

## 論文の内容の要旨

論文題目 確認的再答法による新評価法の研究

氏名 廣瀬 英子

コンピュータが社会に広く普及してきたことにより、これまで紙筆式 (Paper & Pencil Test) の形で行われてきた適性テストや学力テストをコンピュータ式 (Computerized Test) で行うことが、技術的に可能となってきた。コンピュータ・テストについては、いくつかの視点から研究が積み重ねられ、実用化に向けての提言がなされてきている。

コンピュータ・テストの特徴のひとつに、被験者が答えをコンピュータに直接入力することがあげられる。本研究は、コンピュータと被験者の間で行われる情報交換の双方向性と即時性を生かして、被験者の反応をテスト中に採点・評価し、必要に応じてただちにコンピュータから確認の問いかけを行い、被験者がそれをうけて自らの考えを確認しながら再答する、という一種の対話型評価法のモデルを提案した。これは、紙筆式テストでは実現が難しく、テストのコンピュータ化によって可能となる独特のテスト改善法といえる。

心理・教育測定の中で扱われるテストは、大きく分けると、被験者特性の最高値 (Maximum Performance) を観測するためのテストと、被験者の典型反応 (Typical Performance) を観測するためのテストの2種類に分類できる。本論文ではその両方を取り上げ、それぞれのテストで考えられる確認的再答法について個別に検討した。本論文は、項目反応理論の多値型モデル

を理論的基盤として展開されている。

本論文の第1章では、コンピュータ・テストに関するこれまでの研究動向を概括し、確認的再答法の研究の必要性を明らかにした。

第2章では、被験者の典型反応を観測する心理測定尺度における確認的再答法について取り上げた。項目反応理論にもとづいて作成されている尺度では、あらかじめ定められたモデルに従うように項目が精選されており、項目パラメタの値も求められている。そのようなテストを実施した場合、すべての被験者が必ずしもモデルが期待するような回答をするとは限らない。その原因としては、テストを受験する際の被験者自身の動機づけが乏しく、真剣に回答しないといった受験態度要因 (Test Faultiness) や、被験者の持つ特性が、本来その尺度が測定しようと考えている特性に適合しないという本態性要因 (Traitendness) があり得る。本研究では、その尺度への回答としてモデルが予想する回答パターンに従わない回答パターンが発見された場合に、それをモデルに対する不適合回答 (Aberrant Response) としてとらえ、被験者に再度回答してもらおうという手続きを提案した。不注意等の受験態度要因からくる回答が、新しい答えに置き換えられた場合には、不適合回答パターンの一部が解消され、不適合のまま機械的に算出された値で評価してしまう危険を防ぐことができると考えられる。また、再回答の結果、それでも回答パターンがモデルに適合しない場合については、その尺度を使用して被験者の特性を測定することが不相当であると考え、その被験者にふさわしいと考えられる別の方法による測定に切り替える手がかりとすることができる。仮定されたモデルにその尺度が適合しており、かつ、項目パラメタも適切に推定されているという前提条件を伴うものの、確認的再答法は、より正確な診断につながり、教育的にも大きな意味を持つと考えられる。

従って第2章では、まず不適合回答パターンを検出するために定義された適合度指標  $I_z$  を多値型の項目反応モデル (評定尺度モデル: Muraki, 1990) に適用する場合の統計的性質を検討した。項目パラメタ、項目数、選択カテゴリ数を変えた項目セットを用意して  $I_z$  値を算出し、 $I_z$  値の大きさと項目パラメタ、項目数、選択カテゴリ数の関係をまとめた。その結果、識別力の大きい項目が、項目の位置が広範囲に散らばるように配置されると不適合回答パターンが検出されやすいこと、項目数もカテゴリ数も多い方が良いことが示された。ここで、不適合解答パターンのなかの特に不適合な項目を見いだすための項目単位の適合度指標  $I_{z-i}$  を提案した。この  $I_{z-i}$  についても、不適合項目の検出に必要な項目パラメタの条件を検討した。その結果、

位置パラメタが両端に近い項目ほど、 $l_{z-i}$ の最大値と最小値の開きが大きくなることなどが示された。なお、適合度指標  $l_z$  にせよ  $l_{z-i}$  にせよ、尤度に基づく指標であるので、求められた値は適合度の大きさを相対的に示すものであるという限界をとまなう。

第3章では、実際の尺度を用いて不適合回答パタンの検出と、回答の修正プロセスを検討した。具体例として用いた尺度は、大学への適応問題について自己効力の観点から捉えるために著者らが開発した「大学への進路適応に対する自己効力測定尺度」(College Adjustment Self-Efficacy Scale)である。尺度構成後に新規に得られた3,670件のデータから不適合回答パターンを検出し、不適合の原因と考えられる項目を選び出し、被験者の実際の場面での反応として想定される再回答パターンをシミュレーションし、最終的に不適合回答パターンがなくなるまで回答の修正を試みた。その結果、尺度に含まれる項目群のうち2項目修正することで、不適合回答パターンはほとんど見られなくなった。また、識別力パラメタの大きい項目は、 $l_{z-i}$  が極端な負の値になりやすく、不適合項目として検出されやすいなど、効果的な尺度作成の指針が得られた。

第4章では、被験者の最高値を観測するテストの中で代表的な多肢選択式テストにおける確認的再答法について取り上げた。多肢選択式テストでは、普通、ひとつの項目につき正答候補の選択肢が数個提示されており、その中で、問題に対する答えとして最適な選択肢が正答とされ、他の選択肢は誤答として扱われる。しかし、被験者の選んだ選択肢が誤答の場合には再度答えさせ、正答に達するまで解答を続けさせる方法も考えられる。紙筆式では実施することが難しいこの達成式解答法 (Answer-Until-Correct) も、コンピュータ・テストでは可能である。最初の解答が正答でなかったなら、その時どの選択肢を選んでいたのか、2回目の試みでどの選択肢に考えを改めたのか、といった正答にたどりつくまでの選択の道筋は、被験者の学力に関する貴重な付加情報となり得る。これは、すべての選択肢から得られる情報を最大限に利用することになり、より精度の高い被験者パラメタの推定を期待できる。被験者にとっても、個々の課題について考えたその場で正答を確認できれば、内容に対する理解が深まり、教育的にも好ましい効果が期待できる。

そこで従来の項目反応モデルを拡張し、被験者パラメタを推定するための「達成式項目反応モデル」を提案した。ここでは被験者は正答に達するまでに、すでに選んだ誤答選択肢は再度選ばないものとし、項目パラメタの値は選択によって変化せず、また、被験者の能力値も、

先行する項目への解答行動によって大幅には変化しないものとする。これらの限定条件を入れることにより、従来の名義項目反応モデルを拡張し、各選択肢の解答経路情報ごとにその特性関数、情報関数の一般式を求めることができた。

第5章では、達成式項目反応モデルの持つ効果を確認するために、3つのシミュレーション実験を行った。[研究1]では、項目パラメタが既知である3肢選択項目のテストを、項目数を変えて5種類用意し(9, 11, 13, 17, 21個)、生成された解答パターンに対して達成式項目反応モデルの場合と、名義反応モデルの場合とそれぞれについて被験者パラメタの推定を行い、達成式解答法を用いれば、1回限りの解答によるテストより高い精度の推定、また同じ精度なら項目数を減らすことによって同等の推定が可能であることが確認された。[研究2]では、21項目の3肢選択項目で構成されるテストへの10,000件の人工データに対して、達成式項目反応モデルをあてはめた場合と、正答に達するまでの5通りの解答を5つのカテゴリとみなして従来の名義反応モデルをあてはめた場合とで比較を行った。その結果、達成式項目反応モデルで推定された項目パラメタから得られる最大項目情報量は、5カテゴリの名義反応モデルの場合よりも多いか、もしくは同程度であることが示された。項目の選択肢数が増えると、推定する項目パラメタ数が急速に増えるが、達成式項目反応モデルでは、それが少なく済むという利点がある。[研究3]では、4肢選択項目、10,000件の人工データについて、解答回数に制限のある場合について、達成式項目反応モデルによる被験者パラメタの推定を行った。そして、解答回数に制限のある場合の推定値は、完全に正答に達するまで解答する場合の推定値と、使用した項目数が21項目と少ないにも関わらず0.9以上の相関があり、誤差の増加も効果比にして6%程度で、近似として十分役立つことが確かめられた。また、1回限りの解答情報を利用した名義反応モデルや2-パラメタ・ロジスティックモデルによる推定値、さらに、古典的な採点法で正答数から導かれる得点とも比較の結果、それらの値より推定精度が高く、解答回数が制限されても達成式解答法が優位であることが確認された。

第6章では、再回答情報を取り入れた場合の被験者パラメタの推定に関する総括的結論をまとめた。本論文は、心理測定尺度における再回答の手続きの提案と、多肢選択式テストにおける達成式項目反応モデルの提案を行い、その測定論的有効性を示した。ただ、モデル自身に含まれている前述の仮定条件が、実際の適用状況においてどの程度無理なく働くかという点については、今後、実施環境が整ったときに検討していきたい。