}(1 - P_{ij}))} \quad (5)$$

この二種類モデル適合度指標の妥当性については、1970年代から1980年代初頭にかけて多数の識者により批判がなされている (Smith and Hedges⁵⁾, 1982, pp. 44-47、参照)。最近では、その標準化残差の線形性欠如に係わるこの指標の限界性が Karabatsos⁶⁾ (2000) により例証・詳述されており (井澤⁷⁾, 2007, p. 2、参照)、Rasch モデル適合度指標が標準化残差にその数理基盤を置く限りにおいて、その絶対的な信頼性は保証され得ないとの結論が Karabatsos⁶⁾ (2000) により下されている (p. 170)。唯、Rasch モデル測定 (以降、Rasch 測定) 分析プログラム適用上で、通常 Outfit・Infit 統計量、並びに、項目群・受験者群についての分離信頼性指標値以外は出力されないために、Rasch 測定適用者が Outfit・Infit 統計量を各項目と各受験者についてのモデル適合度指標として参照するというのが現状である。

『如何なるモデル適合度指標も必要十分条件を満たすものではない』との Rating Scale モデル考案者 Andrich⁸⁾ (2004) による言葉も道理であり、Outfit・Infit 統計量の本質性理解が Rasch 測定適用者に必要とされる。それに資する重要な論述が Linacre⁹⁾ (2000) であり、それを次節で参照・精査する。三節においては、Outfit・Infit 統計量のその本質性理解補完を目的として、項目反応理論二母数モデル項目弁別力との関係上での人工模擬データへの両モデル適用出力についての査察を与える。

II. Infit・Outfit 統計量と項目反応理論二母数モデル項目弁別力の関係

項目に関する Infit・Outfit 統計量が項目弁別力の変化によってどのような影響を受けるのかを調べてみる。項目 Infit 統計量を Rasch モデル上の正答確率期待値 (以降、期待値と表現する) と項目弁別力母数が変動するとした現実の正答確率 (以降、観測値と表現する) から計算すると、

$$\text{infit統計量} = \frac{\sum_{\text{全ての受験者}} (1 - \text{期待値})^2 \text{観測値} + \text{期待値}^2 (1 - \text{観測値})}{\sum_{\text{全ての受験者}} \text{期待値}(1 - \text{期待値})} \quad (6)$$

となり、この分母は期待値のみで決定され、観測値の違いによって変化しない。分子の総和の中の式における項目弁別力の変化による観測値の違いによって、項目 Infit 統計量の値が変化することが分かる。式 (6) の分子の \sum 中の式 (これを個別値と呼ぶ) は、

$$(1 - \text{期待値})^2 \text{観測値} + \text{期待値}^2 (1 - \text{観測値}) = \text{観測値}(1 - 2 \times \text{期待値}) + \text{期待値}^2 \quad (7)$$

となり、期待値が 0.5 を境界として、次のように変化する。なお、上下矢印は、それぞれ、値の増加と減少を表している。

$$\left\{ \begin{array}{ll} \text{期待値} < 0.5 & \text{観測値} \uparrow \text{で個別値} \uparrow, \text{観測値} \downarrow \text{で個別値} \downarrow \\ \text{期待値} = 0.5 & \text{観測値の変化は個別値に影響を与えず個別値は一定値} \\ \text{期待値} > 0.5 & \text{観測値} \uparrow \text{で個別値} \downarrow, \text{観測値} \downarrow \text{で個別値} \downarrow \end{array} \right.$$

ここで、項目弁別力母数(SLP と書く)と観測値の関係は、

$$\left\{ \begin{array}{ll} \text{期待値} < 0.5 & \text{SLP} \uparrow \text{で観測値} \downarrow, \text{SLP} \downarrow \text{で観測値} \uparrow, \\ \text{期待値} = 0.5 & \text{SLPの変化は観測値に影響を与えず観測値は一定値}, \\ \text{期待値} > 0.5 & \text{SLP} \uparrow \text{で観測値} \uparrow, \text{SLP} \downarrow \text{で観測値} \downarrow \end{array} \right.$$

となるから、結果的に、

$$\left\{ \begin{array}{ll} \text{期待値} < 0.5 & \text{SLP} \uparrow \text{で個別値} \downarrow, \text{SLP} \downarrow \text{で個別値} \uparrow, \\ \text{期待値} = 0.5 & \text{SLPの変化は個別値に影響を与えず個別値は一定値} \\ \text{期待値} > 0.5 & \text{SLP} \uparrow \text{で個別値} \downarrow, \text{SLP} \downarrow \text{で個別値} \uparrow \end{array} \right.$$

となる。つまり、式 (6) の分子で総和される個別値が、項目弁別力母数の大きさによって完全に大小関係が決定されることとなり、この個別値を総和している式 (6) の分子の大きさは、項目弁別力母数の大小に呼応することが分かる。従って、項目弁別力母数が大きくなればなるほど Infit 統計量の値は小さくなるのである。

Outfit 統計量は、

$$\text{outfit統計量} = \frac{\sum_{\text{全ての受験者}} \left(\frac{(1 - \text{期待値})^2 \text{観測値} + \text{期待値}^2 (1 - \text{観測値})}{\text{期待値}(1 - \text{期待値})} \right)}{\text{受験者数}} \quad (8)$$

とあらわすことができる。式(8)の分母は観測値とは関係がなく、分子の総和中の分母も観測値によって影響されない。式(8)分子の総和中の分子のみが観測値に影響されるので、これは、Infit 統計量の式(6)中での観測値が影響を与える部分、つまり、個別値と同様となる。従って、Infit 統計量と同様に、項目弁別力母数が大きくなればなるほど、Outfit 統計量の値は小さくなることが分かる。

また、項目弁別力母数が1の場合は、Infit 統計量、及び、Outfit 統計量の値は、

$$\begin{aligned} \text{infit統計量} &= \frac{\sum_{\text{全ての受験者}} (1 - \text{期待値})^2 \text{期待値} + \text{期待値}^2 (1 - \text{期待値})}{\sum_{\text{全ての受験者}} \text{期待値}(1 - \text{期待値})} \\ &= \frac{\sum_{\text{全ての受験者}} \text{期待値}(1 - \text{期待値})}{\sum_{\text{全ての受験者}} \text{期待値}(1 - \text{期待値})} = 1 \end{aligned} \quad (9)$$

$$\begin{aligned} \text{outfit統計量} &= \frac{\sum_{\text{全ての受験者}} \left(\frac{(1 - \text{期待値})^2 \text{観測値} + \text{期待値}^2 (1 - \text{観測値})}{\text{期待値}(1 - \text{期待値})} \right)}{\text{受験者数}} \\ &= \frac{\sum_{\text{全ての受験者}} \left(\frac{\text{期待値}(1 - \text{期待値})}{\text{期待値}(1 - \text{期待値})} \right)}{\text{受験者数}} = \frac{\text{受験者数}}{\text{受験者数}} = 1 \end{aligned} \quad (10)$$

で以って、共に1となる。項目弁別力が1でない場合の具体的な Infit・Outfit 統計量の値は、項目困難度、受験者数、及び、各受験者能力値に依存するので、これらが定まらなければ Infit・Outfit 統計量の値も決まらない。しかし、上記のように、どのような場合であっても、Infit・Outfit 統計量の値は、項目弁別力母数が1のときに1であり、項目弁別力母数が小さくなればなるほど大きくなり、項目弁別力母数が大きくなればなるほど小さくなる。言い換えれば、任意の項目困難度に対する全ての受験者能力値において、項目弁別力母数と Infit・Outfit 統計量個別値は負の相関関係にあるということである。その結果、Infit・Outfit 統計量個別値を総和して計算される Infit・Outfit 統計量も、項目弁別力母数と負の相関関係を持つことが明らかとなる。

ここで、Infit・Outfit 統計量個別値と項目弁別力母数の関係を図1と図2に例示する。図1は Infit 統計量個別値についてのものであり、(受験者能力-項目困難度)を横軸に、項目弁別力を1とした時の個別値との差を縦軸にしている。図2においては、Outfit 統計量個別値を期待値×(1-期待値)で除した値について、(受験者能力-項目困難度)、及び、弁別力を1とした時の個別値との

差をそれぞれ横軸と縦軸にしている。どのような場合においても、Infit・Outfit 統計量の大小関係が前述のように保たれていることが分かる。

次に一例として、項目困難度 0、受験者能力値 -2.5~2.5 とする 0.01 刻みの上で受験者が一様に配置されている場合の Infit・Outfit 統計量の値を図 3 に示す。Infit 統計量、及び、Outfit 統計量の値は、負の相関関係で以って、項目弁別力母数の変化によって大きな影響を受けることが示されている。

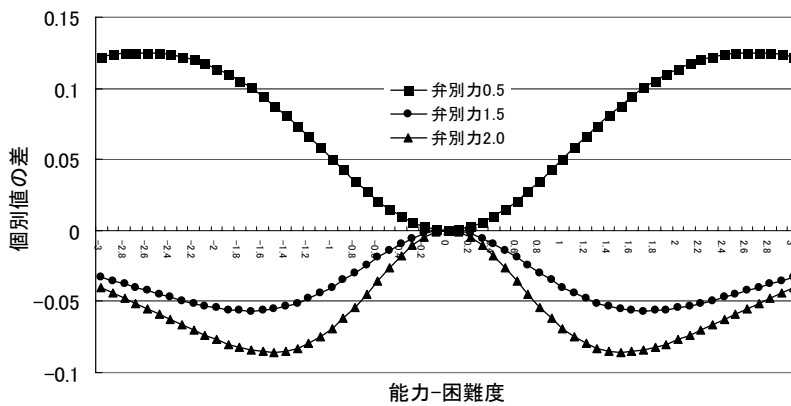


図 1 Infit 統計量計算分子（個別値）の差（基準は弁別力=1）

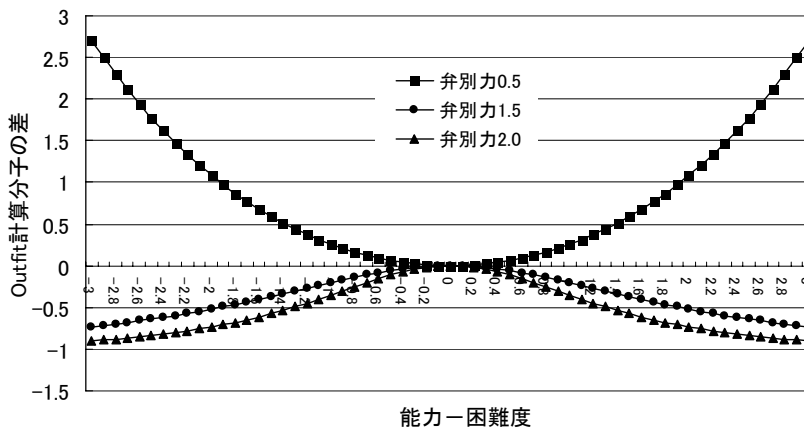


図 2 Outfit 統計量計算分子（個別値 / {期待値 × (1 - 期待値)}）の差（基準は弁別力=1）

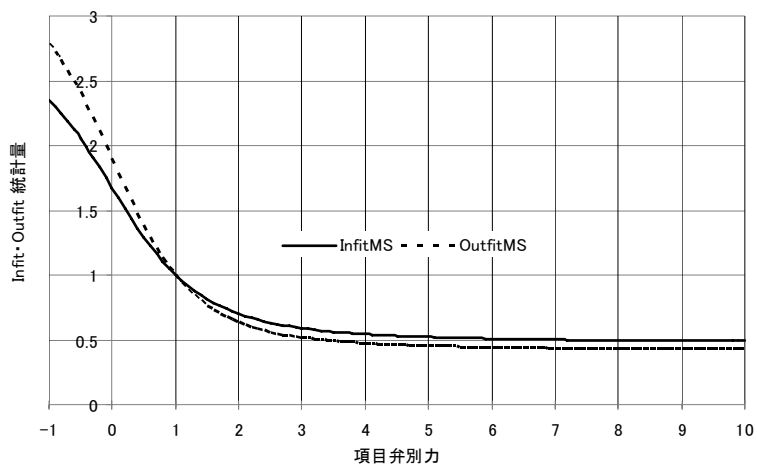


図3 Infit・Outfit 統計量と項目弁別力との各期待値としての負の曲線相関¹⁰⁾

Linacre⁹⁾ (2000) は、図3と同様な図で以って、Rasch 測定による項目 Infit 統計量は項目反応理論二母数モデル上での項目間弁別力異同程度を検知するものであると説明している。なお、図3には、Linacre⁹⁾ (2000) によって提示されていない項目 Outfit 統計量期待値と項目弁別力期待値の負の曲線相関も与えている。上に述べたように、Rasch 項目分析モデル規定としての項目弁別力期待値を 1.0 とし、二母数モデル上での項目弁別力期待値が 1.0 から大小乖離する程度につれて、Rasch 測定上での項目 Infit・Outfit 統計量期待値は負の曲線相関を示すということである。結語として、『現実データがかなりの程度に同質的な項目群で構成されているならば、二母数モデル適用により観察される項目間弁別力異同は測定上無意味である』と Linacre⁹⁾ (2000, p. 743) は述べている。従って、Rasch 測定 Infit・Outfit 統計量は、その本質として、データ次元性ではなく、データにおける項目群と受験者群の弁別力同値 1.0 からの各項目と各受験者の乖離程度を示す指標であるとの示唆が Linacre⁹⁾ (2000, p. 743) の言葉に窺われる。但し、期待値としての項目弁別力が 0 から負の値となる Infit 統計量と Outfit 統計量それぞれの期待値として 1.67 と 1.89 よりも大きな値となれば、Infit・Outfit 統計量はデータ次元性からの乖離度指標に重なることが図3から観察される。又、期待値としての項目弁別力が 1.0 より大きくなるにつれて、Infit・Outfit 統計量は項目間従属性 (response dependence) 増大傾向の指標と重なることも図3に目視される (Andrich¹¹⁾, 2006、参照)。Linacre⁹⁾ (2000, p. 743) はその項目間従属性増大傾向に関する期待値としての Infit 境界値を 0.8 としている。

Ⅲ. 人工模擬データにおける Infit・Outfit 統計量、二母数モデル項目弁別力、及び、データ次元性の連関性査察

Outfit 統計量は、テストの項目数と受験者数、並びに、項目困難度分布と受験者能力分布によって影響される。特に、テストに含まれる項目数についてはテスト時間に制約される限度がある故に、受験者群に付される Outfit 統計量は項目数の大きさに鋭敏に反応する。Outfit 統計量上での 1.4 以上をはずれ値とみなすならば、テストの項目数が大きくなるにつれて、受験者群についてのそのはずれ値生起の割合は Rasch モデル適合データ群において小さくなることが窺われる。一方、1,000 名以上の大規模な受験者群により付される項目群 Outfit 統計量については、その値 1.3 以上をはずれ値とすれば、50 項目以上のテストにおけるその生起割合は Rasch モデル適合データ群において微小であると示されている（井澤・平越¹²⁾、2009、表 4-6 (p. 36) 及び表 4-7 (p. 38)、参照）。

上記観察により、本節での目的にかなう人工模擬データを以下の作成要領とした。

- 1) 50 項目×1,000 名のテスト受験様態。
- 2) 項目困難度設定値と受験者能力設定値はいずれも平均値を 0 とする正規分布より抽出。
- 3) 困難度設定値分布と能力設定値分布の標準偏差を同一として、その値を 1) 0.5、2) 1.0、3) 2.0 の三種類とする。
- 4) 五種類の設定値分布形態の変容を、A) Rasch モデル適合データ群、B) 0.3~2.0 の一様乱数 50 個発生に基づく二母数モデル上での項目弁別力細工データ群、C) 0.3~5.0 の一様乱数 50 個発生に基づく二母数モデル上での項目弁別力細工データ群、D) 二次元データ群、及び、E) 二次元性+項目弁別力 (0.3~2.0) データ群とする。なお、データ次元性の作成は、A) Rasch モデル適合データ群における低い困難度から高い困難度に渡る偶数番号項目群についての受験者能力の無作為移動によるものである。

以上の要領により作成された各データに FACETS¹⁾ 上での Rasch 項目分析モデルと BILOG-MG¹³⁾ 上での項目反応理論二母数モデルを適用した上で、上記 3) での各種 3 標本あるいは 5 標本の出力指標値平均を与えるものが表 1、2、3、4、5 である。なお、その五つの表は、本文中での提示煩雑性を考慮して本稿末尾に補遺として掲載している。

Rasch モデル適合性の観点からは、A、B、C、D、E それぞれ 1)、2)、3) の順に Rasch 測定妥当性が増大しており、それは、各表での項目群と受験者群についての「01 データピアソン主成分分析第 1 主成分寄与率」・「母数推定値分離指標値」・「母数推定値分離信頼性係数」・「母数推定値不変性」の各指標によって明示されている（平越・井澤¹⁴⁾、2008、表 4-1、p. 32、参照）。なお、項目群と受験者群についての「母数推定値不変性」は、Rasch 測定標準化残差の受験者群と項目群それぞれについての主成分分析第 1 主成分負荷量正負分別に基づく各折半データの Rasch 測定の上で、同一項目群と同一受験者群についての母数推定値間に観測されるピアソン相関係数とケ

ンドールタウ b 順位相関係数である。そのデータ折半法が、下限信頼性の点で他の折半法に優るとの著者経験則である。又、Rasch 測定の上で受験者群に付される「母数推定値分離信頼性係数」は、テスト項目群信頼性係数として伝統的に使用されている Cronbach のアルファ係数と、著者のその観察上においても、ほぼ同値となる (Fox and Jones¹⁵, 1998, p. 37)。天井効果によるその「母数推定値分離信頼性係数」1.0 の値の不備を補う指標が「母数推定値分離指標値」であり、その意味内容は、前者と同じく、母数推定値の再現性指標である (Linacre¹⁶, 1997, pp. 580-581)。

本稿の研究査察動機は、『項目 Infit χ^2 平均平方統計量と項目反応理論弁別力との相関はまれにしか -0.8 を下回らない』との Wright (1996¹⁷), p. 6) による言葉であり、併せて、『Infit・Outfit 統計量は程度の強い多次元性には鋭敏ではない』との Linacre¹⁸ (2009, p. 1225)、並びに、Smith¹⁹ (2000, p. 216) による指摘である。前節での数理精査上での項目 Infit・Outfit 統計量と項目反応理論二母数モデル弁別力それぞれの期待値間での負の曲線相関により、前者 Wright (1996¹⁷), p. 6) の言葉は「Infit・Outfit 統計量は、本質的に、データ一次元性からの乖離指標ではなく、Rasch 項目分析モデル規定としての項目群と受験者群の弁別力同値 1.0 からの乖離指標である」と理解される。後者 Linacre¹⁸ (2009, p. 1225)、並びに、Smith¹⁹ (2000, p. 216) による指摘は、「項目群の多次元性による一律的な項目群弁別力の低さが、何等かの形で Infit・Outfit 統計量の変動を微小にしている」ことに起因していると推測される。本節での焦点は以上の二点である。

以降、A)、B)、C)、D)、E) の比較は、補遺にて掲載している表 1、2、3、4、5 におけるそれぞれ三種類対応データ間についてのものである。表 2 での B)、及び、表 3 での C) は、表 1 での A) Rasch モデル適合群データに項目反応理論二母数モデル上での項目弁別力について正方向への変異増大性のみを付加したものであるから、A) に比して B) と C) でのデータ一次元性は同程度以上になっていることが観察される。但し、C) でのその項目弁別力の正方向変異増大性が非常に大きなものであり、C) は A) と B) よりはるかに上回る高いデータ一次元性を示している。これは、C) において非常に高い弁別力を持つ項目群の数が過度に多いということであり、その結果として C) における項目間正誤反応相関が過度に高いものであるとの示唆である。従って、C) での三種データはそれぞれ項目間従属性 (response dependence) が過度に高いと判断される。一方、A) に比べて B) での項目弁別力の正方向変異増大性は異常な程度に高いものではなく、B) における BILOG-MG¹³ 出力上での項目弁別力分布の平均値と標準偏差は A) での各対応値を少々上回っているのみである。然しながら、B) での A) よりも大きい項目弁別力ばらつき程度に呼応して、B) での A) よりも大きい Infit・Outfit 統計量のばらつき程度が観察される。項目弁別力と Infit・Outfit 統計量のこのばらつき程度の呼応増大が、結果として、B) での A) に優る両者相関の高さに示されていると判断される。A) と比しての B) での対応データ間での一次元性 (Rasch 測定妥当性) 充足度はほぼ同程度であり、前節内容、及び、上に引用した Wright (1996¹⁷), p. 6) の言葉に関連して、一例ではあるが、「Infit・Outfit 統計量は、本質的に、データ一次元性からの乖

離指標ではなく、Rasch 項目分析モデル規定としての項目群と受験者群の弁別力同値 1.0 からの乖離指標である」と示されている。なお、A) は理想的な Rasch モデル適合データ群であり、Wright (1996¹⁷⁾, p. 6) の言葉は、Rasch モデルへの高い適合度に併せて、項目間従属性に偏向していない現実データ取得の難しさを暗示している。又、B) と C) の比較により、過度に高い弁別力を持つ項目群が集積すれば、項目間正誤反応相関が過度に高まり、その過度に高い項目間従属性と大変に高いデータ次元充足度の連関性に注目される。これは、Andrich¹¹⁾ (2006) の論述内容、及び、伝統的なテスト理論における‘attenuation paradox’としての過度に高いアルファ係数・テスト信頼性係数と低いテスト妥当性の連関性に関係しており、過度に高い項目間従属性と低いテスト情報量・過度に高いデータ次元性の連関性についての顕示一端である。更に、興味ある生起事象として観察されることは、C-3) での項目困難度と受験者能力の推定値分布いずれもの標準偏差が、A-3) と B-3) でのその設定値に同等な推定値分布標準偏差それぞれ 2.0 に比べて、その値を大きく超える 3.2 に増大している。前記の傾向は、A-1)、B-1) に比して C-1)、並びに、A-2)、B-2) に比して C-2) の場合にも同様に観察される。これは、本来的に大きな弁別力を持つ項目群の集積度が高くなれば、項目群と受験者群の母数真値分布と比べて Rasch 測定上での両群母数推定値分布が広まるとの示唆である。

表 4 での D) 二次元データ群についての「Rasch 測定項目 Infit MnSq 統計量分布」、並びに、「Rasch 測定項目 Outfit MnSq 統計量分布」に付されたその値 1.0 (両者期待値) 無変動傾向への視認により、『Infit・Outfit 統計量は程度の強い多次元性には鋭敏ではない』(Linacre¹⁸⁾, 2009, p. 1225; Smith¹⁹⁾, 2000, p. 216) との一例が顕示されている。D) の「BILOG 測定項目 Slope 分布」での平均値は、A) Rasch モデル適合データ群における三種対応値よりも低くなっており、A) に比べて D) での低い弁別力を持つ項目群集積度が高くなっている。相対的に低い項目弁別力は相対的に高い項目独自性・多(ないしは異)次元性を意味する (Andrich¹¹⁾, 2006, p. 1057)、従って、A) に比して D) でのその低い弁別力を持つ項目群集積度の大きさ、つまりは、「項目群の多次元性による一律的な項目群弁別力の低さ」が、「何等かの形で Infit・Outfit 統計量の変動を微小にしている」と推測される。唯、その起因についての論理・数理は著者に不明である。興味あることは、D) に比して、E) 二次元性+項目スロープ細工 (0.3~2.0) データ群では、項目弁別力の平均値が高くなっていること加えて、項目弁別力と Infit・Outfit 統計量いずれもの標準偏差が大きくなっている。上記 B) と A) との関係と同じく、項目弁別力と Infit・Outfit 統計量に関するこのばらつき程度の呼応増大が、結果として、E) の D) に優る両者相関の高さに示されていると思われる。一方、D) に比しての E) のデータ次元性は強度に同程度であることから、前記との関連で以って、A)、B)、C) においてと同様に E) と D) に関しても、Infit・Outfit 統計量はデータ次元性とは高い程度に無連関であるとの一端が観察される。又、A) と B) に比して C) での項目群と受験者群の母数推定値分布幅拡大とは反対方向に、D) と E) での項目群と受験者群の母数推定値標準偏差は A)

での対応データ間でのその対応値に比べてかなり小さな値となっている。併せて、A) に比して、D) と E) での BILOG-MG¹³⁾出力項目弁別力が小さくなっている。これにより、本来的に小さな弁別力を持つ項目群の集積度が高くなれば、Rasch 測定上での項目群と受験者群の母数推定値分布がそれぞれ母数真値の分布よりも狭小となるとの傾向が窺われる。

図4が本節査察内容の簡明直截なまとめである。なお、当然ながら、図4は、本節で提示された著者作成による人工模擬データのみに関する関係要因生起傾向を俯瞰するためのものである。

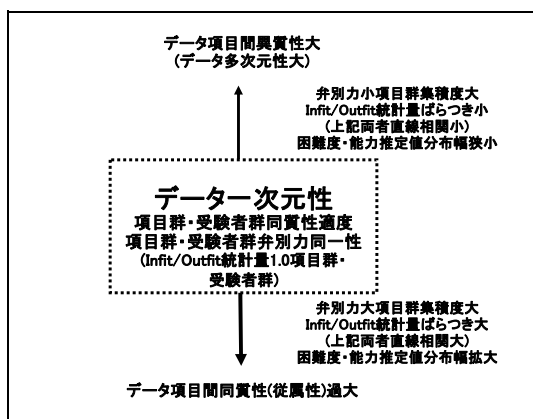


図4 本節査察まとめ

IV. おわりに

第二節において、Linacre⁹⁾ (2000, p. 743) による論述と提示図に基づいてその数理解釈と複製図を与えた。これにより、FACETS¹⁾上での Rasch 項目分析モデル適用で以って出力される Infit・Outfit 統計量の指標としての本質性は、項目反応理論二母数モデル上での項目弁別力との理論期待値としての負の曲線相関であると理解される。一次元数理確率モデルとして認知されている Rasch モデルであるが故に、Rasch 測定適用者は Infit・Outfit 統計量をその「モデル適合度指標」、つまりは、「データ一次元性乖離度指標」と一様短絡的に把握する可能性がある。然しながら、データ一次元性と Rasch 項目分析モデル規定としての項目群と受験者群の弁別力同値 1.0 は実質的な基底において同一ではない。後者の弁別力同値 1.0 が満たされていないなくとも、その値が正である限り、データ一次元性には影響を与えない。データ一次元性と項目群弁別力同一性についてのその実質内容に関する相違が、前節での人工模擬データ群についての Rasch 測定と BILOG-MG¹³⁾出力による両者充足度の上での観察一例として顕示されている。それ故に、Rasch 測定者においては、取得正誤反応データの Rasch 項目分析モデル適合度判断に際しては、Rasch 測定上での Infit・Outfit 統計量のみならず、補遺での表掲載に与えたデータ一次元性充足度指標値への視認が必要とされる。更に、可能であれば、「Rasch 測定尺度構成体系の達成度指標値」(平越・井澤¹⁴⁾、

2008, p. 31、参照)、及び、BILOG-MG¹³⁾出力上での項目弁別力分布への参照も望まれる。

謝辞

本稿第二節記述は、著者取得依頼に対して Dr. John M. Linacre により御送付頂いた BASIC プログラムへの参照に基づいている。彼の御厚意に心よりお礼を申し上げる次第である。

【参考文献及び注】

- 1) J. M. Linacre. 1989-2001. Rasch measurement computer program. Chicago: Winsteps.com
- 2) 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.)
- 3) B. Wright & N. Panchapakesan. 1969. A procedure for sample-free item analysis. *Educational and Psychological Measurement*, 29, pp. 23-48.
- 4) B. D. Wright & G. N. Masters. 1982. *Rating scale analysis*. Chicago, IL: MESA Press.
- 5) R. M. Smith & L. V. Hedges. 1982. A comparison of likelihood ratio χ^2 and Pearsonian χ^2 tests of fit in the Rasch model. *Educational Research and Perspectives*, 9, pp. 44-54.
- 6) G. Karabatsos. 2000. A critique of Rasch residual fit statistics. *Journal of Applied Measurement*, 1, 2, pp. 152-176.
- 7) 井澤廣行 2007. 「Karabatsos の Rasch モデル遊離度線形残差と Rasch 測定標準化残差の異同程度」『流通科学大学論集 — 人間・社会・自然編』第 19 巻、第 3 号、pp. 1-14.
- 8) D. Andrich. 2004. rasch@wu-wien.ac.at, May 26.
- 9) J. M. Linacre. 2000. Item discrimination and infit mean-square. *Rasch Measurement Transactions*, 14, 2, p. 743.
- 10) 注： 上記 9) Linacre (2000) における提示図の著者による作成法教示依頼への Linacre による返答 (2010 年 3 月 29 日) 送付 BASIC プログラムに基づく作図。
- 11) D. Andrich. 2006. Item discrimination and Rasch-Andrich thresholds revisited. *Rasch Measurement Transactions*, 20, 2, pp. 1055-1057.
- 12) 井澤廣行・平越裕之 2009. 「正誤反応データ Rasch モデル測定妥当性についての一考察」『流通科学大学論集 — 人間・社会・自然編』第 22 巻、第 1 号、pp. 21-41.
- 13) M. Zimowski, E. Muraki, R. Mislevy, & D. Bock. 2002. Item response theory computer program. Scientific Software International, Inc.
- 14) 平越裕之・井澤廣行 2008. 「テストの正誤反応行列から得られる次元性情報に関する考察」『流通科学大学論集 — 経済・経営情報編』第 17 巻、第 1 号、pp. 9-33.
- 15) C. M. Fox & J. A. Jones. 1998. Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45, 1, pp. 30-45.
- 16) J. M. Linacre. 1997. KR-20/Cronbach alpha or Rasch reliability: Which tells the “truth”? *Rasch Measurement Transactions*, 11, 3, pp. 580-581.
- 17) B. D. Wright. 1996. Comparing Rasch measurement and factor analysis. *Structural Equation Modeling*, 3, 1, pp. 3-24.

- 18) J. M. Linacre. 2009. Rasch lessons from Netflix® prize competition. *Rasch Measurement Transactions*, 23, 3, pp. 1224-1227.
- 19) R. M. Smith. 2000. Fit analysis in latent trait measurement models. *Journal of Applied Measurement*, 1, 2, pp. 199-218.

補遺：

表1 三種類のA)人工模擬 Rasch モデル適合データ群についての指標値平均

A-1) Raschモデル適合データ群： 困難度設定値 s.d.=0.5 対 能力設定値 s.d.=0.5 データ3標本 指標値平均																							
項目群	7.3%	7.09	0.98	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布							
				平均値	標準偏差	最小値	最大値	有効数	50	ピアノソ	.458	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
				0.00	0.49	-1.10	1.03					ケンドールタウb	.317	1.000	0.000	1.0	1.0	1.000	0.000	1.0	1.0		
受験群	6.7%	1.69	0.74	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布							
				0.00	0.59	-1.84	2.48	有効数	999.7	ピアノソ	.472	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
										ケンドールタウb	.347	1.000	0.068	0.8	1.2	1.000	0.084	0.7	1.4				

A-1) 困難度設定値 s.d.=0.5 対 能力設定値 s.d.=0.5 データ3標本 平均値	BLOG測定項目Slope分布				BLOG項目SlopeとRasch項目Infit MnSqの相関	
	平均値	標準偏差	最小値	最大値	平均	
	0.517	0.057	0.395	0.674	ピアノソ	*
					ケンドールタウb	*
				BLOG項目SlopeとRasch項目OutfitMnSqの相関		
				平均		
				ピアノソ	*	
				ケンドールタウb	*	

A-2) Raschモデル適合データ群： 困難度設定値 s.d.=1.0 対 能力設定値 s.d.=1.0 データ3標本 指標値平均																							
項目群	17.1%	12.53	0.99	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布							
				平均値	標準偏差	最小値	最大値	有効数	50	ピアノソ	.799	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
				0.00	0.96	-2.10	2.06					ケンドールタウb	.608	0.999	0.029	0.9	1.1	0.993	0.059	0.9	1.1		
受験群	15.6%	2.97	0.90	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布							
				0.00	1.08	-3.67	3.80	有効数	989	ピアノソ	.757	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
										ケンドールタウb	.593	1.000	0.114	0.7	1.4	0.996	0.212	0.4	2.6				

A-2) 困難度設定値 s.d.=1.0 対 能力設定値 s.d.=1.0 データ3標本 平均値	BLOG測定項目Slope分布				BLOG項目SlopeとRasch項目Infit MnSqの相関	
	平均値	標準偏差	最小値	最大値	平均	
	0.982	0.095	0.797	1.237	ピアノソ	-.586
					ケンドールタウb	-.390
				BLOG項目SlopeとRasch項目OutfitMnSqの相関		
				平均		
				ピアノソ	-.775	
				ケンドールタウb	-.617	

A-3) Raschモデル適合データ群： 困難度設定値 s.d.=2.0 対 能力設定値 s.d.=2.0 データ5標本 指標値平均																							
項目群	32.0%	19.93	1.00	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布							
				平均値	標準偏差	最小値	最大値	有効数	50	ピアノソ	.928	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
				0.00	1.96	-4.27	4.23					ケンドールタウb	.783	0.999	0.057	0.9	1.1	0.999	0.179	0.6	1.7		
受験群	31.0%	4.46	0.95	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布							
				0.00	2.04	-5.43	5.44	有効数	952	ピアノソ	.882	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
										ケンドールタウb	.740	0.998	0.200	0.5	1.8	0.992	0.762	0.1	8.7				

A-3) 困難度設定値 s.d.=2.0 対 能力設定値 s.d.=2.0 データ5標本 平均値	BLOG測定項目Slope分布				BLOG項目SlopeとRasch項目Infit MnSqの相関	
	平均値	標準偏差	最小値	最大値	平均	
	1.890	0.149	1.561	2.214	ピアノソ	-.833
					ケンドールタウb	-.669
				BLOG項目SlopeとRasch項目OutfitMnSqの相関		
				平均		
				ピアノソ	-.565	
				ケンドールタウb	-.469	

表 2 三種類の B) 人工模擬スロープ細工セット 1 (0.3~2.0) データ群についての指標値平均

B-1) スロープ細工(セット1: 0.3~2.0): 困難度設定値 s.d.=0.5 対 能力設定値 s.d.=0.5 データ3標本 指標値平均																			
項目群	01 データ ピアソン 主成分分析 第1主成分 寄与率	母数 推定値 分離 指標値	母数 推定値 分離 信頼性 係数	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布			
	9.5 %			8.16	0.99	平均値	標準偏差	最小値	最大値	有効数	50	ピアソン	.530	平均値	標準偏差	最小値	最大値	平均値	標準偏差
受験群	8.0 %	1.89	0.78	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布			
				平均値	標準偏差	最小値	最大値	有効数	999.3	ピアソン	.526	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
B-1) 困難度設定値 s.d.=0.5 対 能力設定値 s.d.=0.5 データ3標本 平均値																			
				BILOG測定項目Slope分布				BILOG項目SlopeとRasch項目Infit MnSqの相関											
				平均値	標準偏差	最小値	最大値	平均	ピアソン			-.897							
				0.612	0.235	0.227	1.082	ケンドールタウb			-.788								
								BILOG項目SlopeとRasch項目OutfitMnSqの相関											
								平均	ピアソン			-.920							
								ケンドールタウb			-.817								
B-2) スロープ細工(セット1: 0.3~2.0): 困難度設定値 s.d.=1.0 対 能力設定値 s.d.=1.0 データ3標本 指標値平均																			
項目群	20.0 %	13.88	0.99	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布			
	18.0 %	3.11	0.91	平均値	標準偏差	最小値	最大値	有効数	50	ピアソン	.775	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
受験群	18.0 %	3.11	0.91	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布			
				平均値	標準偏差	最小値	最大値	有効数	988	ピアソン	.770	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
B-2) 困難度設定値 s.d.=1.0 対 能力設定値 s.d.=1.0 データ3標本 平均値																			
				BILOG測定項目Slope分布				BILOG項目SlopeとRasch項目Infit MnSqの相関											
				平均値	標準偏差	最小値	最大値	平均	ピアソン			-.959							
				1.165	0.469	0.344	2.016	ケンドールタウb			-.897								
								BILOG項目SlopeとRasch項目OutfitMnSqの相関											
								平均	ピアソン			-.954							
								ケンドールタウb			-.887								
B-3) スロープ細工(セット1: 0.3~2.0): 困難度設定値 s.d.=2.0 対 能力設定値 s.d.=2.0 データ5標本 指標値平均																			
項目群	32.9 %	20.05	1.00	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布			
	31.1 %	4.43	0.95	平均値	標準偏差	最小値	最大値	有効数	50	ピアソン	.892	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
受験群	31.1 %	4.43	0.95	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布			
				平均値	標準偏差	最小値	最大値	有効数	961	ピアソン	.875	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
B-3) 困難度設定値 s.d.=2.0 対 能力設定値 s.d.=2.0 データ5標本 平均値																			
				BILOG測定項目Slope分布				BILOG項目SlopeとRasch項目Infit MnSqの相関											
				平均値	標準偏差	最小値	最大値	平均	ピアソン			-.927							
				2.238	0.914	0.599	4.151	ケンドールタウb			-.911								
								BILOG項目SlopeとRasch項目OutfitMnSqの相関											
								平均	ピアソン			-.851							
								ケンドールタウb			-.848								

表3 三種類のC) 人工模擬スロープ細工セット2 (0.3~5.0) データ群についての指標値平均

C-1) スロープ細工(セット2: 0.3~5.0): 困難度設定値 s.d.=0.5 対 能力設定値 s.d.=0.5 データ3標本 指標値平均																							
項目群	20.8 %	13.61	0.99	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布							
				平均値	標準偏差	最小値	最大値	有効数	50	ピアソン	.707	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
				0.00	1.07	-2.29	2.33			ケンドールタウb	.503	0.996	0.175	0.8	1.4	0.981	0.285	0.6	1.7				
受験者	17.6 %	3.07	0.90	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布							
				平均値	標準偏差	最小値	最大値	有効数	986.7	ピアソン	.730	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
				0.03	1.13	-3.92	3.50			ケンドールタウb	.569	1.011	0.129	0.7	1.4	0.982	0.212	0.5	2.3				
C-1) 困難度設定値 s.d.=0.5 対 能力設定値 s.d.=0.5 データ3標本 平均値																							
				BLOG測定項目Slope分布				BLOG項目SlopeとRasch項目Infit MnSqの相関															
				平均値	標準偏差	最小値	最大値	平均	ピアソン			-.948											
				1.233	0.618	0.205	2.509	ケンドールタウb			-.908												
								BLOG項目SlopeとRasch項目OutfitMnSqの相関															
								平均	ピアソン			-.939											
								ケンドールタウb			-.907												
C-2) スロープ細工(セット2: 0.3~5.0): 困難度設定値 s.d.=1.0 対 能力設定値 s.d.=1.0 データ3標本 指標値平均																							
項目群	33.3 %	19.49	1.00	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布							
				平均値	標準偏差	最小値	最大値	有効数	50	ピアソン	.840	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
				0.00	1.87	-4.26	3.88			ケンドールタウb	.664	0.967	0.316	0.6	1.9	1.054	0.824	0.3	3.8				
受験者	30.0 %	4.36	0.95	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布							
				平均値	標準偏差	最小値	最大値	有効数	953	ピアソン	.844	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
				0.03	1.96	-5.45	5.20			ケンドールタウb	.697	1.014	0.208	0.5	1.9	1.053	0.731	0.1	8.1				
C-2) 困難度設定値 s.d.=1.0 対 能力設定値 s.d.=1.0 データ3標本 平均値																							
				BLOG測定項目Slope分布				BLOG項目SlopeとRasch項目Infit MnSqの相関															
				平均値	標準偏差	最小値	最大値	平均	ピアソン			-.922											
				2.406	1.206	0.371	4.969	ケンドールタウb			-.930												
								BLOG項目SlopeとRasch項目OutfitMnSqの相関															
								平均	ピアソン			-.838											
								ケンドールタウb			-.852												
C-3) スロープ細工(セット2: 0.3~5.0): 困難度設定値 s.d.=2.0 対 能力設定値 s.d.=2.0 データ3標本 指標値平均																							
項目群	41.4 %	25.61	1.00	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布							
				平均値	標準偏差	最小値	最大値	有効数	50	ピアソン	.932	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
				0.00	3.27	-7.90	6.42			ケンドールタウb	.845	0.939	0.443	0.5	2.5	1.452	2.283	0.1	9.0				
受験者	40.9 %	5.72	0.97	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布							
				平均値	標準偏差	最小値	最大値	有効数	876	ピアソン	.886	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値
				0.08	3.24	-7.95	7.05			ケンドールタウb	.789	0.984	0.308	0.4	2.8	1.337	1.901	0.0	9.0				
C-3) 困難度設定値 s.d.=2.0 対 能力設定値 s.d.=2.0 データ3標本 平均値																							
				BLOG測定項目Slope分布				BLOG項目SlopeとRasch項目Infit MnSqの相関															
				平均値	標準偏差	最小値	最大値	平均	ピアソン			-.880											
				4.282	2.068	0.671	8.207	ケンドールタウb			-.927												
								BLOG項目SlopeとRasch項目OutfitMnSqの相関															
								平均	ピアソン			-.707											
								ケンドールタウb			-.751												

表 4 三種類の D) 人工模擬二次元データ群についての指標値平均

D-1) 二次元データ群: 困難度設定値 s.d.=0.5 対 能力設定値 s.d.=0.5 データ3標本 指標値平均																															
項目群	5.0 %	6.83	0.98	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布															
				平均値	標準偏差	最小値	最大値	有効数	50	ピアソン	.181	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値								
受験群	6.5 %	1.11	0.55	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布															
				平均値	標準偏差	最小値	最大値	有効数	1000	ピアソン	-.059	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値								
				0.00	0.44	-1.61	1.45			ケンドールタウb	-.044	1.000	0.067	0.8	1.2	1.000	0.078	0.8	1.3												
D-1) 困難度設定値 s.d.= 0.5 対 能力設定値 s.d. = 0.5 データ3標本 平均値				BILOG測定項目Slope分布				BILOG項目SlopeとRasch項目Infit MnSqの相関																							
				平均値	標準偏差	最小値	最大値	平均	ピアソン																						
				0.376	0.068	0.244	0.530		ケンドールタウb																						
								平均	ピアソン																						
								BILOG項目SlopeとRasch項目OutfitMnSqの相関																							
												平均				ピアソン															
																平均				ケンドールタウb											
																				平均				ケンドールタウb							
D-2) 二次元データ群: 困難度設定値 s.d.=1.0 対 能力設定値 s.d.=1.0 データ3標本 指標値平均																															
項目群	9.8 %	12.07	0.99	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布															
				平均値	標準偏差	最小値	最大値	有効数	50	ピアソン	.402	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値								
受験群	14.8 %	2.07	0.81	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布															
				平均値	標準偏差	最小値	最大値	有効数	994	ピアソン	.022	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値								
				-0.01	0.74	-2.48	2.26			ケンドールタウb	.020	1.001	0.109	0.7	1.4	0.997	0.163	0.5	1.8												
D-2) 困難度設定値 s.d.= 1.0 対 能力設定値 s.d. = 1.0 データ3標本 平均値				BILOG測定項目Slope分布				BILOG項目SlopeとRasch項目Infit MnSqの相関																							
				平均値	標準偏差	最小値	最大値	平均	ピアソン																						
				0.649	0.231	0.324	1.064		ケンドールタウb																						
								平均	ピアソン																						
								BILOG項目SlopeとRasch項目OutfitMnSqの相関																							
												平均				ピアソン															
																平均				ケンドールタウb											
																				平均				ケンドールタウb							
D-3) 二次元データ群: 困難度設定値 s.d.=2.0 対 能力設定値 s.d.=2.0 データ5標本 指標値平均																															
項目群	18.1 %	17.09	1.00	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布															
				平均値	標準偏差	最小値	最大値	有効数	50	ピアソン	.421	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値								
受験群	27.1 %	2.93	0.90	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布															
				平均値	標準偏差	最小値	最大値	有効数	951	ピアソン	-.055	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値								
				0.00	1.12	-4.27	4.25			ケンドールタウb	-.045	1.000	0.198	0.6	1.7	0.967	0.368	0.2	3.2												
D-3) 困難度設定値 s.d.= 2.0 対 能力設定値 s.d. = 2.0 データ5標本 平均値				BILOG測定項目Slope分布				BILOG項目SlopeとRasch項目Infit MnSqの相関																							
				平均値	標準偏差	最小値	最大値	平均	ピアソン																						
				1.082	0.869	0.174	2.310		ケンドールタウb																						
								平均	ピアソン																						
								BILOG項目SlopeとRasch項目OutfitMnSqの相関																							
												平均				ピアソン															
																平均				ケンドールタウb											
																				平均				ケンドールタウb							

表5 三種類のE)二次元性+項目スロープ細工セット1(0.3~2.0)
データ群についての指標値平均

E-1) 二次元性+項目スロープ細工(0.3~2.0)データ群: 困難度設定値 s.d.=0.5 対 能力設定値 s.d.=0.5 データ3標本 指標値平均																			
項目群	01 データ ピアソン 主成分分析 第1 主成分 寄与率	母数 推定値 分離 指標値	母数 推定値 分離 信頼性 係数	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布			
5.9%	8.14	0.98	平均値	標準偏差	最小値	最大値	有効数	50	ピアソン	.261	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	
			0.00	0.56	-1.17	1.52			ケンドールタウb	.145	1.002	0.018	1.0	1.1	1.001	0.032	0.9	1.1	
受験者	8.1%	1.28	0.62	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布			
			平均値	標準偏差	最小値	最大値	有効数	1000	ピアソン	-.031	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	
			-0.04	0.49	-1.66	1.57			ケンドールタウb	-.022	1.000	0.076	0.8	1.2	0.998	0.092	0.7	1.3	
E-1) 困難度設定値 s.d.=0.5 対 能力設定値 s.d.=0.5 データ3標本 平均値				BLOG測定項目Slope分布				BLOG項目SlopeとRasch項目Infit MnSqの相関											
				平均値	標準偏差	最小値	最大値	平均	ピアソン	-.288									
				0.431	0.151	0.206	0.787		ケンドールタウb	-.247									
								BLOG項目SlopeとRasch項目OutfitMnSqの相関											
								平均	ピアソン	-.550									
									ケンドールタウb	-.420									
E-2) 二次元性+項目スロープ細工(0.3~2.0)データ群: 困難度設定値 s.d.=1.0 対 能力設定値 s.d.=1.0 データ3標本 指標値平均																			
項目群	11.0%	13.13	0.99	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布			
			平均値	標準偏差	最小値	最大値	有効数	50	ピアソン	.420	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	
			0.00	0.98	-2.02	2.50			ケンドールタウb	.268	0.995	0.065	0.9	1.1	0.985	0.099	0.8	1.2	
受験者	16.9%	2.12	0.82	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布			
			平均値	標準偏差	最小値	最大値	有効数	995.3	ピアソン	.000	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	
			-0.05	0.76	-2.82	3.16			ケンドールタウb	-.001	1.003	0.121	0.7	1.4	0.990	0.187	0.5	2.1	
E-2) 困難度設定値 s.d.=1.0 対 能力設定値 s.d.=1.0 データ3標本 平均値				BLOG測定項目Slope分布				BLOG項目SlopeとRasch項目Infit MnSqの相関											
				平均値	標準偏差	最小値	最大値	平均	ピアソン	-.437									
				0.720	0.512	0.177	1.850		ケンドールタウb	-.435									
								BLOG項目SlopeとRasch項目OutfitMnSqの相関											
								平均	ピアソン	-.533									
									ケンドールタウb	-.478									
E-3) 二次元性+項目スロープ細工(0.3~2.0)データ群: 困難度設定値 s.d.=2.0 対 能力設定値 s.d.=2.0 データ5標本 指標値平均																			
項目群	18.2%	17.34	1.00	FACETS出力項目困難度推定値分布				Rasch測定項目困難度推定値不変性				Rasch測定項目Infit MnSq統計量分布				Rasch測定項目Outfit MnSq統計量分布			
			平均値	標準偏差	最小値	最大値	有効数	49	ピアソン	.410	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	
			0.00	1.51	-2.94	3.75			ケンドールタウb	.298	1.000	0.091	0.9	1.2	0.954	0.175	0.6	1.4	
受験者	27.8%	2.89	0.89	FACETS出力受験者能力推定値分布				Rasch測定受験者能力推定値不変性				Rasch測定受験者Infit MnSq統計量分布				Rasch測定受験者Outfit MnSq統計量分布			
			平均値	標準偏差	最小値	最大値	有効数	967	ピアソン	-.038	平均値	標準偏差	最小値	最大値	平均値	標準偏差	最小値	最大値	
			-0.05	1.11	-3.62	4.78			ケンドールタウb	-.033	1.004	0.208	0.6	1.8	0.958	0.385	0.3	4.2	
E-3) 困難度設定値 s.d.=2.0 対 能力設定値 s.d.=2.0 データ5標本 平均値				BLOG測定項目Slope分布				BLOG項目SlopeとRasch項目Infit MnSqの相関											
				平均値	標準偏差	最小値	最大値	平均	ピアソン	-.400									
				1.283	1.243	0.156	4.064		ケンドールタウb	-.292									
								BLOG項目SlopeとRasch項目OutfitMnSqの相関											
								平均	ピアソン	-.484									
									ケンドールタウb	-.384									