

論文の内容の要旨

論文題目 認知診断モデルの理論的および 経験的検討

氏名 山口一大

本論文では、テスト理論モデルにおいて脚光を浴びてきている、認知診断モデル (cognitive diagnostic models, CDM) について、モデルの性質の理論的検討と実際のデータを用いた経験的な検討を行った。CDM は認知診断のための有望なモデルではあるが、種々提案されたモデルの間について理論的な整理がまだ十分になされておらず、どのような場合にどのモデルを使用するのが適切かといったモデル比較についての研究も不足しているのが現状である。本研究は、CDM に関するこれらの問題に理論的・経験的にアプローチし、CDM の応用に資する知見を得ることを目的とした。具体的には、(1) これまでに提案されている CDM をパラメタ設定の観点から理論的に検討・整理し、(2) CDM で重要な役割を担う、各項目への正答にどの能力 (アトリビュート) が必要かを示す Q 行列を誤って設定した場合の影響をシミュレーション実験により検討し、(3) TIMSS (Trends in International Mathematics and Science Study; 国際数学・理科教育動向調査) 2007 の小学校 4 年生の算数データを材料として CDM のモデル比較を経験的に行う、という 3 点を主要な目的とした。

第 1 章の序論で CDM の基本的な考え方を述べ、CDM の応用研究の現状、およびモデル比較研究の現状を概観した後、第 2 章では近年数多く開発されている CDM について、理論的な整理を行った。とくに、統合的モデルとその下位モデルに注目し、新たに儉約モデル、主効果モデル、飽和モデルの 3 つの分類を提案した。儉約モデルは項目パラメタの数が最も

少なく、制約が強いモデルを指す。主効果モデルはアトリビュートが加法的に問題正答確率に影響を与えるモデルである。飽和モデルはアトリビュート間の交互作用を許容するモデルであり、最もパラメタ数が多い統合的モデルである。さらに、これらのモデルが適した使用状況について検討を行った。これらのモデルは制約の強さが異なっており、制約が厳しいモデルほどそれに対応してアトリビュートに関する事前知識が求められると考えられる。したがって、最も制約の強い儉約モデルは、アトリビュートを明確に定義して、そのアトリビュートに関する診断を目的としてテストを開発した場合など、アトリビュートについて詳細な想定がある場合に適していると考えられる。逆に最も制約が弱く、アトリビュートの交互作用をすべて含む飽和モデルは、アトリビュートの組み合わせの効果について事前に予測ができない場合に探索的に使用することに適している。最後に制約の強さが儉約モデルと飽和モデルの中間に当たる主効果モデルは、想定したアトリビュートの一部が各項目の正答確率に影響を与えていると考えられる場合に、その項目パラメタを参照することで、どのアトリビュートがどの項目に寄与しているかについて Q 行列を確認し修正するための簡便な方法として利用するのに適している。

第 3 章では現実のテストにみられる、あるアトリビュートの習得のためには他のアトリビュートの習得が前提となるといったアトリビュート間の階層構造を考慮した場合の Q 行列の誤設定の影響をシミュレーション実験により検討した。とくに基本的なアトリビュート階層構造である、直線型、分岐型、収束型の 3 種類の階層構造に注目して検討した。その結果、分岐型で誤設定の影響が最も強くみられ、直線型や収束型では誤設定の影響はあるものの相対的に見て小さいこと、全体の項目数に対して誤設定の割合が小さい場合は誤設定の影響が緩和できる可能性があるといったことが明らかとなった。診断テストを作成する際には、当該の階層構造が誤設定に頑健であるかどうかという観点からも検討する必要性について述べた。

第 4 章と第 5 章では第 2 章での CDM の理論的な検討を踏まえて、モデル比較に関する経験的な検討を行うとともに、モデルを適用して得られる診断結果について、詳細な分析を行った。第 4 章では、TIMSS2007 の日本人データを用いて、第 2 章で検討したモデルのうちどのようなモデルが実データに最も適合するのかを検討した。モデル比較の軸として、これまで教育測定モデルとして重要な地位を占めてきたロジスティック項目反応理論モデルと一連の CDM 群の比較を行った。そして、それらのモデルを適用することで得られるアトリビュート習得パターンや項目パラメタの推定結果を通してモデルの評価を行った。その結

果、ロジスティック項目反応理論モデルよりも CDM の適合が良好で、その中でも主効果モデルの 1 つである R-RUM モデルの適合が良かった。また、項目反応モデルで推定された特性値と CDM で推定されたアトリビュートとの相関分析の結果から、CDM で仮定するアトリビュートが一般的な算数能力を十分に反映したものであることが示された。また、使用する CDM によって、推定されるアトリビュート習得パターンにはかなりの違いが生じる可能性が示唆された。項目ごとの分析では、項目パラメタの推定結果から、複数のアトリビュートが必要な項目と単一のアトリビュートのみを必要とする項目が混在することが示された。

第 5 章では、同じく TIMSS2007 の算数データについて、対象とする国と地域を拡張して、第 4 章で日本人データについて得られたモデル比較の知見がそれらのデータにも一般化可能かどうかの検討を行った。対象としては TIMSS の公式到達度スコアに注目し、それが高い国・地域と低い国・地域から選択し、そのような広がりのある解答者の解答データに対し、どのような CDM が広く有用であるのか検討した。その結果、いずれも主効果モデルである R-RUM や A-CDM, LLM が多くの国・地域のデータにおいて適合が良い傾向がみられた。主効果モデルの適合が良いという結果は、第 4 章の日本人データの場合と同様であり、少なくとも TIMSS データに関しては、国・地域の平均的な到達度水準などの特性によらず、一般的な傾向であることが推測される。また日本人データと同様に、項目反応理論モデルによって推定された特性値と CDM で推定されたアトリビュートとの相関分析から、アトリビュートが一般的な算数能力を反映していることが示された。更に、日本人データと同様に、項目パラメタの推定値から、複数のアトリビュートが正答に必要な項目と、単一のアトリビュートのみで正答可能な項目の相違が示された。また、国・地域によらず共通のアトリビュートが必要である項目と、国・地域によって異なったアトリビュートが必要な項目が存在することが明らかになった。各国のアトリビュート習得パターンについては、TIMSS の平均的な到達度の違いを反映して、習得しているアトリビュート数の違いがみられた一方、どのアトリビュートの習得割合が高いかについて、国・地域によって違いがみられた。

第 4 章と第 5 章のモデル比較では、R-RUM, A-CDM, LLM といった主効果モデルに比べ、簡約モデルとして分類された DINA モデルや DINO モデルは概して適合が悪かった。また、最もパラメタ数が多い飽和モデルは、そのパラメタ数のために対数尤度の値では他のモデルより高かったものの、情報量規準 AIC, BIC の観点からは適合が相対的に悪いという結果となった。多くの国の TIMSS データにおいて主効果モデルの適合が良かったことは、TIMSS の解答者がそれぞれのアトリビュートを個別に適用することで問題に解答している

可能性を示唆している。なお、TIMSS はもともと算数・数学と理科の到達度を測定する目的で作成されたものであり、アトリビュートの診断のためのテストではないが、これらの分析を通して、そのようなテストに対して CDM を適用することでアトリビュート習得パターンや項目パラメタの推定値といった新たな情報を抽出できるという可能性が示された。

第 6 章では、第 2 章から第 5 章までの理論的・経験的な検討の結果を総括し、本研究の限界と今後の展望について論じた。本研究の経験的検討で用いたモデルは G-DINA モデルとその下位モデルに限定されているという点が限界の 1 つである。また、実データとして TIMSS データのみを検討の対象としており、本研究の結果の一般化には注意が必要である点を述べた。また、第 4 章・第 5 章で得られた項目パラメタの推定値からは、先行研究で利用された Q 行列について、より少数の重要なアトリビュートに限定した上で、不要な要素を削除するほうがデータにより適合する可能性が示唆された。そこで、今後その方向で Q 行列を再構成して分析し、本研究で得られた結果と照合する必要性を議論した。さらに、最近の理論研究から、CDM における項目パラメタやアトリビュートパタンの識別性の問題、すなわち、データからこれらのパラメタやパタンの推定が一意に決まらない可能性が指摘されていることから、本研究で使用したモデルについて、この観点から今後検討する必要があることを述べた。モデル比較の方法については、本研究で用いた基準に加え、近年発展してきているベイズ統計学に基づく情報量規準を用いた方法についても検討していく必要があることを述べた。最後に、発展的な課題として、理論的には今後の CDM のモデル開発について、1 時点における静的なアトリビュートの習得パターンのみならず、解答者の問題解決プロセスにより即し、アトリビュート自体の変化を考慮した動的なモデルの開発に取り組むこと、そして応用的には、第 5 章の分析を拡張して、解答者のアトリビュート習得パターンと各国の教育制度や教育内容との関連について検討することを述べた。