

正誤反応データ Rasch モデル測定妥当性についての一考察

The Validity Level of the Binary Response Data in the Application of the Rasch Item Analysis Model

井澤 廣行^{*}、平越 裕之^{**}

Hiroyuki Izawa, Hiroyuki Hirakoshi

This paper describes the validity level, i.e., the level of the parameter estimates' invariance, of the binary response data in the application of the Rasch Item Analysis model. The validity level is determined by the degree of equivalence between the distributions of the ability and difficulty estimates measured by the Rasch model. The greater the standard deviations' equivalence of both estimates, the higher the validity level of the data in the model application. The data-model fit depends on the deviation size. This paper also touches upon the uselessness of the separate calibration t value and investigates the Outfit statistics' characteristics.

Key words : Rasch model, validity, estimates' standard deviation, separate calibration t value.

I. はじめに

平越・井澤¹⁾(2008a)による「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」において、その中心論点連関性のまとめとして与えたものが表 4-1(p. 32)であり、それを本稿の序として下に再掲する。

表 1 - 1 平越・井澤¹⁾(2008)所載表 4-1(p. 32)再掲

項目群	困難度推定値 標準偏差		受験者群	能力推定値 標準偏差	
	小	大		小	大
項目群第一主成分寄与率	大	小	受験者群第一主成分寄与率	大	小
受験者群第一主成分寄与率	小	大	項目群第一主成分寄与率	小	大
困難度推定値不変性成立度	小	大	能力推定値不変性成立度	小	大
困難度推定値妥当性・信頼性	小	大	能力推定値妥当性・信頼性	小	大

* 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

**流通科学大学情報学部、〒651-2188 神戸市西区学園西町 3-1

その表は、正誤反応データ(以降、01 データ)上での項目群と受験者群に関するピアソン相関係数に基づく主成分分析第一主成分寄与率(以降、第一主成分寄与率)は相互にトレードオフの関係にあることを明示した(平越・井澤¹⁾、2008a, p. 32)ものであり、各母数推定値分布の標準偏差は単一の制御変数とされている。その表に示された要因関係についての理解を深めるためには、両者の標準偏差を二つの制御変数として同時に取り扱う方がより現実的であると思われる。その認識で以って、第一主成分寄与率と母数推定値分布標準偏差それぞれにおける項目群と受験者群の間での均衡性、並びに、01 データの Rasch 項目分析モデル(Rasch²⁾、1960)測定(以降、Rasch 測定)への適用に関する妥当性の上でのデータ固有限界性が重要事項として意識される。それが本稿執筆の動機であり、その現象把握が本稿の目的である。併せて、Rasch 測定上で有用として一般的に参照されるモデル適合度指標値の分布傾向に関する数量的観察をおこなう。

II. 第一主成分寄与率と母数推定値分布標準偏差における項目群・受験者群間での均衡性、並びに、Rasch 測定妥当性達成度に関するデータ固有限界性

表 2-1 母数設定値変更上での人工模擬 150 項目対 150 名データ Rasch 測定妥当性程度の相違

		01データ ピアソン 主成分分析 第1主成分 寄与率	Rasch測定 標準化残差 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	Rasch測定項目困難度推定値不変性		折半 データ間 t ≥ 2の 百分率	各項目と 各受験者 の組み合わせ 上での標準化 残差 z ≥ 2の 百分率
						有効数	ピアソン		
1) 困難度設定値 -1.00~1.00 対 能力設定値 -1.00~1.00	項目群	7.8 %	2.6 %	3.22	0.91	150	.581	26.0 %	1.8 %
	受験者群	7.9 %	2.6 %	3.21	0.91	150	.435		
2) 困難度設定値 -3.00~3.00 対 能力設定値 -1.00~1.00	項目群	5.8 %	2.7 %	7.34	0.98	148	.898	20.3 %	5.5 %
	受験者群	39.9 %	2.5 %	2.68	0.88	150	.483		
3) 困難度設定値 -5.00~5.00 対 能力設定値 -1.00~1.00	項目群	4.4 %	2.7 %	6.54	0.98	124	.814	19.4 %	4.3 %
	受験者群	57.8 %	2.7 %	2.14	0.82	150	.403		
4) 困難度設定値 -3.00~3.00 対 能力設定値 -3.00~3.00	項目群	30.7 %	2.6 %	7.64	0.98	150	.911	20.7 %	4.4 %
	受験者群	30.8 %	2.5 %	7.62	0.98	150	.755		
5) 困難度設定値 -5.00~5.00 対 能力設定値 -3.00~3.00	項目群	21.4 %	2.6 %	8.90	0.99	143	.952	18.2 %	3.8 %
	受験者群	52.0 %	2.8 %	6.64	0.98	150	.735		
6) 困難度設定値 -5.00~5.00 対 能力設定値 -5.00~5.00	項目群	40.2 %	2.7 %	10.39	0.99	150	.950	18.7 %	3.3 %
	受験者群	40.9 %	2.5 %	10.33	0.99	150	.830		

表 2-1 が、第一主成分寄与率と母数推定値分布標準偏差それぞれにおける項目群と受験者群の間での均衡性、並びに、両群の Rasch 測定妥当性達成度に関するデータ固有限界性の一端を顯示するものである。本節での用語としての「Rasch 測定妥当性」の高低は、項目群と受験者群いずれも母数推定値不変性成立度高低に準ずると著者によりみなされている。その要点は、人工模擬 6 データ間での項目群・受験者群母数設定値分布幅均衡度異同に連関する第三列「01 データ第 1 主成分寄与率」均衡度異同、並びに、第七列の項目群・受験者群「Rasch 測定母数推定値不変性」均衡度異同への参照により明確とされる。その二列以外に与えた指標値は、次節以降での論述内容との関連性で以ての参考付加である。なお、各 01 データ作成法は静³⁾(2007, pp. 188-191)によって記述されているものであり、各データの項目群・受験者群母数設定両端値を分布幅とする等間隔母数設定として、Rasch 項目分析モデル(以降、Rasch モデル)正答確率式(平越・井澤¹⁾、2008a, p. 11、参照)に基づく Excel 上での RAND 関数適用無作為 01 生起である。

本来的には、各データにおける統計量について少なくとも 5 標本以上の平均値が取られるべきであるが、本節での主要な論点を明示する目的の上では、表 2-1 での各データ 1 標本統計量生起傾向の観察で事が足りると思われる。何故ならば、データ番号 1、4、6 における項目群と受験者群の間での 01 データ第 1 主成分寄与率と Rasch 測定母数推定値不変性成立度(項目群・受験者群 Rasch 測定妥当性程度)それぞれにおいて均衡性が観察されるからである。その均衡性は、当然に、同数の等間隔同一設定値から成る平行対応としての項目群と受験者群の母数設定によるものであり、両群母数設定値分布標準偏差が最大であるデータ番号 6 において Rasch 測定妥当性程度が最大となっている。表 2-1 全体としての教示は、平越・井澤¹⁾(2008a, p. 32)に示された項目群と受験者群のトレードオフ関係にある一次元充足度への認知と共に、1) 能力推定値分布標準偏差に均衡する困難度推定値分布標準偏差を持つ項目群から成るテスト設定の重要性、及び、2) 受験者群能力推定値分布によって規定される現実データ属性としての Rasch 測定妥当性達成度固有限界性への認識である。前者 1)は、素点上での受験者群能力測定誤差減少への心理意図の上で常識として従来から理解されている。一方、後者 2)についてはモデル識者認知の筈であるにもかかわらず、著者取得関連文献においてその簡明直截な指摘に欠けているものである。それが本稿執筆の動機でもある。

Rasch 測定妥当性達成度において受験者群能力分布に規定されるデータ固有限界性の一端が表 2-1 でのデータ番号 1、2、3 への参照により明確とされる。要するに、現実に 150 項目対 150 名の 01 反応データが与えられたとして、その Rasch 測定の上で仮に受験者群能力推定値分布幅が -1.00~1.00 であるとすれば、データ番号 1、2、3 での項目群困難度推定値分布幅に準ずる Rasch 測定妥当性達成度が表 2-1 に与えた指標上でほぼ最大となる。従って、その値と現実データにおける項目群と受験者群それぞれについての母数推定値不変性成立度との比較の上で、Rasch 測定妥当性達成度ないしは「一次元尺度構成体系」達成度が測られる(平越・井澤¹⁾、2008a、参照)。

なお、表 2-1 での Rasch 測定妥当性程度参照への第三列と七列以外に与えた付加指標値については以下の観察がなされる。Rasch 測定標準化残差の項目群・受験者群第 1 主成分寄与率はモデル識者によっては参照されていない指標であるが、データ次元充足度を観る上での 01 データ第 1 主成分寄与率参照の有用性との比較意図による付加である。その値に 6 データ間での相違がなく、その指標の無用性が認知される。母数推定値分離指標値と分離信頼性係数は Andrich⁴⁾(1982) に由来しており、項目群と受験者群の母数推定値がそれぞれいかなる程度に分離されているかを示す指標である。それは、モデル識者によって、データの受験者群と項目群それぞれの母集団における困難度推定値と能力推定値それぞれ一群の安定・再現程度を示す指標として有用とされている(Andrich⁴⁾, 1982; Fox and Jones⁵⁾, 1998, pp. 35-37; Linacre⁶⁾, 1989-2001, p. 77; Wright and Masters⁷⁾, 1982, pp. 105-106)。表 2-1 において、この母数推定値分離指標値についての項目群と受験者群の間での均衡度が、01 データ第 1 主成分寄与率における受験者群と項目群の間での均衡度、及び、Rasch 測定母数推定値不変性成立度における項目群と受験者群の間での均衡度に連関していることに注目される。「折半データ間 $|t| \geq 2$ の百分率」に関しては次節で更に言及するが、表 2-1 において項目群と受験者群に付された 5% をかなりの程度に上回るその 20% 前後の値に大きな変動が観察されない。「各項目と各受験者の組み合わせ上での標準化残差 $|z| \geq 2$ の百分率」については、『データが Rasch モデルに適合しているならば、Rasch 測定における標準化残差の約 5% 並びに約 1% が、それぞれ $|z|$ 並びに $|z|$ の範囲外の値となる』と Linacre⁶⁾ (1989-2001, p. 52) により指摘されている。表 2-1 での 6 データすべてにおいて $|z| \geq 2$ の百分率は 5.5% 以下であり、その指標値無変動性の一端が窺われる。又、01 データ第 1 主成分寄与率、Rasch 測定母数推定値不変性成立度、並びに、母数推定値分離指標値の上で、データ番号 1 は項目群・受験者群総体としての Rasch 測定妥当性においてデータ番号 4 と 6 に劣る。しかし、データ番号 1 についての「標準化残差 $|z| \geq 2$ の百分率」が最小値 1.8% となっている。各項目と各受験者の組み合わせ上での標準化残差 z についての項目群と受験者群それぞれに渡る平方和を項目数と受験者数それぞれで除した指標値が各受験者と各項目に付される Outfit カイ二乗平均平方統計量である。データ番号 1 の項目群と受験者群いずれに関しても、そのモデル適合度指標値の上でデータ番号 4 と 6 に優ると示されており、1.2 以上の値が付された項目群と受験者群は皆無である。これは、「項目群困難度推定値分布と受験者群能力推定値分布が、双方の狭い分布幅ないしは小さな標準偏差で以って、全体的に双方近接するに連れて、各項目と各受験者に付される Outfit カイ二乗平均平方統計量はその期待値 1.0 に近づく」(平越・井澤⁸⁾, 2008b, pp. 75-76、参照) ことによると理解される。

ちなみに、前記の受験者群能力推定値分布幅に均衡する項目群困難度推定値分布幅の重要性に着目して、表 2-1 でのデータ番号 3 と 5 それぞれについて 150 項目からそれぞれの能力設定値幅範囲外にある Rasch 測定困難度推定値を持つ項目群の削除の上で作成されたデータ 3' と 5' の Rasch 測定妥当性程度を示すものが表 2-2 である。

表 2-2 3') 28 項目 (母数推定値幅-1.00~1.00) x150 名 (母数設定値幅-1.00~1.00)、及び、5') 88 項目 (母数推定値幅-3.00~3.00) x150 名 (母数設定値幅-3.00~3.00) の Rasch 測定妥当性程度

		01データ ピアソン 主成分分析 第1主成分 寄与率	Rasch測定 標準化残差 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			折半 データ間 $ t \geq 2$ の 百分率	各項目と 各受験者 の組み合わせ 上での標準化 残差 $ z \geq 2$ の 百分率
3') 困難度推定値 -1.00~1.00 (28項目) 対 能力設定値 -1.00~1.00 (150名)	項目群	10.4 %	7.2 %	3.29	0.92	Rasch測定項目困難度推定値不変性 有効数 28		53.6 %	2.2 %
	受験群者	11.2 %	7.0 %	1.35	0.65	ケンドールタウb .253			
						Rasch測定受験者能力推定値不変性 有効数 148			
							ケンドールタウb .264		
						ケンドールタウb .197			
5') 困難度推定値 -3.00~3.00 (88項目) 対 能力設定値 -3.00~3.00 (150名)	項目群	30.6 %	3.4 %	7.58	0.98	Rasch測定項目困難度推定値不変性 有効数 88		21.6 %	4.2 %
	受験群者	30.7 %	3.3 %	5.80	0.97	ケンドールタウb .729			
						Rasch測定受験者能力推定値不変性 有効数 150			
							ケンドールタウb .896		
						ケンドールタウb .739			

データ 3' と 5' において、01 データと Rasch 測定標準化残差のいずれにもおける項目群と受験者群についての第 1 主成分寄与率同等性、及び、母数推定値不変性成立度同等性が観察される。その事象は、項目群と受験者群の母数推定値分布幅同等性によりもたらされていると思われる。更に、表 2-2 での 5' と前掲表 2-1 での 4 との指標値比較の上で、項目群と受験者群いずれについても、01 データ第 1 主成分寄与率、母数推定値不変性成立度、並びに、母数推定値分離指標値・分離信頼性係数において大きな違いのない同等な値であることに注目される。それは、項目数が 150 から 88 に減少されたにもかかわらず、項目群と受験者群の母数推定値分布幅をほぼ同様な範囲-3.00~3.00 とされた 88 項目対 150 名データと 150 項目対 150 名データの比較一例の上で、前者が後者に比肩する Rasch 測定妥当性を保持しているということである。又、表 2-1 での 5 における項目群と受験者群の間での不均衡な Rasch 測定妥当性程度が表 2-2 での 5' において均衡化されて、5' における 01 データ第 1 主成分寄与率が上昇していることに併せて、5' における Rasch 測定受験者能力推定値不変性成立度が 5 におけるその程度とほぼ同等となっている。上記の標示により、データ 5 において能力推定値分布幅-3.00~3.00 の範囲外にある困難度推定値を持つ項目群は、データ 5 での受験者群 150 名の能力測定に関して妥当性の低い余剰項目群であると理解される。データ 5' の作成要領は現実データへの有効利用を示唆しており、特に、item banking(テスト受験者群母集団への有用項目群集積)の上でその方法が必須とされる筈である。多くの受験者数を対象とする実力テストにおいては Rasch 測定能力母数推定値は一般的に正規分布となるけれども、-4.00~4.00 をその一般的な分布範囲幅とする故に、-4.00~4.00 の範囲内での困難度推定値等間隔性を満たす項目群、すなわち、高い困難度推定値不変性成立度を属性とする項目群(平越・井澤¹⁾、2008a, pp. 30-31、参照)の集積がテストの受験者群母集団を考慮した item banking への最重

要指針になると思われる。

Ⅲ. Rasch 測定妥当性判断における「折半データ間 $|t| \geq 2$ の百分率」への不信任

「折半データ間 t 値」は、‘separate calibration t statistic’ (Smith⁹⁾, 1996, p. 407)の和訳であり、母数最尤推定値分布の漸近正規性(豊田¹⁰⁾, 2002, pp. 64-65、参照)に基づく折半データ間での Rasch 測定母数推定値同等性検証法として Wright¹¹⁾ (1967)による研究報告での使用に由来する。Rasch 測定出力プログラムの上では、一般的に項目群困難度推定値平均が 0 と設定されることにより、受験者群折半データ AB 間での各項目 i に関する t_{iAB} 値、及び、項目群折半データ CD 間での各受験者 n に関する t_{nCD} 値はそれぞれ以下の様に算出される(McNamara¹²⁾, 1991, pp. 153-154; Wright and Stone¹³⁾, 1979, p. 95、参照)。

$$t_{iAB} = (d_{iA} - d_{iB}) / (e_{iA}^2 + e_{iB}^2)^{1/2} \sim N(0, 1)$$

d_{iA} : 受験者群折半下位データ A の Rasch 測定による項目 i の困難度推定値

d_{iB} : 受験者群折半下位データ B の Rasch 測定による項目 i の困難度推定値

e_{iA} : 受験者群折半下位データ A の Rasch 測定による項目 i の困難度推定値標準誤差

e_{iB} : 受験者群折半下位データ B の Rasch 測定による項目 i の困難度推定値標準誤差

$$t_{nCD} = [(b_{nC} - b_{nD}) - (C_{mean} - D_{mean})] / (e_{nC}^2 + e_{nD}^2)^{1/2} \sim N(0, 1)$$

b_{nC} : 項目群折半下位データ C の Rasch 測定による受験者 n の能力推定値

b_{nD} : 項目群折半下位データ D の Rasch 測定による受験者 n の能力推定値

C_{mean} : 項目群折半下位データ C の Rasch 測定による受験者群能力推定値平均

D_{mean} : 項目群折半下位データ D の Rasch 測定による受験者群能力推定値平均

e_{nC} : 項目群折半下位データ C の Rasch 測定による受験者 n の能力推定値標準誤差

e_{nD} : 項目群折半下位データ D の Rasch 測定による受験者 n の能力推定値標準誤差

Smith⁹⁾ (1996, p. 409)により、50 項目と 1,000 名をそれぞれ一対とする 10 種類の Rasch モデルに適合している人工模擬データの分析上で項目群に関するこの t 値の分布について $|t| \geq 2$ と付された項目数の全数における百分率が平均として 4.6%であり、Rasch モデル適合データにおいては t 値分布の漸近正規性に準じる傾向が報告されている。但し、Smith⁹⁾ (1996, p. 409)によるその 10 種類 1,000 名対 1,000 名データ間の 50 項目 t 値比較についての正確なデータ作成法が筆者には不詳である。この折半データ間 t 値は Smith and Suh¹⁴⁾ (2003)においても有効有用とされている。能力上位群・下位群それぞれ 1,000 名による高校数学多肢選択式 80 問への応答の Rasch 測定の上で、 $|t| \geq 2$ が付された項目数 60(75%)、並びに、その受験者群折半間での 80 項目困難度推定値散布図で以って、テスト 80 問全体としての Rasch モデル不適合が言明されている(pp. 158-159)。

なお、Rasch モデル適合度の低いデータについては折半データ間 t 値のデータ受験者群折半法への依存性が顕著であり、その様な潜在特性一次元性欠如データ上での t 値の項目個別 Rasch モデ

ル適合度指標としての有用性への疑問が井澤¹⁵⁾ (2006, pp. 48-50)によって付されている。又、著者の使用するデータ折半法は、受験者群と項目群それぞれについての Rasch 測定標準化残差の第 1 主成分負荷量正負記号分別によるものである。その折半法使用は、受験者群能力上位・下位群分別及び項目群困難度上位・下位群分別を含む他のデータ折半法に優って信頼性最上位を示唆する母数推定値不変性成立度最下限の値を与えるとの経験則に基づいている。

ここで、前節掲示表 2-1 と 2-2 を再び参照する。表 2-1 での Rasch 測定妥当性程度に明らかに相違のある 6 データに関する各受験者群と各項目群の折半データ間 $|t| \geq 2$ の百分率すべてが 5% を大幅に超えてしかもほぼ一律に 20% 前後の値である。又、表 2-2 において Rasch 測定能力推定値不変性成立度の非常に低いデータ 3' の項目群折半データ間での $|t| \geq 2$ の百分率が 20% の 5 割としての 10.8% である。その $|t| \geq 2$ の百分率 10.8% 算出におけるデータ 3') 148 名の t 値出力上での分子、分母、及び、表 2-1 でのデータ番号 1 にて対応する $|t| \geq 2$ の百分率 24.0% 算出における 150 名の t 値出力上での分子、分母についての基本統計量を比較したものが表 3-1 である。なお、データ 3' での 28 項目折半、及び、1 での 150 項目折半は、Rasch 測定標準化残差第 1 主成分負荷量上でのそれぞれ正記号 15 項目対負記号 13 項目、及び、正記号 74 項目対負記号 76 項目である。

表 3-1 データ 3' と 1 の受験者群能力に関する t 値算出上での分子と分母の比較基本統計量

	3') 分子	3') 分母	3') t 値	1) 分子	1) 分母	1) t 値
最小値	-2.666	0.799	-2.545	-1.713	0.339	-4.551
平均値	0.006	0.873	0.049	0.000	0.356	-0.002
最大値	3.184	1.200	3.336	1.847	0.412	4.711
標準偏差	1.142	0.083	1.249	0.619	0.016	1.695

3') 148 名 t 値出力上での分母が全体的に 1) 150 名 t 値出力上での分母の 2 倍強となっているために、1) においてよりも 3') において t 値が全体的に小値となっている。その結果として、データ 3' の 148 名に関する $|t| \geq 2$ の百分率 10.8% が、データ 1 の 150 名に関する $|t| \geq 2$ の百分率 24.0% よりも小さくなっていると理解される。すなわち、折半データ間 t 値は、折半された標本の大きさ大小、上の例では項目数大小、の影響によってもたらされる対象一対母数推定値標準誤差小大への傾向が全体としての t 値大小につながるという属性を持っているのである。

上記の折半データ間 t 値の属性を簡明直截に例示するものが表 3-2 であり、折半標本の大きさに大なる影響を受ける折半データ間 $|t| \geq 2$ の百分率増大傾向が顕著に示されている。A と B いずれにおいても受験者群 2,500 名折半による下位データにおける 1,250 名前後の大なる受験者数の影響により、各下位データ Rasch 測定上での 100 項目困難度推定値標準誤差が大変に小さな値となる。その結果として、100 項目 t 値算出における全体的に小さな分母が全体的に大きな t 値をもたらしているのである。それは、AB 各一対比較の上で確認済みである。AB 作成要領により、それ

それぞれにおいて 100 項目対 2,500 名データが 100 項目対 150 名データ以上に項目群と受験者群のいずれについても Rasch 測定妥当性において上回っていることは、表 3-2 に付加した母数推定値不変性成立度への参照の上で明白である。然しながら、AB それぞれでの 100 項目対 2,500 名データにおける 100 項目についての $|t| \geq 2$ の百分率が示唆していることは、受験者群折半データ間での 100 項目困難度推定値不変性の大きな否定である。それは明らかに誤謬である。これにより、Rasch モデル識者(Bond and Fox¹⁶⁾, 2001, pp. 54-60; McNamara¹²⁾, 1991, pp. 153-154; Wright and Stone¹³⁾, 1979, p. 95)、特に、Smith⁹⁾(1996, p. 409)、及び、Smith and Suh¹⁴⁾(2003)が母数推定値不変性成立度確認法として推奨する折半データ間 $|t| \geq 2$ の百分率への参照は事実を反する側面を持つと言明される。又、表 3-2 での参考付加指標「各項目と各受験者の組み合わせ上での標準化残差 $|z| \geq 2$ の百分率」については、第二節掲示表 2-1 での 6 データすべてにおけるその百分率 5.5%以下との無変動性が、同様に、A と B、並びに、受験者数 150 名と 2,500 名のデータ間にほぼ同値 4.7%として観察される。従って、この指標は Rasch 測定妥当性程度検知鋭敏性に欠けると判断される。

表 3-2 折半下位データにおける大なる受験者数の影響による項目群 t 値増大の例示

A) 項目群母数設定値-2.00~2.00範囲の100項目x150名、並びに、 100項目x2500名データ5標本Rasch測定妥当性指標値の平均値						
					折半データ間	
100項目 x 150名 5標本 平均値	各項目と 各受験者 の組み合わせ 上での標準化 残差 $ z \geq 2$ の 百分率 4.7%	Rasch測定項目困難度推定値不変性		$ t \geq 2$		
		有効数	100	ピアソン	.824	26.6 %
				ケンドールタウb	.673	
		Rasch測定受験者能力推定値不変性		$ t \geq 2$		
有効数	150	ピアソン	.768	19.2 %		
		ケンドールタウb	.568			
100項目 x 2500名 5標本 平均値	各項目と 各受験者 の組み合わせ 上での標準化 残差 $ z \geq 2$ の 百分率 4.6%	Rasch測定項目困難度推定値不変性		$ t \geq 2$		
		有効数	100	ピアソン	.927	69.4 %
				ケンドールタウb	.771	
		Rasch測定受験者能力推定値不変性		$ t \geq 2$		
有効数	2498	ピアソン	.866	7.3 %		
		ケンドールタウb	.696			
B) 項目群母数設定値-4.00~4.00範囲の100項目x150名、並びに、 100項目x2500名データ5標本Rasch測定妥当性指標値の平均値						
					折半データ間	
100項目 x 150名 5標本 平均値	各項目と 各受験者 の組み合わせ 上での標準化 残差 $ z \geq 2$ の 百分率 4.7%	Rasch測定項目困難度推定値不変性		$ t \geq 2$		
		有効数	97	ピアソン	.930	19.4 %
				ケンドールタウb	.800	
		Rasch測定受験者能力推定値不変性		$ t \geq 2$		
有効数	150	ピアソン	.704	18.4 %		
		ケンドールタウb	.513			
100項目 x 2500名 5標本 平均値	各項目と 各受験者 の組み合わせ 上での標準化 残差 $ z \geq 2$ の 百分率 4.8%	Rasch測定項目困難度推定値不変性		$ t \geq 2$		
		有効数	100	ピアソン	.978	58.2 %
				ケンドールタウb	.884	
		Rasch測定受験者能力推定値不変性		$ t \geq 2$		
有効数	2500	ピアソン	.822	7.5 %		
		ケンドールタウb	.641			

表 4-2 B) 項目群母数-4.00~4.00 範囲等間隔設定値対受験者群母数標準正規分布設定値
6 データの標本の大きさに準ずる Rasch 測定妥当性程度の相違

	項目群	折半データ間			項目群	折半データ間			
		Rasch測定項目困難度推定値不変性	有効数	$ t \geq 2$		Rasch測定項目困難度推定値不変性	有効数	$ t \geq 2$	
1) 20 項目 x 150 名 (No. 1 データ)	項目群	Rasch測定項目困難度推定値不変性			2) 20 項目 x 2500 名 (No. 1 データ)	項目群	Rasch測定項目困難度推定値不変性		
		有効数	20	35.0 %			有効数	20	55.0 %
	受験者群	Rasch測定受験者能力推定値不変性			受験者群	Rasch測定受験者能力推定値不変性			
		有効数	147	17.7 %		有効数	2490	9.1 %	
3) 100 項目 x 150 名 (No. 1 データ)	項目群	Rasch測定項目困難度推定値不変性			4) 100 項目 x 2500 名 (No. 1 データ)	項目群	Rasch測定項目困難度推定値不変性		
		有効数	96	20.8 %			有効数	100	61.0 %
	受験者群	Rasch測定受験者能力推定値不変性			受験者群	Rasch測定受験者能力推定値不変性			
		有効数	150	19.3 %		有効数	2500	7.0 %	
5) 250 項目 x 150 名 (No. 1 データ)	項目群	Rasch測定項目困難度推定値不変性			6) 250 項目 x 2500 名 (No. 1 データ)	項目群	Rasch測定項目困難度推定値不変性		
		有効数	242	14.1 %			有効数	250	49.2 %
	受験者群	Rasch測定受験者能力推定値不変性			受験者群	Rasch測定受験者能力推定値不変性			
		有効数	150	24.0 %		有効数	2500	8.6 %	

本節の序として、表 4-1 と 4-2 において、項目群と受験者群の標本がそれぞれ大きくなるにつれて、Rasch 測定母数推定値不変性成立度も高くなることを示す。その提示意図は、平越・井澤¹⁾ (2008a)における表 3-2(p. 22)での同一受験者数に対する項目数増大の上での傾向観察の敷衍である。これは、要因が一定の場合において、『選抜者集団では常に相関係数が低くなる』(村上¹⁸⁾、2007, p. 216)ことによる項目数大小、及び、受験者数大小と Rasch 測定母数推定値不変性成立度高低との連関であると思われる。現実データへのその示唆は、言わずもがなである。

各データ作成法は、前節掲示表 3-2 においてと同じく、それぞれ-2.00~2.00 範囲、-4.00~4.00 範囲とする該当項目数等間隔母数設定値、及び、標準正規分布 $N(0, 1)$ により発生された該当受験者数母数設定値に基づく Excel 上での Rasch モデル正答確率式と RAND 関数の適用による無作為 01 生起である。それ故に、表 4-1 と 4-2 でのそれぞれ平均として 1.2 と 2.4 前後の値を標準偏差とする A)項目群-2.00~2.00 範囲困難度設定値分布幅、及び、B)項目群-4.00~4.00 範囲困難度設定値分布幅と標準偏差を同一値 1.0 とする受験者群能力設定値分布との相対関係により、困難度推定値不変性成立度高低と能力推定値不変性成立度低高のトレードオフ関係が観察される。項目群と受験者群の総体としてのデータ Rasch 測定妥当性の観点からは、両群の母数推定分布標準偏差が近接している A)項目群-2.00~2.00 範囲困難度設定値分布幅各データがその対応 B)-4.00~4.00 範囲データよりも優っている。なお、各データに付された No. 1 の意味は、以降に観察する 10 標本資料作成上での第 1 番データとのことである。参考として、折半データ間 $|t| \geq 2$ の百分率を付加しており、それは折半データ間 t 値について前節で論述されたその分布属性を再視認するためである。その特徴が再度観察されて、折半データにおける項目数と受験者数を無視したその $|t| \geq 2$ の百分率への一律的な参照による受験者群と項目群についての Rasch 測定母数推定値不変性成立度の確認は、信頼性に大きく欠けると断言される。折半間 t 値についてのこれ以上の付言は不要

である。

Rasch 測定汎用コンピュータプログラム Facets (by Linacre¹⁹, 1989-2001, Winsteps.com)により、各項目と各受験者に関するモデル適合度指標 Outfit・Infit 統計量、並びに、項目群と受験者群それぞれ全体としての Rasch 測定妥当性程度を示す分離指標値・分離信頼性係数が出力される。前者の Outfit・Infit 統計量がモデル適合度検知において不完全な指標であることは識者においては周知の事実である(平越・井澤⁸、2008b, pp. 72-74、参照)。然し、その指標値参照の上で、永年に渡って筆者に査察確認への関心を抱かせた実証考察は Smith, Schumacker, and Bush²⁰ (1998)によるものである。それは、受験者標本の大きさに影響を受ける度合いが各項目に付される Outfit・Infit Z 値(標準化統計量)よりも、モデル識者認知として、一般的に小さいとされているその対応 Outfit・Infit カイ二乗平均平方統計量(McNamara²¹, 1996, p. 129; Wilson²², 2005, p. 129)への Smith, Schumacker, and Bush²⁰ (1998)による反証事例考察である。その要約は、受験者数変動に対して項目群に付される Outfit・Infit Z 値が、そのはずれ値個数の割合($\% > 2$)において Outfit・Infit カイ二乗平均平方統計量によるそのはずれ値個数の割合($\% > 1.1$)に優ってより安定した分布属性を持っているとのことである。特に、Outfit Z 値のはずれ値個数の割合($\% > 2$)が各テスト設計上での平均値 2.5%前後と示されて(p. 76)、その Outfit Z 値参照の有用性が示唆されている。

Smith, Schumacker, and Bush²⁰ (1998)によるモデル母数設定値、テスト設計、及び、「複製」数に関する説明は以下の通りである(p. 70)。但し、正確なその模擬データ作成法は筆者に不詳である。

- 1) 受験者群能力は標準正規分布 $N(0, 1)$ とされている。
- 2) 項目群困難度は-2.00 から 2.00 に渡って「一様に分布された」とされている。
- 3) テスト設計は、20 項目対 150 名、50 項目対 150 名、20 項目対 500 名、50 項目対 500 名、20 項目対 1,000 名、50 項目対 1,000 名の 6 種類とされている。
- 4) 上記 6 種類各テスト設計の 100「複製」とされている。

上記の Smith, Schumacker, and Bush²⁰ (1998)による考察事例との比較も意図された上で、Rasch モデル適合度指標値の標本の大きさに準ずる変動性について概観するために、以下の母数設定値、テスト設計、及び、標本数とした。

- 1) 受験者群能力を標準正規分布 $N(0, 1)$ により発生されたものとする。
- 2) 項目群困難度分布について、A) -2.00 から 2.00 に渡る等間隔設定、及び、B) -4.00 から 4.00 に渡る等間隔設定の 2 種類とする。
- 3) 項目数と受験者数の変動を次のものとして、テスト設計を各組み合わせすべてとする。

項目数： 20、50、100、150、200、250

受験者数： 150、300、500、1,000、2,500、5,000、10,000、30,000

- 4) 標本数を資料作成時間制限により上記の各テスト設計について各 10 標本とする。

なお、各テスト設計の上での、01 データ 1 標本取得の方法は、該当項目数等間隔母数設定値、

及び、標準正規分布該当受験者数母数設定値に基づく Excel 上での Rasch モデル正答確率式と RAND 関数の適用による無作為 01 生起である。参照として、表 4-3 に、A) -2.00~2.00 範囲等間隔母数設定値、及び、B) -4.00~4.00 範囲等間隔母数設定値とされた各項目数標本設計における母数設定値分布標準偏差、並びに、各受験者数標本設計における標準正規分布能力設定値の標準偏差、最小値、最大値についての 10 標本平均値を与えている。

表 4-3 項目群母数設定値等間隔分布の標準偏差、並びに、受験者群母数設定値標準正規分布の標準偏差、最小値、最大値についての 10 標本平均値

項目群困難度等間隔設定値標準偏差						
	20 項目	50 項目	100 項目	150 項目	200 項目	250 項目
A) 設定値範囲 -2.00~2.00	1.25	1.19	1.17	1.17	1.16	1.16
B) 設定値範囲 -4.00~4.00	2.49	2.38	2.34	2.33	2.33	2.32

受験者群能力標準正規分布設定値における標準偏差、最小値、最大値の10標本平均値								
分布幅 統計量	150 名	300 名	500 名	1,000 名	2,500 名	5,000 名	10,000 名	30,000 名
標準偏差	0.98	0.99	0.98	1.01	1.00	1.00	1.00	1.00
最小値	-2.60	-2.96	-3.01	-3.37	-3.52	-3.77	-3.92	-4.01
最大値	2.74	2.74	3.02	3.13	3.48	3.83	3.78	4.09

表 4-4 に、A) -2.00~2.00 範囲項目群母数設定値データ群、及び、B) -4.00~4.00 範囲項目群母数設定値データ群についての 10 標本 Rasch 測定項目群・受験者群母数推定値分布幅統計量平均値を与える。なお、Excel 上での行数と列数の制限により、100 項目対 30,000 名、150 項目対 30,000 名、200 項目対 10,000 名、200 項目対 30,000 名、250 項目対 10,000 名、250 項目対 30,000 名の各 01 データ作成が不可能であることから、表 4-4、及び、以降表提示での平均値算出の上でそれらの標本を除外している。前掲表 4-3 での項目群と受験者群の母数設定値に関する分布標準偏差に比べて、表 4-4 での AB それぞれにおける 10 標本 Rasch 測定項目群・受験者群母数推定値分布標準偏差平均値は、項目群と受験者群のいずれについてもデータ群全体に渡って A よりも B においてその値の増大性が僅かながらも大きくなっていることが観察される。但し、項目群についての A に対する B での母数設定値分布標準偏差二倍の関係は、Rasch 測定困難度推定値分布においてもデータ群全体に渡ってほぼ同様に保持されている。

表 4-4 A) -2.00~2.00 範囲項目群母数設定値データ群、及び、B) -4.00~4.00 範囲項目群母数設定値データ群に関する 10 標本 Rasch 測定項目群・受験者群母数推定値分布幅統計量平均値

A) -2.00~2.00 範囲項目群母数設定値データ 項目群10 標本 Rasch 測定困難度推定値分布幅統計量平均値							
標 本	分布幅 統計量	20 項目	50 項目	100 項目	150 項目	200 項目	250 項目
平均値	標準偏差	1.31	1.22	1.19	1.18	1.18	1.18
	最小値	-2.11	-2.11	-2.13	-2.13	-2.14	-2.16
	最大値	2.13	2.09	2.11	2.09	2.16	2.16

A) -2.00~2.00 範囲項目群母数設定値データ 受験者群10 標本 Rasch 測定能力推定値分布幅統計量平均値									
標 本	分布幅 統計量	150 名	300 名	500 名	1,000 名	2,500 名	5,000 名	10,000 名	30,000 名
平均値	標準偏差	1.05	1.06	1.04	1.07	1.07	1.07	1.09	1.14
	最小値	-2.85	-3.26	-3.34	-3.65	-3.88	-4.06	-4.21	-4.06
	最大値	2.97	3.05	3.29	3.43	3.84	4.15	4.19	4.06

B) -4.00~4.00 範囲項目群母数設定値データ 項目群10 標本 Rasch 測定困難度推定値分布幅統計量平均値							
標 本	分布幅 統計量	20 項目	50 項目	100 項目	150 項目	200 項目	250 項目
平均値	標準偏差	2.70	2.47	2.39	2.37	2.36	2.36
	最小値	-4.37	-4.22	-4.27	-4.28	-4.37	-4.39
	最大値	4.38	4.27	4.26	4.27	4.38	4.41

B) -4.00~4.00 範囲項目群母数設定値データ 受験者群10 標本 Rasch 測定能力推定値分布幅統計量平均値									
標 本	分布幅 統計量	150 名	300 名	500 名	1,000 名	2,500 名	5,000 名	10,000 名	30,000 名
平均値	標準偏差	1.07	1.09	1.07	1.10	1.09	1.10	1.12	1.21
	最小値	-2.92	-3.38	-3.35	-3.73	-3.96	-4.16	-4.49	-5.14
	最大値	3.04	3.05	3.41	3.52	3.89	4.26	4.38	5.07

ここで、付言が必要となる。それは、Smith, Schumacker, and Bush²⁰⁾ (1998)の論考においては観察されていない項目群と受験者群それぞれの標本の大きさ大小とデータの Rasch 測定妥当性程度高低の連関である。その明示意図で以って、本節冒頭に、その観察上での資料作成時間制限により 6 データ各 1 標本提示に限られてはいるが、表 4-1 と 4-2 を与えたのである。又、項目群母数設定値幅として、Smith, Schumacker, and Bush²⁰⁾ (1998)での-2.00~2.00 範囲データ群に加えて、-4.00~4.00 範囲データ群についての提示は、項目群困難度と受験者群能力の間での Rasch 測定母数推定値分布均衡・乖離性によるモデル適合度指標値の変動相違を観察するためである。

上記の項目群と受験者群それぞれの標本の大きさ大小とデータの Rasch 測定妥当性程度高低の

付記として、第二節冒頭提示表 2-1 において注目されたことは、Rasch 測定母数推定値分離指標値についての項目群と受験者群の間での均衡度が、01 データ第 1 主成分寄与率における受験者群と項目群の間での均衡度、及び、Rasch 測定母数推定値不変性成立度における項目群と受験者群の間での均衡度に連関していたことである。

ちなみに、受験者群能力推定値についての分離指標値と分離信頼性係数は以下の様に算出されている(静³⁾、2007, p. 295; Wright and Masters⁷⁾, 1982, pp. 105-106、参照)。

能力推定値分離指標値(Person Separation Index)

$$= [\text{Var}(\hat{\theta}) - \text{Var}(e)]^{1/2} / [\text{Var}(e)]^{1/2}$$

$$= [\text{能力推定値分散} - (\text{能力推定値誤差分散総和}/N)]^{1/2} / [(\text{能力推定値誤差分散総和}/N)]^{1/2}$$

能力推定値分離信頼性係数(Person Separation Reliability)

$$= [\text{Var}(\hat{\theta}) - \text{Var}(e)] / \text{Var}(\hat{\theta})$$

$$= [\text{能力推定値分散} - (\text{能力推定値誤差分散総和}/N)] / \text{能力推定値分散}$$

なお、N は受験者数。

従って、大きさを同一とする一対受験者標本の比較において、能力推定値標準偏差がほぼ同じである場合には、各受験者に付される能力推定値標準偏差が受験者群全体に渡って大きいならば、能力推定値分離指標値とその分離信頼性係数は小さくなる。これにより、表 4-5 での A) -2.00~2.00 範囲項目群母数設定値データ群と B) -4.00~4.00 範囲項目群母数設定値データ群における同一項目数対同一受験者数標本の上で、A と比べて B での受験者群全体として平均的に大きい能力推定値標準偏差が、A と比べて B での能力推定値分離指標値・分離信頼性係数を小値としている。それは、同一項目数対同一受験者数標本の上で、データ全体として受験者群能力水準に相応する困難度を持つ項目群が B では A 以下の少数となっていることを意味する(静³⁾、2007, p. 296、参照)。つまり、例えば受験者群の能力推定値分布標準偏差が 1.0 である場合に、その標準偏差に近接する項目群困難度推定値分布を持つデータのその受験者群能力に関する Rasch 測定妥当性が高いということである。これは、本節冒頭提示表 4-1 と 4-2 での同一項目数対同一受験者数標本の上で、A) -2.00~2.00 範囲項目群母数設定値データ群が B) -4.00~4.00 範囲項目群母数設定値データ群に比べて高い Rasch 測定能力推定値不変性成立度を持つということに呼応している。

項目群困難度推定値分離指標値・分離信頼性係数も上記と同様に算出されるが、その分子に大きく影響する困難度推定値分布標準偏差が、同一項目数対同一受験者数標本の上で B でのその値が一律に A でのその対応値の二倍となっている。それが、A と B での一対標本比較上での項目群全体としての平均的な推定値標準偏差の相違を無効にして、A に比べて B での困難度推定値分離指標値・分離信頼性係数を大きな値としている。更に、同一項目数対同一受験者数標本の上で A に比べて B での困難度推定値間隔が一様に大きいことが、前掲表 4-1 と 4-2 にて観察される様に、B の A に優る Rasch 測定困難度推定値不変性成立度に連関している。但し、前段落記述との関連

の上で、受験者群能力分布に均衡していない項目群困難度分布については、項目群の Rasch 測定妥当性は低いことに留意される。要するに、受験者群一団への「難し過ぎる」、あるいは、「易し過ぎる」項目群の実施テストへの混入は、その受験者群能力と項目群困難度いずれにも関する Rasch 測定妥当性低位を意味する。この点において、平越・井澤¹⁾(2008)所載表 4-1(p. 32)での用語としての「妥当性」に関して、項目群困難度と受験者群能力との分布相互均衡性という重要視点が欠如している。

表 4-6 A) -2.00~2.00 範囲項目群母数設定値データ群、及び、B) -4.00~4.00 範囲項目群母数設定値データ群に関する 10 標本項目群 Outfit 統計量はずれ値百分率平均値

A) -2.00~2.00範囲項目群母数設定値データ群

項目群 Outfitカイニ乗平均平方統計量10標本はずれ値百分率平均値

標本	20項目				50項目				100項目				150項目				200項目				250項目			
	%>1.0	%>1.1	%>1.2	%>1.3	%>1.0	%>1.1	%>1.2	%>1.3	%>1.0	%>1.1	%>1.2	%>1.3	%>1.0	%>1.1	%>1.2	%>1.3	%>1.0	%>1.1	%>1.2	%>1.3	%>1.0	%>1.1	%>1.2	%>1.3
150名	29.0	13.0	6.0	2.5	31.4	9.4	4.0	2.4	30.3	10.8	4.1	2.2	30.5	11.5	4.4	2.1	29.1	12.7	4.9	2.0	30.3	12.4	4.8	2.2
300名	30.0	11.5	3.5	1.0	27.6	8.8	3.0	1.6	27.6	7.1	2.2	0.6	24.6	5.7	1.3	0.5	26.1	6.5	1.7	0.1	26.1	6.4	1.2	0.6
500名	25.5	5.0	2.0	1.5	20.6	3.6	0.4	0.0	22.3	2.9	0.6	0.2	21.1	3.5	0.8	0.3	21.5	3.3	0.7	0.1	22.2	3.1	0.7	0.2
1,000名	19.5	4.5	1.5	0.5	19.8	1.8	0.2	0.0	16.8	0.8	0.0	0.0	16.7	1.3	0.1	0.0	15.9	1.0	0.1	0.0	15.6	1.1	0.1	0.0
2,500名	16.0	1.0	0.0	0.0	8.2	0.0	0.0	0.0	8.1	0.0	0.0	0.0	7.2	0.1	0.0	0.0	7.7	0.1	0.0	0.0	6.5	0.0	0.0	0.0
5,000名	10.5	0.0	0.0	0.0	4.8	0.0	0.0	0.0	3.3	0.0	0.0	0.0	2.6	0.0	0.0	0.0	2.4	0.0	0.0	0.0	2.4	0.0	0.0	0.0
10,000名	3.5	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.5	0.0	0.0	0.0	*	*	*	*	*	*	*	*
30,000名	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
平均値	16.8	4.4	1.6	0.7	14.2	3.0	1.0	0.5	15.5	3.1	1.0	0.4	14.7	3.2	0.9	0.4	17.1	3.9	1.2	0.4	17.2	3.8	1.1	0.5

項目群 Outfit Z値10標本はずれ値百分率平均値

標本	20項目				50項目				100項目				150項目				200項目				250項目			
	%>0	%>1	%>2	%>3	%>0	%>1	%>2	%>3	%>0	%>1	%>2	%>3	%>0	%>1	%>2	%>3	%>0	%>1	%>2	%>3	%>0	%>1	%>2	%>3
150名	8.0	2.5	0.0	0.0	8.8	1.8	0.4	0.0	10.9	1.8	0.2	0.1	10.6	1.5	0.3	0.1	12.3	2.0	0.2	0.0	11.2	1.5	0.3	0.0
300名	7.0	1.5	0.0	0.0	12.2	2.2	0.6	0.0	12.8	1.3	0.2	0.0	11.1	1.6	0.4	0.0	11.7	1.3	0.2	0.0	11.6	1.7	0.2	0.0
500名	11.5	3.0	1.0	0.0	10.4	1.8	0.0	0.0	10.8	1.6	0.0	0.0	12.8	2.0	0.3	0.1	12.3	1.1	0.1	0.0	11.7	1.8	0.2	0.0
1,000名	12.5	1.5	0.5	0.5	13.8	1.0	0.0	0.0	12.1	1.3	0.4	0.1	12.4	1.5	0.4	0.1	12.8	1.4	0.1	0.0	12.5	1.7	0.2	0.0
2,500名	16.5	2.5	0.0	0.0	11.0	1.0	0.0	0.0	12.6	1.6	0.2	0.0	11.9	1.1	0.1	0.0	12.4	1.6	0.1	0.0	12.4	1.2	0.1	0.0
5,000名	17.0	2.0	0.0	0.0	15.0	1.8	0.2	0.0	13.5	0.9	0.1	0.0	12.0	1.7	0.1	0.0	11.8	1.2	0.1	0.0	12.0	1.4	0.0	0.0
10,000名	16.0	2.0	0.0	0.0	13.0	0.8	0.0	0.0	12.7	1.7	0.2	0.0	12.9	1.5	0.0	0.0	*	*	*	*	*	*	*	*
30,000名	18.5	3.5	0.5	0.0	12.2	2.0	0.4	0.0	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
平均値	13.4	2.3	0.3	0.1	12.1	1.6	0.2	0.0	12.2	1.5	0.2	0.0	12.0	1.6	0.2	0.0	12.2	1.4	0.1	0.0	11.9	1.6	0.2	0.0

B) -4.00~4.00範囲項目群母数設定値データ群

項目群 Outfitカイニ乗平均平方統計量10標本はずれ値百分率平均値

標本	20項目				50項目				100項目				150項目				200項目				250項目			
	%>1.0	%>1.1	%>1.2	%>1.3	%>1.0	%>1.1	%>1.2	%>1.3	%>1.0	%>1.1	%>1.2	%>1.3	%>1.0	%>1.1	%>1.2	%>1.3	%>1.0	%>1.1	%>1.2	%>1.3	%>1.0	%>1.1	%>1.2	%>1.3
150名	36.5	20.5	13.5	9.5	32.2	19.4	12.6	8.2	32.7	17.9	11.6	7.8	31.3	18.3	11.1	8.1	31.7	17.1	10.5	7.1	31.0	17.6	10.8	7.6
300名	32.5	20.0	13.0	8.0	27.8	15.6	10.2	6.2	31.7	15.7	9.1	5.0	28.7	13.9	7.4	4.3	30.3	15.5	8.8	5.0	30.4	14.9	8.2	5.1
500名	33.0	17.5	10.0	6.5	29.8	14.4	9.4	5.8	28.5	11.4	6.3	3.4	26.9	11.1	5.6	3.7	27.1	10.3	5.1	3.1	28.2	11.4	5.1	2.7
1,000名	25.0	16.0	10.5	7.0	26.4	10.4	5.4	2.2	26.2	9.1	4.0	2.3	23.5	8.2	3.0	1.4	24.5	7.5	2.9	1.3	24.1	8.5	3.4	1.4
2,500名	33.0	17.5	8.0	5.0	21.6	5.4	2.6	1.4	17.5	4.3	1.2	0.3	18.3	3.5	1.1	0.3	17.1	4.2	1.3	0.4	16.4	2.7	0.7	0.2
5,000名	32.5	8.0	4.0	2.0	17.0	2.6	0.6	0.0	13.0	2.0	0.1	0.0	13.5	2.3	0.3	0.0	11.4	1.2	0.2	0.1	12.5	1.8	0.2	0.0
10,000名	28.5	9.0	5.0	1.5	14.6	0.8	0.0	0.0	10.2	0.7	0.0	0.0	8.1	0.7	0.0	0.0	*	*	*	*	*	*	*	*
30,000名	29.0	6.0	0.5	0.0	7.6	0.2	0.0	0.0	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
平均値	31.3	14.3	8.1	4.9	22.1	8.6	5.1	3.0	22.8	8.7	4.6	2.7	21.5	8.3	4.1	2.5	23.7	9.3	4.8	2.8	23.8	9.5	4.7	2.8

項目群 Outfit Z値10標本はずれ値百分率平均値

標本	20項目				50項目				100項目				150項目				200項目				250項目			
	%>0	%>1	%>2	%>3	%>0	%>1	%>2	%>3	%>0	%>1	%>2	%>3	%>0	%>1	%>2	%>3	%>0	%>1	%>2	%>3	%>0	%>1	%>2	%>3
150名	8.0	1.5	0.5	0.5	9.2	1.6	0.4	0.0	8.4	1.1	0.1	0.0	9.0	1.3	0.2	0.0	8.5	1.3	0.2	0.0	9.4	1.4	0.3	0.0
300名	10.0	1.5	0.0	0.0	8.0	0.6	0.0	0.0	10.4	2.3	0.2	0.1	9.6	1.5	0.2	0.0	10.3	1.5	0.3	0.1	9.4	1.5	0.0	0.0
500名	7.0	1.0	0.0	0.0	11.8	2.2	0.4	0.0	10.4	1.4	0.2	0.0	10.1	1.3	0.2	0.1	10.0	1.1	0.1	0.1	10.4	1.0	0.1	0.0
1,000名	10.5	3.0	0.5	0.0	12.0	1.8	0.4	0.2	12.2	1.1	0.0	0.0	10.7	1.3	0.1	0.0	11.2	1.4	0.0	0.0	11.3	1.4	0.1	0.0
2,500名	21.5	4.5	1.0	0.0	12.8	2.2	0.0	0.0	11.4	1.6	0.1	0.0	10.4	0.8	0.1	0.0	11.8	1.1	0.1	0.0	10.4	0.8	0.0	0.0
5,000名	24.5	2.5	0.5	0.0	13.0	1.4	0.2	0.0	11.8	1.0	0.1	0.0	11.5	1.4	0.1	0.0	10.9	0.9	0.0	0.0	11.6	1.3	0.1	0.0
10,000名	23.0	7.5	2.0	0.5	14.4	1.0	0.0	0.0	12.1	1.0	0.1	0.0	11.3	1.1	0.1	0.0	*	*	*	*	*	*	*	*
30,000名	38.0	12.5	5.0	1.0	18.4	2.2	0.2	0.0	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
平均値	17.8	4.3	1.2	0.3	12.5	1.6	0.2	0.0	11.0	1.4	0.1	0.0	10.4	1.2	0.1	0.0	10.5	1.2	0.1	0.0	10.4	1.2	0.1	0.0

本節の締めとして、表 4-6 と 4-7 に項目群と受験者群それぞれについての Outfit 統計量上でのはずれ値出現百分率の 10 標本平均値を与える。なお、各表に対応する Infit 統計量に関するその平均値の上で、項目群と受験者群それぞれについての対象標本の大きさに連関する漸減傾向が各表に準じており、そのはずれ値百分率が全般的に小値となっているとの違いのみが観察される。従って、その提示は余計で煩雑となることから省いている。

表 4-6 において、Smith, Schumacker, and Bush²⁰⁾ (1998)により報告された項目群に関する Outfit カイ二乗平均平方統計量、及び、Outfit Z 値の受験者標本の大きさとの関係が視認される。それは、受験者標本の大きさによる影響の点において、Outfit Z 値の Outfit カイ二乗平均平方統計量にはるかに優る安定性(Smith, Schumacker, and Bush²⁰⁾, 1998)である。但し、表 4-6 での Outfit Z 値のその安定性に関して、Smith, Schumacker, and Bush²⁰⁾ (1998)によっては観察されていない事象として、A) -2.00~2.00 範囲項目群母数設定値データ群と B) -4.00~4.00 範囲項目群母数設定値データ群の間においても、20 項目標本以外の同一項目数対同一受験者数標本上での Outfit Z 値に関するはずれ値百分率同等性、つまりは、その安定性が顕著に観察される。従って、Outfit Z 値の Outfit カイ二乗平均平方統計量に優る受験者標本の大きさに影響を受けない安定性(Smith, Schumacker, and Bush²⁰⁾, 1998)のみならず、Rasch 測定妥当性に影響する項目群困難度推定値分布異同への Outfit Z 値無変動性の傾向が窺われる。一方、Outfit カイ二乗平均平方統計量に関しては、A と比べて B でのそのはずれ値百分率が全般的にかなり大きなものであり、同一項目数対同一受験者数標本の上で A と B の間に存在する困難度推定値分布標準偏差の相違に反応して、項目群についての A の B に優る Rasch モデル適合性を示している。

表 4-6 での項目群 Outfit Z 値はずれ値百分率のその安定性(Smith, Schumacker, and Bush²⁰⁾, 1998)、ないしは、その無変動性が、表 4-7 での受験者群 Outfit Z 値についても同様に観察される。一方、Outfit カイ二乗平均平方統計量については、各受験者数標本における 20 項目標本以外での各項目数標本の上で、上記の項目群においてと同じく、受験者群についての A の B にはるかに優る高い Rasch モデル適合性が顕示されている。

識者一般の認知として、Outfit・Infit 統計量はモデル適合度検知において不完全な指標である(平越・井澤⁸⁾、2008b, pp. 72-74、参照)。然しながら、Outfit・Infit カイ二乗平均平方統計量の上で、A が B に比べて項目群と受験者群いずれに関しても Rasch モデル適合性に優っていることに注目される。それは、多分に、項目特性曲線と受験者特性曲線の同形性異同に鋭敏な指標であるカイ二乗平均平方統計量の特質(井澤¹⁵⁾、2006、参照)によるものと思われる。項目群困難度推定値分布と受験者群能力推定値分布の各標準偏差が近接するにつれて、項目特性曲線と受験者特性曲線の均衡的な同形性がデータ全体として満たされる度合いが高くなると理解されるからである。それは、又、両群母数推定値分布標準偏差同等性がデータの Rasch 測定妥当性を高める条件であることを示唆している。

表4-7 A) -2.00~2.00 範囲項目群母数設定値データ群、及び、B) -4.00~4.00 範囲項目群母数設定値データ群に関する10 標本受験者群 Outfit 統計量はずれ値百分率平均値

A) -2.00~2.00範囲項目群母数設定値データ群								B) -4.00~4.00範囲項目群母数設定値データ群									
受験者群Outfitカイニ乗平均平方統計量10標本はずれ値百分率平均値								受験者群 Outfit Z値10標本はずれ値百分率平均値									
標本	20項目	50項目	100項目	150項目	200項目	250項目	平均値	標本	20項目	50項目	100項目	150項目	200項目	250項目	平均値		
150名	% > 1.3	18.5	7.8	4.1	1.7	1.6	0.8	5.8	150名	% > 3	0.0	0.0	0.0	0.0	0.0	0.0	
	% > 1.2	23.1	14.4	7.8	5.4	4.3	2.5	9.6		% > 2	0.0	0.1	0.2	0.0	0.3	0.0	0.1
	% > 1.1	29.3	22.9	17.1	14.1	9.6	7.9	16.8		% > 1	1.2	1.5	2.1	2.3	2.1	1.1	1.7
	% > 1.0	36.4	35.4	34.0	33.8	31.7	30.0	33.6		% > 0	13.3	11.9	12.2	13.0	13.5	12.8	12.8
300名	% > 1.3	17.6	9.4	3.4	1.3	1.3	0.7	5.6	300名	% > 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	% > 1.2	22.6	15.0	7.4	3.8	3.2	2.6	9.1		% > 2	0.1	0.1	0.1	0.1	0.1	0.1	0.1
	% > 1.1	29.1	25.8	17.7	13.0	10.5	8.3	17.4		% > 1	1.6	1.8	1.4	1.4	1.5	1.6	1.6
	% > 1.0	36.1	38.6	35.4	32.3	31.2	31.0	34.1		% > 0	12.0	13.0	12.9	13.0	13.0	13.6	12.9
500名	% > 1.3	18.1	8.6	3.8	1.5	0.9	0.6	5.6	500名	% > 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	% > 1.2	23.1	14.3	7.9	4.4	3.1	2.0	9.1		% > 2	0.0	0.1	0.1	0.1	0.1	0.1	0.1
	% > 1.1	29.3	23.7	17.0	13.0	10.6	8.1	17.0		% > 1	1.4	1.7	1.7	1.2	1.6	1.7	1.6
	% > 1.0	36.4	36.8	34.9	33.8	30.8	29.6	33.7		% > 0	12.2	12.5	12.8	13.1	13.8	13.2	12.9
1,000名	% > 1.3	18.6	9.4	3.8	1.9	1.0	0.7	5.9	1,000名	% > 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	% > 1.2	23.3	15.2	7.8	4.8	2.9	2.1	9.4		% > 2	0.0	0.1	0.1	0.0	0.1	0.0	0.1
	% > 1.1	29.2	24.4	17.6	13.4	10.5	8.3	17.2		% > 1	1.5	1.5	1.3	1.5	1.5	1.4	1.5
	% > 1.0	36.9	37.1	35.2	33.6	31.6	29.4	34.0		% > 0	12.7	12.5	12.9	13.1	13.2	12.9	12.9
2,500名	% > 1.3	18.5	8.9	3.7	1.9	1.0	0.7	5.8	2,500名	% > 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	% > 1.2	23.4	14.9	8.1	4.8	2.9	2.2	9.4		% > 2	0.0	0.1	0.1	0.1	0.1	0.1	0.1
	% > 1.1	29.7	24.2	17.7	13.5	10.1	8.3	17.3		% > 1	1.6	1.5	1.6	1.6	1.5	1.6	1.6
	% > 1.0	36.7	37.7	35.2	33.2	31.6	30.1	34.1		% > 0	12.5	12.8	13.1	13.4	12.9	13.2	13.0
5,000名	% > 1.3	18.2	9.2	3.7	1.8	1.0	0.7	5.8	5,000名	% > 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	% > 1.2	23.2	15.0	8.1	4.8	3.0	2.2	9.4		% > 2	0.0	0.1	0.1	0.1	0.1	0.1	0.1
	% > 1.1	29.3	24.6	17.9	13.6	10.3	8.3	17.3		% > 1	1.5	1.4	1.5	1.5	1.5	1.6	1.5
	% > 1.0	36.6	37.6	35.8	33.4	31.2	29.6	34.0		% > 0	12.7	12.8	12.9	13.3	13.2	13.3	13.0
10,000名	% > 1.3	18.1	8.9	3.6	1.8	*	*	8.1	10,000名	% > 3	0.0	0.0	0.0	0.0	*	*	0.0
	% > 1.2	23.0	15.0	8.0	4.6	*	*	12.7		% > 2	0.0	0.0	0.1	0.1	*	*	0.1
	% > 1.1	29.0	24.5	17.6	13.4	*	*	21.1		% > 1	1.6	1.5	1.5	1.5	*	*	1.5
	% > 1.0	36.5	37.5	35.6	33.4	*	*	35.8		% > 0	12.6	12.9	12.9	13.2	*	*	12.9
30,000名	% > 1.3	18.1	9.0	*	*	*	*	13.6	30,000名	% > 3	0.0	0.0	*	*	*	*	0.0
	% > 1.2	23.1	15.1	*	*	*	*	19.1		% > 2	0.0	0.1	*	*	*	*	0.1
	% > 1.1	29.1	24.5	*	*	*	*	26.8		% > 1	1.5	1.5	*	*	*	*	1.5
	% > 1.0	36.6	37.7	*	*	*	*	37.2		% > 0	12.4	12.9	*	*	*	*	12.7
150名	% > 1.3	18.9	20.0	17.3	14.1	10.4	10.1	15.1	150名	% > 3	0.0	0.1	0.1	0.0	0.0	0.1	0.1
	% > 1.2	20.8	22.7	21.4	18.9	16.4	15.3	19.3		% > 2	0.1	0.3	0.3	0.3	0.2	0.3	0.3
	% > 1.1	22.5	25.6	26.6	26.7	24.3	23.2	24.8		% > 1	1.3	1.5	2.4	2.3	1.8	2.5	2.0
	% > 1.0	25.5	30.5	34.1	35.5	35.6	34.2	32.6		% > 0	5.8	7.9	11.3	11.6	10.8	11.8	9.9
300名	% > 1.3	19.6	19.5	16.1	13.5	12.5	10.6	15.3	300名	% > 3	0.0	0.0	0.0	0.1	0.1	0.0	0.0
	% > 1.2	21.7	23.0	21.0	18.9	18.2	15.9	19.8		% > 2	0.2	0.2	0.2	0.2	0.4	0.5	0.3
	% > 1.1	24.2	27.1	26.7	25.2	26.5	24.1	25.6		% > 1	1.3	1.8	2.4	2.1	2.6	2.7	2.2
	% > 1.0	26.6	32.1	34.6	33.5	35.9	34.7	32.9		% > 0	5.9	9.8	10.7	11.7	13.0	13.0	10.7
500名	% > 1.3	18.7	20.0	16.8	14.5	12.1	10.0	15.4	500名	% > 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	% > 1.2	21.1	23.7	21.2	19.5	17.8	15.8	19.9		% > 2	0.1	0.2	0.3	0.2	0.4	0.1	0.2
	% > 1.1	23.3	27.7	26.9	26.7	24.9	24.6	25.7		% > 1	1.4	2.2	2.3	2.4	2.6	2.3	2.2
	% > 1.0	26.0	33.7	34.0	35.7	35.9	36.0	33.6		% > 0	5.3	10.4	12.0	12.6	12.9	13.4	11.1
1,000名	% > 1.3	18.9	19.5	16.9	14.1	12.2	10.2	15.3	1,000名	% > 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	% > 1.2	21.2	22.7	21.5	19.5	17.5	16.3	19.8		% > 2	0.2	0.2	0.2	0.3	0.2	0.2	0.2
	% > 1.1	23.6	27.1	27.6	26.7	25.1	24.7	25.8		% > 1	1.5	2.2	2.3	2.3	2.5	2.2	2.2
	% > 1.0	26.7	32.4	35.2	35.7	35.8	36.4	33.7		% > 0	6.0	10.2	12.0	12.5	13.0	13.5	11.2
2,500名	% > 1.3	19.6	19.6	16.6	14.7	12.5	10.0	15.5	2,500名	% > 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	% > 1.2	21.6	23.0	21.6	20.1	17.8	15.9	20.0		% > 2	0.2	0.1	0.2	0.2	0.2	0.2	0.2
	% > 1.1	24.3	27.3	27.5	27.2	25.5	24.2	26.0		% > 1	1.5	2.2	2.3	2.5	2.6	2.4	2.3
	% > 1.0	27.1	32.6	35.0	36.0	36.3	35.9	33.8		% > 0	6.5	10.2	11.8	13.1	13.6	13.4	11.4
5,000名	% > 1.3	19.6	19.2	16.7	14.3	12.0	10.4	15.4	5,000名	% > 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	% > 1.2	22.0	22.8	21.4	19.7	17.7	16.1	20.0		% > 2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	% > 1.1	24.6	27.0	27.1	26.7	25.7	24.7	26.0		% > 1	1.4	2.1	2.5	2.3	2.3	2.4	2.2
	% > 1.0	27.4	32.5	34.5	35.9	36.2	36.6	33.9		% > 0	6.5	9.9	12.0	13.0	13.3	13.7	11.4
10,000名	% > 1.3	19.4	19.4	17.0	14.1	*	*	17.5	10,000名	% > 3	0.0	0.0	0.0	0.0	*	*	0.0
	% > 1.2	21.9	23.0	21.6	19.4	*	*	21.5		% > 2	0.3	0.2	0.2	0.2	*	*	0.2
	% > 1.1	24.5	27.4	27.5	26.6	*	*	26.5		% > 1	1.6	2.2	2.4	2.4	*	*	2.2
	% > 1.0	27.2	32.6	35.1	35.8	*	*	32.7		% > 0	6.8	10.1	12.2	12.8	*	*	10.5
30,000名	% > 1.3	19.2	19.5	*	*	*	*	19.4	30,000名	% > 3	0.0	0.0	*	*	*	*	0.0
	% > 1.2	21.7	23.0	*	*	*	*	22.4		% > 2	0.2	0.2	*	*	*	*	0.2
	% > 1.1	24.5	27.4	*	*	*	*	26.0		% > 1	1.4	2.2	*	*	*	*	1.8
	% > 1.0	27.3	32.6	*	*	*	*	30.0		% > 0	6.5	10.1	*	*	*	*	8.3

なお、上記用語としての「モデル適合性」について留意すべき付言がここで必要である。A と B での各模擬データ作成法により、各項目数と各受験者数から成る 1 標本上での各データは、その項目数と受験者数の枠内において誤差変動を含めた上でのデータ次元充足度がほぼ最高となっている。従って、表 4-6 と 4-7 における A と B での項目群と受験者群に関する Outfit 統計量はずれ値出現百分率についての比較は、各項目数と各受験者数の制約の下で「モデル適合性」がほぼ最高度に充足されているデータ群についてのものである。Outfit・Infit 統計量はモデル適合度指標ではあるが、本稿での論理の上では、上記の「モデル適合性」は、データにおける項目数と受験者数、並びに、両群母数推定値分布標準偏差の異同程度に制約される Rasch 測定妥当性と同義である。

V. おわりに

本稿での考察により、平越・井澤¹⁾(2008)所載表 4-1(p. 32)の改訂が部分的に必要とされ、「Rasch 測定妥当性」についての但し書きが、概念としてのデータ「Rasch モデル適合性」への本質的な理解に貢献する。それを表 5-1 に与える。「Rasch 測定妥当性」のここでの定義は、受験者群と項目群の両群に観察される母数推定値分布標準偏差の同等性を条件とする両群 Rasch 測定母数推定値の受験者群能力分布認知の上での有意義・有用性である。その認知は、当然に、その受験者群能力の母集団への認知像を含む。01 データの項目数と受験者数の制限下における Rasch 測定妥当性程度は、両群母数推定値分布標準偏差の異同程度に準ずるとの理解であり、受験者群と項目群それぞれ、及び、各受験者と各項目へのモデル適合度個別参照に優先する最重要確認事項である。

表 5-1 平越・井澤¹⁾(2008)所載表 4-1(p. 32)の改訂

項目群	困難度推定値 標準偏差		受験者群	能力推定値 標準偏差	
	小	大		小	大
項目群第一主成分寄与率	大	小	受験者群第一主成分寄与率	大	小
受験者群第一主成分寄与率	小	大	項目群第一主成分寄与率	小	大
困難度推定値不変性成立度	小	大	能力推定値不変性成立度	小	大
困難度推定値分離指標値	小	大	能力推定値分離指標値	小	大

項目群と受験者群の母数推定値分布標準偏差同等性がデータの Rasch 測定妥当性を高める条件である。

注:

- 1) 項目数と受験者数のそれぞれが増大するにつれて、データの Rasch 測定妥当性程度は高くなる。
- 2) データ Rasch 測定妥当性達成度は、受験者群能力推定値分布標準偏差の大きさによって制約される。
- 3) 受験者群能力推定値分布を既知と仮定すれば、その標準偏差と同等の困難度推定値分布標準偏差を持つ項目群のテスト設定がデータの Rasch 測定妥当性を高める。

前節末尾に述べた様に、本稿で提示されたすべての模擬データはその項目数と受験者数の枠内における 01 生起確率性の上で次元 Rasch モデルに適合している。それらのモデル適合データに

において Rasch 測定妥当性程度高低に影響を与える表 5-1 での要因は、現実 01 データ上での Rasch モデル一次元適合性に関するものとなる。現実データ Rasch 測定上での一般的な能力推定値分布標準偏差を考慮すれば、困難度推定値分布標準偏差との乖離状況、及び、その項目数と受験者数の制約も踏まえて、表 5-1 でのモデル適合度指標値がすべて大きな値になることはないと思われる。従って、前節冒頭で言及した現実データの「一次元尺度構成体系」達成度への参照が必須である。

【参考文献】

- 1) 平越裕之・井澤廣行 2000a. 「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.
- 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) 静 哲人 2007. 『基礎から深く理解するラッシュモデリング』 大阪： 関西大学出版部
- 4) D. Andrich. 1982. An Index of Person Separation in Latent Trait Theory, the Traditional KR. 20 Index, and the Guttman Scale Response Pattern. *Educational Research and Perspectives*, 9, pp.95-104.
- 5) C. M. Fox & J. A. Jones. 1998. Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45, 1, pp. 30-45.
- 6) J. M. Linacre. 1989-2001. *A user's guide to Facets*. Chicago: Winsteps.com.
- 7) B. D. Wright & G. N. Masters. 1982. *Rating scale analysis*. Chicago: MESA Press.
- 8) 平越裕之・井澤廣行 2008b. 「Rasch 項目分析モデル測定母数推定値分布幅、テスト次元性、並びに、Rasch 測定適用妥当性」『流通科学大学論集 — 流通・経営編』第 20 巻、第 2 号、pp. 61-85.
- 9) R. M. Smith. 1996. A comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modeling*, 3, 1, pp. 25-40.
- 10) 豊田秀樹著 2002. 『項目反応理論[入門編]』 東京： 朝倉書店
- 11) B. D. Wright. 1967. Sample-free test calibration and person measurement. *Proceedings of the 1967 ETS Invitational Conference on Testing Problems at Princeton, NJ (MESA research memorandum No. 1, MESA Psychometric Laboratory)*.
- 12) T. F. McNamara. 1991. Test dimensionality: IRT analysis of an ESP listening test. *Language Testing*, 8, 2, pp. 139-159.
- 13) B. D. Wright & M. H. Stone. 1979. *Best test design*. Chicago: MESA Press.
- 14) 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.
- 15) 井澤廣行 2006. 「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.
- 16) T. G. Bond & C. M. Fox. 2001. *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

- 17) 芝祐順・渡辺洋・石塚智一編 1984. 『統計用語辞典』 新曜社： 東京
- 18) 村上宣寛 2007. 『IQ ってホントはなんだ? 知能をめぐる神話と真実』 日経 BP 社： 東京
- 19) J. M. Linacre. 1989–2001. *Rasch measurement computer program*. Chicago: Winsteps.com
- 20) R. M. Smith, R. E. Schumacker, & M. J. Bush. 1998. Using item mean squares to evaluate fit to the Rasch model. *Journal of Outcome Measurement*, 2, 1, pp. 66–78.
- 21) T. F. McNamara. 1996. *Measuring second language performance*. Longman: London.
- 22) M. Wilson. 2005. *Constructing measures: An item response modeling approach*. Mahwah, N.J.: Lawrence Erlbaum Associates, Inc.