

令和 2 年 5 月 20 日現在

機関番号：10101

研究種目：若手研究(B)

研究期間：2017～2019

課題番号：17K12940

研究課題名(和文) グループ内ピアアセスメントに基づく能力測定のための実的手法の提案

研究課題名(英文) A proposal of a practical method for ability estimation based on group peer assessment

研究代表者

岩間 徳兼 (Iwama, Norikazu)

北海道大学・高等教育推進機構・准教授

研究者番号：70608900

交付決定額(研究期間全体)：(直接経費) 1,100,000円

研究成果の概要(和文)：近年、高等教育において利用されるようになったパフォーマンス評価について、グループでの評価体制に基づいて、Uto and Ueno (2016) で提案された項目反応モデルによる分析における能力推定精度を検討した。不完備ブロックデザインによって、あるグループのメンバーが一部のグループのメンバーを限定して評価する状況を想定して行ったシミュレーション研究の結果、被評価者数が40や60の条件では、欠測率が0.5であっても精度よく被評価者の能力を推定できることが分かった。

研究成果の学術的意義や社会的意義

パフォーマンス評価データを項目反応モデルによって分析する研究はこれまでに存在したがそれらは理論的な検討の意味合いが強かったため、本研究では、高等教育における現実的な授業場面として、グループで評価を行うという状況を想定して研究を行った。また、従前の研究では、モデルの推定にはベイズ推定が主に採用されてきたが、統計モデルに詳しくない者にとってはそれを実践で利用することは難しい。そこで今回、項目反応モデルの推定で一般に利用される最尤推定のプログラムを作成し、いくつかの項目を指定すれば比較的容易に分析ができる環境を整えた。

研究成果の概要(英文)：I examined ability estimation with the Item Response Model proposed in Uto and Ueno(2016) that can be used for analyzing performance assessment data. In this study, I assumed that students in a course assessed one another through group activities and conducted a simulation study. As a result of the simulation study, in which I set a limit on the number of assessments in an unbalanced incomplete block design, it was revealed that under the condition of 40 or 60 students, abilities of students could be estimated accurately even when 50 percent of assessments were missing.

研究分野：心理統計学

キーワード：グループ学習 項目反応モデル

## 1. 研究開始当初の背景

近年の高等教育においては、アクティブ・ラーニングの重要性が意識され、アクティブ・ラーニングの一学習形態としてグループ活動を取り入れた授業が運営されてきている。アクティブ・ラーニングにおける評価は学習者の行為(パフォーマンス)に対してなされることが多く、通常、行為者とは別の者が評価する。

パフォーマンス評価を能力測定の観点から考えるといくつかの問題が存在している。多数の課題を与えにくい状況で妥当性(評価対象の能力を的確に捉える程度)や信頼性(偶然の要因による影響の少なさ)をどう確保するのか、評価者特性(評価者個人の見目の厳しさや鋭さなど)をどのように扱うのかといった問題点が指摘される。

そのような問題を解決する方法として項目反応モデルが近年利用されている(植野・ソナム・岡本・永岡; 2008, 宇佐美; 2010, Uto & Ueno; 2016 など)。パフォーマンス評価データを項目反応モデルによって分析することで(1)各評価者の持つ評価基準が互いに異なる場合にも、同一尺度(ものさし)上での一貫性のある評価を行うことができる、(2)各評価者の信頼性を考慮した学習者の評価を行うことができる、(3)欠測値をもつデータからモデルの各母数を容易に推定できる、といった利点(植野ら; 2008)が享受でき、質の高い能力測定の実現が試みられている。

ただし、先行研究ではグループは考慮せず、各評価者が数十人から数百人の学習者を評価する状況の検討に留まっていた。グループ活動の状況において精度よく能力を測定するには、複数のグループ、教員等も含めた評価体制を適切に用意して評価データを得る必要がある。

## 2. 研究の目的

本研究では高等教育におけるグループ活動の現実的な制約を考慮した上で、グループ活動に基づく評価データを項目反応モデルによって適切に分析するために必要となるグループ構成や評価体制について検討し、実用的な方法として提案する。具体的には、以下の項目について明らかにすることを目指した。

1. グループを基にした評価データに利用可能な項目反応モデル
2. グループを基にした評価による能力測定精度への影響
3. 授業への適用を通じた、実用面での手法の評価、改良点

1については、グループを基にした評価の現実的な制約を考慮し、既存の項目反応モデルの特徴に基づいて、適用可能性の高いモデルを同定する。2については、グループに対してどれくらいの評価機会を提供すれば各参加者の能力を精度よく測定できるかを示す。なお、グループに加えて教員等の評価への関与についても考慮する。3については、1と2を通して最も適用の可能性が高いモデルとグループによる評価体制に基づいて授業を展開し、実用にあたっての手法の評価を行い、問題点とその解決のための改善法を提示する。

## 3. 研究の方法

研究の目的に掲げた三つの項目のうち、1については、パフォーマンス評価に関連して提案されたモデルを中心に、項目反応モデルに関する文献を調べ、それらの特徴(扱える値の形式、モデル推定の容易さなど)の整理を行う。また、並行して高等教育において実施されているグループ活動の実践報告について調べ、グループ活動におけるピアアセスメントに関する実践上の制約(平均的な授業参加者数、グループ人数など)を明らかにする。そして、両者の結果を統合してグループを基にした評価データの分析において利用可能な項目反応モデルを同定する。

2については、文献研究を通して決定された項目反応モデルを利用して、コンピュータによるシミュレーション研究を行う。まずは、項目モデルを推定するためのシステムを実装する。実装には統計解析ソフト R を用いる。その後、グループ活動の実践上の制約を考慮して、授業参加者数やグループ人数について少数の条件を設定する。そのもとでさらに、グループに基づく評価の仕方や教員等のアセスメントへの関与、評価者の特徴等について複数の条件を設定する。なお、グループの構成には実験計画法を参考とする。そして、条件ごとに多数のシミュレーションデータを発生させて項目反応モデルによって推定を繰り返す。多数の推定結果について、データ発生のためにあらかじめ設定された各種母数の値との一致度や乖離度の観点からどの程度の精度を持って推定できるかを評価する。

3については、シミュレーション研究によって明らかになった評価体制を実際の授業運営に適用してデータ分析を行い、その振り返りにより実際上の問題点を洗い出す。

## 4. 研究成果

研究の目的に掲げた三つの項目のうちの 3 については今後の課題として残されており、ここでは 2 についての成果を中心に報告する。

まず、本研究において主に取り上げる項目反応モデルは、Uto and Ueno(2016)で提案された、段階反応モデルに評価者の厳しさと一貫性の母数を組み込んだモデルとした。なお、決定には以

下の点を考慮した。

- ・大学での評価ではルーブリック等を使用した3から5段階の評価が行われやすいこと
- ・評価者は評価にそこまで習熟していない学生が中心であり、評価者における評価の厳しさと評価の一貫性の二つの母数を扱えること
- ・評価者が比較的多い状況を想定するため、推定する母数がそこまで多くなり、評価者の増加に伴って推定精度が大きく低下しないこと

研究の目的に掲げた三つの項目のうちの2の実現のため、まず、項目反応モデル推定のためのシステムをRによって構築した。当初はMCMC法に基づくベイズ推定を想定したものの、モデリングの複雑さ、推定ごとの推定値の変動、欠測が生じた場合の扱いにくさを避けるために、項目反応モデルの推定で一般に利用される最尤推定のプログラムを作成することとした。厳密には、EMアルゴリズムによる最適化を用いた周辺最尤推定法であり、完全情報最尤推定によって欠測データに対応できるようにした。

能力測定精度への影響の検討にあたってはシミュレーション研究を行った。具体的には、

- (a) 被評価者数 (20, 40, 60)
- (b) 欠測率 (0, 0.25, 0.50)

の条件を設定し、12個の課題に対してグループ（一グループ5名）単位で相互に5段階での評価が行われるという状況を想定した。なお、評価者数は教員一名を加えた21、41、61で、教員は被評価者全員を評価することとした。また、欠測のさせ方はグループに対する不完備ブロックデザインによるものとした。

発生させた300個のシミュレーションデータを開発プログラムによって分析し、モデルが問題なく推定できた比率（収束率）や各種母数の推定精度について検討を行った。まず、収束率の結果（表1）を見ると、被評価者数の水準によらず0.25ほどの欠測でも十分であり、欠測率0.5でも欠測率0の場合から大きく低下するものではないことが分かった。

表 1. 収束率

被評価者数 \ 欠測率	0	0.25	0.5
20	98.333	98.333	95.333
40	97.667	96.667	96.333
60	97.000	97.333	94.667

続いて、被評価者の能力母数の推定精度については、300個のデータについて求めた推定値と真値の相関係数を求めて、それらの最小値、平均値、最大値の点から比較検討した。表2より、評価者数20の条件では欠測率0.5で相関係数が著しく小さくなってしまふ場合があるため、欠測率0.25のほうが精度よく推定が可能であることが分かった。また、被評価者数40および60条件では、欠測率0.5程度でも推定精度を大きく下げずに能力値を推定できると言える。

表 2. 能力母数の推定

被評価者数 \ 欠測率	最小値			平均値			最大値		
	0	0.25	0.5	0	0.25	0.5	0	0.25	0.5
20	0.9587	0.9364	0.8487	0.9842	0.9796	0.9676	0.9972	0.9968	0.9914
40	0.9835	0.9726	0.9699	0.9924	0.9894	0.9849	0.9975	0.9964	0.9942
60	0.9910	0.9881	0.9803	0.9948	0.9932	0.9900	0.9975	0.9970	0.9959

能力母数の推定にあたっては、項目母数および評価者母数を予め推定する必要がある。今回の項目反応モデルでは、項目母数として項目数分の識別力母数、各項目につき（段階数-1）個の困難度母数が、また、評価者母数として、評価者数分の一貫性母数および厳しさ母数が含まれる。項目母数については、300個のシミュレーションデータから計算された推定値の平均値と真値との乖離をまとめた。

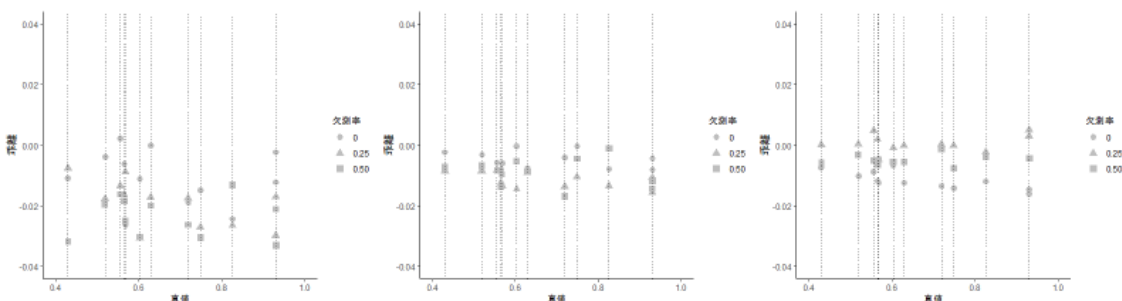


図 1. 識別力母数のシミュレーション結果（左から被評価者 20、40、60）

識別力母数については、被評価者数 20 条件は他条件に比べて真値からの乖離が大きいことは明確であるが、欠測率の違いは推定精度において一貫した傾向として現れていない(図 1)。そして、困難度母数については、被評価者数が大きくなるほど真値からの乖離が明らかに小さくなるものの、こちらについても欠測率の違いは推定精度に反映されていない(図 2)。

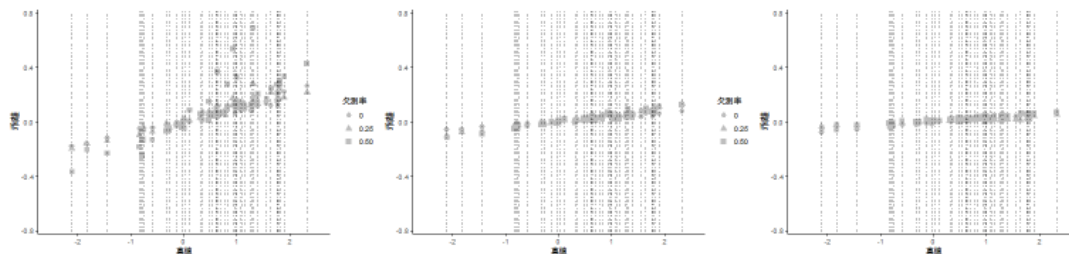


図 2. 困難度母数のシミュレーション結果 (左から被評価者 20、40、60)

次に、評価者母数については、最大 61 人分(被評価者数 60 のとき、評価者数は教員一名分を加えた 61)の母数を扱うことになることから結果の見にくさを避けるため、300 個のシミュレーションデータから計算された推定値の平均値と真値との相関係数を算出した(表 3)。その結果、評価者母数はどちらの母数についてもいずれの条件でも精度高く推定できることが分かった。

表 3. 評価者母数のシミュレーション結果

被評価者数 \ 欠測率	一貫性			厳しさ		
	0	0.25	0.5	0	0.25	0.5
20	0.9987	0.9985	0.9975	0.9957	0.9931	0.9908
40	0.9991	0.9991	0.9990	0.9964	0.9964	0.9956
60	0.9995	0.9995	0.9991	0.9985	0.9979	0.9967

以上のことを総合すると、今回の設定でのシミュレーション結果に関して言えば

- ・被評価者 20 名(5 名×4 グループ)の場合  
一つの課題につき、あるグループは別の 3 グループ(15 名分)を評価する
- ・被評価者 40 名(5 名×8 グループ)の場合  
一つの課題につき、あるグループは別の 4 グループ(20 名分)を評価する
- ・被評価者 60 名(5 名×12 グループ)の場合  
一つの課題につき、あるグループは別の 6 グループ(30 名分)を評価する

ことで、グループ構成に基づきながら、評価の労力を押さえて精度を大きく損なわずに能力母数を推定できるということになるだろう。しかしながら、被評価者数 40、60 の場合、能力母数の推定結果から見れば、欠測率をさらに増加させても実際場面で許容される精度での推定が可能だと思われる。この点に関するさらなる検討および授業場面への適用は今後の課題としたい。

## 文献

- ピアアセスメントにおける評価者特性を考慮した項目反応理論、植野真臣・ソナムアンポク・岡本敏雄・永岡慶三、電子情報通信学会論文誌、J91-D、377-388、(2008)。  
採点者側と受験者側のバイアス要因の影響を同時に評価する多値型項目反応モデル-MCMC アルゴリズムに基づく推定-、宇佐美慧、教育心理学研究、58、163-175、(2010)。  
Item Response Theory for Peer Assessment, M. Uto and M. Ueno, IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, 9, 157-170, (2016)。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計1件

1. 著者名 川端 一光、岩間 徳兼、鈴木 雅之	4. 発行年 2018年
2. 出版社 株式会社オーム社	5. 総ページ数 448
3. 書名 Rによる多変量解析入門	

〔産業財産権〕

〔その他〕

研究成果（グループ内ピアアセスメントに基づく能力測定のための実用的手法の提案） <a href="https://psychometrics-edu.jimdofree.com">https://psychometrics-edu.jimdofree.com</a>
--

6. 研究組織

氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
---------------------------	-----------------------	----