

【10】

氏 名 (本 籍)	はつ 服	とり 部	たまき 環	(群馬県)
学 位 の 種 類	教 育 学 博 士			
学 位 記 番 号	博 甲 第 4 1 3 号			
学 位 授 与 年 月 日	昭 和 62 年 3 月 25 日			
学 位 授 与 の 要 件	学 位 規 則 第 5 条 第 1 項 該 当			
審 査 研 究 科	心 理 学 研 究 科			
学 位 論 文 題 目	Rasch モデルに基づくテストの等化			
主 査	筑波大学教授	文学博士	金 子 隆	芳
副 査	筑波大学教授	教育学博士	福 沢 周	亮
副 査	筑波大学教授	教育学博士	松 浦 義	行
副 査	筑波大学助教授		渡 辺 光	雄
副 査	筑波大学助教授		黒 川 哲	宇
副 査	筑波大学助教授	教育学博士	海 保 博	之

論 文 の 要 旨

テストの等化とは、異なる尺度上で定義された母数値を共通の尺度上へ変換することである。本論文は Rasch モデルを用いて、テストの等化にまつわる諸問題の数値実験的解決、ならびに実用に供する等化された語彙理解力テストの作成について報告するものである。

論文全体は、大きく2部に分かれる。前半の1章から6章までは、コンピュータによる数値実験による等化の諸手法についての比較が行なわれ、後半の7章から9章まででは前半の結果をふまえて、Rasch モデルに基づいて実際のテストの作成を試みている。各章の概要は次のとおりである。

1章では、テストの等化のしくみと、等化がなぜ必要であるかについて述べ、本論文の基本的な問題の所在と、その教育評価上の意義とをあわせて提示している。

2章では、Rasch モデルの導出過程と Rasch モデルを含めたロジスティック・モデルの母数推定法、さらに、Rasch モデルの適合度検定について概観している。

3章では、はじめに Jackknife 推定法と Bootstrap 法による母数推定法を提案している。人工データと実際のテスト・データを用いた数値実験を通して推定精度と頑健性について検討し、

Jackknife 推定値と Bootstrap 推定値は、双方とも推定精度は高いものの、通常の最尤推定法を越える推定精度でなかったことを確かめている。また、Bootstrap 推定値の偏りは最尤推定値と Jackknife 推定値の偏りよりも大きく、また、Jackknife 推定値と最尤推定値の偏りはほぼ等しいことも確認している。この結果から、推定精度と計算コストを考慮するなら通常の最尤推定法が最も望ましいことを示唆している。

さらに、2つの外れ値検出法を提案している。人工データによる数値実験を行ない、既存の検出法の併用が必要であることを述べている。

4章では、2組のテストの等化法を回帰法、主軸法、項目特性曲線法、反応パターン法に分類した上で概観し、さらに、Rasch モデルに準拠して作成された2つのテストを共通項目デザインに基づいて、結合定数と共通項目困難度を階層ベイズ推定する方法を提案している。

5章では、テスト・ネットワーク特有の問題を説明したあと、次のようなテストの同時等化法を提案している。すなわち、テスト k と l において、それぞれ推定された共通項目の平近困難度の差異（結合定数 λ_{kl} の実測値）を(1)式のように分解する。

$$\lambda_{kl} = \delta_k - \delta_l + w_{kl}e_{kl} \quad (1)$$

ここで、 δ_k と δ_l はテスト k と l の相対的な困難度、 e_{kl} は等化誤差である。等化に必要なテスト困難度は、 λ_{kl} と w_{kl} を所与として e_{kl} の自乗和を最小にする値として推定する。実データを用いて既存の等化法と比較検討したところ、既存の等化法よりも良好な推定値が得られたとしている。

6章では、必ずしも困難度推定値の精度と偏りが等しくなるとは限らないことから、共通項目の困難度推定値の間に等化モデルを立て、それを困難度推定精度と外れ値を考慮する等化モデルへ展開し、次の4つの等化モデルを提案している。第1の等化誤差平方和最小法は

$$d_{ijk} = d_{ijk}^* + \delta_i - \delta_j + e_{ijk} \quad (2)$$

とおく。ここで、 d_{ijk} と d_{ijk}^* はテスト i と j の共通項目 k がテスト i および j で尺度化されたときの項目困難度、 e_{ijk} は d_{ijk} と d_{ijk}^* の推定誤差の差異である。テスト困難度 δ_i と δ_j の推定値は d_{ijk} と d_{ijk}^* を所与として等化誤差平方和を最小にする値である。第2の誤差分散を考慮する等化法は、モデル式を

$$d_{ijk} = d_{ijk}^* + \delta_i - \delta_j + w_{ijk}e_{ijk} \quad (3)$$

とおき、 d_{ijk} 、 d_{ijk}^* 、 w_{ijk} を所与として、誤差平方和を最小にする δ_i と δ_j を加重最小自乗推定法によって求めるものである。 w_{ijk} は e_{ijk} の標準誤差の関数である。また、第3の等化法は重みとして Tukey の重みを利用して、外れ値の影響を抑えてテスト困難度を推定するものである。第4の等化法は、項目の困難度推定値の誤差分散と外れ値を考慮して、2つの基準関数を組み込んだものである。数値実験によると、外れ値に対しては第4の外れ値を考慮する等化法が最も頑健であるが、外れ値のないときは5章の等化法と6章の4つの等化法はほぼ等しい推定精度

を示していることを確認している。また、頑健性と推定精度に関しては、既存の等化法は本研究のすべての等化法よりも劣っていることも、あわせて確かめている。

7章では、本研究で提案した等化法を用いて中学生の誤彙理解力を測定する項目プールを作成する手順、およびその結果を報告している。まず、中学校の国語科検定教科書のなかから重要語句を選んで5肢択一形式の問題を作成し、204項目を選んで無作為に17シートへ配分して、6組のテストからなるテスト・ネットワークを構成する。中学校1年生と2年生、合わせて717名に実施して、不良項目を除いた上でテストごとに最尤推定法により項目困難度を推定する。さらに、8通りの等化法によりテスト困難度を推定し、6章の等化誤差平方和最小法による推定値を採用して項目困難度を等化し、最終的に169項目の項目プールを作成している。

8章では、中学2年生133名から収集された項目反応を用いて、誤彙理解力検査のコンピュータに支援された調整テスト方式（adaptive testing）の開発をシミュレーション実験を通して試みている。第1実験として、全員に30項目を提示する調整テスト方式をシミュレートする試みを行なっている。正答率、推定値の偏り、調整テストによる能力推定値と160項目による能力推定値との相関係数などを検討し、第2実験は最大提示項目数を30とする条件の下で、能力推定値の標準誤差の大きさから終了基準を設定している。調整テストによる能力推定値と160項目による能力推定値との相関係数は0.873から0.938、また、平均正答率は0.50をピークに分布していた。各事例を検討したところ、やや偏りの大きい事例が数例見つかったので、第3実験では第1段階において易しい項目と難しい項目を含む7項目のテストを実施する2段階の最尤方略を実施している。その結果、偏りが小さくなった事例と、逆に大きくなった事例があり、正答率と提示項目数から通常の最尤方略が望ましいとの判断を下している。第4実験では交互確認のために、項目プールを折半してシミュレーション実験を行なっている。ここでも調整方式の目的が達成されていたことを確認している。

9章では、3段階方略による誤彙理解力検査の調整テストを行なった結果について述べている。Raschモデルでは受験するテストの組み合わせごとに、テスト得点と能力値推定値の対応表を用意しておけば、教室内でも調整テストが可能である。3段階方略は、この特長を利用したものである。中学3年生152名を折半して、一方には3段階テストを、他方には通常のテストを実施している。受験者はテスト得点に応じて第1段階で3つのテストへ、また、第2段階で7つのテストへ分類された。3段階テスト以外のテストによる能力推定値を外的基準とみなして3段階テストの難易度と能力値の関係、さらに、能力推定値の精度を吟味している。9通りのパスの困難度と、そのパスを通過した受験者の能力値は近く、しかも、能力推定値の偏りは概ね等しかった。また、3段階テストは能力推定値の偏りを通常のテストよりも小さく抑えることができた。以上の結果から3段階テストの本来の目的が達成されたとしている。

審 査 の 要 旨

良質の選抜試験への要求が一般と高まるなかで、テストの等化は、テスト理論の側から、その要求に答える一つの有力な方策として関心が寄せられている。この論文では、等化についての過去の手法を整理、かつ、それらの有効性の比較吟味をコンピュータによる数値実験によって包括的に行なっただけでなく、そこでの知見を実際のテスト作成に適用しても、十分な成果が得られることを確認している。

数値実験への比重がやや大きすぎるきらいがあるが、そのような吟味が後半の7, 8, 9章の実用的なテストの作成のところで十分に生かされているので、問題はない。

よって、著者は教育学博士の学位を受けるに十分な資格を有するものと認める。