教育ネットワークセンター年報, 2018, 18, 49-53（研究報告｜大学院生プロジェクト型研究）

大学院生プロジェクト型研究・研究成果報告書

研究代表者：坂本 佑太朗（教育設計評価研究コース）

■研究題目

テストデータ分析における bi-factor モデルの応用可能性について

■研究代表者・分担者 氏名

坂本 佑太朗（教育設計評価研究コース）（代表者）

■研究成果概要（目的、実施内容、結果、今後の課題など）

1. 目的

本研究の目的は、テストデータ分析における bi-factor モデルの応用可能性を明らかにすることである。

近年の大学入学者選抜を巡る議論の中で、項目反応理論（IRT）というキーワードが見られるように、わが国における心理測定領域への関心は高まっている。従来の IRT 分析では、テストは一つの構成概念（たとえば「数学力」など）を測定しているという仮定が置かれていた。最近では、IRT を多次元に拡張した理論モデルが提唱され、その一種である bi-factor モデルを応用した研究が蓄積されている。bi-factor モデルとは、テスト全体が測定する因子（一般因子）と、下位領域（subscales）の因子（グループ因子）を想定するモデルである。先行研究ではこの特徴を活かし、「測ろうとした構成概念が測れているのか」というテストの妥当性検証のための有効な手段の 1 つであると指摘されている。

しかしながら、bi-factor モデル自体は 1930 年代に提案されているものの、それから約 70 年の間、テスト理論研究において注目されてこなかった。先行研究は主に 2000 年代以降、欧米を中心に増えてきているものの、わが国においては研究の蓄積はほとんどない。わが国においても、「テストで評価する」だけではなく「テストを評価する」ことが求められていることを鑑みれば、bi-factor モデルに関する測定論的研究は喫緊の課題であると判断できる。

そこで、本研究では bi-factor モデルのテストデータ分析における応用可能性について、理論的系譜の整理と実データ分析という観点から明らかにする。具体的には、その理論的系譜について、文献研究を中心に行う。その後、実データを使って、従来のテス

トデータ分析では明らかにできなかったテストの下位領域に関する詳細な情報を抽出することを試みる。さらに、下位領域特有の学力を定量的に表現し、その状況を確認する。

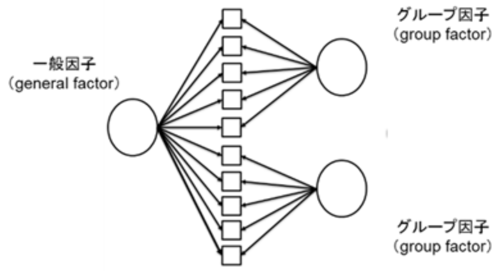

図１　bi-factor モデルのイメージ

2. 研究の経過

本研究は、下記のようなスケジュールで進められた。

なお、期間に重複があるのは、文献研究とデータ分析が同時並行的に試行錯誤しながら進められたためである。

①　文献研究（2017 年 6 月～2017 年 11 月）

心理測定、テスト理論的な観点から bi-factor モデルの理論的な変遷について整理した。その成果は心理測定学の専門誌に投稿中である。

②　データ分析（2017 年 7 月～2017 年 12 月）

実際のテストデータを使用し、bi-factor モデルを適用しながら、その有効性を確認した。その成果は、2017 年統計関連学会連合大会（南山大学）で発表（依頼有）され、専門誌にも掲載された。

③　研究のまとめと今後の展望（2018 年 1 月～2018 年 2 月）

文献研究とデータ分析結果を振り返り、今後の課題を設定した。

次節では、データ分析結果（上記②に相当）の一部を報告する。

3. 方法と結果

データ

平成 18 年度新潟県全県学力調査における中学校第 2 学年数学データ（N=9, 102, 項目数は 25）を使用する。新潟県教育委員会（2007）ならびに当時の学習指導要領（文部科学省, 2006）によれば、項目 1 から項目 10 は「数と式」、項目 11 から項目 16 は「数量関係」、そして項目 17 から項目 25 は「図形」を測定する項目群であり、これらが下位領域となる。

方法

本研究では、Reise et al. (2007)、 Rindskopf & Rose (1988)、坂本（2016）を参考に

して以下の 4 つのモデルを仮説として設定した。まずモデル A として 1 次元性を仮定した通常の IRT モデル（2 パラメータロジスティックモデル）を設定する。このとき、テスト全体は 1 つの「数学力」を測定していることを仮定することになる。次に、新潟県教育委員会（2007）ならびに文部科学省（2006）から、当該の項目が測定すると定められていない領域については項目識別力パラメータを 0 とする確認的多次元 IRT モデルをモデル B として設定する。このとき，テストの下位領域間には何らかの相関関係があると考えるのが自然であることから、下位領域間の因子間相関を認めることにする。次に、テスト全体に「数学力」因子、それに加えて下位領域特有の影響を認める双因子モデルを設定するが、このときモデル B と同様に下位領域間の関係を考慮することが可能である。そこで本研究では下位領域間の相関を許容しないモデルと相関を許容するモデルをそれぞれモデル C とモデル D とする。

結果

すべての仮説モデルについて項目パラメータを推定した結果、モデル D がもっともあてはまりがよかった。そこで、下位領域ごとの正答数得点と、テスト全体が測定する「数学力」を統制した後の、下位領域特有の能力を表す潜在特性尺度値との関係は表 1、図 2 から図 4 に示す通りである。このように、bi-factor モデルを用いることで、下位領域特有の能力を定量的に表現することができる。

表 1 から「素点_図形」と「図形」特有の学力を示す「D_図形」の相関が、他の領域と比べて相対的に低いことがわかる（相関係数 0.515)。このことから、下位領域としての「図形」は他の領域と比較して、やや異なる要素を測定している可能性が示唆される。

下位領域の素点合計点に対応するモデルDで推定された潜在特性尺度値の箱ひげ図を参照すると、図 4 の「図形」領域のみ、その中央値が素点合計点の値にかかわらずそれほど変化がないことがわかる。他の下位領域では、素点合計点の値にしたがって、中央値が変化する傾向がある。このことから、モデル D で推定された「図形」領域の潜在特性尺度値の解釈には注意が必要であり、さらなる検証が必要だと指摘できる。

表 1　素点合計点とモデル D の下位領域ごとの潜在特性尺度値の相関

素点合計点	モデル D の潜在特性尺度値	相関係数
素点_数と式	D_数と式	0.682
素点_数量関係	D_数量関係	0.713
素点_図形	D_図形	0.515

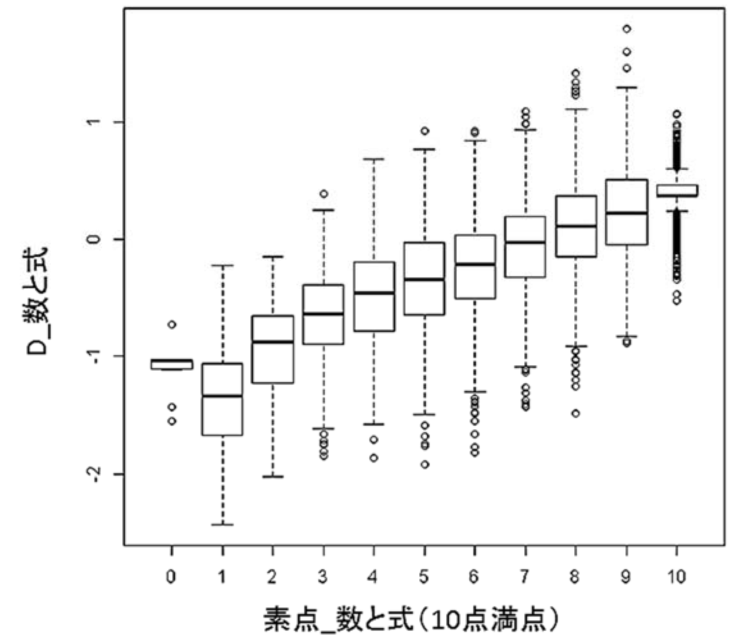

図 2　素点合計点に対応するモデル D で推定された
潜在特性尺度値の箱ひげ図（「数と式」領域」

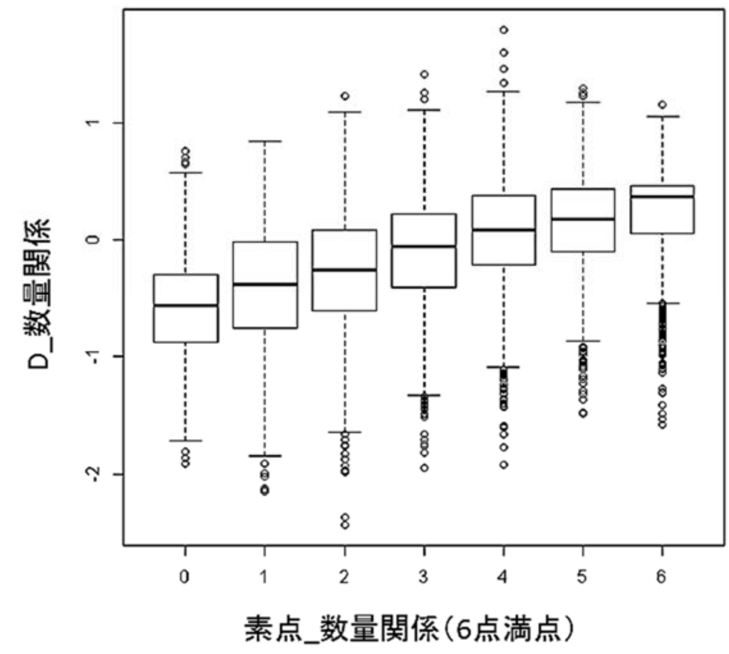

図 3　素点合計点に対応するモデル D で推定された
潜在特性尺度値の箱ひげ図（「数量関係」領域」

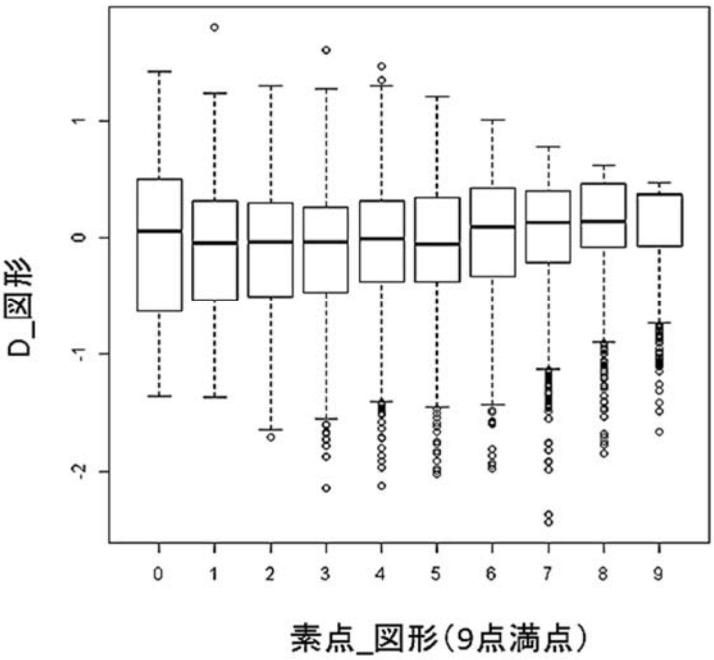

図 4　素点合計点に対応するモデル D で推定された
潜在特性尺度値の箱ひげ図（「図形」領域」

4. 今後の課題

本研究では、bi-factor モデルの歴史的系譜の整理と実データ適用を通して、わが国で実際に行われたテストデータにおいても、その有効性が確認できた。しかしながら、「図形」領域の潜在特性尺度値についての検証（その推定値自体の妥当性検証）は十分にできなかったことは本研究の限界である。

また、今回はテスト項目への反応データのみ焦点を当てており、bi-factor モデルを通して表現される受検者の能力（潜在特性尺度値）と、その規定要因となりうる変数（親の学歴や家庭の経済状況など）との関連性まで検討することができなった。

今後は、bi-factor モデルを通して表現される下位領域特有の潜在特性尺度値と、そのような学力の規定要因となる変数との関連性を探ることが主な課題となる。