

## 博士学位申請論文概要

項目特性図作成方法の精緻化による項目分析の新たな  
展開

秋山 隆

## 第1章 問題と目的

第1章では、項目分析 (item analysis) について概説した後に、本稿を通じて扱っていく項目特性図 (item characteristic chart) を用いた項目分析の方法について概説した。

### 項目分析

項目分析はテストを構成し、実施、運用するために欠かすことのできない作業工程である。項目分析の重要性は池田 (1992), 植野・荘島 (2010), 豊田 (2012) において指摘されている。また、項目の正答比率や、選択肢の振る舞いの分析の重要性は、Brennan (2004, p.329, p.340) において指摘されている。項目分析を行うことで、テスト運用者は、当該項目が測るべき事柄を適切に、どれだけの精度で測ることができているのかを確認することが可能となる。

**多肢選択式項目** 多肢選択式項目は、主として細目積み上げ方式 (池田, 1992) のテストにおいて採用される項目形式である。多肢選択式項目は、出題内容について、用意された複数の選択肢の内から正答となる選択肢を回答することを受験者に対して求める。

これ以降、本稿においては多肢選択式の項目を扱うこととする。

### 項目反応の整理

受験者の項目に対する解答を項目反応 (item response) という。得られた項目反応を元に採点、分析を行うため、項目反応の整理を行う。

任意のテストにおける受験者は添え字記号  $i$  を用いて表すこととする。 $i$  は1から  $N$  までの範囲をとる。また、テストを構成する項目数は添え字記号  $j$  を用いて表現される。 $j$  は1から  $J$  までの範囲をとる。項目ごとに受験者が異なりうる場合には、 $N_j$  と表現する。

任意の  $j$  番目の多肢選択式項目の選択肢をカテゴリとも呼び、添え字記号  $c_j$  を用いて表す。 $c_j$  は1から  $C_j$  までの範囲で変化する。

**テスト得点** ここで、正答 = 1, 誤答 = 0 として、任意の受験者の反応パターンをベクトル表記を用いて  $\mathbf{x}_i$  と表現する。例えば  $\mathbf{x}_1 = (1, 1, 1, 1, 1, 1, 1, 1, 1, 0)'$  である。テスト全体の反応パターンは行列表記を用いて  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]'$  と表される。

$\mathbf{X}$  の行に関して、 $x_i = \sum_{j=1}^J w_j x_{ij}$  を計算することで、受験者個人のテスト (和) 得点を算出可能である。 $w_j$  は項目ごとに定められた配点の重みであり、 $x_{ij}$  は  $i$  番目の受験者の  $j$  番目の項目における受験結果を表している。 $w_j$  はテスト

の採点者が任意に定めることが可能であるものの、一般的には  $w_j = 1$  とされ、その場合には  $w_j$  を省略して表記することもある。例えば1番目の受験者のテスト得点は  $x_1 = \sum_{j=1}^{10} x_{1j} = (1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 0) = 9$  となる。

**選択肢の選択率** 多肢選択式項目の受験結果は、選択肢ごとの被選択数によって、受験者にどの程度選ばれたのかを確かめることができる。項目選択肢の選択率は  $p_{jc} = \frac{n_{jc}}{N_j}$  として表される。ここで  $N_j$  は第  $j$  番目の項目における受験人数、 $n_{jc}$  は項目  $j$  の  $c$  番目のカテゴリにおける、選択人数を表している。カテゴリごとの選択率  $p_{jc}$  を参照することで、正答選択肢以外に選ばれやすい誤答選択肢を確認することができる。

選択率を参照する際には、受験者を任意の数の群に分けることが、更なる詳細な分析を行うために便利である。群に分ける際には、受験者をテスト得点順に並べ、低得点群から高得点群となるように分割する。この際、低得点群、中得点群、高得点群となるように受験者を3群に分割する場合や、5群とする等、群の数は任意に定めることが可能である。一般的には5群に分割されることが多い。群ごとの選択率は  $p_{jgc} = \frac{n_{jgc}}{N_{jg}}$  と表される。添え字  $g (= 1, 2, \dots, g, \dots, G)$  は、 $g$  番目の群を表している。つまり、 $n_{jgc}$  は  $j$  番目の項目の、 $g$  番目の群における、カテゴリ  $c$  の選択人数を示している。

**選択率を用いた項目分析** 選択率を用いた選択肢ごとの被選択傾向を表形式で表現し、分析することで、各群で、どの選択肢が選択されやすかったのかを値の大小から分析することが容易となる。特に魅力のある選択肢に注目し、出題文と特に魅力的であった誤答選択肢について誤答分析を行うことで、誤答者への教科教育を再検討することが可能となる。

## 項目特性図を用いた項目分析

テストを構成している設問項目の性質を評価する際に、当該設問項目における受験者の正答率（通過率）を用いて項目の難しさを調べるための有用な道具として項目特性図（豊田, 2012）を挙げることができる。

**項目特性図の作成方法** 本項で項目特性図の作図方法を述べる。まずテストの受験結果を整理し、受験者のテスト得点を算出する。算出したテスト得点を昇順に並べ、それらを任意の  $G$  個の群（赤根・伊藤・林・椎名・大澤・柳井・田栗（2006）では  $G = 5$  としている）に分割する。次に各群の、各項目の正答率を計算する。最後に群を横軸に、確率を縦軸にとったグラフに正答率をプロットし、各プロットを直線で結ぶ。

誤答選択肢についても興味がある場合は、更に誤答選択肢の選択率についても、正答率と同様にプロットし、プロット間を線で結ぶ。なお、項目特性図作

図における受験者の群への分割方法では、各群の所属人数を等人数とする場合と、各群間の得点間隔が等しくなるように受験者を振り分ける場合の2通りが主要な方法として存在する（豊田, 2012, p.4）。

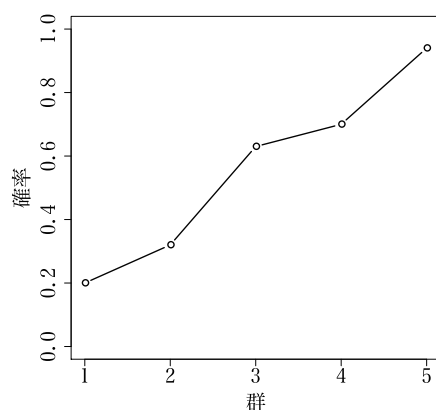


図 1 2 値で整理した場合の項目特性図の例

**項目特性図の解釈例** 図 1 に項目反応を 2 値データとして整理した場合の、項目特性図を例示した。群内における全選択肢の選択率の和が 1 となる制約から、2 値反応の場合の誤答に関する項目特性曲線は、正答選択肢の項目特性曲線と対称の関係となるため、描画が省略されることが一般的である。図 1 では、テスト得点最下位群においては 2 割程度が正答できており、より上位の群になるにつれ正答率が上昇し、最上位群においてはほとんどの受験者が正答していたことが、また第 3 群から第 4 群にかけては他の群間に比べて正答率のグラフの上昇が鈍化傾向にあることがわかる。

**正答分析** 項目特性図の利用方法として、正答分析に用いる方法が挙げられる。項目特性図を観察することで、群ごとに正答選択肢の選択傾向を確認することが可能となる。テスト得点が高い受験者群になるほど、正答選択肢の選択率が高くならなければ、当該テスト項目は何らかの問題を有しているものと解釈することが可能である。

**誤答分析** 項目特性図は誤答分析に用いることも可能である。この場合には、正答選択肢以外の選択肢、すなわち誤答選択肢の選択率もプロットし、プロット間を直線をつなぐ。誤答分析を行うことで、どの得点群の受験者がどの誤答選択肢に魅力を感じ、選択したのかを確認し、教科教育を改善する手立てとすることが可能となる。

## 項目特性図作成時の問題点

項目特性図は非常に有用な項目分析のための道具ではあるが、その作成方法については、注意すべき点が3つ存在する。

**群数決定方法の恣意性** 項目特性図の作成時に受験者を群に分ける際、群数  $G$  の値を決定するための明確な基準は知られていない。現状では分析者によって経験的に決定されており、一般的には  $G = 5$  とされるが、その根拠は明確には示されていない。

**項目特性図の視覚的特徴と解釈の検証可能性** 選ばれ易かった誤答選択肢が存在した場合、その理由を考察することで、受験者の更なる学習の発展に繋げることが期待できる。しかし、選択肢数が増えると、特性曲線は互いに重なり合い、視認性が失われてしまう。このとき、同様の方略、知識によって導かれる複数の誤答選択肢が存在している場合に、それらをまとめることが考えられる。この状況において、その妥当性が確認可能であれば、教科内容についての更なる理解の促進が期待できるであろう。しかし、作成した項目特性図を解釈することは可能であっても、そこから得られた知見に基づいて仮説を構成し、それについて検証する方法は未だ確立されていない。

**同一項目に関する重複した項目特性図の解釈の恣意性** テストの実施方法の1つとして、ブックレット方式による項目の出題が挙げられる。ブックレット方式は多くの受験者を対象とする場合、同一日時、場所でテストを実施することが困難であるような状況において、有用な実施方法である。項目群は複数のブックレットに分割され、各ブックレットが、各受験者集団に割り当てられることになる。このとき、後の分析において比較可能となるように、ブックレット間で共通の項目を含ませしておく。つまり、共通項目については含まれるブックレットの数だけ項目特性図を作成する<sup>1</sup>ことが可能となる。

例え同一項目であっても、項目特性図の表現はブックレット間で異なる場合がある。このとき、共通項目の項目特性を表現する項目特性図として、何れを採用して項目分析を行うべきか、基準は広く知られていない。

本論文では、これらを改善するための方法を提案した。

## 項目特性図の作成方法における基準の導入

本研究では統計モデルを選択するために用いられる、情報量規準 (information criterion) と呼ばれるモデル評価規準を用いることとした。情報量規準を用いた選択を行うことで、群数選択基準や、項目特性図の統合基準、同一項目にお

<sup>1</sup>和得点としてブックレット得点を用いる。

る複数の項目特性図の選択といった判断をモデル選択の観点から行うことができるようになる。本論文では赤池情報量規準とベイズ情報量規準を用いてモデル選択を行う。

**赤池情報量規準** 我々はある観測データが得られた場合に、そのデータがどのような構造の下で発生したのか、ということを考える。しかしながら、多くの分野において、真の構造は未知である。データが得られたモデルを完全に再現しようとする、多くのパラメタを想定しなければならなくなる。しかしそうして構成したモデルは、手元のデータに対してのみ、有効なモデルとなってしまう。赤池情報量規準 (Akaike's Information Criterion, AIC; Akaike, 1973) を用いることで、手元のデータにおいて正しいモデルよりも、より予測精度のよいモデルを選ぶことが可能となる (例えば赤池・甘利・北川・樺島・下平, 2007, p.62)。

**ベイズ情報量規準** Schwarz(1978) はベイズ統計学の立場からベイズ情報量規準 (Bayesian Information Criterion, BIC; Schwarz, 1978) を提案した。BIC も AIC と同様に相対的に値の小さいモデルをよりよいモデルとして選択可能である。

## 第2章 研究 I: 項目特性図の群分け基準

第1章で述べたとおり、項目特性図の作成時には受験者を得点に基づいていくつかの群へと分ける作業が求められる。この際、群数は任意に設定可能である。受験者数が比較的少数である場合には、群数の変更によって、項目特性図の表現が大きく変化する。

群数  $G$  の値を決定するための明確な基準は知られておらず、現状では分析者によって経験的に決定されており、一般的には  $G = 5$  とされるが、その根拠は明確には示されていない。項目特性図の作図の際に、群の分割数の選択について、統計的な基準や根拠を与えることができれば、項目特性図を用いて項目の性質を調べる上で便利である。

### 研究 I 目的

本研究では情報量規準 (AIC, BIC) を利用した項目特性図の群数選択法を提案する。本研究における群数選択法には二つの利点が挙げられる。一点目は、項目特性図を描く際に、受験者の多寡に応じて、安定した特性の表現を行うために利用可能であるという点である。

分割する群の数が少なすぎると正答確率の変化に関する直線の表現が大味になり過ぎ、大域的考察のみが可能となる。例えば2群に分割した場合、項目特性図は1本の直線で表されることになり、当該直線の傾きのみが考察の対象となる。受験者が多い場合には、群数を多くしてもグラフ上の表現は安定するため、項目の性質を細やかに考察することが可能となる。一方で、受験者の全人数が少数である場合は、群数を増やそうとすると各群に属する受験者数が少なくなるために、グラフ上の表現が安定しなくなってしまう。この場合、無理に群数を増やそうとはせず、むしろ減らすことによって、安定した項目の性質を考察することが期待できる。

二点目は、近年我が国において広く用いられている項目反応理論 (item response theory, IRT; Lord, 1980) が適用できない項目に関しても項目特性図は作図可能であり、その特性表現が有する情報は、決してIRTにおけるICCと比較しても見劣りがするものではない、という点である。

### 情報量規準を用いた群数選択法の提案 (方法 I)

ある項目  $j$  における第  $g$  群に配分された受験者の、正誤反応パターンベクトル  $\mathbf{x}_{jg}$  を観測する確率は  $p(\mathbf{x}_{jg}|p_{jg})$  となる。 $p_{jg}$  の最尤推定量を回答人数と正答人数を用いて算出することで、最大対数尤度  $\log l_{\max}$  を導出する。最大対数尤度を用いることで、ある項目  $j$  について、受験者を任意の  $G$  個の群に分割した場合の、情報量規準  $AIC_{j(G)} = -2 \times \log l_{\max} + 2 \times G$ ,  $BIC_{j(G)} = -2 \times \log l_{\max} + G \times \log(N_j)$  を算出可能である。ここで  $N_j = \sum_{g=1}^G N_{jg}$  である。

テスト項目全体で同じ群数を用いて項目特性図を描く場合、テスト全体のAICとBICはそれぞれ  $AIC_{(G)} = \sum_{j=1}^J AIC_{j(G)}$ ,  $BIC_{(G)} = \sum_{j=1}^J BIC_{j(G)}$  で算出する。ただし、項目間の独立性が本質的に否定される場合には上式は成立しないことに注意する必要がある。

### 研究 I 結論

提案手法について、群数と人数条件を変化させた場合のシミュレーションと実データへの適用を行い、手法の有用性に関して検討を行った。

これらの適用を通じて、情報量規準を用いた項目特性図の群数選択法は、一般に本試験よりも受験者数が遥かに少なくなる予備試験における項目分析において特に有用であると結論づけられた。また、群数の変更によって項目特性図の表現が変わり、項目特性の考察に影響し得る場合にも、情報量規準を傍証として、群数を選択することが可能である。

項目特性図の作成において受験者を5群に分けるという経験的判断は、これまでの実務場面において効果を上げている。特に受験者が非常に多いテストに

関しては情報量規準を用いた群数選択法の射程にはなく、経験的に5群に分割することで十分に項目特性考察の目的を達成可能である。

本提案手法はテストにおける受験者数が少ない場合に特に効果を発揮する。例えば受験者数200名程度の予備試験と5000名程度の本試験があるような場合、予備試験の項目特性図に関しては提案手法を用いて作成することが有効であろう。本手法は、項目特性図作成における分割数の決定に傍証を与えることが可能となる。特に同一項目において分割数ごとに表現される項目の特性が大きく異なるような場合には、分割数を選択する際の指針とすることが可能である。

### 第3章 研究II: 誤答分析における項目特性図の作成方法

本章では誤答分析時に、項目特性図へ分析者の仮説を反映し、仮説を比較する方法に関して提案した。

#### 研究II 目的

誤答傾向に興味がある場合は、項目特性図上に正答曲線に加えて誤答曲線も描き、誤答分析を行うことが可能である。誤答分析も正答分析と同様に重要な分析であり(吉村, 2009; 豊田, 2012), 正答分析によって項目の測定性能が担保された後は、誤答分析を行うことで、より詳細に、各選択肢と具体的な受験者群の特性を分析することが可能である。受験者の反応が複数のカテゴリ反応として整理される項目の場合、受験者がどの部分に躓き易かったのか、ということが誤答選択肢の選択確率という形で明示化される。赤根ら(2006)は選択肢別の選択率が、誤答が多い場合に受験者の理解度や項目の適否を検討するために有用であることを指摘している。選択率を視覚化した項目特性図もまた、項目の性質や、受験者の傾向を知るために有用な道具となる。選ばれ易かった誤答選択肢が存在した場合、何故、当該誤答選択肢が選ばれたのか、受験者は如何なる誤った知識を適用してしまったのかを考察することで、テスト後の受験者の誤った知識の修正、更なる学習の発展に繋げていくことが期待できる。しかしながら、作成した項目特性図を解釈することは可能であっても、そこから得られた知見に基づいて仮説を構成し、それについて検証する方法は未だ確立されていない。

例えば同様の方略、知識によって導かれる複数の誤答選択肢が存在している場合に、それらをまとめることが考えられる。あるいは、学習の特定の段階において誤用される方略、知識によって導かれる誤答に焦点を当てるため、それ以外の部分をまとめたい場合も想定可能である。これらの状況において、その妥当性が確認可能であれば、教科内容についての更なる理解の促進が期待できるであろう。本章において提案する手法を用いることで、分析者が項目特性図



を考察した結果得られた仮説を検証し、なおかつ図の視認性を確保することが可能となる。

## 項目特性図を用いた誤答分析の精緻化の提案（方法Ⅱ）

誤答選択肢の特性曲線をまとめるヒューリスティックについて記述する。まず、統合前のオリジナルの項目特性図を作図する。次に、描かれた折れ線から解釈される、出題内容やワーディングの特徴の可能性を見出す。各誤答選択肢の特徴に加え、教科内容的な側面からも検討し、誤答選択肢をまとめる。

最後に、オリジナルの項目特性図やまとめ方の異なる複数の項目特性図に関して、見出された特徴の違いが標本誤差によるものか、それ以上のものであるのかを本章において提案した情報量規準算出方法に基づいて、情報量規準を算出し、モデル比較を行う。

## 研究Ⅱ 結論

実データへの適用例とシミュレーションを通じて、提案手法が誤答分析において、併合可能であると判断可能なモデルを適切に推奨し得ることが確認された。また BIC は保守的な傾向にあることも同時に示唆された。AIC と BIC は必ずしも同一の項目特性図を推奨するわけではないことが適用例において確認された。このことは情報量規準を用いたモデル選択の限界として不可避の事態であり留意する必要があるが、今回の適用例においては例え異なるモデルが推奨されたとしても、解釈可能な範囲内であり、手法の有用性に大きく影響を与えるものではないと考えられる。

提案手法を用いることで、描画する特性曲線を選択するのではなく、すべての選択肢について考慮した状態で特性曲線の本数を減らしつつ、より特徴的な誤答選択肢を抽出して項目特性図による誤答分析を行うことが可能となる。また、モデルを情報量規準を用いて比較検討することで、併合の基となった仮説についても規準に基づいて検討することが可能である。結果として整理された状況で、細やかな項目分析、誤答分析が可能となり、受験者、特に低得点者への効果的な学習指導へと繋げることが期待できる。

## 第4章 ベイズ統計学とモンテカルロサンプリング

本章では、研究Ⅲにおいて提案する手法で用いられる、ベイジアン階層モデリングとマルコフ連鎖モンテカルロ法への理解を深めるため、ベイズ統計学およびマルコフ連鎖モンテカルロ法のアルゴリズムについて概説した。

## 第5章 研究Ⅲ: 同一項目に複数の項目特性図が作成可能な場合の分析方法

### 研究Ⅲ 目的

テスト実施会場が広い地域に分かれている場合や、多数の受験者数が見込まれるために、同一時刻、会場でのテストの実施が困難となる場合には、ブックレット方式によってテストが実施されることがある。ブックレット方式とは、同一の特性を測定するための項目群を複数の小冊子（ブックレット）に分割して配置し、複数の受験者群にブックレットを配布し、項目に対して回答を求める形式である。項目内容の漏洩からテストを守るために、ブックレットごとの構成項目異なる場合が一般的である。つまり、配布されるブックレットが異なる受験者は異なる項目集合に対して回答を行う。

これら構成項目が異なり得るブックレット間の受験者群の特性（能力）を比較可能とする方法の1つは、各ブックレットに共通の項目を含ませる共通項目法である。共通項目法では複数のブックレット内に、共通の項目を配置することとなるが、項目分析を行う場合に、ブックレットに対応した数の分析指標が算出される。これは項目特性図も例外ではなく、共通項目についてはブックレット数分の項目特性図が描画可能となる。

同一項目の項目特性図であっても、受験者群が異なれば、特性曲線の表現も異なるものとなり得る。共通項目が含まれるブックレットが多い場合には、1つの項目についてすべてのブックレットの表現を確かめることは煩雑である。更に、ブックレット間で共通項目の項目特性図の表現が大きく異なるような場合には、項目特性の安定性が疑われることとなり、安定性それ自体を吟味することも求められるであろう。この場合に、どのブックレットの表現を項目分析の対象として採用すべきか、明確な基準は広く知られてはいない。

理想的な状況では、同程度の特性を有する複数の受験者集団に対して、共通項目を含んだブックレットを用いてテストを実施すると、例えブックレットが異なっても、ブックレットが割り当てられた受験者群間の背後に仮定される特性分布が互いに同一であるならば、受験者数が増加するにしたがい、項目特性図の表現も互いに似通うようになるはずである。すなわち、この場合にはブックレットごとの、項目特性図の表現の違いは標本誤差のみによるものといえることができる。

しかしながら、実際の状況ではブックレットの横軸、つまり群分けはブックレット得点によって定められ、ブックレット得点は、ブックレット間で互いに異なり得る（項目特性図作成対象となっている当該項目以外の）他の項目群に対する正誤反応およびその背後に想定される項目特性によって定められる。よって、項目特性図の横軸にはそれぞれ異なる項目特性の影響が含まれていると考えられるために、各ブックレットごとの項目特性図の表現の異なりは、純粋に

標本誤差のみの影響とは見なすことが難しい状況にある。

上記の理由から、ブックレットごとの項目特性図の表現は互いに完全に異なったものを表しているとは一概に見なすことは難しい。一方で、ブックレットごとの項目特性図における表現は、標本誤差以外はすべて同じであると見なすこともまた困難である。研究Ⅲでは共通項目について複数の項目特性図が作成され得る場合に、異なる受験者集団に対する、ブックレット別項目特性図の表現の違いを加味しつつ、同時に統一的な観点から項目特性図を作成し、項目分析に用いる方法を提案する。

### 階層ベイズモデルを用いた統一的な項目特性図作成法の提案（方法Ⅲ）

同程度の特性を有する複数の受験者集団に対して、共通項目を含んだブックレットを用いてテストを実施すると、例えブックレットが異なったとしても、同じ特性分布となるという仮定の下、階層ベイズによる群ごとの2値反応データを用いたモデル化を行い、超パラメタを用いた解釈方法について提案を行う。なお、本研究では正答分析に焦点を当てることとした。

本研究では、前述の階層ベイズモデルについて、ハイブリッドモンテカルロ（HMC; Duane, Kennedy, Pendleton, & Roweth, 1987）法によるパラメタ推定を行った。HMC法の実行にはソフトウェア Stan (Stan Development Team, 2014b) および統計解析環境 R (Ihaka & Gentleman, 1996) 上における Stan のインターフェースパッケージ RStan (Stan Development Team, 2014a) を用いた。

連鎖の構成数は4とした。また、各連鎖の更新期間は2200回とし、このうち初めの200回をバーンイン (burn-in) 期間として破棄した。なお標本の間引きは行わないものとした。以上の設定より、各連鎖におけるバーンイン期間後のサンプル8000 ( $4 \times 2000$ ) 個を用いて事後統計量を構成し、これらを元に考察を行った。

### 研究Ⅲ 結論

本研究Ⅲでは、複数ブックレット間に含まれる共通項目について、ブックレット間の項目特性図の表現の違いを加味した項目分析を可能とする方法について提案した。階層ベイズモデリングを用いた分析によって、観念的な項目特性図を作成し、項目分析の対象とすることで、ブックレットごとの項目特性図の比較分析のみからは得ることのできない知見を得られることが示唆された。本提案手法は、共通項目の項目特性図を用いた項目分析のために有効な方法となるであろう。

## 第6章 総合考察

研究ⅠからⅢを通じて、多肢選択形式項目の項目特性図作成方法に関して、3つの観点から精緻化を行う方法を論じた。

研究Ⅰでは、項目特性図の群数選択に関する統計的な基準を提案した。項目分析を視覚的に行うことが可能となる項目特性図は、統計分析を専門的に行う者のみならず、必ずしも統計分析に通暁しているわけではない、各科目の専門家、項目作成者にとっても、重要な道具となる。

一般的には5群分割で作成されてきており、有効に活用することができている。一方で、項目特性図による特性表現は必ずしも一意に定まるものではなく、受験者数と受験者の群数次第で、表現は変化し得る。しかしながら、何れの表現で正答分析を行うべきかの統計的な基準は広く知られておらず、ある表現での解釈を決定した際に、何故そのような判断を下したのかは、経験的に説明することしかできなかった。

受験者数が少ない場合には、群数が多いとき、項目特性図の表現は不安定になるため、本テストに項目特性を適切に反映することができていない可能性が残る。この場合には群数を減らし、各群の所属人数を増やすことで安定性を増やすことが期待できるが、この場合にもどのように群数を選択すべきかの基準は知られていない。

研究Ⅰによる方法は、項目特性図を用いた正答分析を行う際の群数選択の一助となることが期待される。特性表現を安定させつつ、全体データの傾向を表現するための群数を決定する際の傍証として用いることが可能である。

研究Ⅱでは誤答分析のための項目特性図精緻化の方法を提案した。正答分析により、基本的な測定性能が保障された項目については、誤答分析を行うことで、受験者の教科教育を行うための知見を得ることができる。

一方で、誤答選択肢の特性曲線は、互いに似通いやすいという性質があり、図による視覚的な把握性が阻害され得るという問題が存在する。また、誤答分析によって解釈された誤答傾向が標本誤差による影響であるのかどうかを検証する方法も広く知られていない。このため、分析結果に基づいて、教科教育方針を立てたととしても、基準に照らし合わせて有効であるか否かを議論することはできなかった。

このとき、特異な傾向が認められる誤答選択肢以外の特性曲線について平均選択率を計算することでまとめ、改めて項目特性図を作成することで、誤答選択肢に関して効果的に特性を示す図を作成することが可能となる。また、まとめ方について、研究Ⅱによる提案手法を用いて情報量規準を算出し、比較することで、特性曲線の併合に用いた仮説が妥当であるかどうかを検証するための傍証とすることが期待される。

研究Ⅲでは、ブックレット形式によるテスト共通項目に関する、項目特性図作成方法について提案した。ブックレット形式のテストでは、ブックレット間に

共通する項目を含めることで、統計的に互いに比較可能な状態へと導くことができる。このとき、共通項目に関してはブックレットの数だけ、項目特性図が作成可能である。例え同一項目であっても、標本誤差およびブックレット構成項目の違いから、項目特性は図ごとに変化し得る。このような場合に、何れの表現を項目特性として採用すべきか参照可能な基準は広く知られてこなかった。

研究Ⅲにおいて提案した手法は、互いに全く異なるとも、全く同一であるとも見なすことの難しいこれらのブックレット間項目特性図に関して、階層ベイズモデルを用いることで複数のブックレットの背後に共通して仮定される観念的な項目特性図を作成することを可能とする。同時に、当該項目特性図の安定性について分析することも可能となる。

観念的な項目特性図を解釈対象とすることで、ブックレット間項目特性図選択の恣意性を回避し、統一的な視点から項目分析を行えることが期待される。

上記の研究における提案手法は本稿を通じて個別に項目特性図に適用されてきた。しかしながらこれらの方法は状況に応じて組み合わせて適用することも考えられ、より柔軟に項目特性図を用いた項目分析を行う状況において役立つことが期待される。

## 引用文献

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csaki (Eds.), *2nd International Symposium on Information Theory.*, pp.267-281., Akademiai Kiado.
- 赤池弘次・甘利俊一・北川源四郎・樺島祥介・下平英寿(著) 室田一雄・土谷隆(編) (2007). 赤池情報量規準 AIC—モデリング・予測・知識発見— 共立出版
- 赤根敦・伊藤圭・林篤裕・椎名久美子・大澤公一・柳井晴夫・田栗正章 (2006). 識別指数による総合試験問題の項目分析. 大学入試センター研究紀要, **35**, pp.19-47.
- Brennan, R. L.(Eds.) (2006). *Educational Measurement, Fourth Edition*. Praeger.
- Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, **195**(2), pp.216-222.
- Ihaka, R. & Gentleman, R. (1996). R: a language for data analysis and graphics. *J. Comp. Graph. Stat.* 5:299-314. Available via <http://www.R-project.org>.
- 池田央 (1992). テストの科学—試験にかかわるすべての人に— 日本文化科学社
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.
- Stan Development Team (2014a). RStan: the R interface to Stan, Version 2.4. <http://mc-stan.org/rstan.html>.
- Stan Development Team (2014b). Stan: A C++ Library for Probability and Sampling, Version 2.4. <http://mc-stan.org>.
- 豊田秀樹 (2012). 項目反応理論 [入門編] 【第2版】 朝倉書店
- 吉村幸 (2009). 大学入試センターにおけるテストデータベースによる項目分析 植野真臣・永岡慶三(共編). e テスティング (第8章, pp.167-190) 培風館