流通科学大学論集—人間・社会・自然編－第 20 巻第 2 号，39-62(2008)

Rasch の理論と理念、及び、Wright の Rasch 測定展開

Rasch's Theory and Philosophy,
and Wright's Contribution to the Model Application

井澤　廣行*
Hiroyuki Izawa

This paper describes Georg Rasch's theory and philosophy, and Benjamin D. Wright's contribution to the expansion of the model application, referring primarily to their original writing in that the Japanese translation is not readily available. Rasch's theory is characterized unitarily by parameter separability and Rasch's philosophy psychometrically by 'specific objectivity.' The fundamental principles have been attempted to be specified as accurately as possible within the author's capacity. His intention of this investigation is based on an idea that the righteous application of a scientific model presupposes a reasonably firm foundation of its theoretical understanding.

Key words: Georg Rasch, Benjamin Wright, parameter separability, specific objectivity.

1.　はじめに

Rasch モデルの原典は、1958 年にデンマーク教育研究機構(The Danish Institute of Educational Research)が 1945 年以降の Georg Rasch によるテスト分析成果を'On the use of a general measurement principle to bridge-building between similar psychological tests'としてデンマーク語で謄写版印刷にまとめたものである。それはその機構で 1957 年に開催された Rasch による一連の講演口述内容を彼の娘である Lotte が筆記したものに基づく製版であった(Andrich 録音 [1], in Wright, 1980[2], p. xiv; 原典標題英語訳は、Andersen and Olsen[3], 2001, p. 24、参照)。その著作に Rasch が構築したテスト分析理論の母数分離定理証明を中心とする数理統計基盤を新たに加えた上で(Andersen and Olsen[3], 2001, p. 18)、デンマーク教育研究機構が Rasch の業績を G. Leunbach の英語訳により(Rasch[4], 1960, p. xxii) 1960 年に出版した図書が ***Probabilistic Models for Some Intelligence and Attainment Tests*** である。その再版が、米国での Rasch モデル研究及びその適用における第一人者である Benjamin D. Wright による前書きと後書きを加えて、シカゴ大学より 1980 年に刊行された。

*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

(2007 年 8 月 31 日受理)

その日本語訳が、内田良雄監訳により『心理テストの確率モデル』と題されて、1985 年に名古屋大学出版会から発刊された(芝編 [5]、1991, p. 224、参照)とのことであるが、2001 年の時点で既に廃刊となっていた。筆者は無念にもそれを所有していない。そのこともあり、Rasch モデルの根本的理解を目的として、***Probabilistic Models for Some Intelligence and Attainment Tests*** を中心に参照しつつ、Georg Rasch の理論と理念について考察する。更に、シカゴ大学での Benjamin D. Wright を中心とする心理測定研究者達への Rasch 項目分析モデル伝播上での智見をモデル理解補強のために概観する。

2. Georg Rasch の理論と理念

1925 年にコペンハーゲン大学で数学の修士号取得後も、Rasch は指導教授 N. E. Nørlund の下で働いていた。その教授の依頼で Jensen's Inequality in Probability Theory で以って有名とされるデンマークの数学者 J. L. W. V. Jensen の論文精査をおこなった(Andersen and Olsen[3], 2001, p. 5)とあり、Rasch が確率論に通じていたことは間違いないと想像される。更に、Rasch は 1930 年にコペンハーゲン大学より純粋数学分野において博士号を取得しており(Andersen and Olsen[3], 2001, p. 5)、1934 年(Andrich 録音 [1], in Wright[2], 1980, p. xi)には'On the theory and application of product integrals'との英語標題が付された(Wright[6], 1980, p. 197)ドイツ語論文を発表している。従って、Rasch が 1960 年の著作で頻繁に使用する'multiplicative law'の概念応用に洞察していたことは自然の流れである。その叡智所産として、Rasch[4] (1960)は、Maxwell (1876, cited in Rasch[4], 1960, p. 110)の *Matter and Motion* を参照して、ニュートンの運動第二法則 Force = Mass × Acceleration に基づいて次の帰結に到る尺度基準化概念を展開している(pp. 110−114)。$F_0 = 1$ の基準値を設定すれば、$M_v = 1 / A_{v0}$ が導出されて、

> 『ある物体の質量は基準値 $F_0 = 1$ を持つ力を受けることにより生ずる加速度の逆数に等しい』(Rasch[4], 1960, p. 114)。

ここで、物体の質量、力、加速度をそれぞれ受験者能力母数、音読テスト困難度母数、受験者読み誤りの確率に置き換えれば、以下に参照される Rasch の乗法ポアソン確率モデルにおけるそれら三要素の因子分解関係に相似する。従って、受験者読み誤り単語数の等倍数の仮定、尺度の基準化概念、並びに、確率論応用が Rasch による理論・理念確立の礎である。

1935 年春から 1936 年夏に渡る一年間余り、ロックフェラー財団奨学金授与により Rasch はロンドンにおいて R. A. Fisher の下で研究をする機会が与えられた(Andersen and Olsen[3], 2001, p. 6)。Fisher が確立した最尤法理論と十分統計量の概念は Rasch を魅了するものであった(Andrich 録音 [1], in Wright[2], 1980, pp. xi−xii)。Rasch が十分統計量について考究した(Andersen and Olsen[3], 2001, p. 6)ことから、その存在が母数分離を可能にせしめるとの彼の確信に到ったか、あるいは、最尤推定法との関係の上で確率モデルにおけるその重要性を Rasch が意識していたと想像される。

この滞英期間中に、Julian Huxley との交友によりカニの甲羅の形状がカニ個別の成長度合いを示すものと知らされて、Rasch は団体よりもむしろ個体に関する統計量の意義と重要性を実感した(Andrich 録音 [1], in Wright[2], 1980, p. xii)。この認識が Rasch 理論構築の結実として 1960 年著作の序文(Rasch[4], 1960)において次の様に表明されている。

> 『個体中心の統計技法は個体が団体から分離して特徴付けられる理論を必要とする。その理論は、妥当なデータが与えられたならば、個体に関する母数が推定され得るものであらねばならない。更に、個体間の比較が個別の測定器具から独立してなされ得ることが理想である』(p. xx)。

Rasch がテスト分析に初めて従事したのは 1945 年から 1948 年に渡ってであった。その目的は自衛官選抜のために防衛省で用いられる知能テストの標準化であり(Rasch[4], 1960, p. xxi)、そのテストはコペンハーゲン大学心理学部教授 Rubin と Rasmussen が開発したものであった(Andersen and Olsen[3], 2001, p. 8)。その分析を通して、項目困難度と受験者能力の分離の問題を Rasch は意識することとなった(Rasch[4], 1960, p. xxi)。但し、1947 年に初期分析結果が心理学部に報告されたが、Andersen and Olsen[3] (2001, pp. 8−9)によれば、その内容はプロビット変換手法を含めての実証記述に止まるものであり、Rasch 理論の萌芽は見られなかった。

1952 年に Rasch は音読テストの分析に携わることとなった。これは、読解に障害のある小学校児童群に与えられた特別授業の効果を追検証するために社会福祉省が Rasch に助力を求めたことによるものであった(Rasch[4], 1960, p. xxi)。1940 年代に読解特別授業を受けた小学生読解障害児童群への音読テスト結果と彼等が二十歳前後となっていた 1951 年に追試された音読テストの結果を比較するものであった。然しながら、彼等が小学生時代に受けた音読テストは同一のものではなく、それら数種のテストの難度標準化がなされておらず、データ収集の質も不完全なものであったという問題が存在した(Andersen and Olsen[3], 2001, p. 10)。これらの事由により、読解に障害のあった生徒群に与えられたテスト類中の五つのテストが、小学校 2 年生、3 年生、4 年生、5 年生、6 年生、7 年生の一般児童群に対して、3 年生以上の児童群は二つか三つのテストを受ける要領で、1952 年 1 月に実施されて新たに大きなデータ類が収集された(Andersen and Olsen[3], 2001, p. 5; Rasch[4], 1960, p. 5)。

その 1952 年収集データ類について、Rasch は、先ず、二つのテストを受けた小学生児童群の読み誤り単語数をそれぞれ縦横軸に配して共通座標上に布置した(Rasch[7], 1977, p.59)。Rasch が観察したことは、二つの音読テストを受けた小学生児童群の読み誤り単語数の全般大枠的な誤差変動範囲内での比例関係であった(Rasch[7], 1977, p. 60)。そのグラフ上での目視が、受験者読み誤り単語数の等倍数仮定の上で、読み誤り単語数が微小であることからその確率分布にポアソン分布適用を Rasch に想起させた。かくして、Rasch が目指したことは、ある一人の小学生がいずれの音読テストを受験しても同一の音読能力値を測定できるモデル構築であった(Andersen and Olsen[3], 2001, p. 10)。それは、受験者能力とテスト困難度をどの様に分離してそれぞれ独立に推

定するかを模索することであった。Rasch[4] (1960)による次の言葉にその意図が示されている。

『乗法ポアソンモデルにおけるこの[受験者得点の能力母数とテスト得点のテスト困難度母数への]因子分解はデータ分析上の簡潔さをもたらす数学的利便性として意識されたものであった。それは、又、テスト難易程度と受験者能力程度の関係をごく自然に解釈できるであろうとの心算によるものであった』(p. 109)。

受験者を v = 1, 2, ..., n、及び、テストの種類を i で表せば、上記の読み誤り単語数の比例関係を示すモデル式は次の様になる(Rasch[7], 1977, pp. 61−62)。

$$\lambda_{1\ \mathrm{test}\ i} / \lambda_{1\ \mathrm{test}\ 1} = \lambda_{2\ \mathrm{test}\ i} / \lambda_{2\ \mathrm{test}\ 1} = ... = \lambda_{n\ \mathrm{test}\ i} / \lambda_{n\ \mathrm{test}\ 1}$$

従って、$\lambda_{v\ \mathrm{test}\ i} / \lambda_{v\ \mathrm{test}\ 1} = \delta_{\mathrm{test}\ i} / \delta_{\mathrm{test}\ 1}$ と表すことが出来て、それは基準テスト 1 における音読困難度 $\delta_{\mathrm{test}\ 1}$ に対しての test i の音読困難度 $\delta_{\mathrm{test}\ i}$ の程度を示すことになる。更に、$\delta_{\mathrm{test}\ 1} = 1$ と置いて、$\lambda_{v\ \mathrm{test}\ i}$ について解けば、$\lambda_{v\ \mathrm{test}\ i} = \lambda_{v\ \mathrm{test}\ 1} * \delta_{\mathrm{test}\ i}$ となる。$\lambda_{v\ \mathrm{test}\ 1}$ は受験者 v が音読困難度が 1 とみなされた基準テスト 1 を受験して読み誤る単語数であるから、それは受験者 v の規準テスト 1 において単語を読み誤るという負の能力を意味しており、受験者 v の正しく音読する能力を ξ_v と置けば $\lambda_{v\ \mathrm{test}\ 1} = 1 / \xi_v$ と表現できることになる。結果として、$\lambda_{v\,i} = (1 / \xi_v) * \delta_i = \delta_i / \xi_v$ と与えられる。テスト i を受験する如何なる生徒 v の読み誤り単語数は、基準テストの音読困難度を 1 と想定した上でのその基準テストと比較されたテスト i の音読困難度を、音読困難度が 1 と見なされた基準テストを生徒 v が受験して読み誤らない程度を示すその生徒 v の正しく音読する能力で割ったものになる。この δ_i / ξ_v は Rasch[4] (1960)により $\theta_{v\,i} = \delta_i / \xi_v$ と表現されて(p. 16)、生徒 v が音読テスト i を受験した際の読み誤りの確率であるとも理解される。

音読テスト i に含まれた単語数の合計が N_i であり、生徒 v がそのテスト i を受験して a_{vi} 個の読み誤り単語数を記録すれば、a_{vi} / N_i は非常に小さな百分率であると考えられるから、読み誤り単語数の生起はポアソン分布に従うと想定され得る。上記の読み誤りの確率を $\theta_{vi} = \delta_i / \xi_v$ とすれば、$N_i\theta_{vi} = N_i\delta_i / \xi_v$ が読み誤り平均単語数となる。従って、正しく音読する能力 ξ_v を持つ生徒 v が単語数 N_i から成り音読困難度 δ_i を持つ音読テスト i を受験して、a_{vi} 個の読み誤り単語数を記録する確率は次式で表される。これが Rasch[4] (1960)による読み誤り単語数の乗法ポアソン確率モデル(pp. 15−16)である。

$$P\{a_{vi} \mid N_i\} = [\exp(-N_i\delta_i / \xi_v)] * (N_i\delta_i / \xi_v)^{a_{vi}} / a_{vi}!$$

ここで生徒 v が単語数 N_i から成る音読テスト i と単語数 N_j から成る音読テスト j を受験した状況の下でそれぞれのテストでの読み誤り単語数が a_{vi} と a_{vj} と想定される。その同時生起確率 $P\{a_{vi}, a_{vj}\}$ とその合計読み誤り単語数$(a_{vi} + a_{vj})$の生起確率 $P\{a_{vi} + a_{vj}\}$ がそれぞれポアソン分布に従うことを利用して、Rasch[4] (1960)は次式で示される条件付確率を導出している(p. 19)。

$$P\{a_{vi} \mid a_{vi} + a_{vj}\} = P\{a_{vi}, a_{vj}\} / P\{a_{vi} + a_{vj}\}$$

$$= [{}_{(a_{vi} + a_{vj})}C_{a_{vi}}] * [N_i\delta_i / (N_i\delta_i + N_j\delta_j)]^{a_{vi}} * [N_j\delta_j / (N_i\delta_i + N_j\delta_j)]^{a_{vj}}$$

その注釈は、生徒 v が二つの音読テスト i と j を受験して合計$(a_{vi} + a_{vj})$個の読み誤り単語数を記

録した上で音読テスト i において読み誤り単語数が a_{vi} 個になる条件付確率は、二項分布 B[a_{vi} + a_{vj}； $N_i\delta_i / (N_i\delta_i + N_j\delta_j)$]に従う。従って、それは、読み誤り単語数 a_{vi} の($a_{vi} + a_{vj}$)を条件とするその生起確率がその生徒 v の音読能力 ξ_v からは分離独立しているという啓示である。これがすべての Rasch 理論を特徴づける母数分離定理として読み誤り単語数の乗法ポアソン確率モデルに係わる本質特性である。

Rasch[4] (1960)により、更に、読み誤り単語総数が($a_{vi} + a_{vj}$) = a..である受験者一群についての二つの音読テスト i と j それぞれにおける読み誤り単語数合計 $a_{.i}$ と $a_{.j}$ の分布が考慮されて、a..を条件とする $a_{.i}$ の生起確率が次式で示される(p. 22)。

$$P\ \{a_{.i} \mid a_{..}\} = (a_{..}\ C\ a_{.i}) * [N_i\delta_i / (N_i\delta_i + N_j\delta_j)]^{a.i} * [N_j\delta_j / (N_i\delta_i + N_j\delta_j)]^{a.j}$$

従って、$N_i\delta_i / (N_i\delta_i + N_j\delta_j) \approx a_{.i} / a_{..}$ と推定され、同様に、$N_j\delta_j / (N_i\delta_i + N_j\delta_j) \approx a_{.j} / a_{..}$ と推定されて(Rasch[4], 1960, p. 22)、最終的にテスト音読困難度母数 δ_i と δ_j の推定値が得られる。更に、二項分布の分散としての $a_{..} * [N_i\delta_i / (N_i\delta_i + N_j\delta_j)] * [N_j\delta_j / (N_i\delta_i + N_j\delta_j)]$に基づいて、($a_{.i} / a_{..}$)の標準誤差が $\sqrt{\{(a_{.i}/a_{..})*[1-(a_{.i}/a_{..})]/a_{..}\}}$ として推定される(Rasch[4], 1960, p. 22)。又、($a_{vi} + a_{vj}$)個の読み誤り単語総数が与えられれば、a_{vi} は受験者音読能力母数 ξ_v に関する直接的な情報を有さず、($a_{vi} + a_{vj}$)が ξ_v を推定するための十分統計量となり(Rasch[4], 1960, p. 21)、ξ_v の最良推定値は$(N_i\delta_i + N_j\delta_j) / (a_{vi} + a_{vj})$により与えられる(Rasch[4], 1960, p. 18)。

Rasch[4] (1960)は音読単語数分析の上で母数分離定理を満たして異なる乗法ポアソン確率モデルにも言及している(pp. 34−49)。上述の読み誤り単語数の乗法ポアソン確率モデルと同様にその導入概念は比率の同一性である。受験者 v = 1, 2, ..., n、テストの種類を i で表せば、ある一定の制限時間内での受験者 v に関する読了単語数のモデル比例式は次の様になる。

$$\lambda_{1\ \mathrm{test}\ i} / \lambda_{1\ \mathrm{test}\ 1} = \lambda_{2\ \mathrm{test}\ i} / \lambda_{2\ \mathrm{test}\ 1} = \ldots = \lambda_{n\ \mathrm{test}\ i} / \lambda_{n\ \mathrm{test}\ 1}$$

従って、$\lambda_{v\ \mathrm{test}\ i} / \lambda_{v\ \mathrm{test}\ 1} = \varepsilon_{\mathrm{test}\ i} / \varepsilon_{\mathrm{test}\ 1}$ と表すことが出来て、それは基準テスト 1 における音読容易性 $\varepsilon_{\mathrm{test}\ 1}$ に対しての test i の音読容易性 $\varepsilon_{\mathrm{test}\ i}$ の程度を示すことになる。更に、基準値 $\varepsilon_{\mathrm{test}\ 1} = 1$ と置いて、$\lambda_{v\ \mathrm{test}\ i}$ について解けば、$\lambda_{v\ \mathrm{test}\ i} = \lambda_{v\ \mathrm{test}\ 1} * \varepsilon_{\mathrm{test}\ i}$ となる。$\lambda_{v\ \mathrm{test}\ 1}$ は受験者 v が音読容易性の程度が 1 と見なされた基準テスト 1 を受験して与えられた一定制限時間内に読了する単語数であるから、それは受験者 v の音読単語数能力を意味することになり、ξ_v と表せる。test i の音読容易性 ε_i の代わりに test i の音読困難性の程度 δ_i を考えれば $\varepsilon_i = 1 / \delta_i$ であるから、$\lambda_{v\,i} = \xi_v * \varepsilon_i = \xi_v / \delta_i$ と与えられる。従って、テスト i をある一定の制限時間内で受験する生徒 v の読了単語数は、生徒 v の音読単語数能力をテスト i の音読困難度で割ったものになる。テスト i の音読制限時間が T_i であるならば、その制限時間内で生徒 v が音読できる単語数 $\lambda_{v\,i}(T_i) = \xi_v * T_i / \delta_i$ となり、生徒 v がそのテスト i で記録する音読単語数を a_{vi} とすれば、Rasch[4] (1960)による次式の制限時間内音読単語数乗法ポアソン確率モデルが与えられる(p. 146)。

$$P\ \{a_{vi} \mid T_i\} = [\exp(-\xi_v * T_i / \delta_i)] * (\xi_v * T_i / \delta_i)^{avi} / a_{vi}!$$

このモデルについても、すべての受験者が二つの一定制限時間内音読単語数テスト i と j を受

けた結果として合計音読単語数(a_{vi} + a_{vj})が記録されれば、その条件の下での a_{vi} の生起確率が与えられる。その二項分布確率式にはξ_v が含まれず、母数分離定理の成り立つことが Rasch[4] (1960) により示されている(p. 147)。その母数推定値は読み誤り単語数の乗法ポアソン確率モデルの場合と同様な方法で得られることになる。Rasch[4] (1960, pp. 38−40)は、更に、受験者 v が単語数 N_i から成るテスト i を受験してその音読終了に受験者 v が費やした時間に関する乗法ポアソン確率モデルをも提案しているが、その説明はここでは割愛する。

二つの音読テスト OR3 と OR4 の使用上での総受験者による読み誤り単語数の乗法ポアソン確率モデルへの適合度は高いと分析され(Rasch[4], 1960, pp. 24−27)、残り三つの音読テストに関しても同様の結果であった(Rasch[4], 1960, p. 27) と記されている。一方、単語書き取りテスト、黙読テスト、及び、聞き取り綴りテストに関しては間違い単語数の散らばりが二項分布の 95%信頼区間内での線形性に欠けており、比率同一性欠如の傾向が見出されて(Rasch[4], 1960, pp. 53−56)、それら三種テストそれぞれの単語群困難度不均質性による乗法ポアソン確率モデルへの不適合であるとの分析がなされている(Rasch[4], 1960, pp. 57−59)。制限時間が与えられた引き算テストの受験者正答項目数の分布(Rasch[4], 1960, p. 52)においても同様のモデル不適合傾向が観察されて、計算と解答導出筆写という応答所要時間に係わる二つの異質な要素による制限時間内での項目群応答時間的困難度不均質性がもたらす正答項目数の乗法ポアソン確率モデルへの不適合との分析も付されている(Rasch[4], 1960, pp. 59−60)。一つのテストに含まれた単語群あるいは項目群の全体的困難度不均質性が乗法ポアソン確率モデル不適合の主たる要因であり、そのテストの困難度不均質性は単語群あるいは項目群における困難度のばらつきそのものに他ならない。又、一般的に受験者能力に大きな異同のある受験者一群に対するテストはその受験者群能力異同の程度に相応する異なった困難度を持つ項目群により構成されていることが望ましい。従って、項目群困難度不均質性は 0・1 テストの本質的要件である故に、Rasch[4] (1960)は個別項目への応答分析の重要性を意識することとなる(p. 61)。これが、後に Rasch model として知られる様になる(Andersen and Olsen[3], 2001, p. 13)項目分析モデル構築へのきっかけとなった。Rasch 自身の言葉によれば、

> 『項目分析モデルの発見は、実際には、音読テスト分析と乗法ポアソン確率モデルの研究に従事した 1952 年における仕事との関連の中での成就であった』(Andrich 録音 [1], in Andersen and Olsen[3], 2001, p. 14)。

1952 年に得られた五つの音読テストデータ上での読み誤り単語数がほぼ比例関係を示したことが乗法ポアソン確率モデルの構築につながり、音読テスト結果へのその適用分析は Rasch に単語群音読困難度均質性を条件としてそのモデルの妥当性を確信させた。テスト項目群全体としての困難度が均質である場合に限って誤答と正答から成る 0・1 データにおける受験者素点得点は間隔尺度に近いものとなるが、0・1 データにおいては項目群困難度均質性が保証され得ず、又、それはテスト要件に背く故に、受験者素点得点は本質的に能力尺度にはなり得ないと Rasch[4] (1960)は意識した(p. 64)。従って、正答率と誤答率の連関性を踏まえた上で(Rasch[4], 1960, p. 64)、

二値反応データにおいても、受験者群と項目群のいずれもが本質的テスト特性規定枠としてそれぞれ一次元的であれば、各受験者能力は各項目困難度に確率的に関与すると想定された(Rasch[4], 1960, p. 116)。Rasch は次の様に述べている。

『正答あるいは誤答の確率はある一つの項目要因に対してある一人の受験者要因がからんで生ずるものであると確信した』(Andrich 録音[1]), in Wright[2], 1980, p. xiii)。

Rasch[4] (1960)の意図は以下の如くである(pp. 74−75)。受験者 $v = 1, 2, \ldots, n$、並びに、項目 $i = 1, 2, \ldots, k$、更に、受験者能力と項目困難度の母数をそれぞれ ξ と δ とすれば、

$$0 < \zeta = \xi_1 / \delta_1 = \xi_2 / \delta_2 = \cdot\cdot\cdot = \xi_n / \delta_k < 1$$

そして、『非常に易しい項目と非常に難しい項目、又、能力が大変に高い受験者と大変に低い受験者が存在することから、正答確率関数 ζ は 0 から無限大の間にあるすべての項目困難度とすべての受験者能力に応える必要がある』(Rasch[4], 1960, p. 74)と意識された。この認識に基づいて、Rasch[4] (1960)はロジット関数を用いて、項目正答確率 P を次の様に定義する(pp. 74−75)。

$$P = \zeta / (1 + \zeta) = (\xi/\delta) / (1 + \xi/\delta) = \xi / (\xi + \delta)$$

従って、項目誤答確率は、$1 − P = 1 − [\xi/(\xi + \delta)] = \delta/(\xi + \delta)$であり、オッズは $P/(1 − P) = \xi/\delta$ となる。又、$\ln[P/(1 − P)] = \ln(\xi/\delta) = \ln\xi − \ln\delta$ である。

Rasch のこのロジット関数使用については、Andersen and Olsen[3] (2001)による以下の言説が興味を引く。

『1950 年代には心理学の分野においてはプロビット関数が統計モデルとして一般的に使用されていた。Rasch 自身は述べていないけれども、私たちの想像では、Rasch のロジット関数使用は母数推定に際して十分統計量の存在を彼が意識したからである』(p. 15)。

Rasch[4] (1960, pp. 174−178)の証明により、テストに含まれた受験者群と項目群のそれぞれについて、各同一素点得点一群が各同一素点得点一群の母数同一推定値算出への十分統計量であり、十分統計量それぞれの存在が受験者群と項目群の母数分離を保証するものとなる。

項目分析モデルに関する母数分離定理に関しては、ここでは、Rasch[4] (1960, pp. 171−172)により与えられた二項目間におけるその導出のみを参照しておく。なお、日本人研究者の服部[8] (1988)が二項目間のみならず全項目に関しての母数分離定理を概括して記している。数式が見やすくなることから、項目 i と j の困難度 δ_i と δ_j をそれぞれの容易度 ε_i と ε_j に変換すれば、$\varepsilon_i = 1/\delta_i$ 及び $\varepsilon_j = 1/\delta_j$ である。受験者 v の項目 i と j への正答確率と誤答確率は次の様になる。

$$P_{vi}\ \{a_{vi} = 1\} = \xi_v\varepsilon_i / (1 + \xi_v\varepsilon_i)$$

$$P_{vi}\ \{a_{vi} = 0\} = 1 − P_{vi}\ \{a_{vi} = 1\} = 1 / (1 + \xi_v\varepsilon_i);$$

$$P_{vj}\ \{a_{vj} = 1\} = \xi_v\varepsilon_j / (1 + \xi_v\varepsilon_j)$$

$$P_{vj}\ \{a_{vj} = 0\} = 1 − P_{vj}\ \{a_{vj} = 1\} = 1 / (1 + \xi_v\varepsilon_j)$$

受験者 v が、項目 i と j への応答独立性で以って、いずれかのみに正答する確率は、

$$P\ \{a_{vi} + a_{vj} = 1\} = [P_{vi}\ \{a_{vi} = 1\} * P_{vj}\ \{a_{vj} = 0\}]$$

$$+ [P_{vi}\{a_{vi} = 0\} * P_{vj}\{a_{vj} = 1\}]$$

$$= \xi_v(\varepsilon_i + \varepsilon_j) / [(1 + \xi_v\varepsilon_i)(1 + \xi_v\varepsilon_j)]$$

従って、$(a_{vi} + a_{vj} = 1)$を条件とする上での $a_{vi} = 1$ の生起確率は次式で示される。

$$P\{a_{vi} = 1 \mid a_{vi} + a_{vj} = 1\} = P\{a_{vi} = 1, a_{vj} = 0\} / P\{a_{vi} + a_{vj} = 1\}$$

$$= [\xi_v\varepsilon_i/(1 + \xi_v\varepsilon_i)]*[1/(1 + \xi_v\varepsilon_j)] / \{\xi_v(\varepsilon_i + \varepsilon_j)/[(1 + \xi_v\varepsilon_i)(1 + \xi_v\varepsilon_j)]\}$$

$$= \varepsilon_i / (\varepsilon_i + \varepsilon_j) = \delta_j / (\delta_i + \delta_j)$$

この条件付確率には ξ_v が含まれておらず、$P\{a_{vi} = 1 \mid a_{vi} + a_{vj} = 1\}$ は受験者能力母数から分離独立していることが分かる。

これにより、受験者群の中で項目 i あるいは j のいずれか一方に対してのみ正答している人数が n であり、その上で項目 i への正答者数が x であれば、条件付確率 $P\{x \mid n\}$ は次式で示される二項分布の確率となる(Rasch[4], 1960, p. 172; Rasch[9], 1966, p. 52)。

$$P\{x \mid n\} = {}_nC_x * [\delta_j / (\delta_i + \delta_j)]^x * \{1 - [\delta_j / (\delta_i + \delta_j)]\}^{n-x}$$

更に、$x / n \approx \delta_j / (\delta_i + \delta_j) = 1 / [(\delta_i / \delta_j) + 1]$である(Rasch[4], 1960, p. 172)ことにより、δ_i / δ_j が推定されることになる(Rasch[9], 1966, p. 52)。この母数推定法は、G. Leunbach により教示されたとの注記が Rasch[4] (1960, p. 172)に付されており、***Probabilistic Models for Some Intelligence and Attainment Tests*** の単なる英語訳者としての評価以上に G. Leunbach の名も Rasch モデルと共に留められるべきである。δ_i / δ_j の推定値の上でその項目 i の困難度推定値を任意に基準値 $d_i = 1$ と置けば、テストに含まれた各項目の困難度推定値が順次に得られる。又、受験者能力母数推定法についても同様である。これが PAIR method と呼ばれているものである(Wright[6], 1980, pp. 188−189)。なお、Rasch[9] (1966)により、受験者数に比べて項目数は一般的に小さな数であるから、この受験者能力母数推定法は推奨されないと指摘されている(p. 52)。

表１　項目グループ単位での１項目あたりの平均正答率（単位：%）

r	Item group					Number of persons with each raw score
	3−4	5−6	7−9	10−11	12−13	
4	83.90	10.70	3.80	1.79	0.45	112
5	83.50	39.60	11.80	6.70	4.30	82
6	85.50	70.40	21.40	7.20	5.30	76
7	86.00	70.10	41.50	23.80	8.50	82
8	90.20	79.40	61.50	27.90	9.30	102
9	94.50	84.00	72.50	45.40	15.50	119
10	95.50	88.00	82.90	59.40	28.20	133
11	97.97	93.90	90.00	76.40	39.00	123
12	99.47	94.70	93.30	86.70	66.00	94
13	99.18	97.54	97.87	90.20	79.50	61

[注：　Table 12　Subtest N of BPP (Rasch[4], 1960, p. 71)の一部抜粋]

項目分析モデル上で Rasch[4] (1960, pp. 76−105)が母数推定値を実際に算出した方法はいわゆる LOG method (Wright[6], 1980, p. 188)であり、その概略を以下に記す。上表１は、1953 年 9

月に 1,094 名の新自衛官に与えられた BPP−N と題された知能テスト(Rasch[4], 1960, p. 62)の結果を Rasch が表にまとめたものの一部抜粋である。項目番号は受験者全体で正答率の高い項目から低い項目の順に並んでおり、1 項目単位の正答率があまりにも小さいものがあるために、統合された項目群についての 1 項目当たり平均正答率が与えられている。 r の列は同一素点得点受験者群別である。例えば、r ＝ 4 の行を見れば、四つの項目に正答した受験者が 112 名であり、項目番号 3 と 4 への正答度数の合計が 188 であったからその度数平均は 94 であり、94 / 112 ≒ 0.839 が統合項目群 3−4 についての 1 項目当たりの平均正答率となる。

表 2　表 1 に与えられた正答率のロジット変換値、並びに、能力母数と困難度母数の推定値算出

r	Item group					Average of Log r	能力母数推定値
	3−4	5−6	7−9	10−11	12−13		
4	0.72	−0.92	−1.40	−1.74	−2.35	−1.14	0.073
5	0.71	−0.18	−0.87	−1.14	−1.35	−0.57	0.272
6	0.77	0.38	−0.56	−1.11	−1.26	−0.36	0.441
7	0.79	0.37	−0.15	−0.51	−1.03	−0.11	0.783
8	0.96	0.59	0.20	−0.41	−0.99	0.07	1.175
9	1.24	0.72	0.42	−0.08	−0.74	0.31	2.051
10	1.33	0.86	0.69	0.17	−0.41	0.53	3.373
11	1.68	1.19	0.95	0.51	−0.19	0.83	6.730
12	2.27	1.25	1.14	0.81	0.29	1.15	14.191
13	2.08	1.60	1.65	0.96	0.59	1.38	23.768
Average of Log i	1.26	0.59	0.21	−0.25	−0.74		
正負符号変換	−1.26	−0.59	−0.21	0.25	0.74		
困難度母数推定値	0.056	0.259	0.621	1.795	5.546		

[注：　Table 5　Subtest N of BPP (Rasch[4], 1960, p. 86)の一部抜粋に
能力母数推定値と困難度母数推定値の算出付加]

表 2 は、前掲表 1 に与えられた同一素点得点受験者群別の各統合項目群における 1 項目平均正答率のロジット変換値を提示しているものであり、その算出法は $\log_{10}$(正答率 / 誤答率)である。前記参照の如く、Rasch 項目分析上でのオッズが $P / (1 - P) = \xi/\delta$ となる故に、$\log_{10}$(正答率 / 誤答率) $= \log_{10}(\xi/\delta) = \log_{10}\xi - \log_{10}\delta$ との数理である。従って、例えば、素点得点 4 の受験者群については、五つの統合された項目群のモデル上での $\log_{10}(\xi/\delta)$の合計は$[(\log_{10}\xi_4 - \log_{10}\delta_{3\text{-}4}) + (\log_{10}\xi_4 - \log_{10}\delta_{5\text{-}6}) + (\log_{10}\xi_4 - \log_{10}\delta_{7\text{-}9}) + (\log_{10}\xi_4 - \log_{10}\delta_{10\text{-}11}) + (\log_{10}\xi_4 - \log_{10}\delta_{12\text{-}13})]$となる。 r 列上での 10 組の同一素点得点受験者群について $(- \log_{10}\delta_{3\text{-}4} - \log_{10}\delta_{5\text{-}6} - \log_{10}\delta_{7\text{-}9} - \log_{10}\delta_{10\text{-}11} - \log_{10}\delta_{12\text{-}13})$は定数である。その基準単位の任意性(Rasch[4], 1960, p. 84)により五つの統合項目群の $\log_{10}\delta$ 平均を 0 と置けば、五つの統合項目群ロジット値の平均値が各同一素点得点受験者群に属する受験者の $\log_{10}$(能力母数)の推定値となる。その値が表 2 での Average of Log r の列に与えられている。同様に、Average of Log i の行には、統合項目群単位別による同一素点得点受験者群数 10 組に関するロジット値の平均値が算出されており、各統合項目群単位での $\log_{10}$(応答容易度母数)の推定値となる。その正負の符号を入れ替えることにより、各統合項目群単位での $\log_{10}$(応

答困難度母数)の推定値に変換される。$\log_{10}$(能力母数)と $\log_{10}$(応答困難度母数)の各値が与えられたことにより、各受験者能力母数推定値と各統合項目群単位での困難度母数推定値が容易に算出される。その算出値を表 2 に含めている。

表 2 において、受験者群と項目群それぞれの母数推定値の上で非常に大きなばらつきが観察される。これについては、Rasch[4] (1960)により、$\mu_0(\theta) = \xi/\delta$(p. 120)として、$\mu(\theta) = \gamma(\mu_0(\theta))^{\alpha}$ への変換は任意であり、α と γ が正の値である限り、$\mu(\theta)$はその測定関数として $\mu_0(\theta)$と本質的に同等であると述べられている(p. 121)。従って、例えば、$\mu(\theta) = (\mu_0(\theta))^{1/2} = (\xi/\delta)^{1/2}$ とすれば、得られる受験者群と項目群いずれの母数推定値もそのばらつきは格段に小さくなる。又、上記の LOG method 自体については、能力推定値と項目困難度推定値が完全には分離されていないという理論上の欠陥を有している(Rasch[4], 1960, pp. 181−182; Wright[6], 1980, p. 188)と指摘されている。但し、Wright[6] (1980)により、PAIR method と共に LOG method が、理論的に最も妥当とされる Fully Conditional method (FCON)と較べて、母数推定値の上で実質的に問題となる差異を生ずるものではないと母数推定法比較研究結果として断言されている(p. 189)。

上記の知能テスト BPP−N データの Rasch 項目分析モデルへの適合度合いを見るには、受験者能力母数推定値に関しては、Average of Log r の列と統合項目群単位各列のロジット値相関の程度、又、統合項目群単位上での困難度母数推定値に関しては、Average of Log i の行と同一素点得点受験者群別各行のロジット値相関の程度を観察することが素朴ながらも最も利便性に富む。Rasch[4] (1960)はそれを図示で与えており(pp. 87−90)、このデータの Rasch 項目分析モデルへの大変に高い適合程度が観察される。ちなみに、下表 3 と 4 におけるピアソン相関係数がその高いモデル適合度を表示するものである。

表 3　受験者能力母数推定値に関する列間ロジット値相関係数一覧

		Average of Log r	3−4	5−6	7−9	10−11	12−13
Average of Log r	Pearson Correlation	1.000	.927	.975	.993	.994	.990
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000
3−4	Pearson Correlation	.927	1.000	.837	.897	.920	.907
	Sig. (2-tailed)	.000	.	.003	.000	.000	.000
5−6	Pearson Correlation	.975	.837	1.000	.974	.959	.971
	Sig. (2-tailed)	.000	.003	.	.000	.000	.000
7−9	Pearson Correlation	.993	.897	.974	1.000	.991	.974
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000
10−11	Pearson Correlation	.994	.920	.959	.991	1.000	.978
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000
12−13	Pearson Correlation	.990	.907	.971	.974	.978	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.

表 4　項目困難度母数推定値に関する Average of Log i と同一素点得点受験者群別各行ロジット値の相関係数

同一素点得点受験者群	4	5	6	7	8	9	10	11	12	13
Average of Log I とのピアソン相関係数	0.970	0.968	0.965	0.994	0.985	0.991	0.985	0.985	0.978	0.965
Sig. (2-tailed)	0.006	0.007	0.008	0.001	0.002	0.001	0.002	0.002	0.004	0.008

上記の知能テスト BPP−N のみならず、他に、BPP−F、BPP−L、BPP−V との名称データについても Rasch[4] (1960)は同様な手法で分析した結果を図示で与えており(pp. 97−105)、BPP−L は項目分析モデルにかなり高い程度で適合していたが、BPP−F と BPP−V については明らかなモデル不適合が観察された。Rasch[4] (1960, p. 100)の説明によれば、BPP−N と BPP−L のテスト制限時間は充分に長いものであったが、BPP−F と BPP−V については、それら二つのテスト制限時間を長くすれば多くの受験者が満点に近い得点を取る様な性質を持つテストであったことから、受験者能力弁別のために恣意的に時間制限が短くされたとのことである。この短いテスト制限時間が受験者問題解答能力の上に問題応答速度能力という能力二次元性を BPP－F と BPP−V への回答に結果として求めたことになり、問題応答速度が純粋な一次元性知能測定への汚染要因として働き、それが項目分析モデルへの不適合として顕現されたとの Rasch[4] (1960, p. 100)による解釈である。これは、データの普遍的尺度構成妥当性への本質条件として、Rasch 項目分析モデル規定上での項目困難度と受験者能力の母数結合一次元性、すなわち、Wright[10] (1991, p. 158)によって定義された'A single invariant conjoint order of item and person parameters'の重要性についての Rasch の認識を示すものである。

興味を覚えさせるのは、Rasch 自身が 1959 年前後までは母数分離定理の重要性をあまり意識していなかった(Rasch[7], 1977, pp. 63−66)ということである。この推測は、Rasch による母数分離定理を意味する'bridge−building'という概念に関するデンマーク語での論考が 1958 年という後年になって初めて発表された(Andersen and Olsen[3], 2001, p. 12)ことにも符合する。これについての彼の言葉(Rasch[7], 1977)は Rasch の理念確立の上で大きな示唆があるものとして、以下にその要約記述を与えておく。

『1959 年にノルウェーの経済学者である Ragnar Frisch がコペンハーゲン大学において名誉博士号を授けられた翌日に、Frisch が新たに築いた統計分析手法を 1935 年の春にオスロでの彼の下で研究したこともあって、旧交を温めるために彼を訪れた。その際に、ある一定の制限時間内での受験者読了単語数に関する乗法ポアソン確率モデルを説明した。その後に、すべての受験者が二つの音読テスト i と j を受けた結果として合計音読単語数(a_{vi} + a_{vj})が与えられた条件の下での a_{vi} と a_{vj} が同時に起こる二項分布確率式には受験者能力母数 ξ_v が含まれないことを導出した。それを見て、Frisch は、「能力母数が除去されて、大変に興味がある」と大きく目を見開いて叫んだ。その後の会話中にも、Frisch はその言葉を数回繰り返し、私はその度にうなずいた。幾日か経てある時突然に、私の説明の何が Frisch にあのような驚きの反応をもたらしたのかをはっきりと理解した。確率モデルのどの種類が母数分離を可能とする乗法ポアソン確率モデルに共通する属性を持っているであろうか。Frisch が驚いたその意味するところは、母数分離を可能とすることが大変に重要な種類の確率モデルに求められる必須の属性であるとのことである』(pp. 63−66)。

かくして、Rasch は確率モデルにおいて母数分離定理が持つ本質的重要性を理念化したことになる。Rasch[4] (1960)による母数分離定理(separability theorem)の概念説明は次の通りである。

『テストの組み合わせあるいは一つのテストに含まれた項目群への受験者一群の応答から、一つのテストあるいは一つの項目の母数のみに依存する一組の母数推定値分布と一人の受験者の母数のみに依存する別の一組の母数推定値分布が得られる様に、観測値を整理することは可能である。得られた二組それぞれの母数推定値群の条件付分布は他のいずれの母数推定値にも依存していない』(p. 122)。

これにより、Rasch が構築した乗法ポアソン確率モデルと項目分析モデルが持つ心理測定尺度構成への根本的な意義と効用を意識化したと思われる。

1935 年春から 1936 年夏に渡る Rasch のロンドン在留時に懇意になった Jerzy Neyman により 1960 年に開催された第四回確率理論と数理統計学バークレー大会における招待講演者として Rasch が招聘された(Andersen and Olsen[3], 2001, p. 18)。その講演内容が'On general laws and the meaning of measurement in psychology'と題されて 1961 年に刊行された大会報告書に掲載された(Andersen and Olsen[3], 2001, p. 19)。三つ以上の順序カテゴリーから成る項目群への被験者群応答に関する多次元ベクトルを仮定した確率モデルの数学的考察で以って一次元性順序尺度データ分析モデルを発表した。そのモデル化は、母数 θ を持つ被験者 v のカテゴリー数 m から成り刺激母数 σ を持つ項目 i に対する順序カテゴリーx への応答確率が次式により表されるものであった(Rasch[11], 1961, p. 333)。

$$P\{x \mid \theta_v + \sigma_i\} = \exp[\phi(x)(\theta_v + \sigma_i) + \rho(x)] / \gamma(\theta_v, \sigma_i)$$

上式右辺分母の $\gamma(\theta_v, \sigma_i)$は、$P\{x \mid \theta_v + \sigma_i\}$, $x \in \{1, 2, \ldots, m\}$の確率合計を 1 とする基準化係数 $\sum_{\mu=1}^{m} \exp[\phi(\mu)(\theta_v + \sigma_i) + \rho(\mu)]$ と理解される(Andersen[12], 1972, p. 43)。ϕ と ρ はそれぞれ得点母数(scoring parameter)とカテゴリー母数(category parameter)との名称が Andersen (1972[12], p. 43)によって付されている。それは、各被験者総点一群と各項目総点一群がそれぞれ十分統計量となって各被験者位置母数と各項目位置母数の分離推定を可能として、両母数の結合一次元性を規定する必要十分な数理モデルとして演繹されたものである(Masters[13], 1982, p. 152、参照)。なお、Rasch[11] (1961, p. 333)により与えられたそのモデル確率式が母数推定十分統計量の存在を可能とするには $\phi_{x+1} - \phi_x = \phi_x - \phi_{x-1}$が必要条件であると Andersen[14] (1977, p. 76)によって示された上で、今日では以下に与える Andrich[15] (1978, p. 569)の Rating Scale モデルとして知られている。

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

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

上記の Rasch[11] (1961)による講演論述において、十分統計量の存在が必要条件として数理的に母数分離定理を可能にせしめると主張され(p. 326)、母数一次元性の重要性についても言及されている(p. 326)。更に、論題に示された如く、Rasch[11] (1961)の関心が心理測定として表出しており、

測定への規定概念を次の様に述べている。

『二つの刺激間の測定比較はその測定のために使用された任意の標本被験者群から独立していなければならない。二つの刺激間の測定比較は、又、そのテストに含まれたそれら二つ以外の他の刺激群あるいは別のテスト上での同質刺激群による測定結果からも独立していなければならない。対称的に、二人の被験者間の測定比較はその測定のために使用された任意のテスト同質刺激群から独立していなければならない。二人の被験者間の測定比較は、又、そのテストでのそれら二人以外の被験者群及びそのテストによる別の測定機会における被験者群からも独立していなければならない』(pp. 331−332)。

Rasch 理論の根幹である母数分離定理により、モデル適合度の高いデータにより尺度化される母数推定値は局所独立性充足の下で誤差変動内での不変性を属性とする。この母数推定値不変性をもたらす母数分離性と局所独立性を心理測定への前提条件とみなして、データの Rasch モデルへの適合が普遍的心理測定尺度構成には不可欠であると断言したのが Rasch による上の言葉である。1960 年代初頭に Rasch に教えを受けた Erling B. Andersen (Andersen and Olsen[3], 2001, p. 9)によれば、Rasch が授業時に自身のモデルを「測定のためのモデル」(models for measurement)と表現したとのことである(Andersen and Olsen[3], 2001, p. 21)。母数分離定理を基底理論とする「測定のためのモデル」との理念は、物議をかもす性質を含んでおり、観察データへの適合度を高めるべく開発されてきた伝統的な統計分析モデルとまさに対極にあると認識される。これは、1960 年の著作においての Rasch[4] (1960)によるモデル適合への以下の言及と比べれば、Rasch モデルが心理測定尺度構成妥当性を評価するものとの強固な Rasch 理念を示すものである。

『全体的に見て、限られた種類のテストに対してではあるが、項目分析モデルによるそれらのテスト分析が示唆していることは、テストに含まれる項目群が、複雑な特性に関するものであるとしても、内容同質性と困難度変化性を考慮して作成されるならば、その項目群が潜在特性一次元性モデルに良好に適合する可能性はかなり高い』(p. 125)。

1961 年に Rasch モデルが普遍的心理測定尺度構成妥当性を規定するものとして理念化されたことにより、人間実証科学上の一般化推論のために不可欠な本質概念として Rasch は「固有客観性」(specific objectivity)を導入することになる(Rasch[7], 1977, p. 66)。固有客観性は Fisher による十分統計量の概念と密接に関係するものとして位置付けられ(Rasch[7], 1977, p. 66)、事物の実証科学的な比較対照への必須概念として捉えられている(Rasch[7], 1977)。刺激を作り出す要因と刺激を受ける対象物に関して、二つの要因及び二つの対象物を比較するためには関係枠の中で他の要因群と他の対象物群からは独立している必要性があるということが客観性とされ、客観性はその関係枠に規定される故に固有とされるとの Rasch[7] (1977, p. 77)による概念論である。

1965 年にデンマーク語で発表された論文において、二値反応データに関しては Rasch 項目分析モデルのみが固有客観性を属性としていると Rasch が示したとの Andersen and Olsen[3]

(2001)による言及(p. 22)がある。又、1966 年に発表された英語論文においても、『二値反応データについて固有客観性が実現されるためにはデータ観察値が Rasch 項目分析モデルに適合しなければならず、同じことが三値以上の反応データにおいても制限付きで多項 Rasch モデルに当てはまる』と Rasch[9] (1966, p. 56)は言明している。ちなみに、三値以上の反応データに対する現今の Rasch モデルが、前記参照の Rating Scale モデル(Andrich[15], 1978)に併せて、Partial Credit モデル(Masters[13], 1982)、及び、Many−Facet Rasch Measurement (Linacre[16], 1989)である。Rasch モデルのみが数理上で母数分離定理を規定とするものであるから、観察データの Rasch モデルへの適合により母数が固有客観的に推定され得るとの Rasch による見解である(Andersen and Olsen[3], 2001, p. 21)。それは、人間実証科学上で普遍性のある推論を固有客観的に実現可能とするためには、順序尺度データは「測定のためのモデル」としての Rasch モデルへの高い適合性を必然的に前提とすると Rasch が断定したことを意味する。かくして、1945 年のテスト分析を端緒にその分析理論の構築を経て、Rasch が到った確信は、Rasch モデルのみが人間実証科学における普遍的尺度構成妥当性を検証し得るものであり、人間実証科学上の概念測定に基づくその一般化推論普遍性のためには Rasch 測定が不可欠のものである、と要約される。

1960 年に出版された Rasch[4]の業績に対する Jane Loevinger[17] (1965)による以下の言葉で以って、当時の計量心理学専門分野での一つの Rasch モデル評価を窺い知り得る。それが本節の格好な結びとして筆者に参照される。

『Rasch (1960)は計量心理学上の問題に対して真に新しい解決法を考案した。 彼は古典的な計量法を一切使用せずに、確率モデルに真新しく代数を当てはめている。ある一人の受験者のある一つの項目に対する正答確率がその受験者のみに属する能力母数とその項目のみに属する困難度母数の積であると仮定されている。各受験者に付される能力推定値と各項目に付される困難度推定値の算出を超越する各受験者能力と各項目困難度の独立性がモデル属性とされる。ある一人の受験者の能力推定値は、他の受験群能力推定値から独立しており、又、テストに含まれた項目群の困難度推定値からも独立している。同様に、ある一つの項目の困難度推定値は、他の項目群困難度推定値、及び、テスト項目群により測定された受験者群能力推定値から独立している。事実、受験者能力と項目困難度のこの分離独立性が絶対的な測定基準としてかって提示された(Loevinger, 1947)。然しながら、あの時には、絶対的な測定への提示内容がその分離独立性の基準を満たし得ないものであった。Guttman 尺度もその基準を満たしてはいない。かくして、計量心理学上の二つの大問題の一つである任意ではない測定法の確立への偉大な功績により、Rasch にその栄誉が授与されなければならない。Rasch が果たしたことは異なる種類の、そして、Cronbach、Rajaratnam、及び、Gleser よりもさらに厳格な種類の普遍化である。Rasch モデル適合上での母数推定値は、[その適合データ受験者群・項目群に関する潜在特性再現上での]ある種の広い制限で以って、特定受験者群と特定項目群から分離独

立していることになる。その[データ潜在特性再現性]限界の下では、普遍性が達成されていると言える』(p. 151)。

3. Benjamin D. Wright の Rasch 項目分析モデル測定展開

Probabilistic Models for Some Intelligence and Attainment Tests の序文において、Rasch[4] (1960)はシカゴ大学の L. J. Savage にその最終原稿を批判的に閲読してもらった(p. xxii)と記している。Savage は統計学の教授であったが、ワシントンで 1959 年に開催された計量生物学学会で彼が以前から交友のあった Rasch に出会った時に、Rasch は測定上の発見を世界に伝えたいと Savage に述べていた(Wright[18], 1987, p. 1)。Savage の友人である Benjamin D. Wright はシカゴ大学の教育学部で統計学を教えていたが、その頃は Rasch と Rasch の研究については無知であった。Savage との会話で Wright が Rasch の研究に僅かばかりの興味を示したことから、Savage が Rasch をシカゴ大学に招聘して社会科学研究者と統計学者向けの連続公開講座を開く段取りをつけて、それが 1960 年春に一週二回の講義で三ヶ月に渡り実現された(Wright[18], 1987, pp. 1−2; Wright[19], 1988, p. 26)。これが、米国での Rasch モデル研究と適用の端緒を開き、その後の Wright 及び彼の友人や門下生を中心としてシカゴ大学での Rasch モデル展開へのきっかけとなった(Wright[19], 1988, pp. 27−29)。以下に、Rasch 項目分析モデルのシカゴ大学での Wright を中心とする心理測定研究者達への伝播における智見をモデル理解補強のために概観する。

1964 年までには、シカゴ大学で Wright と数理物理学修士課程修了のイギリス人学徒 Bruce H. Choppin は Rasch[4] (1960)により理論構築された母数推定値算出法としての LOG method、PAIR method、Fully Conditional method の IBM 上での演算処理プログラム化を終えていた(Wright[19], 1988, p. 27)。彼等は、いずれの方法によってもほぼ同一の母数推定値が得られることを確認して(Wright[19], 1988, p. 27)、1965 年の米国中西部心理学学会で‘Sample−free probability models in psychological measurement’と題して特に PAIR method の有用性を中心とする口頭発表をおこなった(Wright[6], 1980, p. 189)。1965 年にインド人核物理学者である Nargis Panchapakesan がシカゴ大学を訪れ、彼女の多大な助力を得て、Wright は UCON (Unconditional maximum-likelihood estimation)法による Rasch 項目分析モデル母数推定値算出プログラムをフォートラン上で作成した(Wright[18], 1987, p. 3)。LOG method と共に、その無条件同時最尤推定法の数理基盤と計算アルゴリズムを提示した論文が Wright and Panchapakesan[20] (1969)による‘A procedure for sample-free item analysis’である。その本質的意義はコンピュータ上での Rasch 測定を可能とする迅速な自動演算処理プログラム作成への指針を与えたことであり、これが 1976 年に市販化された Rasch 項目分析モデル測定プログラム BICAL の Wright と Mead による製作へと結実することとなった(Wright and Stone[21], 1979, p. 46)。

Wright and Panchapakesan[20] (1969)による Rasch 項目分析モデル確率式は、能力母数 b_n を持つ受験者 n が困難度母数 d_i を持つ項目 i に応答して、正答 $a_{ni} = 1$ あるいは誤答 $a_{ni} = 0$ の生起

確率 $P_r(a_{ni})$が次式で示されるものである。

$$P_r(a_{ni}) = \exp[a_{ni}(b_n - d_i)] / [1 + \exp(b_n - d_i)]$$

(Wright and Panchapakesan[20], 1969, p. 24)

前節で見た様に、Rasch (1960, pp. 74−75)はロジット関数を用いて、項目正答確率 P を次式により定義していた。

$$P = \zeta / (1 + \zeta) = (\xi/\delta) / (1 + \xi/\delta) = \xi / (\xi + \delta)$$

受験者 n と項目 i の添え字を付けて、$\xi_n = \exp(b_n)$、$\delta_i = \exp(d_i)$と置けば、

$$P_r(a_{ni} = 1) = \xi_n / (\xi_n + \delta_i) = \exp(b_n) / [\exp(b_n) + \exp(d_i)]$$
$$= [\exp(b_n) / \exp(d_i)] / \{[\exp(b_n) / \exp(d_i)] + 1\}$$
$$= \exp[a_{ni}(b_n - d_i)] / [1 + \exp(b_n - d_i)]$$

従って、Rasch の受験者能力母数 ξ と項目困難度母数 δ がそれぞれ指数変換されており、算出されるそれぞれの推定値群の標準偏差を Rasch 考案式におけるそれよりも格段に小さくなる様に Wright and Panchapakesan[20] (1969, p. 24)によって設定されているのが分かる。なお、前節で参照された様に、母数推定値算出単位変換によるこの便宜性については Rasch[4] (1960, pp. 120−121)によっても触れられている。

今日での Rasch 項目分析モデル確率式とされる上式の属性としての「固有客観性」は次の様に理解される(Wright and Linacre[22], 1987, pp. 5−6)。

> 仮想同一状況下での二人の受験者による同一項目への多数回応答試行から成る試験実施を想定する。両受験者がその項目に対して共に正答及び誤答する度数は両者のその項目に関する能力比較上での情報を与えない。従って、各受験者を m と n として、n の正答及び m の誤答の生起度数 R_nW_m、並びに、n の誤答及び m の正答の生起度数 W_nR_m が両受験者のその項目に関する能力差判定上での考察対象となる。その項目を i として、受験者 m と n による項目 i への応答に関する潜在的正答確率をそれぞれ P_{mi}、P_{ni} と置けば、生起度数 R_nW_m と W_nR_m の対比は次式で表現される。
>
> $$R_nW_m / W_nR_m = [P_{ni}*(1 - P_{mi})] / [(1 - P_{ni})* P_{mi}]$$
>
> 上記の説明は別項目 j においても原理として同様であり、潜在特性一次元性内容同質項目群 i と j への両受験者による応答が個別に独立しているテストにおいては、上式に関する項目間での比率同等性が想定される。それは次式で示される。
>
> $$[P_{ni}*(1 - P_{mi})] / [(1 - P_{ni})* P_{mi}] = [P_{nj}*(1 - P_{mj})] / [(1 - P_{nj})* P_{mj}]$$
>
> 上式を変換して、
>
> $$P_{ni} / (1 - P_{nj}) = [P_{nj}*(1 - P_{mj})* P_{mi}] / [(1 - P_{nj})* P_{mj}*(1 - P_{mi})]$$
>
> 上式における受験者 m と n、並びに、項目 i と j についての潜在的正答確率の関係は一次元性充足項目群におけるすべての一対比較のみならず、一次元性充足受験者群におけるすべての一対比較の上でも保証されなければならない。更に、測定上での「関係枠」

(frame of reference)を定義するために、一つの測定基準値を持つ受験者と項目がそれぞれ設定される必要性から、上式における受験者 m と項目 j にそれぞれ測定基準値を持つ印として 0 が挿入される。

$$P_{ni} / (1 - P_{nj}) = [P_{n0}*(1 - P_{00})* P_{0i}] / [(1 - P_{n0})* P_{00}*(1 - P_{0i})]$$

$$= [P_{n0}/(1 - P_{n0})]*[P_{0i}/(1 - P_{0i})]*[(1 - P_{00})/P_{00}]$$

測定「関係枠」上で基準値印 0 を持つ受験者の基準値印 0 を持つ項目への正答確率を $P_{00} = 0.5$ と基準化することにより $(1 - P_{00})/P_{00} = 1$ が上式に挿入される。

$$P_{ni}/(1 - P_{nj}) = [P_{n0}/(1 - P_{n0})]*[P_{0i}/(1 - P_{0i})]$$

両辺それぞれの対数を取れば、

$$\ln[P_{ni}/(1 - P_{nj})] = \ln[P_{n0}/(1 - P_{n0})] + \ln[P_{0i}/(1 - P_{0i})]$$

$$= \ln[P_{n0}/(1 - P_{n0})] - \ln[(1 - P_{0i})/P_{0i}]$$

上式において、オッズ $P_{n0}/(1 - P_{n0})$は受験者 n の基準値印 0 を持つ項目への正答確率をその誤答確率で割ったものであり、その対数表現 $\ln[P_{n0}/(1 - P_{n0})]$が受験者能力母数 b_n とみなされる。同様に、オッズ$(1 - P_{0i})/P_{0i}$ は基準値印 0 を持つ受験者の項目 i への誤答確率をその正答確率で割ったものであり、その対数表現 $\ln[(1 - P_{0i})/P_{0i}]$が項目困難度母数 d_i とみなされる。それが、オッズの対数表現上での次式で示される Rasch 項目分析モデル確率式である。

$$\ln[P_{ni}/(1 - P_{nj})] = \ln[P_{n0}/(1 - P_{n0})] - \ln[(1 - P_{0i})/P_{0i}] = b_n - d_i$$

受験者能力母数 b_n と項目困難度母数 d_i が独立分離されており、又、各受験者能力と各項目困難度がそれぞれ他の受験者群と他の項目群からも母数の上で独立分離されており、Rasch 項目分析モデル規定としての「固有客観性」が表現されている。その Rasch 項目分析モデルオッズ対数表現式を指数表現式に変換すれば、Wright and Panchapakesan[20] (1969, p. 24)によって与えられた $P_{ni} = \exp(b_n - d_i) / [1 + \exp(b_n - d_i)]$となる。

Rasch 測定理論の真髄は「不変性」(invariance)である(Andrich, 1988[23], pp. 21−22, and 2005[24], p. 53)。それは、上記に参照された Rasch 項目分析モデル規定としての「固有客観性」に基づくものであり、前節で参照された Georg Rasch の理念として表現された「測定のためのモデル」の根拠となっている。それは、又、Wright[10] (1991, p. 158)によって定義されたモデル規定上での「母数結合一次元性」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データー次元性充足によって顕現される特性でもある。Rasch 項目分析モデル上でのその受験者群能力推定値不変性を易しい項目群と難しい項目群へのデータ折半で以って示した研究報告が、1967 年 10 月 28 日に Princeton で開催された Educational Testing Service 研究者招待大会での Wright[25] (1967)による発表である。Rasch 測定母数推定値同等性検証法としてのカイ二乗検定 (McNamara[26], 1991, pp. 153−154)に際しての受験者群折半下位標本上での項目群に関する t 値としての $t_{iAB} = (d_{iA} - d_{iB}) / \sqrt{(e^2_{iA} + e^2_{iB})} \backsim N(0, 1)$ は項目個別 Rasch モデル適合度指標の一つと

して今日有用視されている(Bond and Fox[27], 2001, pp. 56−57; Smith[28], 1996; Smith and Suh[29], 2003; Wright and Stone, 1979[21], pp. 94−95)。但し、Rasch 項目分析モデル適合度の低いデータについては二分割標本 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[30](2006, pp. 48−50)によって付されている。

Wright and Panchapakesan[20] (1969)において言及された母数推定値法は LOG method(pp. 27−33)及び Fully Conditional method (pp. 34−40)であり、フォートラン上でのそれぞれの演算プログラムが記述されている。LOG method については、前節での Rasch[4] (1960)によるその実施要領説明(pp. 76−105)で以ってその基本的方法論が理解される。Rasch[4] (1960)によっても示された Fully Conditional method(pp. 178−181)は、受験者母数を条件付で除去することにより、受験者集団から完全に分離独立した項目困難度母数推定値の算出を可能とするものであり(服部、1988[8], p. 36, and 1998[31], p. 269; Wright[6], 1980, p. 189)、一致推定量を得る上で最良に望ましいとされている(Andersen[32], 1970; Wright[6], 1980, p. 189)。然しながら、基本対称関数の計算が大変面倒であり、演算時間が過大であり、その過程で丸めの誤差の蓄積という厄介な問題を抱えている(服部、1988[8], p. 36, and 1998[31], p. 273)。しかも、その影響が 15 項目を超えると顕著になり、困難度推定値の精度が極端に低くなる(Wright and Douglas[33], 1977a, p. 580)と指摘されている。但し、Gustafsson[34] (1980)によれば、この問題が解決されており(Gustafsson, 1977, cited in Gustafsson[34], 1980, p. 210)、60 項目までなら困難度推定値は正確であり、演算時間も大してかからない(p. 210)とのことである。Gustafsson によるその計算アルゴリズム使用に基づく市販コンピュータプログラムについては筆者に不詳である。

シカゴ大学において 1966 年以降に開発された Rasch モデル測定コンピュータプログラムには UCON (Unconditional maximum-likelihood estimation)が採用されている(Wright[6], 1980, p. 197)。UCON(無条件同時最尤推定)法に関しては、受験者数が増えるにつれて推定すべき受験者能力母数の数が増大し、そのために項目困難度推定値の偏りがなくならず、一致推定量を得られないという性質が存在する(池田[35]、1994, pp. 77－78)。然しながら、修正 UCON 法で以って、項目困難度母数推定値の偏りは[(項目数－1) / 項目数]を係数として乗算すればほぼ是正されて、Andersen (1970[32] and 1972[12])により推奨され、又 Gustafsson (1977, cited in Gustafsson[34], 1980, p. 210)によりプログラム化された条件付最尤推定法の上での一致推定量に限りなく近づくと報告されている(Wright and Douglas[36], 1977b, pp. 288−289)。従って、Rasch 項目分析プログラム BICAL (Wright and Mead, 1976, cited in Wright and Stone[21], 1979, p. 46))に始まり、Quest (Version 2.1, Adams and Khoo, 1996, The Australian Council for Educational Research Ltd.)、並びに、Facets (Linacre, 1989−2001, Winsteps)を含める多くの市販 Rasch モデル測定コンピュータプログラムが修正 UCON 法を採用している。服部[37](1991)により、修正 UCON 法の頑健性は高く(pp. 74−75)、その推定法は条件付最尤推定法に併せて対比較推定法とベイズ的対比

較推定法にも匹敵する母数推定精度であることが模擬実験の上で示されている(服部[31]、1998)。

以下に、UCON 法の原理要点を服部[37](1991)と Wright and Stone[21] (1979)に参照する。

受験者を n = 1, 2, ..., N、及び、項目を i = 1, 2, ..., L として、受験者 n が項目 i に応答して、正答 $x_{ni} = 1$ あるいは誤答 $x_{ni} = 0$ を所与とする尤度関数は「局所独立性」規定の上で次式により表現される。

$$L(x_{ni}) = \prod_{n=1}^{N}\prod_{i=1}^{L} P_{ni}^{x_{ni}}(1-P_{ni})^{(1-x_{ni})}$$

(服部[37], 1991, p. 72)

上式に、Wright and Panchapakesan[20] (1969, p. 24)によって与えられた Rasch 項目分析モデル確率式 $P(x_{ni}) = \exp[x_{ni}(b_n - d_i)] / [1 + \exp(b_n - d_i)]$を代入すれば、

$$\begin{aligned} L(x_{ni}) &= \prod_{n=1}^{N}\prod_{i=1}^{L}\{\exp[x_{ni}(b_n - d_i)] / [1 + \exp(b_n - d_i)]\} \\ &= \exp[\sum_{n=1}^{N}\sum_{i=1}^{L} x_{ni}(b_n - d_i)] / \prod_{n=1}^{N}\prod_{i=1}^{L} [1 + \exp(b_n - d_i)] \\ &= \exp(\sum_{n=1}^{N}\sum_{i=1}^{L} x_{ni}b_n - \sum_{n=1}^{N}\sum_{i=1}^{L} x_{ni}d_i) / \prod_{n=1}^{N}\prod_{i=1}^{L} [1 + \exp(b_n - d_i)] \end{aligned}$$

更に、受験者 n の得点を r_n とすれば $\sum_{n=1}^{N}\sum_{i=1}^{L} x_{ni}b_n = \sum_{n=1}^{N} r_n b_n$、項目 i への正答者数を s_i とすれば $\sum_{n=1}^{N}\sum_{i=1}^{L} x_{ni}d_i = \sum_{c=1}^{L} s_i d_i$ となり、これらを代入すれば、

$$\begin{aligned} L(x_{ni}) &= \exp(\sum_{n=1}^{N} r_n b_n - \sum_{i=1}^{L} s_i d_i) / \prod_{n=1}^{N}\prod_{i=1}^{L}[1 + \exp(b_n - d_i)] \\ &= [\exp(\sum_{n=1}^{N} r_n b_n) * \exp(-\sum_{i=1}^{L} s_i d_i)] / \prod_{n=1}^{N}\prod_{i=1}^{L}[1 + \exp(b_n - d_i)] \end{aligned}$$

これに対数変換を施せば、

$$\ln L(x_{ni}) = \sum_{n=1}^{N} r_n b_n - \sum_{i=1}^{L} s_i d_i - \sum_{n=1}^{N}\sum_{i=1}^{L} \ln[1 + \exp(b_n - d_i)]$$

(Wright and Stone[21], 1979, pp. 19−20 and pp. 62−63)

上式において、$\sum_{n=1}^{N} r_n b_n$ と $\sum_{i=1}^{L} s_i d_i$ が分離されていることにより、各受験者得点 r_n と各項目得点 s_i のそれぞれが独立に各受験者能力母数 b_n と各項目困難度母数 d_i を最尤推定する上での十分統計量になることが示されている(Wright and Stone[21], 1979, p. 20 and p. 63)。この対数尤度関数を最大にする b_n と d_i が最尤推定値であるが、代数的には処理され得ない(服部[37]、1991, p. 72)。従って、Newton-Raphson 法利用の上で両母数 b_n と d_i それぞれの初期値を $\ln[P_{ni}/(1 - P_{ni})]$と $\ln[(1 - P_{ni})/P_{ni}]$として項目群困難度母数推定値平均を 0 とする両母数一連の交互推定反復計算による一定基準上での収束値がモデル母数推定値とされる(Wright and Stone[21], 1979, pp. 64−65)。それが UCON 法の基本原理であり、修正 UCON 法で以って、項目困難度母数推定値の偏りが[(項目数−1) / 項目数]の係数乗算により是正される(Wright and Stone[21], 1979, p. 65)。欠損値がない場合には、同じ得点数を持つ受験者群には同一能力推定値、並びに、同じ正答者数を持つ項目群には同一困難度推定値が与えられる。

古典的テスト理論においては、Cronbach のアルファ係数あるいはそれと 0・1 データに関して同等な Kuder-Richardson Formula 20 に基づくテスト信頼性係数使用の上でテスト測定標準誤差(= テスト得点標準偏差× $\sqrt{(1\text{-テスト信頼性係数})}$ (Crocker and Algina[38], 1986, pp. 122−123)が参照される。Cronbach のアルファ係数あるいは Kuder-Richardson Formula 20 に

基づくテスト信頼性係数が大きな値になるに従って、テスト測定標準誤差は小さくなる。但し、テスト測定標準誤差はテスト平均得点から高低離れる得点に準じてその受験者群の素点誤差を過大評価している(Smith[39], 2001, p. 283)ことに留意される。項目応答理論の優れた特徴として各受験者能力推定値と各項目困難度推定値に関する標準誤差が個別に出力される。Rasch モデル測定コンピュータプログラムにより算出される最尤推定値は、$\hat{\theta}_{\mathrm{MLE}} \sim N(\theta, 1/I(\theta))$ (as $n\to\infty$)という漸近的性質を有している(豊田[40]、2002, pp. 64−65)ことから、各能力母数推定値と共にその誤差分散 $1/I(\theta)$が容易に推定される。誤差分散の平方根を取れば、各受験者能力母数推定値の標準誤差となり、各項目困難度母数推定値標準誤差についても同様であり、それは各受験者と各項目に付されたそれぞれの母数推定値分布の推定標準偏差である。それぞれの母数推定値標準誤差を定義する式は以下のものである(Wright[41], 1977a, p. 223)。

素点得点 r を持つ受験者の項目 i への正答確率：

$$P_{ri} = \exp(b_r - d_i) / [1 + \exp(b_r - d_i)]$$

素点得点 r を持つ受験者に付された能力母数推定値 b_r の標準誤差：

$$SE(b_r) = 1 / \sqrt{\left[\sum_{i=1}^{L} P_{ri}(1-P_{ri})\right]} \quad \text{(項目 } i = 1, 2, \ldots, L)$$

項目 i に付された困難度母数推定値 d_i の標準誤差：

$$SE(d_i) = 1 / \sqrt{\left[\sum_{r=1}^{L-1} \left(n_r P_{ri}(1-P_{ri})\right)\right]}$$

(受験者素点得点 $r = 1, 2, \ldots, L-1$;　n_r は同一素点得点 r を持つ受験者数)

母数推定値の許容出来る精度と Rasch 項目分析モデルへの適合度の観点により、100 人の受験者に 30 個以上の項目群が与えられれば有用な情報が得られ、受験者数が 400 人であれば項目困難度母数推定値は十分に精度の高いものになると報告されている(Wright[41], 1977a, p. 204; Wright[42], 1977b, p. 106)。モデル適合度が高いと結果として判明するならば、受験数が 200 人であっても項目困難度推定値の 99%信頼区間はほぼ推定値±0.5 ロジット値以内に収まり、それ以上の受験者数を以ってしても現実的には±0.3 ロジット値が得られる最高の精度であると指摘されている(Linacre[43], 1994, p. 328)。当然に、受験者能力推定値の精度も項目数増大により上昇することになる(Wright[41], 1977a, p. 223)。Linacre (1994[43], p. 328)が与える受験者数と項目困難度母数推定値標準誤差の関係から察すれば、潜在特性一次元性充足度の高い 100 個の項目群が与えられれば、能力推定値の 95%信頼区間はほぼ推定値±0.5 ロジット値以内になるかと思われる。現実的なテスト制限時間を考慮すれば、受験者群母集団能力分布に全体的に相応する困難度を持つ項目群が内容妥当性と同質性がほぼ保証されている条件下においては、母集団を代表する 200 名の受験者群に項目数 100 個から成るテストを与えることにより精度の高い母数推定値が得られることになる。

4. おわりに

本稿で筆者は「規定」という用語を頻繁に使用した。それは、Wright[10] (1991)によって記された specification (p. 158)の訳語であり、三規定として「母数結合一次元性」(データ一次元性)、「母数分離性」(十分統計量)、及び、「局所独立性」(Rasch 測定残差項目間・受験者間独立性)が範疇化されている。「規定」が Rasch モデルの本質性を顕示しているとの筆者による認識であり、それは人間科学研究者の普遍的に妥当である順序尺度構成への厳格な要請事項と理解される。一方、人間科学上での曖昧模糊とした多義性により、現実データのその規定充足度の低さを Rasch 測定適用により容易に認知されることが常となる。いみじくも、日本を代表する計量心理学者の一人である豊田 [40](2002)に『理論的には、項目間で識別力が等しい 1 母数モデルを考えることもできるが、経験的には制約が強すぎるようである』(p. 106)との言葉が見出される。筆者の限られた Rasch 測定適用経験で以ってしても、Rasch モデル規定充足度つまりモデル適合度の高いデータ取得は至難の業であると意識される。本考察により、『Rasch 理論は概念構築への不断の研究努力においてデータに要求されることを具体的に顕示するものである故に、データが Rasch モデルに適合しておらずに不備のある尺度構成との判明は、Rasch モデルではなくデータ自体に大きな欠陥があるとの示唆である』(Linacre[44], 1989−2001, p. 107)と刻印される。それは潜在特性一次元構成概念妥当性をデータにあくまでも厳格に要求する Rasch 数理モデルの本質性に由来する。その対極にあるデータ分析法の一つが共分散構造分析上での構造方程式モデリング(豊田 [45]、1998、参照)であり、それは、三件法以上から成る順序尺度データの現実多次元世界を反映する多義性を最適に説明する統計分析手法である。ここに、豊田 [40](2002)による上記の「強い制約」という言葉が Rasch モデルの限界性を意味するものとして往々に付される理由が見出される。事の核心は Rasch モデルの「強い制約」を実証研究上で如何なる程度に必要条件であるとみなすかである。

最後に、Rasch「規定」を実証研究上での強い必要条件とみなす一文への参照で以って本稿を閉じることにする。それは、Rasch モデルが人間実証科学性の向上に寄与する尺度構成と尺度評価に不可欠な‘fundamental measurement’であると主張するものである。

> 『社会科学分野での実証研究に諸種の統計分析手法を用いている多くの研究者達が Rasch 測定を単純過ぎるものとして非難する傾向がある。如何に人間の能力が真面目に一次元であるとみなされ得るのかとの疑義である。. . . それに対する返答は、Rasch 測定が単純過ぎるのではなく、Rasch モデル数理自体が理路整然として単純明快なのである。Rasch 測定の目的は、異なったテストを異なった日に異なった受験者群へ与えることにより得られるデータを単に記述することにあるのではない。その真の目的は、同様に適正な測定機会を通して使用され得る根本的に信頼性と妥当性のある尺度を作り上げることにある。Rasch モデルはテスト結果による個別の観察を超えて受験者と項目に関するそのテスト属性としての普遍性への推察を意図するものである』(Bond and Fox[27], 2001,

p. 106)。

【参照文献及び注】

1) 1978 年に Rasch モデルに属する Rating Scale Model を発表したオーストラリア人研究者 David Andrich が、1979 年 6 月にデンマークの Læsø にて Georg Rasch と面談した際の談話録音内容の筆述記録。

2) 1980 年に再版された ***Probabilistic Models for Some Intelligence and Attainment Tests*** の前書きにおいて Benjamin D. Wright が言及した Andrich(注１)による Rasch 談話録音の筆述記録。

3) E. B. Andersen & L. W. Olsen. 2001. The life of Georg Rasch as a mathematician and as a statistician. In Boomsma, A., van Duijn, M. A. J., & Snijders, T. A. B. (Eds.), *Essays on Item Response Theory* (pp. 3−24). New York: Springer.

4) 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)

5) 芝　祐順編著　1991.　『項目反応理論　―　基礎と応用』　東京大学出版会

6) B. D. Wright. 1980. Afterword. In Rasch, G., *Probabilistic models for some intelligence and attainment tests* (Reprinted in 1980, pp. 185−199). The University of Chicago Preess.

7) 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.

8) 服部　環　1988.　「Rasch モデルにおける母数のＵＣＯＮ推定法」　宇都宮大学教育学部、『教育実践研究指導センター紀要』第 11 号、pp. 33－41.

9) G. Rasch. 1966. An item analysis which takes individual differences into account. *The British Journal of Mathematical and Statistical Psychology*, 19, 1, pp. 49−57.

10) 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.

11) 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.

12) E. B. Andersen. 1972. The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistic Society, Series B*, 34, pp. 42−54.

13) G. N. Masters. 1982. A Rasch model for partial credit scoring. *Psychometrika*, 47, 2, pp. 149−174.

14) E. B. Andersen. 1977. Sufficient statistics and latent trait models. *Psychometrica*, 42, 1, pp. 69−81.

15) D. Andrich. 1978. A rating formulation for ordered response categories. *Psychometrika*, 43, 4, pp. 561−573.

16) J. M. Linacre. 1989. *Many-facet Rasch measurement*. Chicago: MESA Press.

17) J. Loevinger. 1965. Person and population as psychometric concepts. *Psychological Review*, 72, 2, pp. 143−155.
18) B. D. Wright. 1987. Rasch and Wright: The early years. *Rasch Measurement Transactions*, 1, 1 in Linacre, J. M. (Ed.), 1995, *Rasch Measurement Transactions, Part 1* (pp. 1−4). Chicago: MESA Press.
19) B. D. Wright. 1988. Georg Rasch and measurement. *Rasch Measurement Transactions*, 2, 3 in Linacre, J. M. (Ed.), 1995, *Rasch Measurement Transactions, Part 1* (pp. 25−32). Chicago: MESA Press.
20) B. Wright, & N. Panchapakesan. 1969. A procedure for sample-free item analysis. *Educational and Psychological Measurement*, 29, pp. 23－48.
21) B. D. Wright & M. H. Stone. 1979. *Best test design*. Chicago: MESA Press.
22) B. D. Wright & J. M. Linacre. 1987. Rasch model derived from objectivity. *Rasch Measurement Transactions*, 1, 1 in Linacre, J. M. (Ed.), 1995, *Rasch Measurement Transactions, Part 1* (pp. 5−6). Chicago: MESA Press.
23) D. Andrich. 1988. *Rasch models for measurement*. Newbury Park, California: SAGE Publications, Inc.
24) 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.
25) 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).
26) T. McNamara. 1991. Test dimensionality. *Language Testing*, 8, 2, pp. 139−159.
27) T. G. Bond & C. M. Fox. 2001. *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
28) 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.
29) 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.
30) 井澤廣行　2006. 「Rasc 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.
31) 服部　環　 1998. 「Rasch モデルにおける条件付き最尤推定と対比較推定 ― 諸推定方法の推定精度に関する検討」　宇都宮大学教育学部、『教育実践研究指導センター紀要』第 21 号、pp. 268−281.
32) E. B. Andersen. 1970. Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistic Society, Series B*, 32, pp. 283−301.
33) B. D. Wright & G. A. Douglas. 1977a. Conditional versus unconditional procedures for sample-free item analysis. *Educational and Psychological Measurement*, 37, pp. 573−586.
34) J. E. Gustafsson. 1980. Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 33, pp. 205−233.

35) 池田　央　1994.　『現代テスト理論』　東京：　朝倉書店

36) B. D. Wright & G. A. Douglas.　1977b.　Best procedures for sample-free item analysis.　*Applied Psychological Measurement*, 1, 2, pp. 281−295.

37) 服部　環　 1991.　「標本再抽出法による Rasch モデルの母数推定と外れ値検出」　宇都宮大学教育学部、『紀要』第４１巻、第１号、pp. 67−85.

38) L. Crocker & J. Algina.　1986.　*Introduction to classical and modern test theory*.　Belmont, CA: Wadsworth Group/Thomson Learning, Inc.

39) Smith, Jr., E. V.　2001.　Evidence for the reliability of measures and validity of measure interpretation:　A Rasch measurement perspective.　*Journal of Applied Measurement*, 2, 3, pp. 281−311.

40) 豊田秀樹　2002.　『項目反応理論[入門編]』　東京：　朝倉書店

41) B. D. Wright.　1977a.　Misunderstanding the Rasch model.　*Journal of Educational Measurement*, 14, 3, pp. 219−225.

42) B. D. Wright.　1977b.　Solving measurement problems with the Rasch model.　*Journal of Educational Measurement*, 14, 2, pp. 97−116.

43) J. M. Linacre.　1994.　Sample size and item calibration stability.　*Rasch Measurement Transactions*, 7, 4 in Linacre, J. M. (Ed.), 1996, *Rasch Measurement Transactions, Part 2* (p. 328). Chicago:　MESA Press.

44) J. M. Linacre.　1989−2001.　*A user's guide to FACETS.*　Chicago:　Winsteps.com.

45) 豊田秀樹　1998.　『共分散構造分析[入門編]』　東京：　朝倉書店