

CBT 医師国家試験の導入の実現に向けた研究
2025 年度 CBT トライアル問題の項目反応理論による分析と等化の試み

研究分担者 久保 沙織(東北大学教育学研究科 准教授)
岡崎 仁昭(自治医科大学医学教育センター 顧問)
松山 泰(自治医科大学医学教育センター 教授)
浅田 義和(自治医科大学医学教育センター 准教授)
伴 信太郎(愛知医科大学医学教育センター 特命教育教授)

研究要旨

2025 年度の医師国家試験 CBT トライアルでは、項目プールの構築を見据え、2022 年度と 2023 年度に出題した項目を組み合わせて試験問題を構成した。本研究の目的は、2025 年度に出題された 200 問を対象とした項目反応理論による分析結果から、各項目の特徴とテスト全体の測定精度を明らかにすること、および項目母数の等化の一例を示すことである。項目反応理論による分析の結果、各論、必修、総論という各問題内容に応じた困難度・識別力の特徴が確認され、テスト情報関数からは、想定される能力値の範囲で高い測定精度が保たれていることが示された。また、本研究では、共通項目計画によりテストを実施し、等化のプロセスを経ることで、2023 年度の項目母数について 2025 年度の項目母数と同一の尺度上で表現できることを示した。この成果は、項目プールの構築に向けた大きな進展と言える。

A. 研究目的

本研究課題においては、医師国家試験を CBT 化するだけではなく、現在の古典的テスト理論の枠組みでの実施・評価方法から脱却し、項目反応理論(Item Response Theory, IRT)に基づく安定的・持続的なテスト運用を実現することを目指している。IRT では、項目母数(困難度母数や識別力母数など)の推定値が得られていて、かつ、それらが互いに比較可能となるように共通尺度上への等化(equating)がなされている、数多くの項目をあらかじめ用意しておく必要がある。このような項目の集まりを項目プール(item pool)あるいは項目バンク(item bank)と呼ぶ。IRT による試験運用においては、この項目プールの構築が極めて重要となる。本研究では、2025 年度のトライアルで出題された 200 問を対象とした IRT モデルによる分析結果をもとに各項目の特徴とテスト全体の測定精度を明らかにすること、および 2025 年度と 2023 年度の推定結果を利用して項目母数の等化の一例を示すことを目的とする。

B. 研究方法

1. 出題項目の選定

2025 年度の CBT トライアルでは、項目プールの構築を見据え、2022 年度と 2023 年度に出題した項目を組み合わせて試験問題を構成することとした。問題選定では、古典的テスト理論における識別力の指標である I-T 相関(item-total correlation)が 0.20 以上であることを基準とし、問題内容を参照しながら問題作成者と分析者が共に議論することで、どの問題を採用するか検討した。ただし、出題内容のバランスを考慮して項目数を確保するため、I-T 相関の基準を満たさない項目も一部採用した。

2. IRT モデルによる項目母数の推定

2025 年度の CBT トライアルで出題した 200 問について、A 問題(各論, 75 問)、B 問題(必須, 50 問)、C 問題(総論, 75 問)のそれぞれで、古典的テスト理論に基づく項目分析および、IRT モデルによる項目母数の推定を実行した。まず、古典的テ

スト理論に基づき項目困難度(通過率)と項目識別力(I-T 相関)を確認した後に、IRT の 2 母数ロジスティックモデル(two parameter logistic model: 2PLM)を適用して項目母数(困難度母数と識別力母数)を推定した。なお、IRT の分析においては、識別力が極端に低い項目をあらかじめ除いて項目母数を安定的に推定するために、I-T 相関が 0.20 以上の項目のみを用いた。また、テスト情報関数を描画して、各問題の測定精度を確認した。

3. 項目母数の等化

2023 年度と 2025 年度のトライアル問題を対象として、等化係数法による項目母数の等化を試みた。上述のように、2025 年度のトライアルでは、2022 年度と 2023 年度に既に出題している項目で試験問題を構成している。2022 年度、2023 年度、2025 年度の CBT トライアルにおける、受験者集団と問題を構成する項目との関係性を、図 1 に模式的に示した。受験者は各年度で異なるが、項目は 2022 年度と 2025 年度、2023 年度と 2025 年度でそれぞれ重複していることを表している。そのため、同一の項目に対して、異なる受験者集団の回答データに基づく 2 年分の項目母数の推定値が得られる。

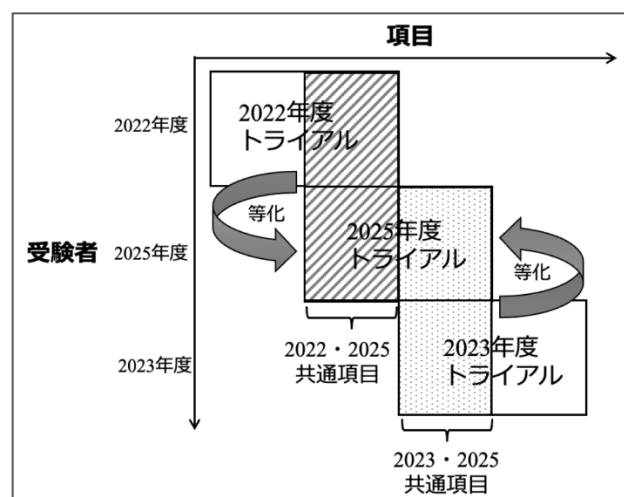


図 1 CBT トライアルにおける受験者と項目との関係

ここで、2025 年度の試験をテスト X 、2023 年度の試験をテスト Y とすると、困難度母数 b について以下の関係が成り立つ。

$$\begin{aligned}\mu_{bX} &= K\mu_{bY} + L \\ \sigma_{bX} &= K\sigma_{bY}\end{aligned}$$

すなわち、傾き K と切片 L を求めることができれば、テスト Y (2023 年度の試験) の受験者の能力値の尺度上で推定された項目パラメタを、テスト X (2025 年度の試験) の受験者の能力値の尺度上に等化することができる。ただし、式中の μ_{bX} と σ_{bX} はテスト X での困難度母数の母平均と母標準偏差を、 μ_{bY} と σ_{bY} はテスト Y での困難度母数の母平均と母標準偏差を表現している。傾き K と切片 L は等化係数よばれ、上式の関係性から、 $K = \sigma_{bX} / \sigma_{bY}$ 、 $L = \mu_{bX} - K\mu_{bY}$ と導かれる。さらに、テスト X において推定された共通項目の困難度母数の平均値と標準偏差、テスト Y において推定された共通項目の困難度母数の平均値と標準偏差をそれぞれ μ_{bX} と σ_{bX} 、 μ_{bY} と σ_{bY} の推定量として用いることで等化係数の推定値が求められる。

C. 研究結果

1. A 問題 各論($N=1645$)の分析結果

全 75 項目の通過率の平均は 0.543、I-T 相関の平均は 0.317 であった。I-T 相関 ≥ 0.20 となった項目数は 68 (90.7%) であり、2021 年以降の CBT トライアルの中で最多となった。I-T 相関 < 0.20 となった 7 項目(2022 年度 4 項目、2023 年度 3 項目)のうち 4 項目は過年度も同様の結果であった。また、クロンバックの α 係数は 0.88 であった。

2PLM による A 問題の項目母数の推定値を図 2 に示した。プロットされている文字は項目番号を表す。もっとも困難度母数の値が大きかった X23.A15 は通過率 0.06、もっとも困難度母数の値が小さかった X23.A34 は通過率 0.89 の項目であった。X22.A23、X23.A26、X23.A45 などは、比較的易しい問題でありながら識別力が高い良問である。

図 3 には、A 問題のテスト情報関数を示した。横軸は能力値 θ 、縦軸は情報量であり、情報量が

きい θ の範囲で、特に推定の精度が高いと解釈することができる。図 3 から、A 問題は、 $-1.5 < \theta < 0.5$ 付近で、受験者の能力の推定精度が高いテストであると言える。

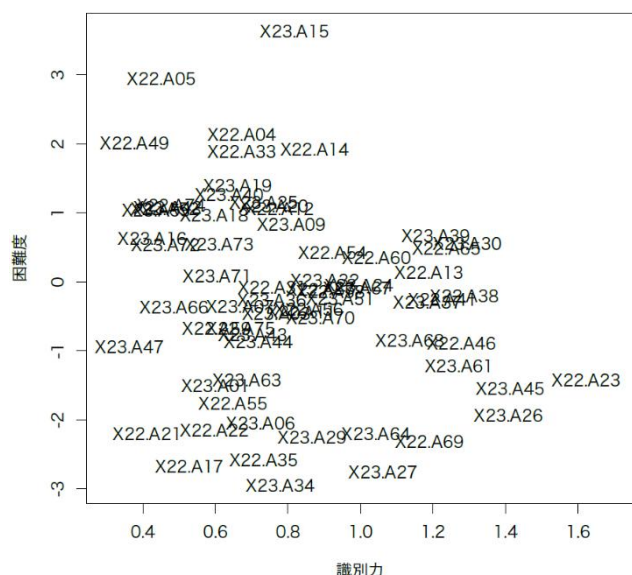


図 2 2025 年度 A 問題の項目母数(68 項目)

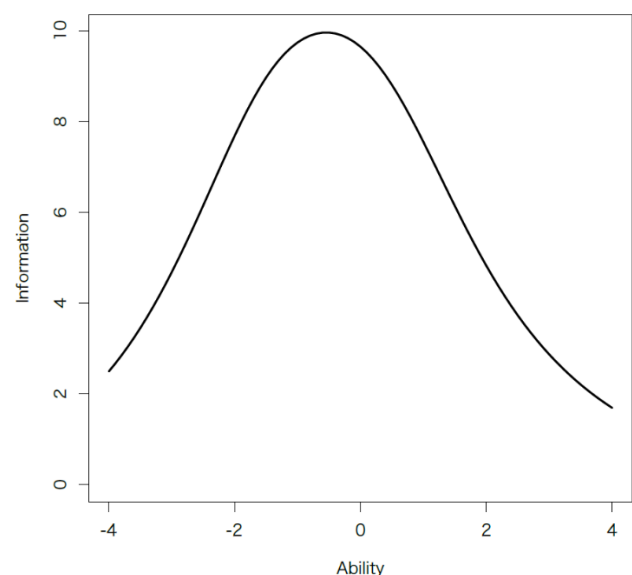


図 3 2025 年度 A 問題のテスト情報関数

2. 必修 B 問題($N=1640$)の分析結果

全 50 項目の通過率の平均は 0.645, I-T 相関の平均は 0.274 であり, I-T 相関 ≥ 0.20 となった項目数は 37 (74.0%) であった。I-T 相関 < 0.20 となった 13 項目 (2022 年度 4 項目, 2023 年度 9 項目) のうち 10 項目は過年度も同様の結果であり, 2025 年の通過率を確認すると, 6 項目で 0.80 以上となっていた。クロンバックの α 係数は 0.75 であった。

2PLM による B 問題の項目母数の推定値を図 4 に示した。分析に用いた 37 項目のうち困難度母数の推定値が正となった項目は 11 (29.7%), 負となった項目は 26 (70.3%) であり, 必修では易しい項目が多く含まれていたことがわかる。もっとも困難度母数の値が大きかった X23.B34 は通過率 0.24, もっとも困難度母数の値が小さかった X22.B05 は通過率 0.86 の項目であった。識別力母数の大きな項目に着目すると, X23.B24, X22.B30 はいずれも $\theta = 0.0$ 付近の能力値をもつ受験者をよく識別できる項目, X23.B25 は $\theta = -2.5$ 付近の受験者をよく識別できる項目である。

B 問題のテスト情報関数を図 5 に示した。B 問題は, $-2.0 < \theta < 1.0$ 程度の受験者の能力の推定精度が高いテストであることがわかる。

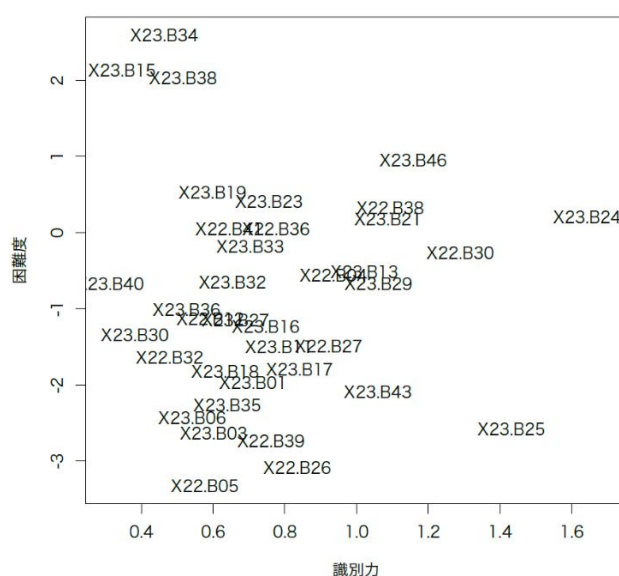


図 4 2025 年度 B 問題の項目母数(37 項目)

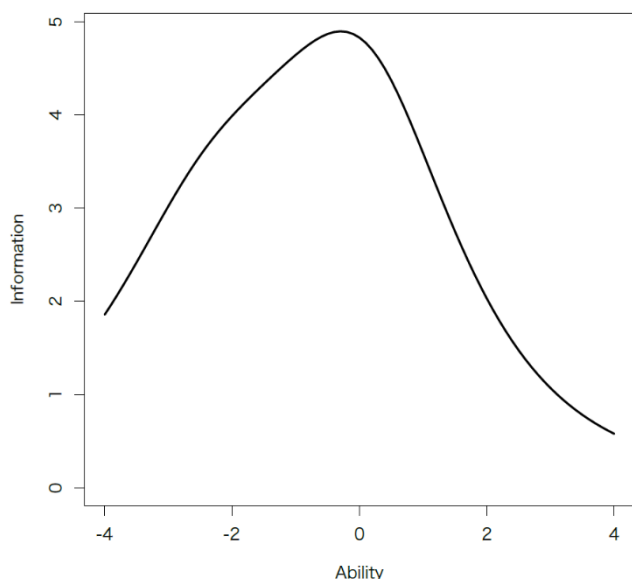


図 5 2025 年度 B 問題のテスト情報関数

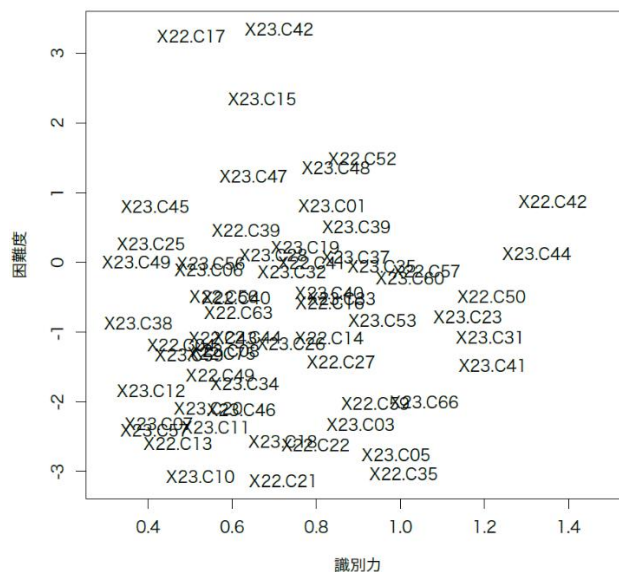


図 6 2025 年度 C 問題の項目母数(63 項目)

3. C 問題($N=1634$)の分析結果

全 75 項目の通過率の平均は 0.608, I-T 相関の平均は 0.280 であった。I-T 相関 ≥ 0.20 となった項目数は 63(84.0%)であり, 2021 年以降の CBT トライアルの中で最多となった。I-T 相関 < 0.20 となった 12 項目(2022 年度 5 項目, 2023 年度 7 項目)のうち 8 項目は過年度も同様の結果であった。クロンバックの α 係数は 0.84 であった。

2PLM による C 問題の項目母数の推定値を図 6 に示した。A 問題, B 問題と比較して, 識別力母数の範囲はやや狭くなっていた一方で, 困難度母数はおおよそ $-3.0 \sim 3.0$ の広い範囲で偏りなく推定された。困難度母数の値が大きかった X23.C42, X22.C17 の通過率はそれぞれ 0.10, 0.18 であり, 困難度母数の値が小さかった X22.C21, X23.C10 の通過率はそれぞれ 0.89, 0.82 であった。識別力母数が相対的に高かった項目について困難度母数も合わせて参照すると, 能力値が $-2.0 < \theta < 1.0$ 程度の受験者をよく識別できる項目が多く含まれていたことが確認できる。テスト情報関数(図 7)でも同様に, $-2.0 < \theta < 1.0$ 付近の受験者の能力母数の推定精度が高いテストであることが示されている。

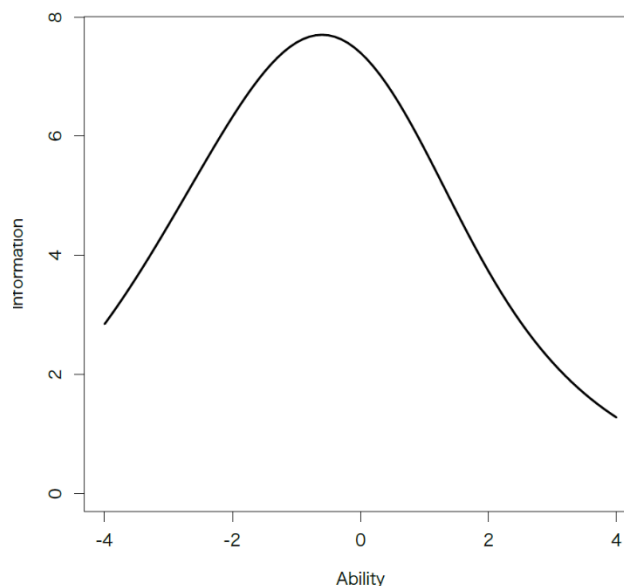


図 7 2025 年度 C 問題のテスト情報関数

4. 等化係数法による項目母数の等化

2025 年度と 2023 年度の CBT トライアルにおいて, IRT による分析に用いられた共通項目は A 問題で 39 項目, B 問題で 26 項目, C 問題で 34 項目であった。共通項目の困難度母数 b の平均値と標準偏差(SD)および, それらを利用して求めた等化係数 $\hat{K}$ と $\hat{L}$ を表 1 に示した。なお, 困難度母数 b の平均値と標準偏差(SD)は, 上段が 2025 年度, 下段が 2023 年度の値である。

表 1 等化係数の算出

	b の平均	b のSD	$\hat{R}$	$\hat{L}$
A 問題	-0.270 -0.563	1.324 1.179	1.124	0.363
B 問題	-0.674 -1.062	1.337 1.087	1.078	0.471
C 問題	-0.535 -0.753	1.407 1.232	1.142	0.324

D. 考察

まず、古典的テスト理論に基づく項目分析と、IRT の 2PLM による分析の結果について考察する。2022 年度と 2023 年度の分析結果に基づいて項目選定を行ったことで、A 問題(各論)と C 問題(総論)では、識別力が極端に低い(I-T 相関<0.20)項目の割合をこれまでより減少させることができた。一方で、B 問題では、I-T 相関が 0.20 未満となる項目を減らすことができなかったが、通過率が 0.90 を超える項目が 1 割含まれていたことや、通過率の最小値も 0.19 と A 問題・C 問題より高かったことがその要因として考えられる。B 問題では必修を扱っていることから、易しい項目を多く含むという特徴はこれまでのトライアル試験においても一貫して見られており、項目選定の段階で、I-T 相関<0.2 という基準を満たしていない項目であっても、項目数確保のためにやむなく採用した項目の割合が多かったという事情がある。

IRT による分析結果から、A 問題と C 問題は、易しい項目から難しい項目まで幅広い困難度の項目を含み、おおよそ-2.0 から 1.0 までの能力値をもつ受験者をよく識別するテストであった。B 問題は、相対的に易しい項目が多いことが明らかとなった。これらの傾向は、A 問題、B 問題、C 問題が、それぞれ各論、必修、総論を扱っているという内容に鑑みて妥当な結果と言える。また、いずれの問題においても識別力母数は 0.30 から 1.65 の間で推定されており、十分な識別力を備えていることが確認できた。

さらに本研究では、項目プールの構築を見据え、等化係数法による項目母数の等化を試みた。

2025 年度のトライアルでは、2022 年度と 2023 年度に出題済みの項目を用いて試験問題を構成したため、同一の項目に対して、それぞれ異なる受験者集団の回答データに基づいて 2 年分の項目母数の推定値が得られる。本研究では、受験者数が 1000 名を超えて推定結果が安定している 2023 年度と 2025 年度の推定結果を利用して、2023 年度の項目母数を 2025 年度の項目母数と同一尺度上で表現するための等化係数の推定値を求めた。推定された等化係数 $\hat{R}$ と $\hat{L}$ を用いることで、以下の関係性により、2023 年度の識別力母数 a と困難度母数 b を 2025 年度の尺度上の母数となるように変換することができる。

$$a^* = \frac{a}{\hat{R}}$$

$$b^* = \hat{R}b + \hat{L}$$

また、2023 年度の等化後の項目母数 a^* と b^* を利用することで、2023 年度の受験者に対して 2025 年度と同一の尺度上で能力母数を算出することも可能となる。

E. 結論

上述した出題項目選定のプロセスを踏むことにより、A 問題と C 問題で、これまでのトライアルと比較して I-T 相関が極端に低い項目が減少したという本研究の結果は、各項目の内容に関する検討のみならず、テスト理論に基づき測定の観点から項目の評価を行うことの意義を示唆している。

IRT の 2PLM による分析結果からは、各論、必修、総論という各問題内容に応じた項目母数の特徴が確認された。また、テスト情報関数を参照することで、テスト全体の測定精度を検討した。問題ごとに、より高い精度での測定が望まれるような受験者の能力値の範囲において、実際に高い測定精度が示された。このように、テスト情報関数では、受験者の能力値のレベルごとに測定精度を評価することができるという利点を持つ。

IRT に基づくテスト運用では、等化の手続きを経た項目の集まりである項目プールをあらかじめ用意する必要がある。そこで本研究では、等化係数法による項目母数の等化の一例を示した。実際

のテスト運用においては、項目プールの中から、回答者や目的に応じて項目を組み合わせてテストを作成することとなる。例えば、テスト全体の難易度および測定精度が同程度となるような複数のテストを構成することができ、それらのテストを受験した受験者の得点を相互に比較することも可能となる。医師国家試験においても、異なる受験者集団が、異なる項目から構成される問題を、異なる日時に、異なる場所で受験したとしても、統一の枠組みで評価可能となるような、IRT に基づくテスト運用は望ましいであろう。なお、項目プールの安定的な維持と効率的な拡大のためには、項目内容の非公開が前提条件となる。

F. 健康危険情報

該当なし

G. 研究発表

1. 論文発表

なし

2. 学会発表

- 1) 久保沙織・浅田義和・松山泰・岡崎仁昭・伴信太郎・中田祐太・上田茂・河北博文，医師国家試験 CBT トライアルの IRT による分析——マルチメディア使用問題の結果を中心に—— 第 57 回日本医学教育学会大会 ポスター発表

H. 知的財産権の出願・登録状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

記載すべき事項なし