

# 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）

<b>■ 研究題目</b>
テストの下位領域に着目した学力の要因分析
<b>■ 研究代表者・分担者 氏名</b>
坂本 佑太朗（教育設計評価研究コース）（代表者）
<b>■ 研究成果概要（目的、実施内容、結果、今後の課題など）</b>
<p><b>1. 問題と目的</b></p> <p>学力テストの結果解釈時に、下位領域（subscale）ごとの得点や平均正答率が注目されることがある。たとえば、PISA2015においては「読解力（reading literacy）」の下位領域（側面別）の平均正答率として、「探求・取り出し」は72%、「統合・解釈」は60%、「熟考・評価」は62%であることが報告されている（国立教育政策研究所，2016）。また、心理測定論的立場から、柴山・佐藤・熊谷・澁谷・板宮・江尻（2018）では、下位領域に属する項目数が少なくなることによる信頼性の低下と正確な学力分布を把握できないという問題に対し、IRTを基盤とした推算値（plausible values）の援用により、下位領域ごとの学力分布を推定する試みを行っている。このような事実から、テストの下位領域をいかに扱うかは、学力テスト研究においても重要なポイントの一つであると判断できる。</p> <p>下位領域は、内容的（content）な項目のまとまり（教科で言えば「単元」等）あるいは認知的（cognitive）な項目のまとまり（「知識」「推論」等）で形成されることが多い。Reise, Cook and Moore（2015）は、下位領域がテストの次元性（dimensionality）に影響を及ぼすことがあるため、テストデータ分析の際には多次元性を考慮した多次元IRT（multidimensional IRT）（Reckase, 2009）を用いた分析結果も同時に報告すべきであると指摘している。最近では、坂本（2016）、坂本・柴山（2017）が、多次元IRTの枠組みで捉えられるbi-factorモデル（Gibbons &amp; Hedeker, 1992）を用いて、テスト全体が測定する学力（一般因子）と下位領域特有の学力（グループ因子）を分離して捉え、下位領域特有の学力の影響を相対的に強く反映する項目を抽出する試みを行っている。また、水本・脇田・名部井（2017）は実際の英語入試問題に対して、下位領域を考慮した多次元</p>

IRT を適用し、テストの構造的な側面の妥当性検証を行っている。このように、最近の心理測定論的テスト研究においては、テストの下位領域に関する精緻な検証が意識された研究が増加してきていると言える。

しかしながら、学力テストのスコアを従属変数にし、その要因分析を行う教育社会学的な研究においては、下位領域を考慮せず、学力は一つの潜在特性尺度値として表現されることが多い。その中でも例外的に、Li, Lei and Pace (2013)、鳶島 (2016) は下位領域ごとの学力に注目した要因分析を行っている。Li, Lei and Pace (2013) は、PISA2009 の上海 (中国) とアメリカの「読解力」データを使って、下位領域 1) (「探求・取り出し」、 「統合・解釈」、 「熟考・評価」) 特有の能力に着目した分析を行っている。具体的には、それらの下位領域のスコアを従属変数に、性別や階層情報、それに加えテスト全体が測定していると仮定する「読解力」スコアを独立変数に設定している。これにより、「読解力」を分析モデル上でコントロールした上で、下位領域特有の学力に対し、独立変数からどのような影響があるかを明らかにする分析を行っている。鳶島 (2016) では、Li, et al (2013) が領域ごとの分析を別々に行っていたために起こりうる、標準誤差が過小に推定される (Snijders & Bosker, 2011) 現象を防ぐため、領域ごとの能力を考慮して、一つのモデルで分析を行っている。

たしかに Li, et al (2013)、鳶島 (2016) とともに下位領域特有の能力を捉えようとしている。しかしながら、下位領域特有の能力を心理測定論的なモデルを通して表現することについては検討されていない。最近適用例の見られる bi-factor モデルを用いることで、通常が多因子構造を仮定する多次元 IRT 分析で推定される潜在特性尺度値とは異なる、下位領域特有の潜在特性尺度値を捉えられることが指摘されている (Demars, 2013)。つまり、下位領域特有の能力を bi-factor モデルで定量的に表現し、それを使った学力の要因分析につなげていくことは依然として課題であると指摘できる。

そこで、本研究では、テストデータの因子的な構造を確認した上で、下位領域特有の学力を IRT 技術にもとづいて定量的に表現する。その上で、下位領域特有の学力に焦点を当てた要因分析を行うことで、心理測定論に立脚した学力研究上の知見を獲得することを目的とする。なお、本研究では授業方法が学力へ与える影響に焦点を絞ることとする。

## 2. 方法

### 2.1. 使用データ

わが国における TIMSS2003 の中学二年生理科 (N=4,856) を用いる。学校数 (教師数) は 146 となっている。本研究では、下位領域に着目した学力の要因分析を行う上で、授業方法が下位領域の学力に与える影響に焦点を絞る。そのため、授業方法について詳細に生徒質問紙また教師質問紙で調査を行っている TIMSS2003 を用い、現時点での最新調査である TIMSS2015 は用いない。なお、授業方法については、各授業方法の頻度について生

徒・教師それぞれの認識に基づくものであることには注意されたい。

TIMSS2003 の中学校二年生理科における下位領域としては大きく認知的領域 (cognitive domain) と内容的領域 (content domain) がある。本研究ではこのうち認知的領域に焦点を当て、それを構成する 3 領域、「知識」(Factual Knowledge)、「概念的理解」(Conceptual Understanding)、「推論・分析」(Reasoning and Analysis) をテスト全体が測定する「理科の学力」の下位領域として扱う。

なお、要因分析のアプローチとしては、授業方法が学力（「理科の学力」ならびに各下位領域の学力）へ与える影響について、マルチレベル分析を使って検証することとする。マルチレベル分析に関しての理論的詳細は、たとえば Raudenbush and Bryk (2002) を参照されたい。

## 2.2. 下位領域に着目した項目分析

下位領域特有の学力を捉えるための方法として多次元 IRT の枠組みにおける bi-factor がある。本研究では、一次元性を仮定する IRT モデル、下位領域のそれぞれを因子とみなし相関を仮定した 3 因子モデル、そして bi-factor モデル 2) の情報量規準を比較し、テストデータの構造を確認する。採用した項目反応モデルは多次元段階反応モデル (Muraki & Carlson, 1995) であり、推定には Metropolis-Hastings Robbins-Monro 法 (MHRM 法, Cai, 2010) を用いた。表 1 にその結果を示す。

表1 モデルごとの情報量規準

	一次元性 モデル	3 因子モデル (因子間相関 あり)	bi-factor モデル
AIC	236848.84	236292.35	232649.29
BIC	241857.55	241320.53	240058.55
AICc	237141.15	236587.15	233352.39
SABIC	239404.41	238857.86	236429.68

この結果から、bi-factor モデルがもっともあてはまりがよいことがわかる。つまり、テスト全体としては一次元の理科に関する学力を測定していると言うよりは、一次元の学力に加えてそれぞれの下位領域の影響も受けた多次元的な学力を測定していることを意味する。bi-factor モデルにおいて、項目パラメータが異常に大きく推定される項目を除外しながら再推定する作業を繰り返し、最終的に分析対象になった項目は 346 項目であった。下位領域としての「知識」「概念的理解」「推論・分析」の内訳はそれぞれ 106 項目、138 項目、102 項目となった。

各下位領域に関する潜在特性尺度値をEAP法（平均0，標準偏差1を仮定）で推定し，そこから5つの推算値（Plausible Values）を算出した．今後の分析にはこの推算値が従属変数となる．

### 3. 結果

マルチレベル分析に先立って，「理科の学力」（推算値）とbi-factorモデルを通して得られた各下位領域の潜在特性尺度値（推算値）を従属変数とした場合の，説明変数を投入する以前（nullモデル）における級内相関係数（ICC）を算出した．その結果，「理科の学力」が9.4%であった．つまり，「理科の学力」の全分散（学校間分散と学校内分散の合計）のうち，学校間分散が9.4%占めているということを意味している．次に，下位領域についても同様に算出したところ，「知識」が0.6%，「概念的理解」が1.4%，「推論・分析」が1.0%であった．言い換えれば，各下位領域特有の学力に関して，学校レベルの要因が寄与するのはおよそ0.1割程度ということの意味する．たしかに下位領域のICCは高いとはいえないが，「理科の学力」また各下位領域の潜在特性尺度値を従属変数とするモデル間の傾向を比較する意味で，下位領域の潜在特性尺度値を従属変数とする場合でもマルチレベルモデルを適用することとする．

要因分析は，独立変数に性別、学校種別、出身階層、授業方法（生徒）等を投入したモデル1、それに加え授業方法（生徒）と出身階層（生徒）の交互作用も加味したモデル2、そしてさらに授業方法（学級）も考慮したモデル3で行う．なお，教育社会学領域における学力研究において，生徒の出身階層と学力との関連はいわば通説となっている（荻谷，2012，2004）．そのため，マルチレベルモデルによって下位領域に関する学力の規定要因を探る際には，個人要因と同時に出身階層に関しては学校差についても考慮しておくべきと判断できる．そこで，先行研究を参考にモデル化したモデル1からモデル3においては，切片と「出身階層（生徒）」において変量効果を仮定したランダム切片・係数モデルを適用することとする．なお，マルチレベル分析におけるパラメータの推定には最尤法を用いた

「理科の学力」については，モデル1において授業方法以外の生徒の属性に関する変数では，「女子ダミー」，「国・私立ダミー」，「出身階層（生徒）」「出身階層（学校）」が有意となっている．授業方法（生徒）と出身階層（生徒）の交互作用を考慮するモデル2，モデル3については「出身階層（生徒）」が非有意となるものの，その他の変数は有意である．「出身階層（生徒）」がモデル2～3で非有意となることは，鳶島（2015）と比較して異なる結果である．次に，授業方法（生徒レベル）については，モデル1からモデル3において一貫して「宿題試験型」が正の効果で有意となっている．モデル2においては「社会日常型」が正の効果で有意，モデル3においては「聴講演習型」が負の効果で有意となっている．特に，一貫して「宿題試験型」が正の効果でありかつ有意であることは，鳶島

(2015) や須藤 (2013) とは異なる結果である。このような先行研究の傾向との違いは、今回の分析モデルにおいてそもそも従属変数が下位領域の影響を取り除いた「理科の学力」となっていること、授業方法の学校平均を投入できていないこと、出身階層（生徒）に変量効果を仮定していることなどが起因していると推察される。

先行研究においては、授業の中で、宿題の答え合わせや確認テストが中心となる授業（「宿題試験型」）は新しいことを学ぶ行為に欠けるため学力に負の効果をもたらすという仮説（須藤，2013）にもとづいて検証が進められてきた。しかしながら、今回下位領域の影響を統制した上で得られた「理科の学力」の要因を探ると、授業の中で宿題の答え合わせや確認テストといういわば「あたりまえ」の行為を行うことによって、「理科の学力」を向上させることが明らかになったこととなる。

下位領域に関する分析では、まず「知識」において、モデル 1～3 まで一貫して「女子ダミー」が負の効果で有意、「課外活動ダミー」が正の効果で有意、「社会日常型」が負の効果で有意となっている。このうち、「課外活動ダミー」は「知識」における分析のみで有意となっている。課外活動とは、学校以外で理科の授業を受けているかを示す変数であり、実質的には通塾しているかどうかを示す独立変数である。塾や予備校では、今回で言えば「実験調査」「社会日常」的な学習というよりは、体系的に知識を教授することが主な内容であると推測できる。そのような「課外活動ダミー」が「知識」に対して正の効果で有意となっていることは整合性が取れる。

「概念的理解」「推論・分析」については、「理科の学力」と「知識」に比べて相対的に有意となる変数が少なくなっている。「概念的理解」では、モデル 1 において「出身階層（生徒）」が正の効果で有意、モデル 3 において「教師主導型」が負の効果で有意となっている。次に「推論・分析」では、モデル 1 において「出身階層（生徒）」が正の効果で有意であることに加え、モデル 1～3 に一貫して「宿題試験型」が正の効果で有意となっている。須藤 (2013) では「宿題試験型」こそ学力への影響がないという仮説を立てた上で検証されていた。しかしながら、今回の分析を通して、この「宿題試験型」が「理科の学力」に加えて「推論・分析」に対しても正の効果が認められることが明らかになった。言い換えれば、理科の授業においてきちんと宿題の確認や定期的に行われる確認テスト（小規模テスト）に取り組む中で、理科の学習における「推論」を行うことや、「分析」を行う取り組み自体が機能しているとも捉えられる。この知見は、下位領域特有の学力に着目してはじめて得られるものであり、本研究の意義と言える。

#### 4. まとめと今後の課題

本研究では、わが国における TIMSS2003 中学校二年生理科データを用いて、下位領域（「知識」「概念的理解」「推論・分析」）に焦点を当てた学力の要因分析を行った。具体的には、マルチレベル分析を用いて生徒レベルと学校レベルの影響を考慮し、授業方法の違

いが bi-factor モデルを通して得られる「理科の学力」（一般因子）と各下位領域特有の学力（グループ因子）に与える影響について検討した。その結果、「宿題試験型」がモデル 1 から 3 において「理科の学力」「推論・分析」に対して正の効果で有意となっていた。須藤（2013）や鳶島（2015）では、この「宿題試験型」の学力への影響は負の効果で有意となっていることから、本研究との結果と異なる。しかしながら、本研究では、テストが測定している学力そのものを多次元 IRT により精緻に表現し、推算値を使って集団統計量として捉え直している。その上で、先行研究での知見を再検証しているということになる。今回の知見は、心理測定技術と教育社会学的なアプローチをハイブリットしてはじめて得られるものであり、本研究の意義であると言える。

直近での「理科」領域を扱った平成 27 年度全国学力・学習状況調査においては、「実験の結果を数値で表した表から、分析して解釈し規則性を見出すこと」という、まさに本研究における「推論・分析」に関する学力は、今後の課題とされている（国立教育政策研究所、2015）。たしかに、最近では「アクティブラーニング」等の新しい学びが重要視されている。しかし、本研究は、日々の宿題に取り組みかつ授業の中でそれを確認する、また定着度を確認テストで把握しながら日々の学習を行うなど、「従来型」の学習も学力向上のために必要であることを示唆しているといえる。心理測定技術により学力を精緻に表現し、かつ集団統計量として捉えた本研究結果から、このような教育的な示唆をも導ける。

## 5. 参考文献

- Cai, L. (2010). Metropolis-Hasting Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, **35**, 307-335.
- Demars, C.E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing*, **13**, 354-378.
- Gibbons, R.D. & Hedeker, D.R. (1992). Full-information item bifactor analysis, *Psychometrika*, **57**, 423-436.
- 荻谷剛彦（2004）。「学力」の階層差は拡大したか 荻谷剛彦・志水宏吉編 学力の社会学—調査が示す学力の変化と学習の課題—（pp.127-151）岩波書店。
- 荻谷剛彦（2012）. 学力と階層. 朝日文庫.
- 国立教育政策研究所（2015）. 平成 27 年度全国学力・学習状況調査中学校理科 報告書.
- 国立教育政策研究所（2016）. OECD 生徒の学習到達度調査（PISA）2015 年調査国際結果の要約.
- Li, H., Lei, P & Pace, C.L. (2013). Reading subskill differences between students in Shanghai- China and the US: evidence from PISA2009. *Educational Research and Evaluation*, **19(6)**, 490-509.
- 水本篤・脇田貴文・名部井敏代（2017）. 関西大学英语入試問題データの分析—テスト理論の活

- 用を目指して— データ分析の理論と応用, **6**, 21-30.
- Muraki, E., & Carlson, E. B. (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement*, **19**, 73-90.
- Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Application and data analysis methods* (2<sup>nd</sup> ed.). Newbury Park, CA: Sage.
- Reckase, M.D. (2009). *Multidimensional Item Response Theory*. New York: Springer.
- Reise, S. P., Cook, K. F., & Moore, T.M. (2015). Evaluating the impact of multidimensionality on unidimensional item response theory model parameters. In Reise, S.P. & Revicki, D.A. (Eds). *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*. (pp.13-40) Routledge.
- 坂本佑太郎 (2016). わが国の TIMSS2011 数学データにおける多次元 IRT を使った妥当性の検証について. 日本テスト学会誌, **12**, 37-53.
- 坂本佑太郎・柴山直 (2017). 学力テストの下位領域に関する多次元 IRT 分析— データ分析の理論と応用, **6**, 31-45.
- 柴山直・佐藤喜一・熊谷龍一・澁谷拓巳・板宮千尋・江尻大亮 (2018). 経年変化分析調査との対応付けによる本体調査の年度間比較の試み—平成 29 年度文部科学省委託研究「学力調査を活用した専門的課題分析に関する調査研究」研究成果報告書.
- 須藤康介 (2013). 学校の教育効果と階層—中学生の理数系学力の計量分析— 東洋館出版社.
- Snijders, T.A.B. & Bosker, R.J. (2011). *Multilevel Analysis: An Introduction to Basis and Advanced Multilevel Modeling. Second Edition*. Sage.
- 鳶島修治 (2015). 学力の階層差と授業方法との関連—マルチレベル分析による検討— 社会学研究, **95**, 1-24.
- 鳶島修治 (2016). 読解リテラシーの社会経済的格差—PISA2009 のデータを用いた分析— 教育社会学研究, **98**, 219-239.