

CBT 医師国家試験の導入の実現に向けた研究
CBT 問題の作成、問題バンクの確立

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

医師国家試験の CBT 化に向けて、問題バンクの大規模整備が求められている。しかし、専門家による作問には時間的制約が大きく、大量の問題を短期間で作成することは困難である。本研究では、主に既存問題をもとに、生成 AI を用いて医師国家試験のフォーマットに準じた派生問題を作成する手法を確立した。また、作成された問題を体系的に整理・格納できる問題バンクを開発し、国試 CBT に向けた問題管理基盤を整備した。

A. 研究目的

医師国家試験(以下、国試)にコンピュータ試験を導入する利点として、以下の2点が挙げられる。第一に、項目反応理論(item response theory: IRT)に基づく出題を行うことで、異なる日時、場所、受験者に対して異なる問題が出題された場合でも、等化された基準で同等に合否判定を行うことが可能となる。これは、試験開催形式に柔軟性をもたせることができ、天災などにより通常どおり試験を実施できない場合の再試験も容易になる。また、隣席の受験者と異なる出題のため、不正防止にもつながる。第二に、静止画だけでなく動画や音声といったマルチメディアにより臨床情報を提示することで、現行の国試よりも多様な出題が可能となる。これにより、臨床現場において医師が動的な視覚情報や聴覚情報を認知し、重要な判断を行う能力に相当する能力評価が可能となる。

IRT に基づく CBT を実施するうえでは、共用試験実施評価機構の臨床実習前 CBT が参考となる。同 CBT では、適切な IRT 運用のために 30,000 問近いプール問題が確保されている¹⁾。国試 CBT に向けても同程度のプール問題を、整理・格納する問題バンクの確立が必要である。そこで我々は、CBT トライアルを行うなかで、1. プール問題のライ

ブラリ化、および2. 生成 AI による大量作問の可能性について実証を行った。

B. 研究方法

1. プール問題のライブラリ化

現行の国試は、「必修の基本的事項」、「医学総論」、「医学各論」(以下、必修、総論、各論)の3領域で構成されており、一般問題と臨床問題とを合わせて、2日間で計400問が出題される。出題基準では、各領域について大・中・小項目でテーマが明示され、さらにブループリント(国試設計表)により、各領域および大項目の出題割合が規定されている²⁾。

本研究における CBT トライアルは、出題割合を遵守しつつ、参加のしやすさを考慮して、1日200問で完遂できる試験として設計している。また、IRT に基づく出題の実証に向けて、問題のプール化を進めている。体系的なプール化を進めるため、Moodle(作問・校正プラットフォーム)と TAO(CBT プラットフォーム)とを連携させた、通称 MooTAO に、国試出題基準の領域および大・中項目で区分したライブラリを設け、新作問題や次節に示す派生問題をプールした(図1)。

C. 研究結果

1. プール問題のライブラリ化

MooTAO をプラットフォームとし、国試出題基準の領域(各論・総論・必修)および大・中項目で区分したライブラリを設け、新作問題や生成 AI による派生問題をプールした。プールされた問題については、出題履歴、問題特性(詳細は久保らの項を参照)、校正のバージョン、マルチメディア素材の有無などをライブラリ上で一覧できるようにし、問題の最終校正や出題後の分析・修正を効率的に行えるようにした(図1)。

2. 生成 AI による大量作問

カスタム GPT による作問数と、専門医が採用可能と判定した問題数を表 1 に示す。すべての問題が採用可能と判定されたわけではなく、全体で65.2%の採用率であった。不採用となった主な理由は、正答を選択するための決定的な臨床情報の不足や、実際の臨床現場では起こりえない設定であった。

表 1 カスタム GPT による作問数と採用可能な問題数

元の問題	カスタム GPT 作問数	採用可能問題数
119 回医師国家試験 単問臨床問題	1,000	526
119 回医師国家試験 公衆衛生一般問題	500	340
2023 年度 CBT トライアル問題	875	586
2024 年度 動画・音声付問題	80	49
新作・公衆衛生一般問題	289	289

カスタム GPT の初期型では、臨床問題文中の患者年齢や性別、検査値のみが変更され、設問や選択肢は元問題とほぼ同一であった。しかし、プロンプトの改良により、元問題と同じ症候や疾患をテーマとしつつ、患者の病歴、現症、検査所見、設問内容(病態、疫学、診断、治療、予後など)、選択肢に多様性を持つ派生問題が作成されるようになった。

公衆衛生領域の一般問題については、派生問題ではなく、国試出題基準の記載用語をキーワー

ドとした新作問題であった場合でも、すべて採用可能な問題を作成できた。

マルチメディアのうち静止画については、元問題の本文中に含まれる臨床情報をカスタム GPT が推測することで、カスタム GPT が静止画そのものを直接読み取らなくても派生問題を作成できた。一方、動画・音声付問題については、元問題の本文中の臨床情報が限られていることに加え、動画や音声そのものをカスタム GPT が読み取ることができなかった。そのため、作問者自身が動画・音声の内容について 100~200 字程度の説明文を作成し、その説明文をカスタム GPT に読み取らせることで、派生問題を作成する必要があった。

本研究に先行した厚生労働科学研究事業「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」から開始された国試 CBT トライアル(2021~2023 年)では少人数の研究分担者が³⁾、本研究の 2024 年には計 33 名の研究分担者・協力者が作問を担当した。これに対し、カスタム GPT を導入することで、それらを上回る問題数を作成することができた(表 2)。

表 2 2021~2025 年度の累積作問数

年度	2021	2022	2023	2024	2025
作問数	200	200	200	367	1,790
累積問題数	200	400	600	967	2,757

D. 考察

今年度の研究により、臨床問題については元問題からの派生問題として、公衆衛生領域の一般問題については元問題を用いない新作問題として、カスタム GPT が多様な設問および選択肢を自ら生成し、作問できることが実証された。しかし、作問数のうち、文章校正のみで出題可能となるものは全体の 65%程度であった。

したがって、まず領域専門医が採用可能か否かを判定し、その後、改めて文章校正を行うという、少なくとも 2 段階のヒトによるクオリティチェックは必須である。一方で、一定期間内に問題を量産するためには、適切な基礎学習とプロンプトによる指示のもとで管理された生成 AI の支援による派生

問題作成が有用であることが示された。

また、現在の国試でも扱われている静止画問題については、静止画から得られる視覚情報の制限を補足するために、臨床情報、設問文および選択肢が詳述されているためか、静止画を直接読み取らなくてもカスタム GPT が作問できていた。一方、動画・音声問題については、カスタム GPT による直接的な読み取りは困難であり、派生問題を適切に作成できない状況であった。

このことは、動画・音声情報を文脈に即して認知する能力が、現時点の生成 AI では対応しにくい領域であることを示している。これは CBT における動画・音声付き臨床問題が、生成 AI 時代におけるヒト臨床医の実践能力を評価するうえで、妥当な形式である可能性を示唆するものである。

E. 結論

国試 CBT に必要な問題バンク構築に向けて、MooTAO 上で国試出題基準に沿った問題ライブラリを整備し、生成 AI による大量作問の可能性を検討した。

問題の体系的な管理基盤を構築でき、カスタム GPT により臨床問題の派生問題および公衆衛生領域の一般問題を短期間に多数作成できることが示された。ただし、採用可能な問題は約65%であり、専門医による採否判定と校正は不可欠であった。また、動画・音声付き問題では生成 AI のみでの作問に限界があり、作業工程にヒトが介することを前提とした取り組みが必要と考えられた。

一方、動画・音声が提示する視覚・聴覚情報の認識は、生成 AI 時代においてもヒト臨床医が引き続き育むべき臨床実践力ともいえる。CBT 導入に際しては、動画・音声付き臨床問題の積極的採用を検討すべきである。

F. 健康危険情報

特記すべきことはない。

G. 研究発表

1. 論文発表

なし

2. 学会発表

- 1) 松山泰. マルチメディアを活用した CBT 問題の作成. シンポジウム 14 臨床実習における医学生 of 適正な評価と医師国家試験 CBT 化に向けた展望. 医学教育 2025; 56(Suppl): 138.

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

特記すべきことはない

I. 研究協力者一覧（50 音順・敬称略）

2026年3月31日時点

伊藤 真人	（自治医科大学 耳鼻咽喉科）
及川 沙耶佳	（秋田大学 大学院医学系研究科 先進デジタル医学・医療教育学講座）
大久保 由美子	（帝京大学 医学部 医学教育学講座）
海津 聖彦	（日本医科大学 思春期・小児医学／付属病院小児科／医学教育センター／臨床研修センター）
加賀 重亜喜	（山梨大学医学部附属病院 臨床教育部臨床実習センター、シミュレーションセンター）
角 総一郎	（自治医科大学 皮膚科学講座）
兼田 裕司	（自治医科大学 メディカルシミュレーションセンター 消化器一般移植外科）
菅野 武	（自治医科大学 医学教育センター医療人キャリア教育開発部門）
木村 友和	（名古屋大学医学部附属病院 泌尿器科）
金城 紀与史	（琉球大学 医学部 医学教育企画室）
黄 世捷	（聖マリアンナ医科大学 医学教育文化部門）
小佐見 光樹	（自治医科大学 地域医療学センター公衆衛生学部門）
後藤 理英子	（東京医科大学 医学・看護学教育推進センター）
才津 旭弘	（自治医科大学 医学教育センター 医療人キャリア教育開発部門）
相良 由佳	（医療法人社団 幸野メディカルクリニック）
笹原 鉄平	（自治医科大学附属病院 感染制御部）
澤幡 美千瑠	（自治医科大学 呼吸器内科）
鋪野 紀好	（千葉大学大学院医学研究院 地域医療教育学）
清水 郁夫	（千葉大学医学部附属病院 医療安全管理部）
高橋 宏典	（自治医科大学 産科婦人科）

田鎖 愛理	(岩手医科大学 衛生学公衆衛生学講座/全学教育推進機構教学 IR センター)
蓮沼 直子	(広島大学 医学部附属医学教育センター)
林 幹雄	(関西医科大学 教育センター)
堀切 海都	(大間病院 内科・総合診療科)
町田 幹	(日本医科大学付属病院 放射線科)
松原 大輔	(自治医科大学 地域医療学部門)
松本 卓子	(東京女子医科大学 統合教育学修センター基礎教育学)
三原 弘	(札幌医科大学 医療人育成センター)
宮道 亮輔	(自治医科大学 メディカルシミュレーションセンター)
山田 和徳	(金沢医科大学 医学教育学)
山本 祐	(自治医科大学 地域医療学センター総合診療部門)
早稲田 勝久	(愛知医科大学 医学教育センター)
渡辺 茅里	(自治医科大学 眼科学講座)
渡会 瑠斗	(山形県立中央病院)

【参考文献】

1. 公益社団法人共用試験実施機構. 共用試験実施評価機構と共用試験. 令和 4 年度第 1 回医道審議会医師分化会医学生共用試験部会参考資料 2-1.
2. 厚生労働省. 医師国家試験出題基準 令和 6 年版.
3. 河北博文(研究代表者). 厚生労働省科学研究費補助金事業「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」令和 3 年度～5 年度総合研究報告書.

