

令和 3 年 5 月 10 日現在

機関番号：12612

研究種目：若手研究(A)

研究期間：2017～2020

課題番号：17H04726

研究課題名(和文) パフォーマンス評価のための項目反応理論とその運用モデルの開発

研究課題名(英文) Item response theory for performance assessment and its applications

研究代表者

宇都 雅輝 (Masaki, Uto)

電気通信大学・大学院情報理工学研究科・准教授

研究者番号：10732571

交付決定額(研究期間全体)：(直接経費) 19,200,000円

研究成果の概要(和文)：人間評価者の主観的な採点を伴う、記述・論述式試験や語学試験などのパフォーマンス評価では、評価者のバイアスの影響を取り除いて受検者の能力を測定することが重要となる。本研究では、そのような能力測定を実現できる手法のひとつとして実用化が期待されている「評価者特性を考慮した項目反応理論」に関する以下のテーマに取り組んだ。1) 異質評価者に頑健な新たなモデルの開発、2) 最先端のMCMCアルゴリズムによるパラメータ推定の効率化、3) 様々な実践場面への適用を想定したモデル拡張と本技術の応用・運用方法の検討、および性能評価。これらの3つの研究テーマに関し、当初研究計画を上回る成果が得られた。

研究成果の学術的意義や社会的意義

本研究で開発する技術は、記述・論述式試験や語学試験、実技試験をはじめ、オンライン学習環境における学習者同士の相互評価や入社試験・人事考課で行われる面接試験など、様々なパフォーマンス評価に活用できる。本技術は、パフォーマンス評価の基礎理論として広く活用される可能性が高く、その学術的・社会的インパクトは大きい。また、本研究の適用範囲は、教育評価分野に限定されない。本技術は、オンラインショップにおける商品のレーティングやクラウドソーシングの品質評価など、評価者を伴う様々な評価データに広く適用可能であり、様々な応用タスクの性能向上に寄与すると期待できる。

研究成果の概要(英文)：In performance assessment where human raters subjectively grade examinees' performances, it is important to estimate examinee ability while removing effects of raters' biases. The purpose of this study is to develop and evaluate item response theory (IRT) models for performance assessment that can estimate examinee ability while removing rater bias effects. Concretely, we conducted the following three studies. 1) Development of a new IRT model incorporating various rater characteristic parameters to improve robustness against aberrant raters. 2) Development of an efficient Markov chain Monte Carlo method using the No-U-Turn sampler algorithm for the proposed IRT model. 3) Extensions and applications of the proposed method for various performance assessment situations.

研究分野：教育工学

キーワード：パフォーマンス評価 項目反応理論 eテスト e テスティング 統計的自然言語処理

## 1. 研究開始当初の背景

近年、入学試験や入社試験を始めとする様々な評価場面において、論述式テストやスピーキングテストなどのパフォーマンス評価が広く活用されている。わが国で将来導入予定の「大学入学希望者学力評価テスト」においても論述・記述式テストの採用が検討されており、パフォーマンス評価の重要性は今後ますます増加すると予測できる。一方で、正誤が一意に定まる客観式テストとは異なり、パフォーマンス評価では、得られる評点が評価者の特性(例えば、評価者の甘さ・辛さ)に強く依存し、これにより能力評価の信頼性が低下する問題がある。そのため、パフォーマンス評価では、評価者のバイアスをできる限り取り除いて受検者の真の能力を精度よく推定することが重要課題となる。

この問題を解決するアプローチの一つとして、評価者特性を考慮して受検者の能力を推定できる数理モデルが開発されてきた。具体的には、数理モデルを用いたテスト理論の一つである項目反応理論を拡張した技術として、評価者特性パラメータを付与した項目反応モデルが多数開発され、評価の信頼性改善に有効であることが示されてきた。しかし、既存モデルには、評価者数が増加するとパラメータ数が急速に増加し、パラメータ推定の精度が低下する問題があった。この問題を解決するために、申請者は、低次元の評価者特性パラメータを採用した新たなモデルを開発し、評価者数が増加する場面において世界で最も高精度な能力評価を実現した。

申請者の手法は様々なパフォーマンス評価場面で実用化が期待されており、すでに医療系大学間共用試験実施評価機構などと共に実用化に向けた取り組みを開始しているが、その本格的な実用化のためには次の課題を解決する必要がある。

- (1) 数百人単位の評価者を要する大規模試験では、異質な採点を行う評価者(例えば、独自の評価基準で採点を行う評価者や、多くの受検者に対して同一の評点を与える評価者など)が混在する可能性が高い。しかし、既存モデルでは評価の甘さ・辛さなどの単純な評価者特性しか考慮できないため、異質性の強い評価者が混在する場合には評価の信頼性が低下する問題がある。
- (2) 既存研究の多くでは、モデルのパラメータ推定法としてマルコフ連鎖モンテカルロ法(MCMC)を採用している。MCMC アルゴリズムとしては、実装が容易な Metropolis Hastings within Gibbs sampling 法が利用されてきたが、このアルゴリズムは推定効率が悪く、評価者数や受検者数が増加する大規模試験に適用すると、実用的な時間で推定を行うことが困難になる。
- (3) 現実の評価場面は、評価者数や受検者数、評価者の質、課題内容、評価基準などがそれぞれ異なり、それらの条件によって本技術の有効性や活用方法も異なる。しかし、これまで、本技術は、いくつかの限定的な場面でしか評価がなされておらず、現実の多様な評価場面における有効性や課題などは明らかにされていない。また、状況に応じた活用方法も整理されておらず、非専門家による効果的な運用は困難である。

## 2. 研究の目的

本研究では、上記の課題を解決することを目標として、次の3つの研究を行う。

### (1) 異質評価者に頑健な項目反応モデルの開発

本研究では、異質な評価者が混在しても頑健な能力評価を実現できる新たな項目反応モデルを開発する。具体的には、異質評価者の反応特性として、1) 評価の甘さ・辛さ、2) 非一貫性、3) 尺度範囲の制限(少数の特定の評点を過度に使用する特性)の3つを想定し、これらに対応する評価者パラメータを付与した項目反応モデルを開発する。これらの評価者特性は、統計的には1) 評価者による評点の平均、2) 真の能力と観測評点との相関、3) 各評点に対する評価者の反応分布の分散、に対応している。一般的な項目反応モデルでは、その反応分布の形状が、最大でも上記の3つの統計量に対応する項目パラメータによって決定されることから、これらの評価者パラメータを付与した提案モデルは、評価者の異質反応の影響を最大限に考慮できる最も頑健な項目反応モデルと解釈できる。

### (2) ハイブリッド・モンテカルロ・アルゴリズムに基づく効率的なパラメータ推定手法の開発

本研究では、提案モデルのパラメータ推定法として、効率の良いMCMCアルゴリズムとして近年注目されているハイブリッド・モンテカルロ法に基づくアルゴリズムを開発し、パラメータ推定の効率改善を目指す。ハイブリッド・モンテカルロ法には複数の種類が存在し、問題に応じてその計算効率や特性が異なるため、本研究では、実験を通して提案モデルのための最適なハイブリッド・モンテカルロ・アルゴリズムを探索し、採用する。

### (3) 様々な実データへの適用を通じた提案技術の性能評価

本研究では、現実の多様なパフォーマンス評価場面を想定して実データを収集し、それらのデータへの本技術の適用を通してその性能と課題を明らかにする。さらに、各評価場面の特徴に応じた本技術の活用法を整理する。

## 3. 研究の方法

上記の3つの研究目標に対して次の方法で研究を行なった。

### (1) 異質評価者に頑健な項目反応モデルの開発

異質評価者に対して頑健な項目反応モデルの開発は、代表的な多値型項目反応モデルとして知られる一般化部分採点モデルの拡張モデルとして定式化した。具体的には、一般化部分採点モデルに、評価者の厳しさ、非一貫性、尺度範囲の制限に対応するパラメータを付与したモデルとして定式化した。また、提案モデルの有効性を評価するために、関連モデルの実装も行った。シミュレーション実験では、実験条件（評価者数、受検者数、異質評価者数、異質反応のパターンなど）を変更しながら、既存モデルとの比較を通して提案モデルの性能を評価した。さらに、実データ適用を通じた性能評価も実施した。実験設定としては、異質評価者が混在しやすい場面を想定し、学習者同士によるピアアセスメントのデータを用いた。

### (2) ハイブリッド・モンテカルロ・アルゴリズムに基づく効率的なパラメータ推定手法の開発

MCMC によるパラメータの推定効率を改善するために、上記の提案モデルのパラメータ推定アルゴリズムとして、ハイブリッド・モンテカルロを用いた MCMC 法を開発した。ハイブリッド・モンテカルロには、1)ハミルトニアン・モンテカルロ、2)その拡張の No-U-Turn サンプラ、3) Metropolis Hastings の拡張である Metropolis-adjusted Langevin アルゴリズム、4) スライスサンプリングなど、複数の種類が存在する。そこで、本研究では、提案モデルのパラメータ推定法として、それぞれのハイブリッド・モンテカルロ・アルゴリズムを検討した。結果として、実装が容易で、モデル拡張にも柔軟に対応できるとともに、サンプリング効率や精度にも優れることから、Stan を利用した No-U-Turn サンプラを採用し、その性能を評価した。

### (3) 様々な実データへの適用を通じた提案技術の性能評価と有効な運用モデルの作成

複数の異なる評価場面を想定し、各場面にあわせたモデル拡張とモデルの運用方法を整理した。また、様々な実践場面で収集した実データを用いて提案技術の性能評価を行なった。具体的には、1) 中・小規模な論述式試験、2) グループワーク形式のパフォーマンス評価、3) e ラーニングにおける学習者同士の相互評価、4) 医療系大学間共用試験における実技試験、で得られたデータへの適用を行なった。また、パフォーマンス評価の実践家と専門家へのインタビューを通して、実践場面の特徴と課題を抽出し、それらの特性に対応した提案技術の活用方法を整理した。

## 4. 研究成果

上記の3つの研究目標に対し、それぞれ以下の成果が得られた。

### (1) 異質評価者に頑健な項目反応モデルの開発

研究計画に従い、異質評価者に頑健な新たな項目反応モデルの開発を行った。提案モデルは、一般化部分採点モデルに対して、評価者の評価の一貫性、厳しさ、尺度範囲の制限の特性を反映したパラメータを付与したモデルとして定式化した。シミュレーション実験と実データ実験から、提案モデルは、様々な特性の評価者が存在する場合に既存モデルより高い性能を示すことが明らかとなった。本研究の成果の一部は、2018年に電子情報通信学会論文誌 D とトップ国際会議の Artificial intelligence in Education (AIED 2018) に掲載され、最終成果が 2020 年に英文論文誌 Behaviormetrika, Springer に掲載された。

### (2) ハイブリッド・モンテカルロ・アルゴリズムに基づく効率的なパラメータ推定手法の開発

上記の提案モデルを含め、これまでに開発してきた評価者パラメータ付きの項目反応モデルのパラメータ推定法として、最先端の MCMC アルゴリズムの一つである No-U-Turn サンプラに基づくハミルトニアンモンテカルロ法を開発し、その性能を評価した。本研究に関連する成果は、英文論文誌 Behaviormetrika, Springer と電子情報通信学会論文誌 D に 2020 年に掲載された。

### (3) 様々な実データへの適用を通じた提案技術の性能評価と有効な運用モデルの作成

本研究では、様々な実評価場面への適用を想定し、モデル拡張とモデルの利用方法の検討を行うとともに、実試験での実証実験を行った。具体的には、次の研究を行った。

ループリックを用いた評価のための拡張モデル開発

言語処理技術を統合した項目反応モデルの開発  
大規模評価環境における評価者割り当て最適化手法の開発  
適切なモデル活用のための試験デザインの設計  
医療系大学間共用試験の実データ適用を通じた実証評価

上記の各細目に対する研究成果は次の通りである。

#### ループリックを用いた評価のための拡張モデル開発

パフォーマンス評価ではループリックと呼ばれる評価基準表を用いて採点を行うことがある。このような場合、評価者はループリックに基づいて複数の採点項目で評点をつけることになる。そこで本研究では、個別の評価項目の特性も考慮できるように(1)で開発したモデルを発展させることで、ループリックを用いた評価データに直接に適用できる新たなモデルを開発した。本研究の成果は、2019年と2020年に電子情報通信学会論文誌Dに採択されるとともに、それを発展させたモデルがトップ国際会議のAIEDに2021年に採択された。

#### 言語処理技術を統合した項目反応モデルの開発

小論文やレポートなどの採点場面を想定し、統計的自然言語処理技術を活用した拡張モデルの開発を行った。具体的には、評価者が与える評点データに加えて、受検者が執筆した小論文の文章情報も同時に活用して能力推定を行うことができる項目反応モデルを開発し、受検者の能力測定精度を向上できることを示した。統計的自然言語処理技術としては、まず潜在ディリクレ配分法を利用したモデルを開発したのち、それを発展させる形で深層学習を活用したモデルを開発した。潜在ディリクレ配分法を利用したモデルの研究成果は、2019年に国際会議AIEDで発表し、同年、電子情報通信学会論文誌Dに掲載された。深層学習を活用したモデルの成果については、2020年に国際会議AIEDに採択され、論文賞を受賞するとともに、2021年に電子情報通信学会論文誌Dに掲載された。

#### 大規模評価環境における評価者割り当て最適化手法の開発

本研究で開発した技術の応用として、項目反応理論と整数計画法を用いた最適評価者選択手法の研究を行なった。具体的には評価者特性パラメータを付与した項目反応モデルを用いて、各評価者が各受検者をどの程度の精度で評価できるかを「情報量」として定量化し、その情報量を最大化するように各受検者に評価者を割り当てる手法を開発した。実験により、本技術を用いることで、少ない数の評価者でより高精度な能力測定が実現できることを示した。本研究の成果は、国際会議AIEDに2017年に採択され、2018年には電子情報通信学会論文誌Dに、2020年にはIEEE Transactions on Learning Technologiesに掲載された。

#### 適切なモデル活用のための試験デザインの設計

評価者特性パラメータを付与した項目反応モデルを用いて異なるパフォーマンステストの結果を比較する場合には「等化」と呼ばれる手続きが必要となる。しかし、どのようにテストを設計すれば高精度に等化できるかについては、これまで十分には明らかにされてこなかった。そこで、本研究では、多様な条件を想定したシミュレーション実験を通して、等化精度に影響を与える要因を明らかにするとともに、望ましいデザインについてのエビデンスを示した。従来、高精度な等化にはテスト間に最低5人の共通評価者と5つの共通課題が必要とされてきたが、本研究の結果、これらより大幅に少ない共通評価者・課題で等化可能である場合も多いことが示された。本研究の成果は、2018年に電子情報通信学会論文誌Dに掲載され、最終成果は2020年にBehavior Research Methods, Springerに掲載された。

#### 医療系大学間共用試験の実データ適用を通じた実証評価

本研究で開発した技術の実用化を目指し、リクルートキャリアや医療系大学間共用試験などでの実証実験を進めた。中でも、全国の医療系大学の学生が受験する医療系大学間共用試験では、試験信頼性向上専門部会が組織されており、研究代表者を含むテスト理論・教育評価の専門家や作問委員、運営委員などを含んで評価の信頼性向上のために実施データを詳細に分析し、今後の改善に活かすPDCAサイクルを実践している。医療系大学間共用試験にはOSCEと呼ばれる受検者の実技を評価者が採点する試験も実施されており、この試験は将来的に国家試験となることが計画されている。研究代表者の技術はOSCEの信頼性向上のための技術としての利用が期待されている。それに伴い研究代表者は、2016年からOSCEの実データを用いて本技術の実証実験を進めるとともに、全国の医療系大学に向けた講演を毎年行い、本技術の普及・浸透や試験デザインの改善に関する提言などを行ってきた。

## 5. 主な発表論文等

〔雑誌論文〕 計13件（うち査読付論文 12件 / うち国際共著 1件 / うちオープンアクセス 12件）

1. 著者名 Masaki Uto	4. 巻 -
2. 論文標題 Accuracy of performance-test linking based on a many-facet Rasch model	5. 発行年 2020年
3. 雑誌名 Behavior Research Methods, Springer	6. 最初と最後の頁 in press
掲載論文のDOI (デジタルオブジェクト識別子) 10.3758/s13428-020-01498-x	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Masaki Uto, Duc-Thien Nguyen, Maomi Ueno	4. 巻 13
2. 論文標題 Group optimization to maximize peer assessment accuracy using item response theory and integer programming	5. 発行年 2020年
3. 雑誌名 IEEE Transactions on Learning Technologies, IEEE Computer Society	6. 最初と最後の頁 91-106
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TLT.2019.2896966	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Uto Masaki, Miyazawa Yoshimitsu, Kato Yoshihiro, Nakajima Koji, Kuwata Hajime	4. 巻 30
2. 論文標題 Time- and Learner-Dependent Hidden Markov Model for Writing Process Analysis Using Keystroke Log Data	5. 発行年 2020年
3. 雑誌名 International Journal of Artificial Intelligence in Education	6. 最初と最後の頁 271 ~ 298
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s40593-019-00189-9	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Masaki Uto, Maomi Ueno	4. 巻 47/2
2. 論文標題 A generalized many-facet Rasch model and its Bayesian estimation using Hamiltonian Monte Carlo	5. 発行年 2020年
3. 雑誌名 Behaviormetrika	6. 最初と最後の頁 469-496
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s41237-020-00115-7	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 宇都雅輝	4. 巻 37/1
2. 論文標題 テスト理論と人工知能に基づくパフォーマンス評価の新技术	5. 発行年 2020年
3. 雑誌名 教育システム情報学会論文誌	6. 最初と最後の頁 8--18
掲載論文のDOI (デジタルオブジェクト識別子) 10.14926/jsise.37.8	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 宇都雅輝・植野真臣	4. 巻 103/05
2. 論文標題 ルーブリック評価における項目反応理論	5. 発行年 2020年
3. 雑誌名 電子情報通信学会論文誌D	6. 最初と最後の頁 459--470
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2019JDP7065	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 八木嵩大・宇都雅輝	4. 巻 J102-D
2. 論文標題 パフォーマンス評価における多次元項目反応モデル	5. 発行年 2019年
3. 雑誌名 電子情報通信学会論文誌D	6. 最初と最後の頁 708~720
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2019JDP7018	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 宇都雅輝	4. 巻 J102-D
2. 論文標題 論述式試験における評点データと文章情報を活用した項目反応トピックモデル	5. 発行年 2019年
3. 雑誌名 電子情報通信学会論文誌D	6. 最初と最後の頁 553~566
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2019JDP7007	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 宇都雅輝	4. 巻 101
2. 論文標題 評価者特性パラメータを付与した項目反応モデルに基づくパフォーマンステストの等化精度	5. 発行年 2018年
3. 雑誌名 電子情報通信学会論文誌D	6. 最初と最後の頁 895 ~ 905
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2017LEP0027	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Masaki Uto, Maomi Ueno	4. 巻 4
2. 論文標題 Empirical Comparison of Item Response Theory Models with Rater's Parameters	5. 発行年 2018年
3. 雑誌名 Heliyon, Elsevier	6. 最初と最後の頁 1 ~ 32
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.heliyon.2018.e00622	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Sebastien Louvigne, Masaki Uto, Yoshihiro Kato, Takatoshi Ishii	4. 巻 45
2. 論文標題 Social constructivist approach of motivation: social media messages recommendation system.	5. 発行年 2018年
3. 雑誌名 Behaviormetrika, Springer	6. 最初と最後の頁 133 ~ 155
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s41237-017-0043-7	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Nguyen Duc Thien・宇都雅輝・植野真臣	4. 巻 101
2. 論文標題 ピアセサメントにおける項目反応理論を用いたグループ構成最適化	5. 発行年 2018年
3. 雑誌名 電子情報通信学会論文誌D	6. 最初と最後の頁 431-445
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2017JDP7040	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 宇都雅輝・植野真臣	4. 巻 101
2. 論文標題 ピアアセスメントにおける異質評価者に頑健な項目反応理論	5. 発行年 2018年
3. 雑誌名 電子情報通信学会論文誌D	6. 最初と最後の頁 211-224
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2017JDP7055	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

[学会発表] 計37件(うち招待講演 5件/うち国際学会 6件)

1. 発表者名 新田森・宇都雅輝
2. 発表標題 ループリック評価における多次元4相型項目反応モデルの提案
3. 学会等名 人工知能学会 第91回 先進的学習科学と工学研究会
4. 発表年 2021年

1. 発表者名 宇都雅輝
2. 発表標題 OSCEにおけるIRT利用について
3. 学会等名 公益社団法人医療系大学間共用試験実施評価機構 試験信頼性向上部会第18回講演会(招待講演)
4. 発表年 2020年

1. 発表者名 宇都雅輝
2. 発表標題 パフォーマンス型試験の課題とその解決に向けた人工知能研究の現在
3. 学会等名 SCATE-21研究会(招待講演)
4. 発表年 2020年



1. 発表者名 Masaki Uto
2. 発表標題 Rater-effect IRT model integrating supervised LDA for accurate measurement of essay writing ability.
3. 学会等名 第19回情報科学技術フォーラム FIT2020, 情報処理学会(トップコンファレンスセッション)(招待講演)
4. 発表年 2020年

1. 発表者名 宇都雅輝・植野真臣
2. 発表標題 ループリック評価のための多相項目反応モデル
3. 学会等名 行動計量学会第48回大会
4. 発表年 2020年

1. 発表者名 宇都雅輝・森本剛・野上康子・内田啓子・吉田素文・片桐瑞希・葛西一貴・川上智史・江藤一洋・齋藤宣彦・仁田善雄
2. 発表標題 OSCEにおける項目反応理論の適用
3. 学会等名 第52回医学教育学会全国大会
4. 発表年 2020年

1. 発表者名 岡野将士・宇都雅輝
2. 発表標題 アノテータのバイアスを考慮した記述・論述式自動採点手法
3. 学会等名 言語処理学会第27回年次大会
4. 発表年 2021年

1. 発表者名 Masaki Uto, Yikuan Xie, Maomi Ueno
2. 発表標題 Neural Automated Essay Scoring Incorporating Handcrafted Features
3. 学会等名 International Conference on Computational Linguistics (国際学会)
4. 発表年 2020年

1. 発表者名 Masaki Uto, Masashi Okano
2. 発表標題 Robust neural automated essay scoring using item response theory
3. 学会等名 International Conference on Artificial Intelligence in Education (国際学会)
4. 発表年 2020年

1. 発表者名 Masaki Uto, Yuto Uchida
2. 発表標題 Automated short-answer grading using deep neural networks and item response theory
3. 学会等名 International Conference on Artificial Intelligence in Education (国際学会)
4. 発表年 2020年

1. 発表者名 Masaki Uto
2. 発表標題 Rater-effect IRT model integrating supervised LDA for accurate measurement of essay writing ability.
3. 学会等名 International Conference on Artificial Intelligence in Education (国際学会)
4. 発表年 2019年

1. 発表者名 Masaki Uto, Maomi Ueno
2. 発表標題 Item response theory without restriction of equal interval scale for rater's score
3. 学会等名 International Conference on Artificial Intelligence in Education (国際学会)
4. 発表年 2018年

1. 発表者名 宇都雅輝
2. 発表標題 レイティングデータとテキスト情報を用いて受験者の能力を推定する項目反応トピックモデルの提案
3. 学会等名 人工知能学会 第85回先進的学習科学と工学研究会 (ALST)
4. 発表年 2019年

1. 発表者名 岡野将士・宇都雅輝
2. 発表標題 アノータタのバイアスに頑健な小論文自動採点手法
3. 学会等名 言語処理学会第26回年次大会
4. 発表年 2020年

1. 発表者名 岡野将士・宇都雅輝
2. 発表標題 評価者バイアスに頑健な小論文自動採点手法
3. 学会等名 第88回 人工知能学会 先進的学習科学と工学研究会
4. 発表年 2020年

1. 発表者名 宇都雅輝
2. 発表標題 OSCEにおけるIRT利用について
3. 学会等名 公益社団法人医療系大学間共用試験実施評価機構 試験信頼性向上部会第16回講演会（招待講演）
4. 発表年 2019年

1. 発表者名 宇都雅輝
2. 発表標題 項目反応理論と機械学習技術を用いた小論文評価手法
3. 学会等名 日本テスト学会第17回大会
4. 発表年 2019年

1. 発表者名 岡野将士・宇都雅輝
2. 発表標題 評価者バイアスを考慮した小論文自動採点手法
3. 学会等名 情報処理学会第241回自然言語処理研究発表会
4. 発表年 2019年

1. 発表者名 八木高大・宇都雅輝
2. 発表標題 パフォーマンス評価における多次元項目反応モデルの提案と評価
3. 学会等名 人工知能学会 第85回先進的学習科学と工学研究会（ALST）
4. 発表年 2019年

1. 発表者名 八木嵩大・宇都雅輝
2. 発表標題 レビューアー特性を考慮した多次元項目反応モデルの提案
3. 学会等名 第21回情報論的学習理論ワークショップ (IBIS2018)
4. 発表年 2018年

1. 発表者名 宇都雅輝
2. 発表標題 OSCEにおけるIRT利用について
3. 学会等名 公益社団法人医療系大学間教養試験実施評価機構 試験信頼性向上部会第14回講演会 (招待講演)
4. 発表年 2018年

1. 発表者名 八木嵩大・宇都雅輝
2. 発表標題 パフォーマンス評価における多次元項目反応モデル
3. 学会等名 日本テスト学会第16回大会
4. 発表年 2018年

1. 発表者名 宇都雅輝
2. 発表標題 ライティング能力を推定する項目反応トピックモデルの提案とエッセイ自動評価への応用
3. 学会等名 日本テスト学会第16回大会
4. 発表年 2018年

1. 発表者名 宇都雅輝
2. 発表標題 レイティングデータとテキスト情報を用いたライティング能力推定のための項目反応トピックモデル
3. 学会等名 行動計量学会第46回大会
4. 発表年 2018年

1. 発表者名 宇都雅輝・宮澤芳光・加藤嘉浩・中島功滋・桑田一
2. 発表標題 キーストロークデータに基づくライティングプロセス推定のための執筆者依存型隠れマルコフモデル
3. 学会等名 教育システム情報学会第43回全国大会
4. 発表年 2018年

1. 発表者名 八木嵩大・宇都雅輝
2. 発表標題 評価者パラメータを付与した多次元反応モデルの提案
3. 学会等名 行動計量学会第46回大会
4. 発表年 2018年

1. 発表者名 堤瑛美子・宇都雅輝・植野真臣
2. 発表標題 学習者の能力の時系列変化を考慮した項目反応モデル
3. 学会等名 行動計量学会第46回大会
4. 発表年 2018年

1. 発表者名 八木高大・宇都雅輝
2. 発表標題 パフォーマンス評価における多次元尺度を構成する項目反応モデル
3. 学会等名 人工知能学会 第83回先進的学習科学と工学研究会 (ALST)
4. 発表年 2018年

1. 発表者名 宇都雅輝・宮澤芳光・加藤嘉浩・中島功滋・桑田一
2. 発表標題 キーストロークデータからライティングプロセスを推定する執筆者依存型隠れマルコフモデルの提案
3. 学会等名 人工知能学会 第83回先進的学習科学と工学研究会 (ALST)
4. 発表年 2018年

1. 発表者名 宇都雅輝
2. 発表標題 項目反応理論のパフォーマンス評価への応用
3. 学会等名 公益社団法人医療系大学間教養試験実施評価機構 試験信頼性向上部会第13回講演会
4. 発表年 2018年

1. 発表者名 堤瑛美子・宇都雅輝・植野真臣
2. 発表標題 ダイナミックアセスメントのための隠れマルコフIRTモデル
3. 学会等名 第32回人工知能学会全国大会
4. 発表年 2018年

1. 発表者名 Masaki Uto, Nguyen Duc Thien, Maomi Ueno
2. 発表標題 Group optimization to maximize peer assessment accuracy using item response theory
3. 学会等名 International Conference on Artificial Intelligence in Education (国際学会)
4. 発表年 2017年

1. 発表者名 Nguyen Duc Thien, Masaki Uto, Maomi Ueno
2. 発表標題 A grouping method for optimizing peer assessment accuracy
3. 学会等名 日本教育工学会第33回全国大会
4. 発表年 2017年

1. 発表者名 宇都雅輝・植野真臣
2. 発表標題 異質評価者の特性を考慮した項目反応モデル
3. 学会等名 行動計量学会第45回大会
4. 発表年 2017年

1. 発表者名 宇都雅輝・Nguyen Duc Thien・植野真臣
2. 発表標題 ピアアセスメントの精度を最適化する自動グループ構成システム
3. 学会等名 教育システム情報学会第42回全国大会
4. 発表年 2017年



1. 発表者名 宇都雅輝・植野真臣
2. 発表標題 評価者特性を考慮した項目反応理論における等化手法
3. 学会等名 日本テスト学会第15回大会
4. 発表年 2017年

1. 発表者名 Nguyen Duc Thien・宇都雅輝・植野真臣
2. 発表標題 ピアアセスメントの精度を最適化するグループ構成手法
3. 学会等名 日本テスト学会第15回大会
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

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
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------