

不正行為を含む回答データが項目特性推定に与える影響 将来の受検者能力測定への影響波及の検討

甲斐 江里¹, 尾崎 幸謙²

¹株式会社リクルートマネジメントソリューションズ, ²筑波大学ビジネスサイエンス系

1. はじめに

項目反応理論 (IRT) に基づくテストの継続的運用において、問題項目の拡充は重要な課題である。新作項目の項目特性を推定するためには事前に十分な回答データが必要であり、その収集手段として本番テストに新作項目を採点除外項目として数問組み込む手法(杉山, 2021)がある。この手法は、テスト本番環境下で回答データを得られるため、予備テストのような別途の実施が不要だけでなく、より実態に即したデータ収集が可能となる利点がある。

しかし、近年普及している Web テストにおいては、特に無監視環境での受検時に不正行為リスクが生じる。Web テスト場面における不正行為とは、主に本人以外が受検する替え玉受検、本人以外が回答に協力する共謀、事前のテスト問題漏洩などにより得た事前知識を参照するカンニングが想定される。これらは本来、厳格な監視体制等により未然防止することが理想的であるが、実施環境の制約により完全な不正対策が困難な場合も存在する。このような不正行為がなされた場合、当該受検者の採点結果の妥当性低下やテスト全体の公平性阻害といった直接的な問題に加え、項目拡充プロセスを通じた項目特性の汚染(Eckerly, 2016)という間接的な影響も懸念される。すなわち、収集した回答データに不正行為者のデータが混入していた場合、その回答データを用いて推定された項目特性にバイアスが生じる。特に、カンニング行為は回答パターンに特異な歪みを生じさせるため、他の不正行為よりも項目特性推定への影響が大きいことが予想され、実際に歌川他(2025a)はカンニングによる項目特性推定への影響を検証し、高困難度項目がカンニングによるバイアスの影響を受けやすいことを示している。しかし、本研究で着目する手法での項目拡充を想定すると、採点除外項目は初出のため事前の項目漏洩リスクは採点項目よりも低い。一方で、項目特性推定時には採点項目と採点除外項目の回答データを合わせて同時推定することから採点項目の漏洩影響を受け、先行研究とは異なる結果となる可能性がある。さらにこのバイアスを有した新作項目が、将来的に不正行為を行わない適正な受検者に採点項目として出題されることで、受検者の能力推定にも系統的なバイアスをもたらす可能性がある。この視点は継続的なテスト運用という実務場面において重要な意味を持つが、先行研究では十分に検討されていない。

そこで本研究では、事前の採点項目の項目漏洩によるカンニング行為を行った回答データの混入が項目特性推定に与える影響を定量的に評価し、ひいては適正な受検者への間接的影響の度合いをシミュレーション研究により検証する。

2. 方法

2. 1 回答データ生成

まず、採点項目 40 項目と採点除外項目(新作項目)10 項目から構成される固定版テストを 4 版設定した。この際、4 版の採点項目は共通とし、採点除外項目のみを版間で変更した。各項目の項目特性真値は、2 パラメタ・ロジスティックモデル(式 1)に従って設定し、識別力パラメタ a を一様分布 $U(0.3, 1.5)$ 、困難度パラメタ b を標準正規分布 $N(0, 1)$ から独立に生成した。

$$P(\theta) = \frac{1}{1 + \exp(-Da(\theta - b))} \quad (1)$$

受検者 10,000 名分の能力特性真値 θ は標準正規分布 $N(0, 1)$ から生成した。カンニング行為の影響を検証するため、項目漏洩割合と不正行為者混入割合を設定し(表 1)、各割合に従って漏洩項目と不正行為者をランダムに選定した。不正行為者の再現には歌川他(2025b 印刷中)の手法を参考とし、以下の手

順で実施した。まず、能力特性真値 θ に基づいて、式(1)により全項目に対する正答確率を算出し、ベルヌーイ試行により回答(正誤)を生成する。次に、不正行為者が漏洩項目に誤答した場合、一定の確率（漏洩項目正答確率）で誤答を正答に変換する。これにより、カンニング行為による正答率の向上を模擬した。

ここで本シミュレーションの項目漏洩割合、不正行為者混入割合、漏洩項目正答確率の設定根拠を以下に示す。不正行為の実態について言及されている先行研究は少ないが、カンニング以外の不正行為を含む不正行為者の混入割合について、歌川他(2025a)では 1.9~5.1%、Cizek & Wollack(2016)では少なくとも 2.8~2.9%、Newton & Essex (2023)では 29.9%(コロナ前)~54.7%(コロナ後)、株式会社サーティファイによる調査(2024)では約 45%と報告されており、研究間で結果に大きな差異がある。この差異は混入割合の推定手法の違いに加え、対象テストの種類や実施環境の違いも影響していると考えられる。項目漏洩割合については参考となる先行研究が確認出来なかった。これらの状況を踏まえ、本研究では先行研究で報告された不正行為全体の混入割合を参考に、カンニング行為はその一部であることを考慮し、混入割合を 20%としたパターン①を含む 3 パターン（表 1）を設定した。パターン①②の項目漏洩割合については、不正行為者混入割合と同値としたが、パターン③については大量の項目漏洩を想定して 80%とした。

表 1 検証条件

パターン	項目漏洩割合	不正行為者混入割合	想定される状況
①	20%	20%	現実的に発生し得るカンニング行為による汚染状況
②	40%	40%	より深刻なカンニング行為による重度の汚染状況
③	80%	20%	一部受検者への大量項目漏洩による汚染状況

漏洩項目正答確率については、カンニングペーパーを使用した場合でも、その活用が不十分で誤答する可能性を考慮し、0.7 に設定した。

2. 2 採点除外項目の項目特性の推定

生成した回答データを用いて、各パターンについて各版の採点除外項目(合計 40 項目)の項目特性を推定した。項目特性の推定には R の `mirt` パッケージを使用し、採点項目の項目特性は 2.1 で設定した項目特性真値で固定の上、最尤推定法により識別力パラメタ a と困難度パラメタ b の値を算出した。推定精度を評価するため、ここで算出した項目特性値を 2.1 で生成した項目特性真値と比較した。

2. 3 適正受検者への影響評価

項目特性推定のバイアスが適正受検者に与える影響を評価するため、2.1 で生成した受検者とは独立に新たな受検者データを 10,000 名分生成した。これらの受検者の能力特性真値は 2.1 と同様に標準正規分布 $N(0, 1)$ から生成し、4 版分の採点除外項目である合計 40 項目に対する回答データを、2.1 で設定した項目特性真値を用いて生成した。次に、この回答データを用いた受検者能力推定において使用する項目特性値を、ケース A : 2.1 で設定した項目特性真値(不正行為データの影響なし)、ケース B : 2.2 で推定した項目特性値 (不正行為データの影響あり)として別々に能力特性を算出し、両者の結果を比較した。

3. 結果

不正行為データの混入が項目特性推定と能力特性推定に与える影響を検証した結果を図 1~3 に示す。各図は左から識別力パラメタ a 、困難度パラメタ b 、能力特性 θ の推定結果を表している。項目特性の推定結果（左・中図）では、横軸を 2.1 で設定した真値、縦軸を不正行為データを含む回答データからの推定値として、採点除外項目 40 項目の結果を示した。異なる形状は版の違いを表している。能力特性の推定結果（右図）では、横軸をケース A での能力推定結果（項目特性真値から算出した能力推定値）、縦軸をケース B での能力推定結果（不正行為の影響を受けた項目特性値から算出した能力推定値）として、10,000 名の結果を示した。また、各散布図には単回帰分析の結果を実線で示し、回帰式および決定係数、真値(横軸)と比較した際の推定値(縦軸)のバイアスと RMSE の値を図中に記載した。なおいずれのグラフにも、単回帰分析結果との比較のために 45 度線（破線）を引いた。

結果の概要は以下の通りである。パターン①（項目漏洩割合、不正行為者混入割合ともに 20%）では、項目特性・能力特性ともに回帰直線（実線）が 45 度線（破線）とほぼ一致しており、不正行為データの影響は軽微であった。しかし、パターン②（同 40%）では影響が顕著に現れた。識別力パラメタ a では回帰直線の傾きが 0.91 と 45 度線より小さくなり、過小推定の傾向が見られた。一方、困難度パラメタ b では傾きが 1.09、切片が 0.29 となり、全体的に真値より高く推定される結果となった。能力特性 θ も困難度パラメタ b と同様の傾向を示した。パターン③（項目漏洩割合 80%、不正行為者混入割合 20%）では、回帰直線の傾きは識別力パラメタ a で 0.71、困難度パラメタ b で 1.25、能力特性 θ で 1.32 と真値からの乖離がさらに拡大した。切片については、識別力パラメタ a で 0.03、困難度パラメタ b で 0.25、能力特性 θ で 0.25 となり、パターン②と同程度であったが、困難度パラメタ b と能力特性 θ では、それぞれ $b=-1.00$ 、 $\theta=-0.78$ 付近で 45 度線と交差した。

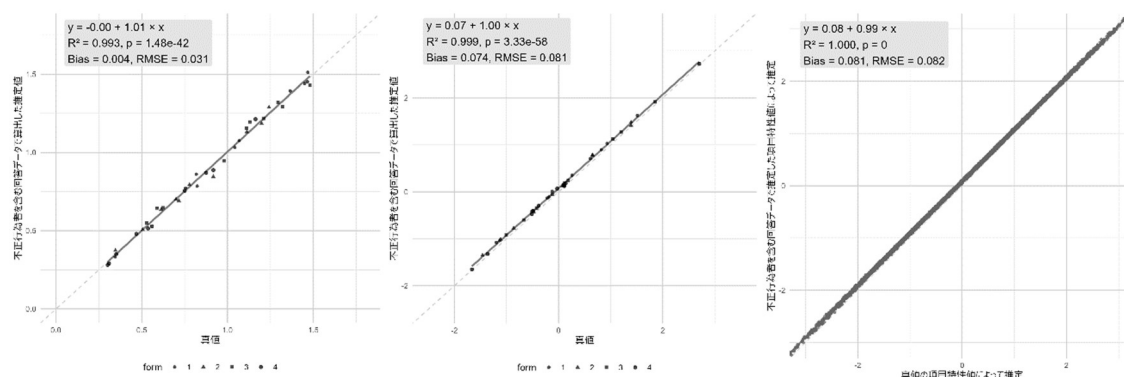


図 1 パターン①の推定結果(左：識別力パラメタ a 、中：困難度パラメタ b 、右：能力特性 θ)

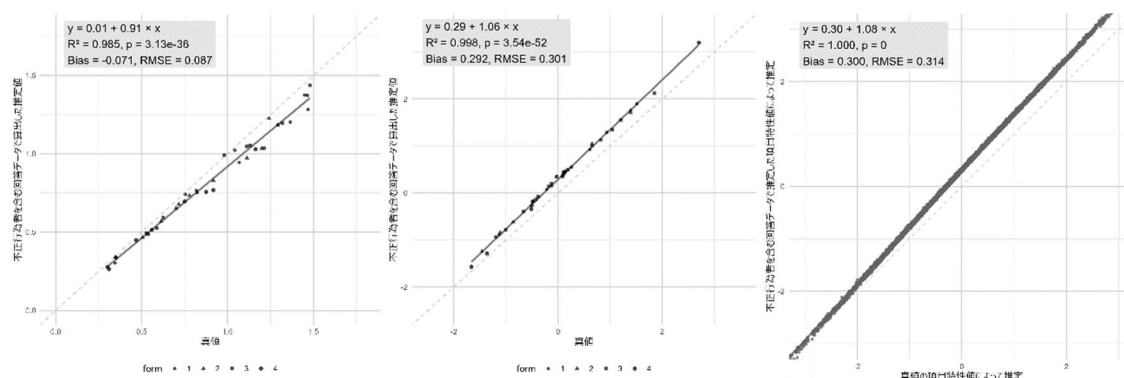


図 2 パターン②の推定結果(左：識別力パラメタ a 、中：困難度パラメタ b 、右：能力特性 θ)

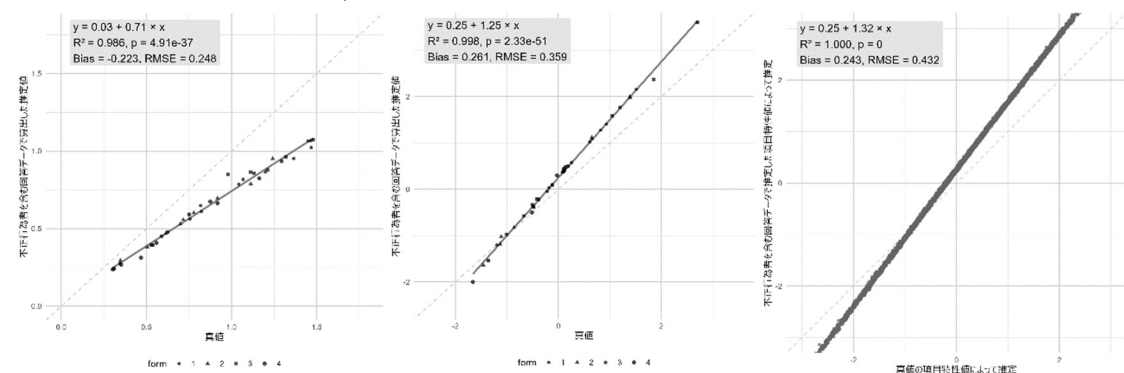


図 3 パターン③の推定結果(左：識別力パラメタ a 、中：困難度パラメタ b 、右：能力特性 θ)

4. 考察

本研究では、不正行為者を含む回答データから推定した項目特性と、将来の受検者の能力推定に与える影響を定量的に検証した。パターン①（不正行為者混入率 20%、漏洩項目率 20%）では、項目特性推定および能力推定への影響は軽微であった。この結果は、完全な監視が困難な実施環境において収集された回答データを採点除外項目の項目特性推定に使用することが、パターン①程度の汚染状況の場合は実用上大きな問題とならないことを示唆しており、継続的なテスト運用における設計指針として有用な知見である。

一方、パターン②③では項目特性推定に系統的バイアスが生じ、能力推定にも影響を及ぼした。パターン②では全受検者の能力推定値が一律に上昇し、平均的能力（ $\theta=0$ ）の受検者で θ が 0.3 上昇した（偏差値換算で 3 点相当）。本研究で用いた 40 項目テストの信頼性係数は 0.82、測定標準誤差は約 4 点（偏差値）であったことを考慮すると、このバイアスは測定精度の観点から看過できない水準である。パターン③では、平均的能力受検者における θ の上昇幅は 0.25（偏差値換算で 2.5 点）とパターン②より軽微であったが、高能力層と低能力層でバイアスの方向が逆転する、パターン②とは異なる影響を示した。

前提として、実際のテスト運用では一定の不正行為対策が実施されており、パターン②③のような深刻な汚染状況は起こりにくいと考えられる。しかし、汚染状況の事前予測は困難であり、このような極端なケースを想定した検討には意義がある。

5. 今後の課題

本研究では、完全な不正対策が困難な実施環境を想定し、不正行為の間接的な影響という観点で検証を行った。ただし、本来的には、不正行為への対応は直接的な影響も含めて検討すべき課題であり、本研究のような事後の影響評価と並行して、効果的な不正防止策の開発・実装についても継続検討が必要である。

本研究に関する今後の課題としては、以下の点が挙げられる。第一に、本研究では漏洩項目が不正行為者全員に等しく漏洩する条件でシミュレーションを実施したが、実際のテスト環境では、特定の受検者に特定の項目が部分的に漏洩するなど、より複雑な不正行為パターンが想定される。今後は、こうした実態により即したモデル設定での検証が必要である。

第二に、本研究は固定版テストを前提としたが、CAT（Computer Adaptive Testing）のように項目プールから受検者の回答状況に応じて問題項目を出題する方式への拡張も重要な課題である。CAT では出題項目が受検者によって異なるため、不正行為者への漏洩項目の出題数にもばらつきが生じ、バイアスの影響が変化する可能性がある。また、将来的な適正受検者への影響についても、系統的誤差が生じた項目の出題有無によって受検者への影響度合いが変わるため、本研究結果のような能力特性との比例関係とは異なる、より複雑な影響パターンが生じると予想される。この点についても実証的検証が求められる。

引用文献

- Cizek, G. J., & Wollack, J. A. (2016). Exploring cheating on tests: The context, the concern, and the challenges. *In Handbook of quantitative methods for detecting cheating on tests*, p. 3-19. Routledge.
- Eckerly, C. A. (2016). Detecting preknowledge and item compromise: Understanding the status quo. *In Handbook of quantitative methods for detecting cheating on tests*, p.101-123. Routledge.
- Newton, P.M., Essex, K. (2023). How Common is Cheating in Online Exams and did it Increase During the COVID-19 Pandemic? A Systematic Review. *Journal of Academic Ethics*, 22, p.323–343. doi:10.1007/s10805-023-09485-5.
- 株式会社サーティファイ(2024). オンラインによる就職活動についての実態調査. PR TIMES.
<https://prtmes.jp/main/html/rd/p/000000070.000071304.html>(参照 2025-08-04)
- 杉山 剛(2021). 適応型テストの回答データを用いた項目特性値推定 ―本番テストの回答情報の利用. *日本テスト学会第19回大会*, p.144-147.
- 歌川 真一郎, 登藤 直弥, 尾崎 幸謙 (2025a). 入社試験における Web テスト不正行為経験割合の推定. *日本テスト学会誌*, 21 巻, 1 号, p.1-15. doi:https://doi.org/10.24690/jart.21.1_1
- 歌川 真一郎, 登藤 直弥, 甲斐 江里, 尾崎 幸謙 (2025b 印刷中). 機械学習による適応型テストの不正行為検出条件の検討, *行動計量学*.