【個人研究】

部分得点モデルにおける同時尺度調整法による垂直的等化の研究

藤森　進*

Research on vertical equalization by concurrent calibration in Fujimori's partial test score model

Susumu FUJIMORI

By using concurrent calibration in the polytomous item response model, this study examines vertical equalization of the averages of two populations through simulation. Simulated polytomous test data was generated using Fujimori's partial test score model with vertical equalization, on the assumption that individual tests were conducted for two populations having a significantly different ability (some anchor polytomous items are included in both tests). The population average and variance were estimated using a method based on Mislevy's research (Mislevy, 1984). Unfortunately, the concurrent calibration in this study showed poor results. For reproduction of difference between the averages of two populations under conditions in which concurrent calibration was used, it was found that the results would improve proportionally to the increase in the number of anchor polytomous items.

Key words: item resoponse model, vertical equating, concurrent calibration, simulation, partial test score model
項目反応モデル、垂直的等化、同時尺度調整法、シミュレーション、部分得点モデル

1　研究の目的

共通項目を持つ複数のテストの得点を、ある共通尺度上に位置付けることは、等化と呼ばれる。近年、テストの等化には、項目反応理論が利用されることが一般的である。同理論では、項目は正答を1、誤答を0とする２値のみをとることが仮定されており、2母数ロジスティックモデルが代表的なものとして知られている。多値の得点を扱うモデルとしては、Samejima(1969)の段階反応モデルgradedresponse modelや評定尺度モデル(rating scale model;Andrich,1978)partial credit model (Masters,1982)などが代表的なものとして知られている。また藤森の部分得点モデル(partial test score model;藤森2002a,2002b)は簡明なモデルであり、実用的なもので実際のテスト結果への適用も行われているものである。部分得点モデルと他の代表的なモデルとの比較については、藤森(2002c)を参照のこと。多値の得点を扱うこれらのモデルは項目反応モデルの一群であるが、テスト得点の分析で本格的実務での運用を目指す場合には、これらのモデルを利用したときの等化が不可欠である。2値テスト得点の等化に関しては、ある程度の結論が得られているように思われるが、多値のモデルを利用した場合の等化については十分わかっている状況ではない。そこで、本研究では、部分得点モデルの垂直的等化が旨く行える基礎的条件(例えば等化情報を持つテストの項目数や人数など)をシミュレーションにより検討することにする。

* ふじもり　すすむ　文教大学人間科学部心理学科

2　方法

藤森の部分得点モデルについて簡単に紹介するため、初めに代表的な項目反応モデルである2母数ロジスティックモデルについて説明し、その後部分得点モデルについて述べる。

2.1.　項目反応モデル

2.1.1.2　母数ロジスティックモデル

項目反応モデルに属するものは数多くあるが、(1)式の2母数ロジスティックモデル(Birnbaum,1968)は、その代表的なものである。

$$P(\theta)=P\left(x_{ij}=1\middle|\theta_i,a_j,b_j\right)=\frac{1}{1+\exp\left(-Da_j\left(\theta_i-b_j\right)\right)} \quad (1)$$

ここで i は受験者、θ はその能力を表す母数、D=1.7の定数、j は項目番号、a_jはその識別力、b_jは困難度を表すモデルの母数である。またx_{ij}は、被験者 i の項目 j に対する正誤を表し、正答のとき1、誤答のとき0となるダミー変数である。

2.1.2.　部分得点モデル

藤森(2002)の部分得点モデルを以下で簡単に紹介する。まず受験者のテスト項目 j の得点が多値の得点 r_jによって表現されることを仮定する。部分的な得点を取り得る各テスト項目 j は、各項目に固有であり潜在的な2値の正誤問題の合計得点(ただし、どの範囲の部分得点も扱えるようにするため平均化して０から１の範囲の部分得点になるようにする。採点の実務上は最低点を０、最高点を１となるようにする。即ち、採点された各得点を配点上の満点で割ればよい。)であることを仮定する。受験者は、テスト項目jの部分得点を、この潜在的な2値の項目に反応しながら得ることになる。

またその潜在的な2値の正誤反応は(1)式の2母数ロジスティックモデルが当てはまることを仮定し、しかもその母数は全て同一であることを仮定する。実際にはこの同一母数の仮定は、類似母数を持つ項目であれば近似的に成立するため、過度にこの仮定を重んじる必要はない。

(1)式の P_j（θ）は、受験者が正答１又は誤答０のいずれか一方の潜在的反応を取り得る潜在的問題 k に正答する確率である。多値項目 j は2値の潜在項目を s_j 回繰り返し受験したときに、受験者が潜在的に取り得る正誤反応パターンに基づいて生じると考える。項目 j に関する尤度は潜在項目の母数が同一であるため

$$L(\theta)=P_1^{x_1}\left(1-P_1\right)^{1-x_1}\times\cdots\times P_s^{x_s}\left(1-P_s\right)^{1-x_s}$$
$$=\left\{P^{\sum_{k=1}^{s}x_k}\left(1-P\right)^{\sum_{k=1}^{s}\left(1-x_k\right)}\right\} \quad (2)$$

となる。潜在的正誤得点の平均を考えるためには(2)式のs乗根をとれば

$$\sqrt[s]{P_1^{x_1}\left(1-P_1\right)^{1-x_1}\times\mathrm{L}\times P_s^{x_s}\left(1-P_s\right)^{1-x_s}}$$
$$=\left\{P^{\frac{\sum_{k=1}^{s}x_k}{s}}\left(1-P\right)^{1-\frac{\sum_{k=1}^{s}x_k}{s}}\right\} \quad (3)$$

となる。(2)(3)式の最尤推定値は一致するので、2値の和得点でなく、平均化した0から1の部分得点 r_j による分析が可能となる。

$$Q_j(\theta)=1-P_j(\theta) \quad (4)$$

$$\frac{\sum_{i=1}^{s}X_i}{s_j}=r_j \quad (5)$$

として全て同一の母数であることを考慮して(3)式の対数をとって、

$$\ell_{j,part}(\theta)=s_j\left(r_j\ln\left(P_j(\theta)\right)+\left(1-r_j\right)\ln\left(Q_j(\theta)\right)\right) \quad (6)$$

となる。即ちテスト全体でθを推定するための尤度関数は

$$\ell_{part}(\theta)=\sum_{j=1}^{n}s_j\left(r_j\ln\left(P_j(\theta)\right)+\left(1-r_j\right)\ln\left(Q_j(\theta)\right)\right) \quad (7)$$

によって表される対数尤度ℓ_{part}（θ）を用いて受験者の能力θが推定される。ここで n はテストの項目数である。

これは一般の2母数ロジスティックモデルの対数尤度

$$\ell(\theta)=\log\left(\prod_{j=1}^{n}P_j^{x_j}Q_j^{1-x_j}\right)=\sum_{j=1}^{n}\left\{x_j\log P+\left(1-x_j\right)\log Q_j\right\} \quad (8)$$

の各項目 j の尤度を(6)式で置き換えたものになる。

ここで注意すべきは、観測可能なものは、受験者が問題 j に対して獲得する０から１までの間の値を取り得る部分得点 r_j であり、潜在的問題に対する受験者の潜在的な2値反応は観測できないという点である。また、同じ母数でなくても類似

母数であるような潜在項目の場合も能力推定は近似的に一致することが示せる。

2.1.3. モデル母数と繰り返し数Sの推定

項目反応理論では、母数の推定を最尤法あるいはベイズ法によるのが一般的である。本研究では、モデル母数を推定するため

$$\ell_{part}(\theta)\varphi(\theta)=\varphi(\theta)\sum_{j=1}^{n}s_j\left(r_j\ln\left(P_j(\theta)\right)+\left(1-r_j\right)\ln\left(Q_j(\theta)\right)\right) \quad (9)$$

を最大とするような能力母数θと(潜在項目の正誤反応を2母数ロジスティックモデルとしたため)項目困難度bと識別力aを項目母数の推定値とする交互同時推定法を利用した。ここで$\varphi(\theta)$は能力母数の事前分布であり、本研究では正規分布を仮定している。なお交互同時推定は問題があることが知られている。能力母数を尤度関数から積分により除外して得られる周辺尤度による項目母数の推定が適当とされるが、例えばEMアルゴリズムによる方法などがある。本研究では、藤森の部分得点モデルに関する尤度の周辺化による項目母数の推定プログラムが準備できなかったため同時推定法を利用した。

なお母数の推定は自作のfortranで行い、繰り返し数の推定は自作のpascalプログラム(delphi6)によった。

2.1.4. 繰り返し数Sの推定

テストが実施された集団の能力分布$\varphi(\theta)$を正規分布と仮定した上で、s_j回の正誤の和の生じる確率を２項分布で表し、正規分布と仮定された能力分布との積

$$f(\theta)\Pr(X=r)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(\theta-\mu)^2}{2\sigma^2}\right){}_sC_r\left\{P^xQ^{1-x}\right\} \quad (10)$$

をθで積分して部分得点の周辺分布を求め、これが実際のデータの部分得点の経験分布と最もよく一致するsを求めるという方法で潜在的な問題の繰り返し回数であるs_jを推定している。(10)式中のpは潜在的な問題の2母数ロジスティックモデルによる反応確率である。s_jは項目jに関係するだけなので、(9)式を項目jの母数に関して最大化する際には無視しても差し支えない。このためs=1として(9)式で最尤推定を行い項目母数の推定値を定め、項目母数を所与として(9)式で能力母数の推定値を求め、これらを所与として(10)式を利用してsの推定値を得ている。この推定に関しては、藤森(2002b)によりシミュレーションによる検討が行われているが、その推定成績は良いものであった。

2.1.5. 等化方法

本研究では、初めに述べたように受験者集団の能力分布に差がある場合のテスト得点の等化、即ち垂直的等化が必要となる場面を問題としている。先に述べたように、本研究で採用した推定方法は、項目母数所与としたときの能力母数の推定と能力母数所与とした時の項目母数の推定の交互同時推定であり、この過程で得られた能力母数の推定値を利用して、テストを受験した受験者集団の平均、標準偏差について推定することも考えられるが、藤森(1998)などの2母数ロジスティックモデルでの同時尺度調整法での垂直的等化の分析結果を見る限り、良好な成績を示すとは期待し難い。このため、本研究でも、藤森(1998)と同様にMislevy(1984)と同様の方法を集団分布の母数の推定で採用することにした。すなわちN人のデータが得られたとき、能力母数θで積分することによって周辺化された尤度を最大にするμ,σを推定値とする方法である。

$$L\left(\mu,\sigma|r_1,r_2,\mathrm{L}\ ,r_N\right)=\prod_{i=1}^{N}\left\{\int_{-\infty}^{\infty}f\left(r_{ij}|\theta\right)\varphi\left(\theta|\mu,\sigma\right)d\theta\right\} \quad (11)$$

ここでiは被験者、nはテストの項目数、r_{ij}は、被験者の各項目での部分得点であり

$$f\left(r_{ij}|\theta\right)=\prod_{j=1}^{n}P_j^{r_{ij}}\left(1-P_j\right)^{1-r_{ij}} \quad (12)$$

である。2.1.3、2.1.4、2.1.5に述べた推定方法の組み合わせで、能力母数の推定、項目母数の推定、母集団分布の推定のサイクルを繰り返すことで、能力母数の推定にベイズの事前分布の情報が反映されるため、結果的に項目母数と能力母数の同時尺度調整的垂直的等化が行えると期待して、本研究は行われた。ただし、この過程には繰り返し数の推定は含んでいない。繰り返し数の推定はpascalであり、モデルの母数や母集団分布の推定はfortranであったため1つのソフトウェアとして統合する余裕がなかったためである。今後の課題である。

2.2.　シミュレーションデータ

等化成績の検証のためのシミュレーションデータは、以下のようにして作成した。今回は部分得点モデルの水平的等化および垂直的等化の基礎的条件を検討する研究目的であるので、テスト及び集団数は2ないし3とする。テスト数が4以上となるケースは機会を改めて検討したい。データA1は、集団数が2で被験者数はいずれも3000人で、集団1の能力分布は、平均0、分散1の正規分布に従っている。また集団1は潜在的正誤項目として２値データ換算で40項目のテストを受験するとした。40項目の識別力母数は、平均0.8、標準偏差0.25、下限0.25、上限2.0の切断正規分布に従って作成された。困難度母数は平均0、標準偏差0.5の正規分布に従って作成され、これらの母数から(1)式の正答確率を求め、0から1の一様乱数と比較して正答確率がより大きいときは、潜在的項目について正答1とし、小さいときは誤答0として2値の正誤データを作成した。部分得点データは40項目の潜在的項目から順に5項目の正誤平均を作成して部分得点データとした。即ち、第1集団は2値40項目のテストであったため8項目の部分得点テストを受けることになる。データA1の集団2は、人数3000であり、集団1と同様の能力分布である。能力分布の平均値に差が無く、水平等化であるが、部分得点で水平等化に問題が生じないことを確認するためのシミュレーションデータである。第2集団に対するテスト項目は、第1集団に対するテストと同様の条件で作成した。

第1テストと第2テストの共通項目は、8項目のうち4項目とした。即ち、テスト1,2を通して考えると12項目の部分得点問題があり、受験者集団1のみが受けるテスト項目が1から4の部分得点項目、5から8の部分得点項目は、受験者集団2も受験する共通項目、9から12の部分得点項目は、受験者集団2のみが受験する。

データA2は、テスト1と2の部分得点共通項目数が3である点だけがデータA1と異なる。

データA3は、テスト1と2の部分得点共通項目数が2である点だけがデータA1と異なる。

データA4は、テスト1と2の部分得点共通項目数が1である点だけがデータA1と異なる。

データBは、第2集団の人数が1000名に少なくなったこと以外は、データAと同様である。

垂直的等化データC1は、第1集団の人数が3000、第2集団も3000であるが、能力母数の平均が第1集団が0.0であるのに対して第2集団が-0.5と、やや低くなっている。共通部分得点項目数は4であり、項目母数や能力母数などの作成方法はデータA1と同様である。

データC2、データC3、データC4は、データC1と共通部分得点項目数がそれぞれ順に3,2,1と減るだけで他の諸点は、データC1と同様である。

データD1は集団数を3とした垂直的等化データである。ターゲットとなるのは、第3集団であり、比較的受験者が多く3000とする。第1集団は標準正規分布であり、第3集団の平均は1.0、SDは1.0であるが、第2集団の平均は0.5、SDは1.0と垂直的等化となっている。第1と第2は共通部分得点項目数として6問、第3と第2も共通部分得点項目数として6問を持つ。第1と第3集団に実施されるテストは共通問題を2問とする。第2集団は、等化情報を得るために第1集団に実施したテストと第3集団に実施したテストから問題を選抜し、共通被験者として実施する場合の検討のため人数は1000とした。1000では、共通受験者としてはやや多いという感じがある。つまり等化のためだけのテスト実施を想定した場合、実用上は500人程度が上限であろうからである。しかし、本研究はシミュレーションであり、十分な成績を示す等化条件は如何なるものであるかを調べるものであり、事前の水平等化の検討である程度目途の出ている1000名規模とした。

なお共通項目数はデータD1では6+2としている。これは集団1のテストと集団2のテストの共通項目数が6、集団1のテストと集団3のテストの共通項目が2ということであり、集団2と集団3の共通項目数は6である。すなわち、このデザインは、集団1と集団3を等化するための集団2を利用するデザインである。なお共通項目は能力的に下位の集団に対するテスト項目から選ばれている。この理由は、教育場面では上位集団に対するテスト項目を下位の集団に実施することが実務上困難なためである。垂直等化では４項目でなく６項目

としたのは、後述するように水平等化条件で４項目は必要と分かったため、それより困難な垂直等化条件では等化項目を増やすことが必要と考えられたためである。

３　結果と考察

項目反応モデルでは尺度の平均と標準偏差の情報に関して自由度があるため本研究では、全ての分析において集団1が標準正規分布となるよう標準化した。シミュレーションの真の分布と一致しているため以下の推定成績を少し割り引いて考える必要があるかもしれない。すなわちシミュレーションデータでは、全て集団1は標準正規分布としたため、第2集団以下の集団平均や標準偏差がシミュレーションの真値近くを再現しているか否かが推定の成否の目安となるが。集団1が真の分布と一致していることは推定成績にポジティブな影響はあるとしても、ネガティブな影響を与えることはないと考えられるからである。まず表 1は水平等化の概要である。表中共通項目数とあるのは、テスト版間の部分得点の共通項目数である。データA1の共通項目数が4とあるのは、集団1に実施したテストと集団2に実施したテストの部分得点の共通項目が4項目という意味である。水平等化であるので集団2も標準正規分布であり平均0.0、標準偏差1.0としてある。表中の推定繰り返し数とは、①能力母数の推定、②項目母数の推定③集団の能力分布の推定を1サイクルとしてこの①から③の推定を何回繰り返したかの回数である。図 1は、データA1の集団2の母集団平均の推定値の繰り返し推定の経過である。図の横軸は繰

表1　水平等化の結果(第2集団の人数3000)

	母集団値	データA1 共通項目数4	データA2 共通項目数3	データA3 共通項目数2	データA4 共通項目数1	推定 繰り返し数
集団2の平均	0.000	0.026	0.076	−0.108	−0.170	10回
集団2のSD	1.000	0.992	1.032	1.075	1.183	
集団2の平均		0.011	0.024	−0.142	−0.301	30回
集団2のSD		1.000	0.987	1.192	1.141	
集団2の平均		−0.005	−0.003	−0.140	−0.115	50回
集団2のSD		0.985	0.970	1.084	1.133	
集団2の平均		0.018	−0.075	0.041	−0.290	100回
集団2のSD		0.947	0.971	1.045	1.087	

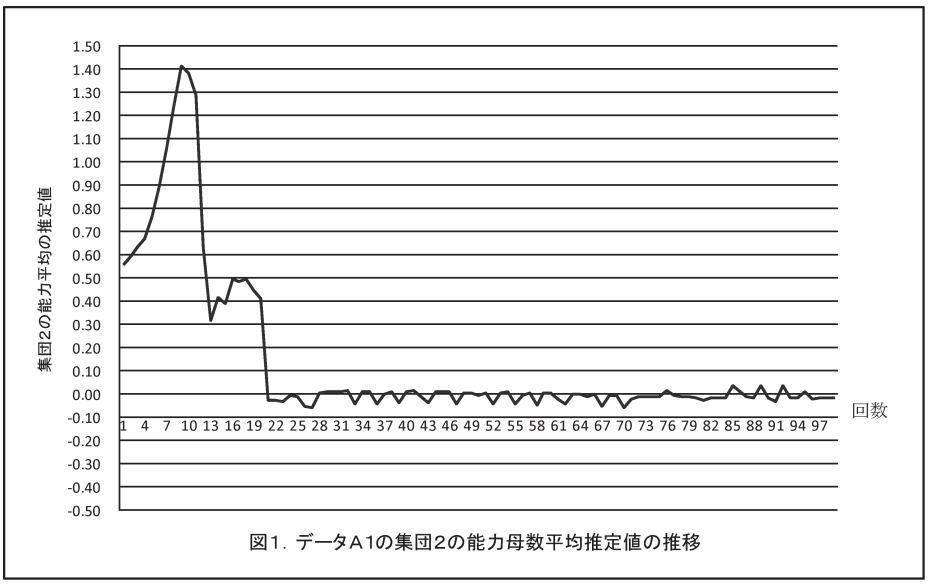

図1. データA1の集団2の能力母数平均推定値の推移

表2　水平等化の結果(第2集団の人数1000)

	母集団値	データB1 共通項目数4	データB2 共通項目数3	データB3 共通項目数2	データB4 共通項目数1	推定 繰り返し数
集団2の平均	0.000	−0.004	0.050	0.115	−0.077	10回
集団2のSD	1.000	1.000	0.945	1.000	1.058	
集団2の平均		0.013	0.063	0.097	−1.021	30回
集団2のSD		1.000	0.949	1.116	1.000	
集団2の平均		0.008	0.007	0.110	−0.018	50回
集団2のSD		1.028	0.968	1.035	0.987	
集団2の平均		0.140	−0.007	0.311	0.046	100回
集団2のSD		1.031	0.961	1.054	1.004	

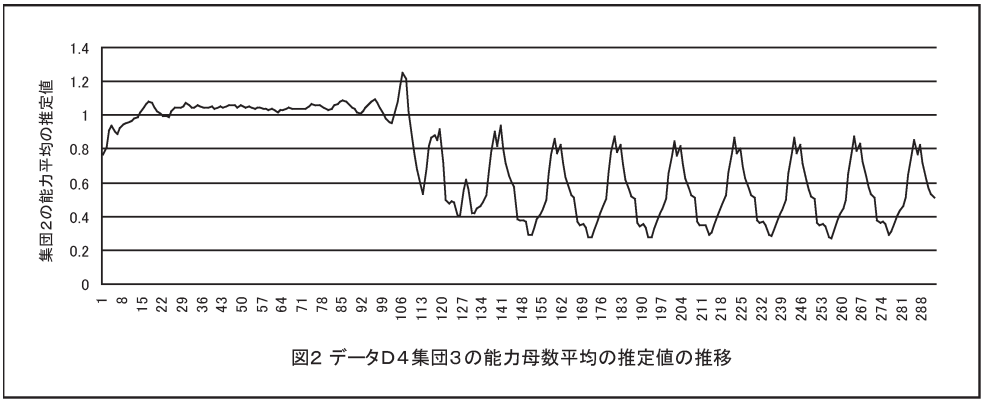

図2 データD4集団3の能力母数平均の推定値の推移

表3　水平等化における能力母数及び項目母数の平均二乗誤差

	データA1 共通項目数4	データA2 共通項目数3	推定 繰り返し数
識別力母数	0.008	0.041	100回
困難度母数	0.003	8.983	
能力母数(集団1)	0.099	3.440	
能力母数(集団2)	0.093	4.050	

り返し数であり、約20回程度の回数で推定結果は安定してくることが分かる。部分得点モデルの垂直等化の推定回数や収束基準をどの程度にするかの目安はないが、後述の垂直等化データD4の結果(図 2)なども含めて検討すると繰り返し数100を超えると推定結果の振動がみられるため、本研究では①から③の繰り返し数を100回までとり、途中10、30、50、100回の結果を検討することとした。表 1のデータA1、A2の第2集団の平均値推定と標準偏差の推定結果をみると、僅か10回の繰り返し推定であっても、第2母集団の平均値は真値の0に近くになっていることが分かる。A3はやや成績が悪く、部分得点の共通項目数が1のデータA4の推定成績は悪い、標準偏差もほぼ同様の傾向であり部分得点による共通項目が少なくなるにつれて推定成績が悪化している。部分得点による共通項目が1項目のデータA4でも第2集団の能力平均値は-0.17であり、水平等化に関しては比較的良好な結果が得られていることが分かる。また同じく表 1より、繰り返し数が増えると50回までは推定成績が順調に向上していることが分かる。しかし100回では必ずしも推定成績は向上していない。

表 2のデータBでは、水平等化に関して人数の影響を検討するため第2集団の人数を1000名と減少させている。このデータでも、3ないし4項目の部分得点共通項目があるときは、第2集団の平均値は0に近く、等化の成績は比較的良好といえるだろう。表 1と表 2を比較すると、人数の減少の影響はそれほど感じられない。等化項目が3,4項目あれば、1000名程度でも水平等化では十分と言えよう。本研究は項目母数の推定も同時に行う、同時尺度調整法による検討なので表 3の項目母数及び能力母数の平均二乗誤差を検討しよ

う。データA1の識別力と困難度、能力母数の平均二乗誤差は、いずれも小さくなっている。2母数ロジスティックモデルの場合の推定成績(藤森,1999)と比較して、遜色ない結果となっている。しかし、部分得点の共通項目数が3のデータA2となると、困難度、能力母数の平均二乗誤差は大きく悪化している。以上を総合すると、水平等化条件での同時尺度調整法による分析では、部分得点の共通項目数は4以上必要ということが結論付けられる。また等化データを担う被験者人数に関しては、検討の余地があるが1000以上と思われる。

続いて集団平均に差がある場合の垂直的等化の検討に移ろう。まず表 4に示した集団数２の垂直的等化では、共通部分得点項目数の減少による等化成績の低下が顕著と言えよう。集団2の能力母数の母集団平均値－0.5と比較して何とか使い物になりそうなのは、共通部分得点項目数が4か3の50回の繰り返し推定結果程度であり、それ以外は実用的な使用に耐えられないと思われる。

では続いて集団数を3とした垂直的等化の結果である(表 5)。集団1の能力は標準正規分布で人数は3000、集団2も同様である。表 1、表 2の水平等化の結果に比較して、データD1からD4まで相当低い成績である。最も共通項目数の多いケースでも垂直的等化の困難さを感じさせる結果である。部分得点の共通項目数が増えるに従って推定成績の向上が認められ、推定の繰り返し回数が50回に向けて増加するにつれて、推定成績は良くなっているが十分とは言えない。推定成績が100回になると成績が悪くなることは水平等化と

表4　垂直等化の結果(集団数2、第2集団の人数3000)

	母集団値	データC1 共通項目数4	データC2 共通項目数3	データC3 共通項目数2	データC4 共通項目数1	推定 繰り返し数
集団2の平均	-0.500	-0.706	-0.575	-0.707	-0.859	10回
集団2のSD	1.000	0.996	0.877	1.001	1.144	
集団2の平均		-0.585	-0.562	-0.738	-0.826	30回
集団2のSD		0.938	1.000	0.951	1.129	
集団2の平均		-0.552	-0.533	-0.760	-0.854	50回
集団2のSD		0.963	0.985	0.955	1.106	
集団2の平均		-0.353	-1.075	-0.614	-1.356	100回
集団2のSD		0.869	1.057	0.719	0.880	

表5　垂直等化の結果(集団数3)

	母集団値	データD1 共通項目数6+2	データD2 共通項目数5+2	データD3 共通項目数4+2	データD4注１ 共通項目数6+2	推定 繰り返し数
集団2の平均	0.500	0.609	0.534	0.694	0.361	10回
集団2のSD	1.000	0.996	0.977	0.989	0.917	
集団2の平均		0.694	1.264	0.810	0.388	30回
集団2のSD		0.982	0.978	0.975	0.945	
集団2の平均		0.702	0.665	0.782	0.400	50回
集団2のSD		0.992	1.000	1.000	0.981	
集団2の平均		0.345	0.302	0.504	0.377	100回
集団2のSD		0.949	0.963	0.928	0.977	
集団3の平均	1.000	1.238	1.264	1.405	0.903	10回
集団3のSD	1.000	0.880	0.978	0.988	1.056	
集団3の平均		1.348	1.357	1.414	1.022	30回
集団3のSD		0.872	0.942	0.954	1.064	
集団3の平均		1.381	1.378	1.436	1.058	50回
集団3のSD		0.929	0.968	0.973	1.084	
集団3の平均		0.845	0.751	0.690	1.066	100回
集団3のSD		0.921	0.938	0.990	1.113	

注1:繰り返し数sの推定あり

表6　垂直等化データD1,D4における
能力母数及び項目母数の平均二乗誤差

	データD4 共通項目数6+2	データD1 共通項目数6+2	推定 繰り返し数
識別力母数	0.041	0.839	
困難度母数	0.609	14.500	
能力母数(集団１)	0.846	10.827	100回
能力母数(集団２)	0.812	14.779	
能力母数(集団３)	0.729	18.875	

同様である。

表 6の項目母数及び能力母数の平均二乗誤差からは、データD4の結果が相対的に推定精度が高いことが分かるが、表 3の成績と比較してみれば分かるように実用の水準に達しているとは言えないであろう。データD4は、データD1と同一のデータに関して、2.1.4の繰り返し数sの推定値を別途計算して分析したものである。データD4の分析は、2.1.4の繰り返し数sの推定値を全て１と固定して、計算している。

以上を総合すると、部分得点モデルの垂直的等化は、実用的な水準に達しているとは言えない。この原因の最大の可能性としては、項目母数の推定における問題点が克服されていないことがあげられるかもしれない。２母数ロジスティックモデルにおける垂直的等化の結果(藤森,1999)との相違は、モデルの違いを除けばこの点だけであるからである。もちろんモデルの違いが垂直的等化の困難さを生みだしていることも否定できないが、今後の検討を待ちたい。2.1.4の繰り返し数sの推定が①から③の垂直等化プログラムに統合されていないことも、小さい可能性として残るかもしれない。また(藤森,1999)の結果と異なり、いずれの結果でも①から③の繰り返し回数の増加が、100回の結果が50回の結果より悪化するなど、必ずしも結果の良化や安定性をもたらさない状況からは、推定結果が安定するようなプログラム上の工夫が必要かもしれないし、適当な時点で①から③の繰り返し推定を終了する収束条件の検討が必要なことが示唆される。

文　献

Andrich,D. A rating formulation for ordered response categories. *Psychometrika*,**43**,561-573.

Birnmaum, A.　1968 Some latent trait models and their use in inferring an examinee's ability. In F.M.Lord & M.R.Novick (Eds.), *Statistical theories of mental test scores* (pp.395-479). Reading,MA:Addison-Wesley.

藤森進　1999　算数・数学学力の到達度水準に関する発達的研究（研究課題番号08610130）　平成８年度～平成１０年度科学研究費補助金（基盤研究（Ｃ）(2)）研究成果報告書.

藤森進　2002a 項目反応理論におけるテストの部分得点の処理方法について　未発表論文

藤森進　2002b 部分得点モデルとその応用　第1回心理測定研究会.

藤森進　2002c 項目反応理論による多値データの分析について－段階反応モデルと部分得点モデル－　文教大学人間科学研究,**24**,21-31.

Masters,G.N. 1992 A Rasch model for partial credit scoring. *Psychometrika*,**47**,2,147-174.

Mislevy,R.J. 1984 Estimating latent distributions. *Psychometrika*,**49**,359-381.

Samejima,F. 1969 Estimation of latent trait ability using a response pattern of graded scores. *Psychometric monograph*,No**17**.

[アブストラクト]

この研究では、多値項目反応モデルに於いて同時尺度調整法による垂直的等化を試み、シミュレーションによって２つの母集団の能力平均の垂直的等化を検討する。藤森の部分得点モデルを利用して、多値テストデータをシミュレーションによって作成し、異なった能力水準の集団が、多値の正答結果を取り得る複数の等化共通項目が含まれている異なったテストを受験し、同時尺度調整法を実施した。母集団平均と分散がMislevyの方法に従って推定された。残念ながら、本研究では、同時尺度調整法による集団間の能力差の再現性は、必ずしも満足いく結果が得られなかった。多値共通項目数の増加に伴って、推定成績は改善傾向が見られた。