

厚生労働科学研究研究費補助金  
医療技術評価総合研究事業

医師国家試験コンピュータ化に関する研究

平成 14 年度 総括・分担研究報告書  
主任研究者 細田 瑳一

平成 15 (2003) 年 4 月

目 次

I. 総括研究報告

医師国家試験コンピュータ化に関する研究 -----	1
細田 瑛一	

II. 分担研究報告

1. 試験問題のコンピュータ化と諸外国の実状視察 -----	4
高林 克日己、福井 次矢、	
2. 試験問題の改良と整理の技術的解決研究 -----	8
神津 忠彦	
(資料) 本ソフトウェアの概要	

III. 研究成果の刊行に関する一覧表  
なし

IV. 研究成果の刊行物・別刷  
なし

厚生労働科学研究費補助金（医療技術評価総合研究事業）

総括研究報告書

医師国家試験コンピュータ化に関する研究

主任研究者 細田 瑛一 財団法人日本心臓血圧研究振興会附属榊原記念病院病院長

研究要旨

医師国家試験の問題作成出題、試験実施及び試験結果の解析にコンピュータシステムを利用することは不可欠となりつつある。昨年度から医科大学の共用試験の出題と結果解析のコンピュータ利用ネットワークシステムが稼働するようになり、医師国家試験でも古くからコンピュータシステムを利用して、プール問題の分類収録や選択、採点と結果の解析が行われている。本研究では①プールされた試験問題の中から受験者の解答結果に対応して、適切に次の問題を選択配布するCATシステム（Computer adaptive test）を試作して、その妥当性の検討のパイロット試験をすること。②従来報告してきた臨床シミュレーション問題（PMP）の作成に続いて、一般問題を予め良い問題の中から領域・テーマ、難易度、選択肢、正解肢、誤答肢、を分類して収録したデータベースから、任意にテンプレートを選擇してコンピュータを用いて、作問する方法を考案し、具体的に検討すること。また一方③米国の試験システムの実状を視察して、今後、わが国にコンピュータシステムを導入する準備に役立てようとした。次年度は試行のフィールドを広げて具体的検討を更に進める予定である。

分担研究者 高林克日己  
千葉大学医学部附属病院  
文部教官助教授  
神津忠彦  
東京女子医科大学教授  
福井次矢  
京都大学大学院  
文部教官教授  
田村光司  
東京女子医科大学助手  
(2002年7月31日まで)  
分担協力者 池田 央  
立教大学前教授  
植村研一  
弘慈会加藤病院

験のコンピュータ化は既にアメリカ合衆国をはじめとして実施されており、また医師以外の分野でもコンピュータ化が進んでいる。本研究を行なうことで実際に本邦でもコンピュータを導入するための準備を整える。

B. 研究方法

①CAT(computer adapted test)の形式で800問の問題を国立大学医学部6年生に解答させ、その結果を分析して具体的問題点を検討する。  
②既出医師国家試験問題を収録分類し、形式、難易度、タクソノミー、領域別、回答数等を取りあげ、問題のテーマ、正・誤肢、選択肢等をデータベース化し、webサーバー上で稼働する登録ソフトウェア（PHP受信）を開発する。  
③米国、NBMEを視察し、USMLEのStep 1、2及び3の資料を収集、実態を調査する。

（倫理面への配慮）

倫理面の配慮として、対象学生の採点

A. 研究目的

本研究は医師国家試験をコンピュータ化することを現実的視野に立って検討し、問題点を分析してその解決法を研究することを目的とする。医師国家試

結果の守秘などが挙げられる。

C. 研究結果及び、D. 考察  
①CATの試行検討(高林、田村)  
CATは、20名の6学年の学生を対象に、2時間以内で100題終了するまで行なわれた。時間内に終了しなかった学生は、67%が紙のテストより大変だったと答え、コンピュータによるCAT形式では当然のことながら前問に戻れないこと、メモを記入できないこと、集中して疲労度が大きいことを問題点とした。長所としては、選択肢の選択が容易で画面がきれいである、こどが挙げられた。CATに対しては22%が不公平さを、22%が違和感を指摘した。全体としては63%が従来紙のテストを希望した。また試行対象の数が少なく次年度も対象を増やして行なう。

②既出問題の分類とデータベース化(田村、神津、細田)  
今年度の研究では、方法論の提示のみを目的としたため、約200問(1,000を選択肢)のみを登録し、実地での試験実施はできなかった。しかし、開発ソフトウェアはWebベースでの評価のため、作成された問題集の登録数はある程度増加した。既出問題を全年度・登録作業の継続を検討する。これにより登録された選択肢プールは20,000選択肢となり、この選択肢プールの精度維持についても具体的に作業を行ない、その作業量評価が可能となると予想される。選択肢を分類する項目(難易度のみならず、履修年限・時期等)を適切に選定することにより、医師国家試験のみならず、学生試験、専門医試験、さらには、看護師試験等への応用が可能であり、コンピュータ化試験にふさわしい方法論と考えられた。適切な数の選択肢プールの完成時点で、100問程度のコンピュータ化試験を実際に施行し、試験の実地データの収集も検討している。

③米国での実状(高林、福井、神津)  
NBMEの主要メンバーとの会談で米国の実状を入手した。

1) 問題のセキュリティ  
少人数で行なうので監視の目が行き届いていない。それぞれが離れたいが、互いに相談はできないし、室内はまた各人の問題の内容が異なる。室内にはビデオで監視しており、入室時には余計な持ち込みなどは禁止している。  
2) 各大学のカリキュラム改革の自由度  
ステップ1、ステップ2はいつでも受けられるため、どの教科をいつまで終わらなくてはならないという制約がなくなる。

### 3) Simulation テスト

紙の試験(paper and pencil test)の判定能力を超えることが可能である。紙では態度や意思決定能力を評価できない。コンピュータでなければできない試験が行える。

4) 試行が容易で日々試験ができ、大がかりでなく、コストがかからない。

Step 3がどのように行なわれているかについては、米国内で300ヶ所のサイトにあり、一ヶ所で大体15題から45題のワークシートをシヨンを設けている。これがサーバーと接続しており、情報はサーバー内に収集され、USMLEのセンターに送られる。300ヶ所の試験会場となるサイトはほとんどはペントナーによって運営されており、各サイトは週6日オプンしており、選んで受験する。受験者はStep 3選んで受験する。さまざまな試験が行なわれている。

5) 試験問題の内容  
試験問題は480問のMCQと9問のCase simulationからなる。試験の評価はまず各サイトから結果をISDNを介してダイヤルインでUSMLEセンターに送り、これを専門家が評価する。採点・評価に4週間を要する。

6) 不合格者  
実際に年間に300名の受験者が失敗する。彼らは60日後から再受験が可能で、これは年に3回まで許可される。受験生の1%は最終的に合格していない。

7) 問題の作成  
問題の作成には40名の医師、310名のNBMEの職員、400名のボランティアを容る。USMLEに關与するのはNBMEとFASBであるが、FASBは各州での医師免許の発行などに關与しており、問題作成の作業は全てNBMEが行なっている。

8) 国際的な動向  
NBMEはこのシステムをパナマ、中国に売り込むことに成功し、さらにフランスやアイルランドと交渉中である。

### E. 結論

①問題の選択配布と解答結果の解析を連動させるCAT法の検討。

国家試験に対するCAT、CBTの実施については受験者の慣れ、時代の變化に合わせて利用する時期が近いが、実施方法については更に検討を行なう必要があると考えられる。

②一般問題作成のためのデータベース構築とその利用法の具体的検討。  
医師国家試験のうち、一般問題に限定されはいるが、今回開発した手法は国家試験問題プールの維持管理を効率的に行え、未試験問題の精度・難易度評価、選択肢数設定等の点において、完成問題のプールを維持するより、有

効な方法論であると考えられた。

③米国の実状

米国の NBME の現状を視察し、USMLE の各 Step の問題についての要項を入手し、わが国での利用について検討した。

F. 健康危険情報

なし

G. 研究発表

なし

H. 知的財産権の出願・登録状況（予定を含む。）

なし

厚生科学研究費補助金（医療技術評価総合研究事業）

分担研究報告書

試験問題のコンピュータ化と諸外国の実状視察

分担研究者 高林克日己 千葉大学医学部医療情報部助教授  
福井次矢 京都大学医学部大学院教授

研究要旨 医師国家試験のコンピュータ化の導入を検討するために MCQ を CAT(Computer Adapted Test)の形式で施行して、問題点などを検討するとともに受験者の反応をアンケート調査した。また米国 NBME(National Board of Medical Examiner)を訪問して、米国におけるコンピュータによる試験方法を視察した。

A 研究目的

医師国家試験をコンピュータで行う是非・可能性についての検討を行う。

B 研究方法

昨年入力した 800 問の試験問題を基礎に、これを三菱電機に依頼して CAT(computer adapted test)の形式に変更して国立大学 6 年生 20 名に実際に施行し、その問題点を検討した。また米国の視察にてこれらのシステムがどのように機能しているかを検討した。

C 研究結果

1) CAT の施行とアンケート

CAT は学生の希望の時間で、2 時間以内に 100 題終了時まで行なった。時間内に終了しなかった学生はいなかった。学生のアンケートで

は 67%が紙のテストより大変だったと答え、前問に戻れないこと、メモを記入できないこと、集中して疲労度が大きいことを理由に挙げた。長所としては選択が容易（マークを塗らなくてよい）、画像がきれいである、ことなどが挙げられた。CAT に対しては 22%が不公平さを、22%が異和感を指摘した。全体としては 63%が従来の紙のテストを希望した。

2) NBME の視察(高林・福井)

NBME の主要メンバーとの会談で米国の状況を入手した。

CBT を Step3 に導入した理由については以下のである

1. セキュリティ (copying, cheating, stealing を防ぐことができる)。少人数で行うので監視の目が行き届いている。それぞれが離れたブースで行うのでお互いに相談はで

きないし、また問題の内容が異なる。室内はビデオで監視しており、入室時には余計な持込などは禁止している。

2. 各大学のカリキュラム改革の自由度を高められる (Step 1、Step2 はいつでも受けられるため、どの教科をいつまでに終わらなくてはならないという制約がなくなる)

3. コンピュータでなければできない試験ができる (Simulation Case など)

紙の試験 (paper and pencil test) の判定能力を超えることが可能である。紙では態度や意志決定能力を評価できない。

4. 日々試験ができ、大掛かりでなくてよく、コストがかからない。

Step 3 がどのように行われているかについて米国内で 300 ヶ所のサイトがあり、一ヶ所で大体 15 台のワークステーションを設けている。(NYC などでは特別に 45 台あるところがある) これがサーバーと接続しており、情報はサーバー内に収集されたうえで、USMLE のセンターに送られる。300 ヶ所の試験会場となるサイトのほとんどはベンダーによって運営されている。各サイトは週 6 日オープンしており、このいつに受けてもよいが、連続する 2 日を選んで受験する。受験者は Step3 だけでなくさまざまな試験が行われている。行われる試験はしたがって 15 名、最も大きい施設のニューヨークで 45 名まで同時受験できる。受験はレジデント 2 年目の後半から 3 年の間に行われる。

#### 試験問題の内容

試験問題は 480 問の MCQ と 9 問の Case simulation からなる。試験の評価はまず各サイトから結果を ISDN を介してダイヤルイ

ンで USMLE センターに送り、これを専門家が評価する。採点・評価に 4 週間を要する。評価する内容は具体的に示されなかったが、MCQ と Simulation は同じ配点比ではないようである。またその評価は時間や選択したもの、選択した順序なども考慮されているが、具体的な診断を記入したりはせず、特定のコースにしたがって選んだかどうかは評価の対象にはない。実際にどのような流れでいくべきかは多数の受験者に行わせてそのデータから解析するようにしたいと話していた。しかしその多数の人の答えを正解としてよいかについては疑問もある。MCQ とシミュレーションの間には 0.4-0.5 の弱い相関があるという。もしこれが全く相関するのであれば、シミュレーションの意味はなくなってしまう。シミュレーションが何を評価するのかについては、画像は入っていないが、読影力をみるものではないと割り切っている。

#### 不合格者

実際に年間に 300 名の受験者が失敗する。彼らは 60 日後から再受験が可能で、これは年に 3 回まで許可される。受験生の 1% は最終的に合格していない。

#### 問題の作成

問題の作成には 40 名の医師、310 名の NBME の職員、400 名のボランティアを要する。

USMLE に関与するのは NBME と FASB であるが、FASB は各州での医師免許の発行などに関与しており、問題作成の作業は全て NBME が行っている。今まで作成した問題数については秘密であるが、それが有限数の問題でないように見せる仕掛けが必要である

と考えられている。1問の作成に5年を要したというが、はじめにどのようなツールを作成するかが問題であると考えられた。

#### ベンダーについて

ベンダーに任せることにより、仕事を効率的にしているという(週6日開けても営業できる体制である)、一方でベンダーには暗号化して内容が見えないような仕掛けがしてあるという。これが具体的にどのような形で現場で作業している彼らに分からないようにできるかは現在問い合わせ中である。また試験会場がきちんと運営されているか調べるために、「罎」の受験者を廻しているとのことである。

#### 実施にあたっての全体的な問題点

Step3のコンピュータ化にどのような問題があったかという質問には、特に大きな問題はないとの答えであった。このように自信をもっているのは、実際にこの3年間大きな問題なく施行できていることによるのかもしれないし、日本への販売を考えての経営戦略的な態度なのかもしれない。

#### 国際的な動向

NBMEはこのシステムをパナマ、中国に売り込むことに成功し、さらにフランスやアイルランドと交渉中である。

#### CATについて

CATがナースの国家試験には導入されているなどアトラクティブな方法であることは認めていて、パイロット的な仕事はしているが、実際には問題が多くて導入は現在考えていないとのことであった。理由として、①医師の場合範囲が広く、一分野ができたからといっ

て次の分野ができるわけではないかもしれない、あるいはある分野が割愛されてしまうかもしれない、②できる人間にはより難しい問題が与えられる結果、できる受験者とできない受験者の差が正確でなくなる(より近似する)、③問題数が違ってくるのは受験生に不公平さを感じさせる、④実際にはCATを導入しても問題数が少なくなるにことはなかった(Peter)⑤短時間で試験を終える受験生がいても、試験場の使用時間をあらかじめ短縮できるわけではない(コストは同じ)、⑥採点がPass or failであればよいが、スコア化するのは難しい、⑦degree of difficultyの調整の方法論はあるが、どれが最適な方法なのかは検討していない、⑧Pilot Studyでは、受験生からの反対が多かった、などの意見が出て、今後もCATは導入する意志はないとのことである。

#### D 考察

CATの導入により各問の解答ごとに受験者のレベルを評価し次問題を選択することで、従来より少ない問題数での判定が期待できる。しかし今回のトライアルでは問題数が少ないために十分に適切な次問題が用意できず、少量の問題数での判定は困難であった。また学生のアンケートでは紙の試験をより好ましいとする意見の方が多かった。今回はCATのためというより準備不足のために思わしい結果でなかったことも考えられ、これらを含め今後CBT、CATの医師国家試験の導入を考えたときに解消されなければならない問題点を検討する必要がある。

また米国の視察の結果でも、CATは医師国家試験には導入されていないが、今後はこうした試験方法が一般化することも十分に考え



られる。ただし第三者機関に委ねるだけの十分な予算がないとなれば、十分なセキュリティを確保するために各大学などに人的な援助を依頼する必要がある。時代の趨勢をみながら検討すべき課題である。PMP やシミュレーションについては学習効果のあることはよく認められるが、これが果たして評価に向くかということに関しては米国においても必ずしも肯定的ではない印象を得、医療システムの相異などを考慮するとわが国独自の判定方法の採用も考えられる。

#### E 結論

国家試験に対する CAT、CBT の実施については受験者の慣れ、時代の変化に合わせて利用する時期が近いが、実施方法についてはさらに検討を行なう必要があると考えられる。

#### F 健康危険情報

なし

#### G 研究発表

なし

#### H 知的財産権出願・登録状況（予定を含む。）

なし

厚生科学研究費補助金（医療技術評価総合研究事業）

分担研究報告書

試験問題の改良と整理の技術的解決研究

分担研究者 神津忠彦 東京女子医科大学医学教育学

研究要旨

国家試験問題の問題文が「XXXX」（以下目的という）について「{正しい|誤っている}（以下正誤という）ものはどれか」という形式に類似している一般問題に着目し、この問題の目的と正誤、選択肢(文)をデータベースに登録し、目的、正誤、形式、正解等を任意に指定することにより、試験問題を作成するソフトウェアを開発した。

このソフトウェアを用いることにより、完成問題（問題文と形式に従った5つの選択肢を持ち正解が設定されている問題）を大量にプールする方法のより効率のよい運用が可能と考えられた。

A. 研究目的

国家試験問題を高品質なものにするために、完成問題（問題文と形式に従った5つの選択肢を持ち正解が設定されている問題）を約20,000題プールする必要性が唱えられている。しかし、このうち半数以上を占める一般問題を定期的に見直し、新作問題を追加していく作業量は膨大であり、不可能といえる作業である。その中には類似問題や関連問題が数多く存在することも予想される。

医師国家試験のうち一般問題の問題プールの維持管理を効率的に行えるツールの開発を目的とする。

B. 研究方法

1) 既出医師国家試験問題の分析とデータベース作成

1984年(第77回)医師国家試験から2000年(第94回)医師国家試験問題までの約6000問の試験問題を詳細に分類し、難易

度、タクソノミー、形式、正解、関連診療科等の各項目をデータベース化した。試験問題、付随する図表もすべてデータベースと関連付けて登録した。

2) 正誤問題のデータベース登録

さらに、一般問題のうち問題文が「XXXX」（以下目的という）について「{正しい|誤っている}（以下正誤という）ものはどれか」という形式に類似している問題に着目し、この問題の目的と正誤、選択肢(文)を200問（1,000選択肢）データベースに登録した。同時に、WebベースのWebサーバー上で稼動する登録ソフトウェア（PHP言語）も開発した。

3) 一般問題作成ソフトウェアの開発

目的、正誤、形式、正解等を任意に指定することにより、一般問題を作成するソフトウェアを開発した。WEBベースで利用可能とするため、データベースをPHP

言語でアクセスし、Web サーバー(apache) 上で稼動するものとした。

### C. 研究結果及び考察

今年度の研究では、方法論の提示のみを目的としたため、約 200 問 (1,000 選択肢) のみの登録であり、実地での試験施行は出来なかった。しかし、開発ソフトウェアは Web ベースであるため、作成された問題の実地での評価はデータベースへ登録された選択肢数がある程度増加した時点で可能であり、次年度以降、既出問題すべてにわたる分類・登録作業の継続を検討する。これにより登録された選択肢プールは 20,000 選択肢となり、この選択肢プールの精度維持についても具体的に作業を行い、その作業量評価が可能となると予想される。

選択肢を分類する項目 (難易度のみならず、履修年限・時期等) を適切に選定することにより、医師国家試験のみならず、学生試験、専門医試験、さらには、看護師試験等への応用が可能であり、コンピュータ化試験にふさわしい方法論と考えられた。

適切な数の選択肢プールの完成時点で、100問程度のコンピュータ化試験を実際に施行し、試験の実地データの収集も検討している。

さらに、未試験問題の難易度の設定について完成問題をプールする方式では、その事前評価は困難であるが、われわれの開発した方法であれば、未試験選択肢を既出問題の選択肢に紛れ込ますことにより難易度評価も可能である。

さらに進んだ、CATをおこなうためには、選択肢ごとの難易度評価が必要であるが、これについても識者によらずとも、また、

形式、正解を変化させることにより識別指数・難易度の設定も可能であると考えている。

### D. 結論

医師国家試験のうち一般問題に限定されてはいるが、今回開発した手法は国家試験問題プールの維持管理を効率的に行え、未試験問題の精度・難易度評価、問題の難易度・識別肢数設定等の面において、完成問題のプールを維持するよりはるかに有効な方法論であると考えられた。

### E. 健康危険情報

なし

### F. 研究発表

なし

### G. 知的財産権の出願・登録状況

なし

図 1

本ソフトウェアの概要

本ソフトウェアのメニュー画面 (図1)

平成14年度厚生科学研究 医師国家試験問題改善研究

項目	実行	説明
データ入力	<input type="button" value="ここから"/>	Roff形式の問題のDBへの取り込み
自動データ入力	<input type="button" value="ここから"/>	Roff形式の問題のDBへの取り込み
問題作成	<input type="button" value="ここから"/>	問題を作成する
問題作成 5以上	<input type="button" value="ここから"/>	問題を作成する
国家試験問題 (Word,roff)	<input type="button" value="ここから"/>	国家試験問題WORD形式なし
国家試験問題オリジナル	<input type="button" value="ここから"/>	国家試験問題のTIFFイメージ
国家試験問題 画像集	<input type="button" value="ここから"/>	国家試験問題の画像集イメージ
領域一覧	<input type="button" value="ここから"/>	東京女子医科大学
国試分類	<input type="button" value="ここから"/>	東京女子医科大学

問題作成

項目欄が「XXXX」について(目的)の部分。最上段は単に「正しいのはどれか」という問題の選択肢が登録されている。既出問題の形式、正解より、それぞれの選択肢の正誤数が右に表示されている。

http://tsukuba.u-t.ac.jp/~kawai/ken/qa/qa.html

num	番号	項目	正誤
0	10		16/24
19	1	1~4歳児の不慮の事故死で多いもの	2/3
45	3	3歳児の発達のチェックポイントとして用いられるもの	3/2
11	11	ボール(球)	4/6
56	56	Crohn病	6/4
57	57	Cushing症候群	6/3
58	58	Falko四徴症	2/8
2	2	インフォームド・コンセント (informed consent)	1/5
7	7	がんの二次予防	4/1

問題の作成にあたり、正誤、形式、正解を指定する。左図では、形式は (A, K2, K3) からランダムに選択。正解もランダムを指定している。

10 285 養育医療は児童福祉法に規定されている。  
 10 58 療養型病床群は終末期患者の緩和ケア施設である。  
 10 93 老人の不眠では入眠困難が特徴的である。  
 10 60 老人保健施設は高齢者の健康増進施設である。  
 10 94 躁病では睡眠障害は少ない。  
 正誤:  正しいのは  誤っているのは  
 形式:  A  K2  K3  K'  X2  X3  
 Random(A,K2,K3,K')  Random(A,K2,K3)  Random(A,X2)  
 正解:  a  b  c  d  e  Random

実際に作成された問題。形式はA、正解はcとなった。正誤の右の欄は、出典(既出医師国家試験施行回、問題、番号、選択肢順)、出題分野(東京女子医科大学学習要項による)

正しいのはどれか。  
 a 母子健康手帳は出生の届出の際に交付される。 282 誤 92-A-017-1 社会環境と保健医療  
 b 特別養護老人ホームは常時介護を要する老人の収容施設である。 61 誤 94-A-002-5 社会環境と保健医療  
 c うつ病では熟眠感が失われることが多い。 96 正 94-A-042-5 精神・神経の機能と病態  
 d 老人保健施設は高齢者の健康増進施設である。 60 誤 94-A-002-4 社会環境と保健医療  
 e Wernicke脳症では作業せん妄が特徴的である。 64 誤 94-A-005-3 社会環境と保健医療  
 RIGHT:c

再度、作成したところ。形式はK2、正解はc。

正しいのはどれか。  
 (1) 躁病では睡眠障害は少ない。 94 誤 94-A-042-3 精神・神経の機能と病態 ( )  
 (2) 左回轉神經は大動脈弓を反回する。 55 正 94-A-014-4 循環器・呼吸器の機能と病態 (胸壁、\*  
 (3) 法的に人工妊娠中絶が可能なのは妊娠満22週未満である。 283 正 92-A-017-2 社会環境と保健医療 ( )  
 (4) 水道水の健康に関する水質基準項目には農薬も含まれている。 307 誤 92-A-022-1 社会環境と保健医療 ( )  
 (5) 特別養護老人ホームは常時介護を要する老人の収容施設である。 61 誤 94-A-002-5 社会環境と保健医療 ( )  
 a (1), (2) b (1), (5) c (2), (3) d (3), (4) e (4), (5)  
 RIGHT:c