

ICT を利用した医学教育コンテンツの開発と活用に向けた研究

研究協力者 久保 沙織 東北大学高度教養教育・学生支援機構 准教授

研究要旨

現在の医師国家試験は、毎年新たな問題を準備し、マークシート方式のPBT(paper-based testing)で実施されている。試験後には全ての問題と正答が厚生労働省の Web ページで公開され、回答データの分析、合格基準の設定など、試験の運用は古典的テスト理論に基づいて行われている。古典的テスト理論では、得られる結果がその時々の受験者集団の性質に依存してしまうという問題を抱えている。すなわち、毎年異なる問題で構成される試験を、それぞれ異なる集団が受験している状況では、各年の結果の変動について、試験に含まれる問題の難易度や識別力の違いによるものなのか、受験者集団の能力分布の違いによるものなのか、区別することができない。

IRT(Item Response Theory)は上述のような古典的テスト理論の問題を克服するものである。IRT では、試験を構成する各問題の難易度や識別力といった項目の持つ特性と、受験者の持つ能力とを分離して考えることができる。将来的に、IRT の枠組みで等化された項目プールを用いてテストを運用することにより、事前に難易度や測定精度を制御して一定の質を満たす問題セットを継続的に作成し、実施日程や会場、受験者集団が異なる場合でも、共通尺度上のスコアとして表現することが可能となる。そこで、IRT による研究に着手した。

A. 研究目的

上述のように、IRT に基づいて試験の運用を行うことにより、実施前に難易度や測定精度を制御して、難易度が同程度となるような問題のセットを複数作成できること、それにより、実施日程や会場、受験者集団が異なる試験でも、受験者の能力を共通尺度上のスコアとして表現可能となることなどのメリットがある。IRT による試験運用を目標とした場合、項目の難易度と識別力を表すパラメタ(母数)の推定値が得られていて、かつそれらが同じ尺度上に等化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。このような項目を集めたものを項目プール(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である。本研究では、その第一歩として、2021 年度および 2022 年度に出題した各 200 問について、IRT による項目母数の推定を行うことを目的とする。

B. 研究方法

2021年度および2022年度に出題した各200問について、A問題(75問)、B問題(50問)、C問題(75問)のそれぞれで、古典的テスト理論に基づく項目分析と、IRTによる項目母数の推定を実行した。まず、各項目の要約統計量を確認した上で、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(item-total correlation: IT相関)を求めた。IRTの分析では、次元性の仮定が満たされていることが前提となるため、IT相関が0.2未満となった項目を除き、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。

C. 研究結果

C-1 2021年度の問題

C-1-1 A問題(N=323)

通過率が低かった項目はA40(0.080), A47(0.118), 高かった項目はA29(0.938), A43(0.907), A54(0.907), A02(0.901)であった。項目識別力が低かった項目はA45(0.044), A65(-0.002), 高かった項目はA50(0.635), A66(0.637)であった。IT相関が負の値となったA65は、画像を用いた問題であった。信頼性を表すクロンバックの α 係数は0.90であった。

2PLMによるA問題の項目母数の推定値を図1に示した。プロットされている文字は項目番号を表す。易しい項目がやや多く、困難度母数の絶対値が大きな項目は識別力母数が大きくなりやすい傾向が見られる。A50, A66, A61, A34など識別力が特に高い項目

目の中には画像を利用した問題が複数見られた。

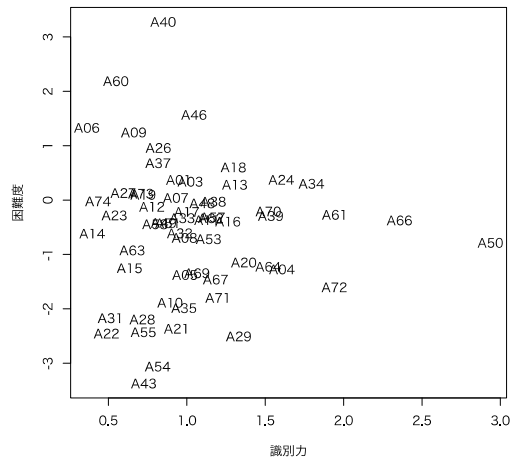


図1 2021年度A問題の項目母数(60項目)

C-1-2 B問題(N=320)

古典的テスト理論に基づく項目分析では、通過率が0.90を超える項目が11項目あり、通過率が最も低い項目(B06)で0.131であった。全項目における通過率の平均は0.714であり、B問題は易しい項目が多かった。項目識別力が特に低かった項目はB42(0.116), B01(0.130), B23(0.142), B09(0.149)などであり、高かった項目はB44(0.546), B03(0.495)であった。クロンバックの α 係数は0.80であった。

2PLMによる各項目の項目母数の推定値を図2に示した。困難度母数の推定値が正の値となった項目は6項目のみであり、やはり易しい項目が多かったことがわかる。一方で、識別力母数の推定値が1.50を超える項目はB44とB07の2項目のみであり、易しい項目が多く項目得点に差がつきにくいことから、高い識別力が得られにくかったと考えら

れる。なお、B44 は 2 連問のうちの 1 問であった。

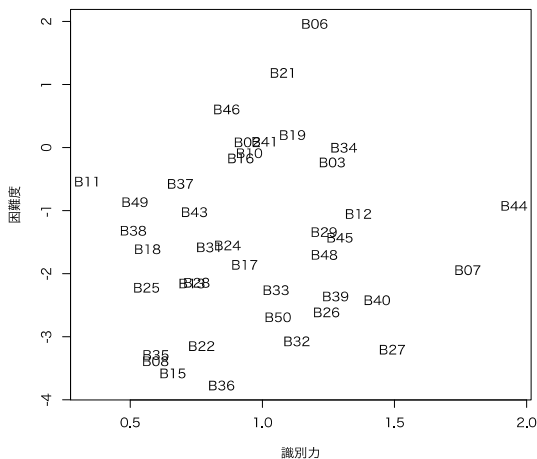


図 2 2021 年度 B 問題の項目母数 (40 項目)

C-1-3 C 問題(N=317)

通過率が低かった項目は C02(0.104), C04(0.136), C11(0.132), 高かった項目は C37(0.972), C67(0.950), C69(0.950), C32(0.943), C01(0.931)であった。項目識別力が低かった項目は C43(-0.154), C68(0.001), C64(0.011), C46(0.013), C73(0.022), 高かった項目は C10(0.564), C24(0.566)であった。IT 相関が負の値となった C43 は、画像を用いた問題であった。クロンバックの α 係数は 0.84 であった。

2PLM による各項目の項目母数の推定値を図 3 に示した。困難度母数に着目すると、易しい項目から難しい項目までバランスよく含まれていることがわかる。C10, C42, C58 など特に識別力が高かった項目の中には画像や音声を利用した問題が複数見られた。

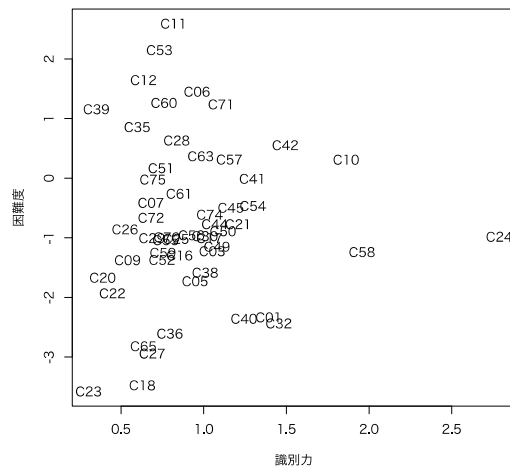


図 3 2021 年度 C 問題の項目母数 (53 項目)

C-2 2022 年度の問題

C-2-1 A 問題(N=449)

古典的テスト理論に基づく項目分析において、通過率が低かった項目は A18(0.082), A03(0.100), A25(0.145), A71(0.147) 通過率が高かった項目は A69(0.969), A07(0.960), A48(0.955), A23(0.911), A26(0.909)であった。項目識別力が低かった項目は A47(-0.084), A36(0.036), A42(0.036), 高かった項目は A40(0.499), A38(0.491)であり、全体として IT 相関が低い傾向が見られた。なお、A47, A36, A38 は画像を用いた問題であった。クロンバックの α 係数は 0.82 であった。

2PLM による各項目の項目母数の推定値を図 4 に示した。困難度については幅広い値が得られているが、識別力の高い項目は少なく、多くの項目の識別力母数が 0.0 から 1.0 の間に集中している。比較的識別力の高かった A69, A23, A35, A38 は画像を用いた問題であった。

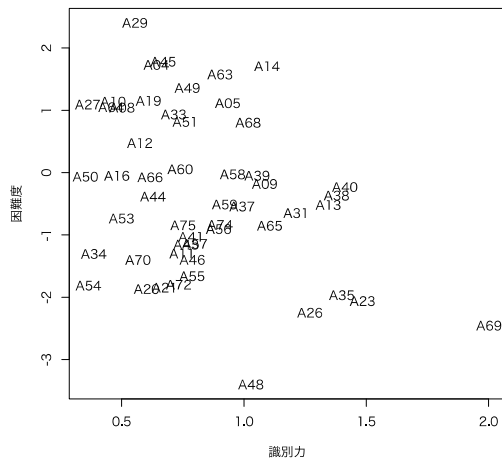


図4 2022年度A問題の項目母数(52項目)

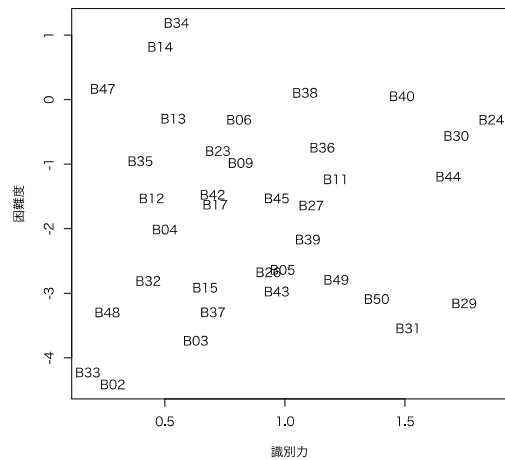


図5 2022年度B問題の項目母数(36項目)

C-2-2 B問題(N=449)

通過率が2.00を下回る項目はB19(0.071), B16(0.145)の2項目であり, 0.90を超える項目は9項目あった。全項目における通過率の平均は0.687であり, 2021年度と同様, B問題は易しい項目が多かった。項目識別力が特に低かった項目はB01(0.054), B08(0.102), B28(0.103)などであり, 高かった項目はB24(0.545), B30(0.524)であった。クロンバックの α 係数は0.73であった。

2PLMによる各項目の項目母数の推定値を図5に示した。困難度母数の最大値は1.189, 最小値は-4.413であった。困難度母数正の値となった項目は5項目のみであり, IRTの分析結果からも易しい項目が多かったことが示された。比較的識別力の高かったB24, B40は画像や動画を用いた問題であり, B44は2連問のうちの1問であった。

C-2-3 C問題(N=449)

通過率が低かった項目はC51(0.065), C09(0.089), C60(0.1114), C40(0.140), 高かった項目はC35(0.935), C66(0.915), C32(0.909)であった。項目識別力が低かった項目はC60(-0.014), C06(0.012), C09(0.058), C03(0.083), C47(0.083), 高かった項目はC33(0.513), C23(0.483), C61(0.468)であった。IT相関が負の値となったC60は, 動画・音声を用いた問題であった。信頼性を表すクロンバックの α 係数は0.84であった。

2PLMによる各項目の項目母数の推定値を図6に示した。困難度母数が負の値として推定される項目が相対的に多いことから, 易しい項目が多かったようだ。また, 識別力の高い項目は少なく, 識別力母数が1.5を上回っていたのはわずか1項目であった。比較的識別力が高かったC74及びC66は3連問の中の1問であった。

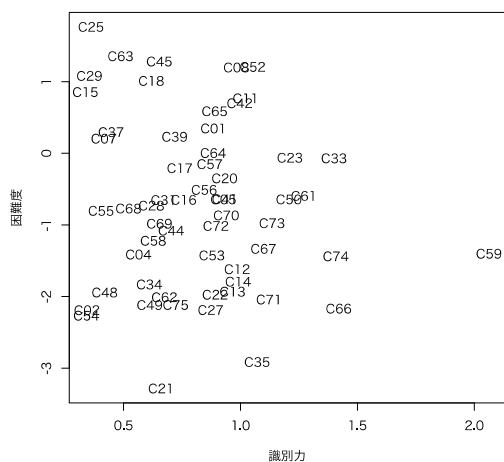


図6 2022年度C問題の項目母数(58項目)

D. 考察

2021年度と2022年度に出題した各200問について、古典的テスト理論に基づく項目分析と、IRTの2PLMによる分析を行った結果、A問題は、困難度、識別力ともに幅広い推定値が得られる傾向が見られるが、必修を扱うB問題は易しい項目が多く、C問題では高い識別力が得られにくい傾向が示された。画像・動画・音声を用いたCBTならではの問題について、一部IT関連がほぼ0となる項目も含まれていたが、識別力や困難度に関してそのような問題に一貫した傾向は見られず、識別力の高い良問も複数あった。

IRTの適用においては本来、次元性と局所独立性が満たされていることが前提となる。ところが、IT関連が下回る項目がA・B・C問題それぞれの中に2~3割程度含まれていることがわかった。また、今回は、連問となる問題についてもそれぞれ独立した問題と見なして2PLMを適用したが、今後は、連問

の局所独立性の検討を含め、医師国家試験という特殊な実施条件の下でIRTによる試験運用を実現するための課題を析出し、解決していくことが必要となる。

E. 結論

研究で実施したCBTトライアル試験問題に関して、IRTの2PLMによる分析を行った結果、A問題は、困難度、識別力ともに幅広い推定値が得られる傾向が見られるが、B問題は易しい項目が多く、C問題では高い識別力が得られにくい傾向が示された。一方で、画像・動画・音声を用いた問題に一貫した傾向は見られなかった。

2023年度には、困難度と識別力だけではなく、誤答選択肢のうちどの選択肢が選ばれやすかったかなど、選択肢に関する詳細な分析も併せて行い、項目プールに含められるような良問と、本試験での実施に適さない問題をそれぞれ精査することを目標としたい。合わせて、難易度が高すぎる、あるいは低すぎる問題、識別力が極端に低かった問題などについては、作問に当たられた先生方と共に問題内容を精査し、どこに問題があったかを検討するとともに、良問については類似の新作問題を作成することで、項目プールの構築に役立てることができるであろう。

F. 健康危険情報

特になし

G. 研究発表

1. 論文発表

なし

2. 学会発表

なし

H. 知的財産権の出願・登録状況

1. 特許出願

なし

2. 実用新案登録

なし

3. その他

なし