

200732013B

厚生労働科学研究費補助金

医療安全・医療技術評価総合 研究事業

## 医師国家試験のコンピューター化に関する研究

平成19年度 総合研究報告書

主任研究者 細田 瑛一

平成20(2008)年 4 月

## 目 次

I. 総括研究報告		
医師国家試験のコンピューター化に関する研究	-----	1
細田 瑳一		
II. 分担研究報告		
1. 医師国家試験のコンピューター化に関する研究	-----	4
高林 克日己		
2. 医師国家試験としてのコンピューター試験の適正化	-----	38
吉岡 俊正		
III. 研究成果の刊行に関する一覧表	-----	43
IV. 研究成果の刊行物・別刷	-----	44

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ー化に関する研究

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 最高顧問

研究要旨

医師国家試験のコンピューター化について、従来米国のフェーズ I にあたる教養試験が行われており、この文字の電子化によるMCQ中心の形式だけでなく問題解決型の臨床試験が必須であり、そのシステムの開発とそれらの形式の問題による学生の模擬試験の実施等からその実用性、可能性について検討してきた。5年前から欧米諸国の国家試験のあり方とコンピューター化について視察し、コンピューター試験の意義を確認した。この3年間は外国の状況に学びながらシステムを試作した。高林はClinical Case SimulationについてWebベースで動く試験システムについて、吉岡は問題解決能力の評価が可能なシステムについて開発、東京女子医科大学、千葉大学の学生に模擬試験を実施したところ、最終的にどちらにおいてもシステム的な問題事項は発生せず、学生のコンピュータリテラシーの向上にも支えられて、円滑に行うことができた。評価方法については、紙ベースの試験との比較を行いながら、コンピューター試験の目的や導入方法を含めてさらなる検討を要すると思われる。なお一定の大規模な研究で検証する必要はあるものの、本邦における国家試験へのコンピューター導入を提言したい。

分担研究者 高林克日己  
千葉大学医学部附属病院  
企画情報部 教授

吉岡俊正  
東京女子医科大学教授  
医学教育学 教授

高林は三菱電機インフォメーションシステムズの協力の下、これまでに作成下問題MCQとPMPのコンピューター化ツールをWeb化し、問題集積とその評価方式を確立するために、サーバ側で問題とプログラムを実施できるWeb Application方式のMCQとPMPを開発した。学生に模擬試験を実施し、特に項目選定順序の比較を行うことで、新たな評価ができるかを検討した。

吉岡は、臨床事例を提示し問題発見・情報検索・結果の判断解釈が順次性を持って行うような臨床推論能力・臨床判断能力評価としてのコンピューター試験のシステム(問題解決能力試験, Problem-solving ability test, P-SAT)を構築し、問題解決能力評価方法を検討した。最終年度に学生への模擬試験を実施し、そのシステムの妥当性、問題形式、共用試験CBT成績との相関などを検討した。

17年にはドイツ、フランス(吉岡はイギリスも視察)、18年にはアメリカ(高林はドイツも視察)を訪れ、試験制度とコンピューター化の状況を視察した。

(倫理面への配慮)

模擬試験においては、学生、および医師の受験者の個人情報取り扱いに留意する。

A. 研究目的

医師国家試験のコンピューター化の可能性を、システムの開発、模擬試験の実施等から検討する。高林はClinical Case SimulationについてWebベースで動く試験システムについて、吉岡は問題解決能力の評価が可能なシステムについて開発・検討する。

あわせて、先進欧米諸国の状況を視察し、本邦における国家試験へのコンピューター導入の可能性や問題点の抽出をする。

B. 研究方法

### C. 研究結果

高林らは平成17年度にWEB上で作動する試験システムを構築した。その後、問題を作成すると共に班として各大学の学生の協力を得て試行し、大きな問題なく比較的順調に実施することが出来た。高林の開発したシステムの模擬試験は平成18年度、19年度に東京女子医科大学、千葉大学医学部の4年生を対象に実施した。平成18年度は4症例の問題を70分で解答させたところ、一症例における所要時間は20分前後であり、それ以上の時間で解答率の上昇はみられなかった。平成19年にはさらに同等量の別の問題を作成するとともに、昨年までの5題のほかインフルエンザウイルスの問題を新たに作成した。これはcommon diseaseの中でも特に代表的な疾患であり、また診療方法も検査ではなく本来問診と所見だけで検査をほとんどせずに結論に導かれる点で、今までの問題とも大きく異なったものとなっているためである。試験の採点は自動的に行なわれ集積される。結果としては、学生と医師では得点は医師の方が高く、また問診、検査選択項目は医師の方が少なく、また得点と選択項目数には負の相関が認められた。これは特に得点と検査選択項目において著明であった。

吉岡の開発した問題解決能力試験(Problem-solving ability test, P-SAT)システムは、平成20年2月6日に医学部学生(第4学年)95名に対して実施した。2時間で58問を出題したが、全問題を解けない学生が約45%いた。同時期同じ対象者に実施した想起的知識評価法である共用試験CBTとの比較を行ったが、有意の相関は認めなかった。

ドイツ・フランス・イギリスでは医師国家試験にcomputer-based testing(CBT)は導入されていないが、様々な形で研究・検討がされている。EUでは医師のEU内での移動を各国が規制できないので、教育レベルの国家間での差を減少させたいとフランス・ドイツでは考えている。医師資格の付与は各国で異なるので、医科大学の共通評価を通じて教育の質を統一しようとしている。このため、ヨーロッパ医学教育連盟は世界医学教育連盟のグローバルスタンダードを批准した。教材開発の電子化が進んでいるので、教材を試験問題として使用することも可能であり、今後教材開発がCBTを促進することも考えられるが、それには初期投資と国家試験としての統一的な規格に基づく電子化が必要であろう。

米国におけるNational Medical Board(NMB)試験は電子化が進んでおり、大学もそれに対応している。コンピューター試験により単なる想起的な多肢選択問題(MCQ)よりも高度な判断・問題解決能力を評価することができるようになってきている。米国の試験会場は互いに相手の画面を見ることができず、また試験官より低い位置で試験をする隔壁の中で行われており、試験官からは観察が容易になっているが、少人数が対象の部屋である。これが全米で150施設ほどあるという。NBMEにはこのようなミニ試験場があるだけでなく、問題作成の

ための広い施設が提供され250名のスタッフが勤務している。ここに全米から集める専門家でCCSを作成しており、一問の作成に10人以上の専門家集団が関わっている。

### D. 考察

高林の今回の検討から、実際の解答結果を詳細に検討すると、それぞれの学生・医師がどのような意図で診断を進めているのかがよく理解でき、MCQの同一問題と比較して、その設問量や個々の設問自身が簡明すぎて試験に適さないことを考えると、common diseaseの管理能力の評価はシミュレーション試験にしかできないといえる。

しかしこうした解法の評価を実際にコンピュータ上で行なおうとなると、解法は一つではありえず、さまざまな正解順序が存在することから、なかなか困難であることが推察される。今回の結果で、正解選択率は医師の方が高かったのにもかかわらず、はじめの5つの問診選択の正解数はむしろ学生の方がよい結果であったことは、診断の浅いルーチンの検索では学生と医師の間に実力差が出ないものの、評価のためにはそれぞれの深いレベルの診断論理に入っていくことを示しており、そうしたことを評価することが全体の得点と比較して別の能力の検定として計算できるほど精緻なものとするためには、相当量の仕事が必要となることが考えられるためである。今後はベクトルの概念でこうしたグルーピング問題が解決できるかどうかを試みる方向もあるが、むしろ得点が低く不合格となる学生が具体的にどのような思考回路であったのかを検証するとき、生データの設問の具体的な選択順序をみることで正確に評価ができるという利用法もあるように思われる。

Webを用いた今回の試験において、トラブルはサーバ側には起こらず、かなりの人数を対象としてもサーバのスペックさえ十分であれば、トラフィック的にもネットワーク環境においても問題なく運営できることが確認できた。学生のコンピュータリテラシーの向上もあり、現在最大の課題としては試験場を通常のスペースより広く、かつ相互の端末画面を見ることができないような試験場の構造を考えなければならないことであろう。

国家試験基準に合わせた基本的な選択項目とその標準解のデータセットがほぼ出来上がっているため、新しい問題の作成は容易になり、大量の問題作成も可能となった。問題はそれを客観的、一般的にみても不整合のないものであるのかの細かい確認にあり、このためには多くの専門家や一般医の参加が不可欠である。現在必要とされている国家試験実施後の問題の開示についても、以降の問題のヒントとならないような方法を考える必要がある。

吉岡の検討は、P-SATの臨床能力評価の新たなコンピューター試験としての実用化の可能性を示唆する。問

題作成サーバーの構築により、問題作成者は自分のいる部署からオンラインで問題作成を行う事ができるとともに、2つの問題点、セキュリティと通信経路上の盗聴も作成ソフトとサーバーの設置場所を工夫することによって解決出来、従来の問題作成システムに比べて利便性・堅牢性が高いものとなった。

また、時間と労力がかかるとされた診療問題解決能力の評価についても、自動的に採点し評価するシステムの構築によって評価時間の短縮、採点ミスの減少が実現した。

P-SATの実証実験は、能力評価が一定の分布を示すことが明らかになったが、一方想起的知識評価としては全国規模で行われ信頼度の高い共用試験CBTとの成績の相関を認めなかったことは今後の検討を要する。この結果は、ただちにP-SATの信頼性がないということではなく、P-SATとCBTでは別の能力を測定している可能性を検討しなくてはならない。臨床医に必要なとされる推論能力・判断能力は従来の評価では測定が難しく、今後P-SAT評価結果が、臨床実習あるいは卒業研修などでの臨床推論・判断能力と相関するかの検討を行わなくては最終的信頼性判定ができない。

海外視察においては、ドイツと米国が日本に近く、CCSという点ではアメリカが一番似ているといえる。そしてcommon diseaseの管理という医師の基本的な能力でかつ今まで評価できなかった分野をCCSの形式で試験に導入することは、医療技術の実技試験を導入するのと同様に意義のあることと考える。コンピューター化試験により、従来の試験よりも多様な医師としての能力評価が可能となっている。本邦でも、経済性、効率性だけでなく、試験自体が従来より高度の能力評価を行うことができる利点を取り入れるべきである。しかし、とくにNBMEの視察の結果、CCS試験の実施方法については試験会場の点だけでも日本に応用するのは現実的には難しい。問題作成のためのエネルギーの高さ、及び評価法につき込む労力は、わが国では実現困難なほどに大きいものであった。とは言え、一方でこのような試験を本邦でどのレベルまでで行うかを考える材料とはなったといえる。特に評価法については米国ではかなり綿密な解析を行っており(資料)、本邦での実施においても、より客観的に評価することが重要であり、むしろこのことについて今後深く検討すべきことが考察された。

## E. 結論

医師国家試験のコンピューター化は今後実現する意義が非常に高い。まだ具体的な目標と問題作成のプロセスなど具体化の前に検討すべき問題は残っているが、本研究によりWeb baseで作動する試験システムを開発し、本格運用の目途を立てた。医師としての基本的能力を評価する重要な試験方法として、一定の試験期間を設け、より大規模な研究で検証し、実際の医師国家

試験の中に導入することを提言する。

## F. 研究発表

### 1. 論文発表

Ishihara S, Matsui K, Sato Y, Tang AC, Suganuma T, Fukui Y, Yamaguchi N, Kawakami Y, Yoshioka T. Self-efficacy achieved through problem-based learning tutorial. 医学教育 2007;38:391-397

### 2. 学会発表

選択項目による症例シミュレーション試験の評価

高林克日己 吉岡俊正 細田瑛一

第40回第日本医学教育学会 2008年

## G. 知的財産権の出願・登録状況

### 1. 特許状況

なし

### 2. 実用新案登録

なし

### 3. その他

なし

### 研究要旨

医師国家試験の **Clinical Case Simulation** について Web ベースで動く試験システムを作成し、模擬試験を行うことでその実用性、適応対象、採点方法について検討した。また諸欧米諸国の国家試験とそのコンピュータ化について視察し本邦の今後の国家試験のコンピュータ化について提言した。

### A 研究目的

医師国家試験におけるコンピュータ試験、特に **clinical Case Simulation (CCS)** の導入についてその蓋然性を実用性、適応対象、評価法の見地から検討する。また諸外国における現状を視察し、コンピュータ導入に関する我が国の進むべき方向を示すことを目的とする。

### B 研究方法

1 Web 上で作動する試験システムを構築する。  
その上で CCS 問題を開発し、作成した問題を実際に学生に与え検定するとともに医師にも試験を受けてもらい、その項目選定順序の比較を行うことで、新たな評価ができるかを検討する。同時に学生には試験に対するアンケート調査を行う。具体的には今まで開発してきた MCQ と PMP のコンピュータ化ツールを Web 化した。また将来の国家試験コンピュータ化に向けて、問題集積とその評価方式を確立するために、試験用の端末毎に問題とプログラムをインストールする **Windows Application** 方式ではなく、サーバ側で問題とプログラムを実施できる **Web Application** 方式の MCQ と PMP を開発した。このことにより、ネットワークを利用した広域での問題の集積と評価・分析を目的としたデータ収集が効率的に実施できることを確認した。ここで求めた仕様は以下のごとくである。

#### 1) 実現方式

ネットワーク品質・速度やセキュリティ、本人確認・認証などの技術的課題は将来の課題とし

て、以下の要件を満たすものとした。

- (1) ネットワークを利用したセンターサーバ方式であること。
- (2) 問題の配信・採点がセンターサーバ側で実施できること。
- (3) 試験結果・履歴がセンター側にて蓄積できること。

なお、実施にあたっては個人情報やセキュリティ・問題の漏洩に配慮して、専用の広域イントラネット利用や ID、パスワードによる管理、URL および TCP ポートの秘匿に留意すること。

#### 2) 動作環境

本システムの動作環境は以下のとおりとした。

##### 2.1 サーバ (Windows2003 サーバ SP1 以上推奨)

- (1) **Internet Information Server(IIS)5.0** 以上が動作すること
- (2) メモリを **512MB** 以上実装していること。
- (3) **40MB** 以上のディスク空容量があること。
- (4) インターネット又はイントラネットに接続できること。

##### 2.2 クライアント(WindowsXP SP2 以上推奨)

- (1) **Internet Explorer 6.0** 以上が動作すること
- (2) ポップアップブロックがサーバの URL に対し解除されていること。
- (3) **NET Framework 1.1** がインストールされていること。
- (4) インターネット又はイントラネット経由でサーバに接続できること。

また採点方法の検討として米国 **NBME** の評価

方式などを参考にしながら、我が国の試験に適した採点方法を全体に重み付けして点数化する方法、選択順序から勘案する方法などを比較した。

2 英米独仏の医師国家試験の現状を視察し、今後の試験のコンピュータ化に関する国際的動向を分析し、我が国の取り組み方を考察した。平成17年にドイツ、フランスを視察し、18年にはアメリカとドイツを再び視察した。

### C 結果結果

1 平成17年度に、Web上で作動する試験システムを構築した。その仕様は以下のごとくである。

H/W: Dell Latitude(インテル Core2 Duo T5500、メモリ 1GB) OS: Windows2000 Server

WebServer: Internet Information Server 5.0

Data Base: Microsoft SQL Server 2000

開発環境: Microsoft Visual Studio .Net 2003

クライアント: Microsoft Windows2000 または

XPでMicrosoft Internet Explorer 6.0以上が動作する環境

ネットワーク: イン트라ネット、インターネットを問わず Webサーバとクライアントがデータを送受信する際に使われるプロトコル HTTP通信が可能な環境をもちいた。実際の実施環境は東京女子医大では構内網のイントラネットを使用、千葉大では広域のインターネットを使用した。

また問題の集積を標準化するために記述形式を統一を行なった。問題の登録は本来、登録者認証と共に暗号化された専用のネットワークを経由することとするが、当面は媒体・メール等により問題を集積し、手動(マクロ)でデータベース登録を行うものとした。以下MCQとCCSの問題の構造について述べる。

#### 1) MCQ

MCQは(Comma Separated Values)形式またはEXCEL形式(推奨)とし1問1行で作成されている。各項目は以下の定義に従うものとした。

番号	形式	問題	選択 解答 1	選択 解答 2	選択 解答 3	選択 解答 4	選択 解答 5	解答 組合	正 解	正解率	識 別 指 数	写 真 1	写 真 2	写 真 3	写 真 4	分 野	作 成 者	解 説
K001	K3	負荷試験に用いている診断薬とホルモンの組み合わせで正しいのはどれか。	(1) デキサメサゾン	(2) LH-RH	(3) フロセミド	(4) ソマトチン	(5) 高張食塩水	a (1)(2) b (1)(2)(3)(4)(5)	c	94.5	0 . 0 7					内分泌・代謝	千葉大学	

< STRONG >治療薬はどれか</ STRONG >

■上付き文字

<SUP>~</SUP>

例) m<sup>2</sup>

m<SUP>2</SUP>

■下付き文字

<SUB>~</SUB>

例) CO<sub>2</sub>

CO<SUB>2</SUB>

例) HCO<sub>3</sub><sup>-</sup>

HCO<SUB>3</SUB><SUP>-</SUP>

■色

<Font Color="#0000ff"> RGBによる指定

<Font Color="red"> 色による指定

①番号 登録時のユニークな番号。データベース登録時に自動採番するため登録時の仮番号となる。(大学コード+分野+SEQ・・・定義が必要)

②形式 問題の形式、A、X、K等

③問題文 問題文。使用できる属性は以下のとおり。

■アンダーライン

<U>~</U>

例) 治療薬はどれか

<U>治療薬はどれか</U>

■太字 (強調)

<STRONG>~</STRONG>

例) **治療薬はどれか**

black #000000	gray #808080	silver #C0C0C0	white #FFFFFF
maroon #800000	red #FF0000	purple #800080	fuchsia #FF00FF
green #008000	lime #00FF00	olive #808000	yellow #FFFF00
navy #000080	blue #0000FF	teal #008080	aqua #00FFFF

④選択文

⑤解答組合せ

⑥正解番号

⑦正解率 初期登録時は想定正解率または実施学内での正解率。これはセンターで実施都度、最新正解率に更新される。

⑧識別指数 初期登録時は想定識別指数または実施学内での識別指数。これはセンターで実施都度、最新識別指数に更新される。

⑨写真 問題により参照する写真・図があれば当該イメージファイル名を設定。イメージファイルは JPEG (300×300) 形式を推奨

⑩分野 問題の分野、以下から選択  
循環器、呼吸器、消化器、肝・胆・膵、血液、腎、リウマチ・アレルギー、内分泌・代謝、神経、消

化器外科、循環器外科、脳神経外科、胸部(肺)外科、運動器・リハビリ、女性生殖器、男性生殖器・泌尿器、耳鼻咽喉、皮膚、形成、小児、精神、救急、麻酔、放射線、衛生・公衆衛生

⑪作成者 作成した大学または作成者

⑫解説 問題の解説文 サーバのデータベースには試験実施時数を母数として記録しておく項目を設ける。実施学内において各問題の母数を記録している場合は母数の項目を追加することを検討する。

2) CCS

CCSはEXCEL形式(推奨)とし1問に対してオープニング文書、解説文書、病名を記述したものと症候、一般検査、特別検査、治療について既定(正常)に対して想定病状における状態を記述す



る。各項目は以下の定義に従うものとする。

<例 症候>

用語番号	用語	既定	本問題
100045	妊娠歴	ない	いままで3回流産している
100044	背部痛	ない	
100043	生牡蠣の摂取	ない	
100042	腹痛の放散	腹痛はない	
100041	腹痛の部位	腹痛はない	
100040	喘鳴	ない	わずかにゼーゼーする
100039	薬物アレルギー	ない	
100038	内服薬	常時服用している薬はない	関節の痛みの薬を服用している
100037	頭痛薬の内容	服用していない	
100036	悪寒	ない	
100034	薬の服用	ない	
100033	息苦しさ	ない	息が苦しい
100032	胸痛の性状	胸はいたくない	3時間前から急に前胸部が痛く呼吸が苦しい
100020	口渇感	ない	息が切れ喉が渇く
100010	体重	変化なし	
100009	便秘	ない	
100008	頭痛	ない	
100007	腰痛	ない	
100006	痰	ない	血のついた痰が出た
99999	咳	ない	出る
99996	カレン徴候	ない	
99994	アルコール	のまない	
99988	むくみ	ない	
99987	呼吸苦	ない	胸が苦しい
99986	発熱	36.2℃	
99976	吐物の性状	なにもない	
99974	吐血	ない	
99973	潰瘍の既往	ない	
99972	検診	異常なし	
99971	海外旅行の既往	ない	昨年ソウルに行った
99970	便秘	ない	
99969	心房収縮性雑音（症候）	ない	
99965	めまい	ない	
99964	呼吸数	12から18	28
99963	心電図の既往	正常	
99962	咳嗽	ない	
35120	沢蟹の摂食	最近ない	
35038	眼底検査	正常眼底	
35037	転倒の既往	最近はない	

- ①用語番号 データベースのキー、変更不可
- ②用語 検索時に検索される用語、変更不可。追加用語がある場合は最下行に追加すること。(番号はシステムで自動採番される)
- ③規定 正常時の値または状態
- ④今回 想定問題時の値または状態

この資料は成果物として、旧バージョンの問題を移行して Web 上で動作することを確認し、本方法で、大量の試験者を対象にして同時に試験を行なうことを可能とした。

## 2 試験の実施

試験は平成 18 年度、19 年度に東京女子医科大学、千葉大学医学部各 4 年生各 47 名、43 名に実施した。新たな問題試験問題を 2 題作成し、従来の問題も We 上で動くようにし併せて試験を行った。平成 18 年度は 4 症例の問題を 70 分で解答させたが、一症例における所要時間は 20 分前後であり、それ以上の時間で解答率の上昇はみられなかった。一方資料の表 1 にあるように、C 大学の成績と T 大学の成績には大きな乖離があり、これは終了時間をみても C 大学では解答するのに十分な時間がなかったことを示している。この検討から一題の所要時間は約 20-30 分であると考え(図 1)、平成 19 年にはさらに同等量の別の問題を作成するとともに、一題 20-30 分を標準として試験を行なった。試験問題は昨年までの 5 題のほかインフルエンザウイルスの問題を新たに作成した。これは common disease の中でも特に代表的な疾患であり、また診療方法も検査ではなく本来問診と所見だけで検査をほとんどせずに結論に導かれる点で、今までの問題とも大きく異なったものとなっている。この作成には高林があたり、1 日で作成された。問診選択項目 498 (うち今回追加分 17 項目)、検査選択項目 540 (うち今回追加分 1 項目)、診断選択項目 (うち今回追加分 1 項目)、診断病名選択項目 903 (うち今回追加分 2 項目)、治療選択項目 427 (うち今回追加分 5 項目)であった。これら東京女子医大 4 年次学生 24 名、千葉大学医学部 4 年次学生 24 名、および総合診療部医師 9 名に 5 問題を実施した。

## 3 試験の採点方法と結果

試験の配点は、問診・所見、検査、診断、治療それぞれの正解と禁忌肢に対して全体を 100 点としたときの重みから判断してそれぞれのプラス点マイナス点を与えた。試験の採点は自動的に行なわれ集積された。試験結果の中で、途中で終了したものなどを排除し、得点、選択項目数、選択項目正解率、問診初期選択 5 項目の正解率を算出した。その結果を実得点と問診、検査の選択項目数を比較して表 2 にまとめた。ここで示すように学生と医師では得点は医師の方が高く、また問診、検査選択項目は医師の方が少なく、また得点と選択項目数には負の相関が認められた。これは特に得点と検査選択項目において著明(図 2)であった。しかし図 3 に示すように個々の学生間における得点と問診、検査項目の相関をみると、得点と選択項目数には必ずしも負の相関は認められなかった。これは点数の極端に低い学生の中に選択数の多いもの少ないものが存在するためである。これらの例を除くと概ね得点と選択数との間には負の相関がある。また選択数の中でどれだけ選択すべき項目を選んだ率が高いか(正解選択率)を比較すると(図 4)、いずれの問題でも医師の方が選択数の中の正解率が高かった。しかしさらに初期 5 選択の正解率で見るとこれは学生の方が高い結果であった(表 3)。

学生のアンケート結果では 2 年前と比し、国家試験への CCS の導入にどちらかといえば賛成までが 19%と増加するとともに、どちらかといえば反対までの反対が 49%と初めて過半数を割った。これとともにコンピュータリテラシーの向上から、計 4 回の試験の中でコンピュータ試験の要領がわからない操作ができないといった訴えは両大学とも皆無となった(図 5)。

## 4 米独仏の医師国家試験の現状視察

欧米の試験制度とコンピュータ化について平成 17 年にはミュンヘン市ドイツ連邦共和国バイエルン州医師国家試験事務局、マインツ市医師国家試験製作機関(IMPP)、ボン市ドイツ厚生省医師国家試験担当局を訪問した。またフランス共和国ではパリ第 7 大学サンルイ病院、およびパリ第 12 大学医学部 アンリ・モンドール病院を訪

問した。平成 18 年には米国フィラデルフィアの NBME とドイツ IMPP を再訪した。これらから得られた情報を次にまとめる。

#### 5 ドイツの医学教育システムと試験制度

ドイツはヨーロッパの中では数少ないわが国に最も類似した医師国家試験方式をとっている国と言えることができる。6 年間の医学部教育の後に医師国家試験を合格することが、医師になるための必須条件である。そしてこの医師国家試験も一斉に行われる試験である。ドイツでも国家試験は改革の中にあり、新しい試験 AappO (Approbationsordnung fuer Artzte)は 2003 年 10 月から施行されることになった。

医学研修は徹底的に改革され、内容が近代化され、新しい教育法や試験が導入され、教育の総括的評価が紹介され、また将来に必須な研修が導入された。こうした大学医学教育の向上により、初期研修教育 AIP(Arzt im Praktikum)が廃止され、医師国家試験を通過すれば直ちに臨床医としての免許を取得できるようになったことは、日本とは反対の方向である。

ドイツのすべての大学で医学教育を受けるには入学時に選抜が行われている。これは Zentralstelle für die Vergabe von Studienplätzen (高等教育配置中央局)が管轄している。ドイツでは私学の医学校は 1 つのみで、あとはすべて国立大学である。

**Abitur**(general certificate of aptitude for higher education)

長い間大学における保健専門職につくための必須要件であった Abitur は州の大学入学資格に置き換えられることとなった。結局州政府は Abitur を持たない人に医学教育職を与える権利を有することになった。

#### Curriculum

医学教育は連邦医学法(1987 年制定、2004 年 7 月最新改定)をもって以下のように規定されている。AappO (医師免許法)は 2002 年に制定され 2004 年 7 月に改定されている。

ここでは以下の教育を必要としている

6 年以上の大学医学教育、うち 48 週以上の連続

した臨床研修を受けること

救急治療

3 ヶ月の看護ケア

4 ヶ月の選択臨床教育

2 回の医学試験

新しい医師免許取得規約では実地研修とくに医師の social skill の教育を推進している。患者を全人的に捕らえ専門性を越えた治療に重きがおかれている。臨床の内容も理論的な要素が統合されることを要求されている。結論としてより専門を超えた教育を行い、テーマ別、患者中心、problem oriented な内容が教えられるべきであるし、試験もそのようなものにすべきである。

#### 試験

医学教育改革に伴って、試験も上記の内容に沿った改革が求められる。将来新たな試験形式の発展と履行の機会が与えられるように、医学試験は今後より大学で計画されるべきである。州の試験の数は 2 回まで減少し、MCQ の比重を下げるべきである。そして case-based examination が臨床の専門能力として求められる。試験は次のように構成される

州の一次試験は 2 年間の学習後に行われ、臨床実習に先立つものとして marked course credits を構成する。この course credit が臨床実習のための条件となる。

二次試験は 6 年間の医学教育後に行われる

#### Transitional provisions

連邦医学規約履行の責任と医師免許交付は州によっていて、厚労省は州の決定に対して影響力をもっていない。

#### Model Clause

改革としてのモデルコースが実施されている。これは Charite Berlin 大学で McMaster や Harvard で行われているものを始めているが、またいくつかの他のパイロットも始まっている。

#### Examination certificates

試験を合格すれば州法により当局より証明書が発行される。一次試験とに二次試験の点数にファクターをかけて加えた上で 3 で割って合計点を算出する。最終スコアは小数点二桁まで記録さ

れる。これは EU 法でも認知され EEA の参加国であれば相互に認められるものである。Dr.med を取得するには医学部の卒業が必要であり、臨床研修やスペシャリスト t レーニングは不要である。

#### 専門医養成

卒業後の研修と専門家としての認知は州の health professional 法で定められる。また研修の規約は州で決められているが、どの州もほぼ同じである。

#### 試験の内容

上記に示したように、医師になるための試験として、従来は 4 つの試験があった。すなわち 2003 年以前は 4Semester(2 年次終了)時に、医学部進学試験、Semester (3 年次終了)のあとに第一次試験 (MCQ290 問)、10Semester (5 年次終了)のあとに第二次試験 (MCQ580 問と 2 分野の口答試験)、そして 12Semester (6 年次終了)のあとに口答試験が行われていた。2003 年以降は 4Semester のあと (2 年次終了時)に第一次試験 (MCQ320 問、解剖学、生化学、分子生物学、生理学の口答試験)、そして 12Semester のあと(6 年次終了)で第二次試験 (MCQ320 問、内科、外科、選択科目の口答試験)の 2 回の試験制度に変わった。

#### 6 ドイツ試験制度視察

##### 1) Landespruefungsaeemter fuer Medizin (州立医師試験事務局)

ルートウィッヒ・マキシミリアン大学 (ミュンヘン大学) 医学部内の建物の一室がこれに当てられている。ここには事務職の女性が一名常駐しているだけである。2 次試験は州の試験として行われるが、実際は国中が同じ試験で行われ、合格点も同じである。視察にあたっては、その試験内容について質問し、2 次試験における口頭試問は 4 人のグループに 3 人の試験官が付き、内科、外科と症例に即した問題が行われる。この case based test は午前中に患者を診察し、その午後症例に関する口頭試問を 60 分行うものである。

##### 2) IMPP Institut für medizinische und pharmazeutische Prüfungsfragen (医学薬学

#### 試験問題局)

ドイツ連邦共和国の医師国家試験の中心的存在。マインツにある建物で共同ビルの 1 階から 4 階を占めている。ここでは Juergen Neuser 所長、および Michael Fischer 医学教育指導者などの職員と会談した。問題作成時にはここに全国からの 75 人の問題作成者が集まり協議をする。日本の厚生労働省の試験局に相当するが、試験問題、とくに MCQ を作ることに主目的であり、数学者 (Dietmar Neumann) や統計学者が常駐している。試験問題は一年中インターネットを介して集められ、修正も可能になっている。このように問題作成に関してはコンピュータの利用が行われているが、試験自体をコンピュータ化しようという構想はまだない。その代わり 320 問の問題は次第に case study に置き換わりつつある。この Step2 (卒業時の試験、わが国の国家試験に相当)は 3 日間で一日 5 時間行われるものである。約半数が case study になっている。この問題ではたとえば心筋炎の症例について、免疫学、薬理学、循環器学、放射線学など、多面的な質問が行われ、ひとつの問題の中に 10-15 の小問が用意されている。試験の合格点としては 60%としている。禁忌肢はない。問題は 8000 問がプールされ 15%は以前に使われたものが再使用される。問題は公開されていない。

##### 3) ドイツ厚労省

ボンにある厚労省医学教育試験局の Heinz Haage 部長と会談を行った。

従来は 4 回の試験を 2 回とし、IMPP の作成した 2 次試験は三回しか受験資格がない。また以前には卒後にあつた 1 年半の Approbation(開業認可の研修)を卒前 6 年の中に包含することとした。この点が日本と大きく異なる点である。この理由としては研修医の給与が低く、このためにインターン制度をやめることになったとのことである。多くの医師はこのあと開業するにしても勤務医になるとしても、specialist を志向する。保険医になるためには specialist になる必要がある。ヨーロッパでの医学教育の必要時間は 5500 時間とされており、ドイツはこれを越えているが

オーストリア、チェコでは6年の上にさらに3年間が必要である。ドイツでは年間10000人の医師の育成が必要であるとされる。最近では女医の比率が増えていることがあり、また15%が医師免許を取れない。一方で国外より1500人以上の医師が流入してきて、この1万人という数をクリアしている。

1987年から Model Clause と呼ばれる新しい医学教育法が導入されベルリン大学などで Problem oriented な学習などが行われているが、国試に OSCE や social skill などの試験を導入する計画はなく、各大学に任せるようである。

#### 7 フランスの医学教育と試験

フランスでは公立の大学病院 (Centre hospitalier universitaire : CHU) で医学教育を受ける。6年間の医学教育のうち、専門医になるのに4年、一般医になるのに2年の研修がさらに必要である。6年間の教育は2年の premier cycle des études médicales:PCEM (first cycle) と4年の deuxième cycle des études médicales:DCEM (second cycle)に分かれる。PCEM1年次の終了時に試験があり、これを通してPCEM2に進級できるのは20%にすぎない。すなわち日本の実質的な入学試験に相当する、あるいは非常に厳しい進級試験を2年次に入るところでしていることになる。PCEM2では解剖学、生理学、生化学、微生物学、統計学などの基礎医学を学ぶ。DCEM1は基礎医学と臨床医学が半日ずつであるが、DCEM2-4までは externat と呼ばれ、臨床医学を学ぶ。午前が臨床実習、午後が講義である。この終わりにあるのが CNCI (Centre National des Concours d'Internat) で、この成績順に全国での専門と病院のマッチングが行われ、internat としての研修を始めることになる。この CNCI を受験してからは Certificat de synthèse clinique et thérapeutique を授与されて診療にあたることができる。このあと一般医になるのに2年、専門医になるのに4年を要する。専門医は DES(diplome de'études spécialisées)を授与される。

#### 8 フランスの試験制度視察

#### 1) パリ Saint Luis 病院

サンルイ大学の内科学、リウマチ病学 Sereni 教授と会談を行った。彼はこの大学の教育担当であり、またフランス内科学会の重鎮である。フランスでは上記のように、ドイツや日本の国家試験に相当するものはない。PCEM1からPCEM2への関門は厳しく、一浪までは許される。PCEM2以降は一学年250名の学生が44医学部に在籍するので、一年に11000人の医学生がいることになる。このうちパリには11医学部があるが、これを現在5医学部に統合しようと考えられている。最終的にフランス全土で80大学の中で35医学部になる予定である。CNCI と呼ばれる ranking examination は字数制限のある記述式の試験であり、9つの症例が各8問の質問で構成され、1問1時間、計9時間で行われる。これによりどの病院のどの診療科に進むかを決定するが、放射線科、循環器科、眼科が人気があり、外科、内科の人気は低く、general medicine はもっとも魅力がないとされる。

専門医としての DESC ((diplome de'études spécialisées complément))は老年医学、救急医学、感染症、腫瘍学、集中医学、脈管学などで、これらはそれぞれの学会から発行される。

#### 2) パリ アンリ・モンドール (Henri Mondor) 病院

腫瘍学 Jean-Paul Le Bourgeois 教授と会談を行った。彼は CNCI 局の副所長である。CNCI はフランス教育省と厚労省によって行われる。4800名の研修の専門科と病院を成績でマッチングさせるのであるが、実際に受験しているのは3800名で、1000名の場所が空いている。とくに general medicine が人気がないが、試験を受けていない600-700名のもも零点としてどこかを選ぶ権利がある。このため全然できなくても研修できるのはおかしいとの批判がでている。設問には医学雑誌の批判的吟味を求めるものもある。今後は9症例を6症例に減じるとともに、script concordant test の導入を考えている。このように CNCI は6年次終了時に行われる全国统一試験としては、本邦の国家試験に似ているが、医師の

資格試験ではなく、*trexie'me cycle des e'tudes me'dicales: TCEM* の研修場所を競う試験である。これをコンピュータ化する方向は認められなかった。

## 9 米国 NBME の試験問題作成と試験場見学

Step 1 から Step3 までの問題作成を担当しているフィラデルフィアにある National Board of Medical Examination の本部を訪問した。米国ではこの Step3 に CCS を導入している。米国の試験会場は互いに相手の画面を見ることができず、また試験官より低い位置で試験をする隔壁の中で行われており、試験官からは観察が容易になっているが、少人数が対象の部屋である。これが全米で 150 施設ほどあるという。NBME にはこのようなミニ試験場があるだけでなく、問題作成のための広い施設が提供され 250 名のスタッフが勤務している。ここに全米から集める専門家で CCS を作成しており、一問の作成に 10 人以上の専門家集団が関わっている。またその評価法についても厳密で、資料に示す 3 つの算定方法 Raw score, Rule-based policy capturing, Regression-based policy capturing を比較して、Rule-based policy が優れているとしている。NBME の考える CCS, MCQ の位置づけ、また CCS の評価法について資料を添付する。

## D 考察

### 1 Web による試験の実施

Web を用いた 3 年間の試験において、トラブルはサーバ側には起こらなかった。同時に大量の受験者に対して試験を行なうときには、一定数の端末側の故障は不可避でありこの準備対策は常に必要であるが、8000 名以上を対象としてもサーバのスペックさえ十分であれば、トラフィック的にも問題なくネットワーク環境においては問題なく運営できることが確認できた。以前より円滑にできるようになったもう一つの要因は学生のコンピュータリテラシーの向上であり、こうしたコンピュータ試験に対する拒絶反応は CBT の影響もあるのか、既にほとんどなくなったといっ

てよい。最大の課題としては試験場を通常のスペースより広く、かつ相互の端末画面を見ることが

できないような試験場の構造を考えなければならない。これを 100 名単位の規模で行なうにはまだ課題が残っている。

## 2 採点方法

実際の解答結果を詳細に検討すると、それぞれの学生・医師がどのような意図で診断を進めているのかがよく理解できる (図 4)。これは MCQ 方式の試験で行なうことにすると大量の設問になることや、設問自身が簡明すぎて試験に適さないことから、シミュレーション試験の最大の特徴ともいえる。しかしこれを実際にコンピュータ上で評価するとなるとそれはなかなか困難であり、いくつかの方法が考えられる。その解法は一つではありえず、さまざまな選択順序が考えられるからである。一般に問診・診察や検査において選択する数は少ないほど効率的に探していることになり、これは臨床能力の基準として評価できるかもしれない。しかし鑑別診断の否定用件として必要な項目も含めると、ただ少ないだけではなくてある一定範囲にあるべきと考えられる。今回の結果で学生と医師では得点は医師の方が高く、また問診、検査選択項目は医師の方が少なく、また得点と選択項目数には負の相関が認められた。しかし図 2 に示すように個々の学生における得点と問診、検査項目間の相関をみると、得点と選択項目数には必ずしも負の相関は認められない。これは得点の特に低い一団に選択項目数が少ないものから多いものまでに分散していることにより、時間がなくてできなかったものから要領を得ないでできなかったものまでが含まれると考えられた。これらを除けば、やはり得点と選択数の間には負の相関が見られることから、もしはじめの得点にさらに選択数の大小から換算した得点を加えることは、より正確に評価するのではなく、二重に計算をすることになってしまう可能性がある。それでは選択数の中でどれだけ選択すべき項目を選んだ率が高いか (正解選択率) を比較すると、いずれの問題でも医師の方が選択数の中の正解率が高かった。一方で初期の問診 5 問の選択での正解数を比べると、むしろ学生の方が高かったが、このことはまず診断はじめの浅く広いルー

チンの検索では学生と医師の間に実力差が出ず、評価のためにはそれぞれの深いレベルの診断に入っていくところをチェックしなければならないことを示している。そうであるとまず想起すべき疾患とそのための鑑別診断としての質問群を整理して、それぞれの重み付けをしていけば評価はできるかもしれないが、そこまでしたときの評価が全体の得点と比較して別の能力の検定として計算できるほど精緻なものとするためには相当量の仕事が必要であろう。例えばひとつには同等に並列に質問すべき内容があるときに、それぞれを選ぶことによりさらに次が深い設問になり、配点を比較することが難しくなることである。今後はベクトルの概念などでこうしたグルーピング問題が解決できるか試みるとともに、選択順位の評価を総合評価に加える方式を継続して検討することも考えられる。またこうした試験の結果は一概に正規分布にはなりえない。これは設問数、あるいは評価項目数にもよるが、プラス点のほかマイナス点も存在するからと考える。このことは検定試験としての医師国家試験の評価上で問題になるかもしれない。しかしこの CCS が、知識体系や知識統合能力が不十分であって医師として相応しくないと考えられる受験者を篩いにかけるための評価法と考えるのであれば、ここで得点の著しく低い受験者をスクリーニングし、その内容を詳細に検討することで十分評価できるものと考えられる (図 4 および PDF)。

### 3 問題の作成

昨年までの研究で国家試験基準に合わせた基本的な選択項目とその標準解がほぼ出来上がっている。したがって新しい問題の作成に当たってはこれにその問題に特異的な項目と鑑別のための設問を加えるだけで作成できるので、問題の素案の作成においてこの基本のデータベース上で 30 分もあれば一題の問題を作ることは不可能ではなく、大量に問題を作成できるところまで来ている。問題はそれを客観的、一般的にみても適切で不整合のないものであるのかの細かい確認にある。このためには一問題の作成には少なくとも数名の専門家が担当する必要があるだろう。また

この問題が国家試験においてその対象を common disease を解くためのものであるとすると、subspecialty の専門家だけでなく、一般医の参加が不可欠である。いわゆる common disease は 30 種類くらいに限られるが、より広い意味での一般の疾患の典型例まで加えるとその問題は 100 問以上作成可能であろう。またそうでないと、疾患が限定されてしまい、鑑別すべき疾患が少なくて試験が簡単になってしまう。現在必要とされている国家試験実施後の問題の開示についても、この内容をすべて露呈するわけにはいかないであろう。オープニングシーンと出題された疾患名と治療くらいでないと、以降の問題のヒントを与えてしまうことになるし、それだけでも翌年以降に残された common disease の数を減らすことになってしまうからである。

### 4 諸外国の試験の視察から

とくに NBME の視察の結果、CCS 試験の実施方法についてはこのような多数の試験会場を日本に応用するのは現実的には難しいと考えられ、日本で行うとすれば一斉試験の形で進めるしかないと思われた。また問題作成のためのエネルギーの高さ、及び評価法につき込む労力はわが国では実現困難なほどに大きいものであった。一方でこのような試験を本邦でどのレベルまで行うかを考える材料とはなったといえる。特に評価法については米国ではかなり綿密な解析を行っており (資料)、本邦での実施においても、より客観的に評価することが重要であり、むしろこのことについて今後深く検討すべきことが考察された。ドイツ IMPP の試験は、一つの症例を単に診断治療をするのではなく、多数の分野、領域から多面的に質問して答えを得ようとするもので CCS とは異なるが、本邦にはない試験形式として大いに参考となるところがある。しかしながらその評価法についてはさらに議論を要するし、このことを含めて今後も交流する中で彼らのコンピュータ試験との比較検討が望まれる。

以上を総合すると、英国では試験官がペアで受験者と一緒に診療行為を見ながら判定してライセンスを与えるし、フランスは順位づけ(コンク

ール)はあっても既に医学部を卒業しただけで医師免許は授与されるので日本の医師国家試験としての存在ではない。ドイツと米国が日本に近く、CCS という点ではアメリカが一番似ているといえる。そして **common disease** の管理という医師の基本的な能力でかつ今まで評価できなかった分野を CCS の形式で試験に導入することは、医療技術の実技試験を導入するのと同様に意義のあることと考える。試験の評価方法として NBME で行なわれるような方法を検討するとすると、われわれの評価に対する研究は米国に比し不十分ということになるが、従来の正しい項目の選択で加点する方法で得点をつけ、著しく点数の低い受験者を検討対象とし、その選択順序などから詳細に検討することで、不適格者を除去する手段にすることは十分適切な判断であると考ええる。医師国家試験のコンピュータ化研究はこれで 9 年になるが、このシステムが評価法まで含めてほぼ完成したことと、この間学生のコンピュータリテラシーの著しい向上から試験自身の導入においての根本的問題はなくなったと考える。そして全国規模に展開し、実際の国家試験の中で試験的に運用していく時期になっている。今後一定試験期間を経て CCS が十分周知できた段階で、MCQ についてもコンピュータすることで試験の完全電子化が可能になるであろう。

## E 結論

Web base で作動する CCS 問題の試験システムを開発し、本格運用の目途を立てた。 **Common disease** 管理という医師としての基本的能力を評価する重要な試験方法として、一定の試験期間を設け、より大規模な研究者で検証し、実際の医師国家試験の中に導入することを提言する。

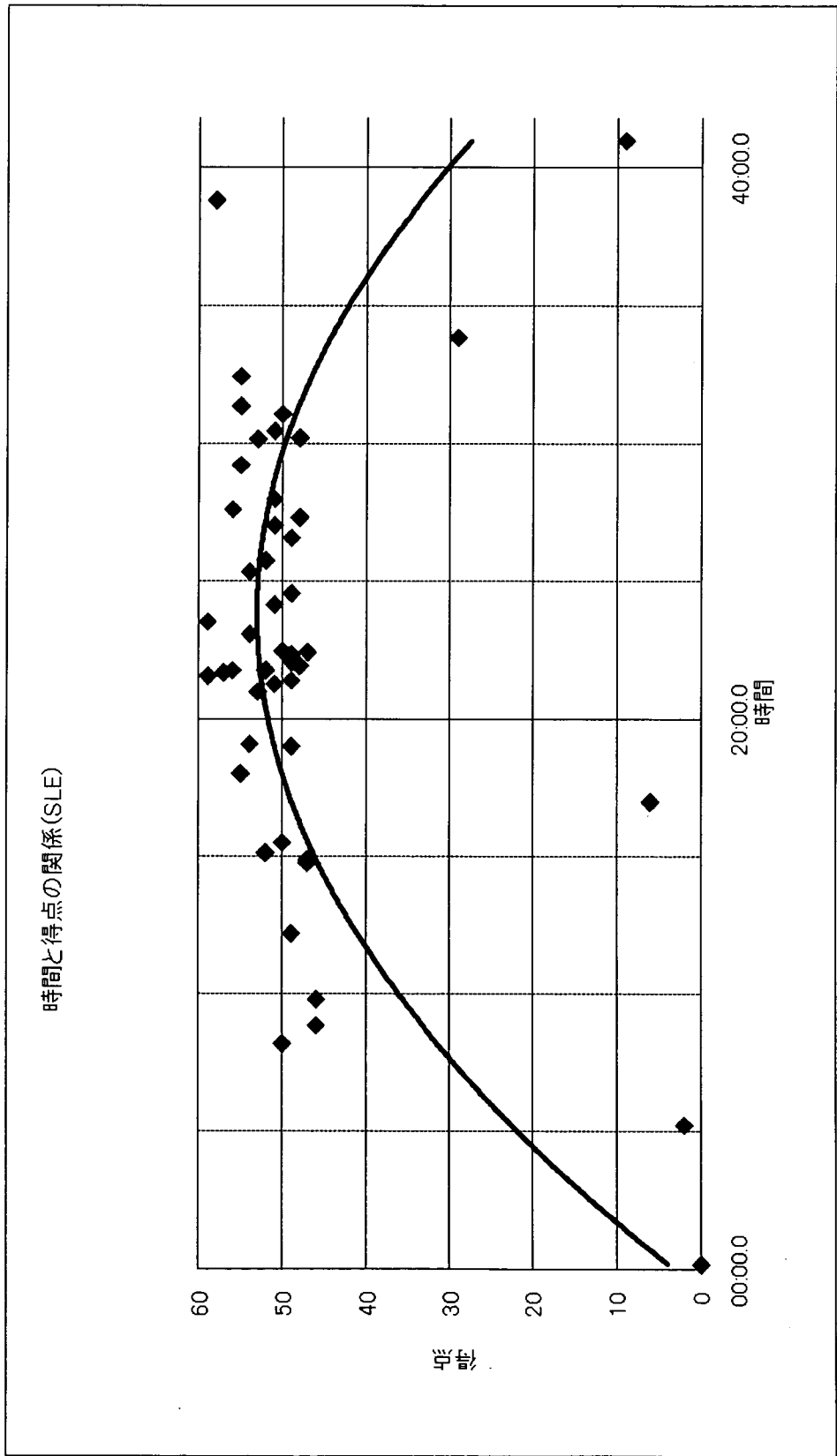


# 表1 H18年度CCS結果

腎盂腎炎	2大学	C大	T大
被験者数	88	43	45
平均点	31.18	21.49	40.44
標準偏差	18.09	14.61	16.25
最高点	79	52	79
最低点	0	0	2
平均時間	15:59.9	09:41.7	22:01.4
最大時間	38:16.0	24:53.0	38:16.0
最小時間	00:09.0	00:09.0	03:33.0

SLE	2大学	C大	T大
被験者数	80	33	47
平均点	35.80	19.82	47.02
標準偏差	22.01	21.57	14.03
最高点	59	54	59
最低点	0	0	0
平均時間	15:51.9	06:37.0	22:21.1
最大時間	40:56.0	21:15.0	40:56.0
最小時間	00:06.0	00:06.0	00:11.0

# 図1 所要時間と得点の関係 SLE T大学



# 図1 所要時間と得点の関係 SLE C大学

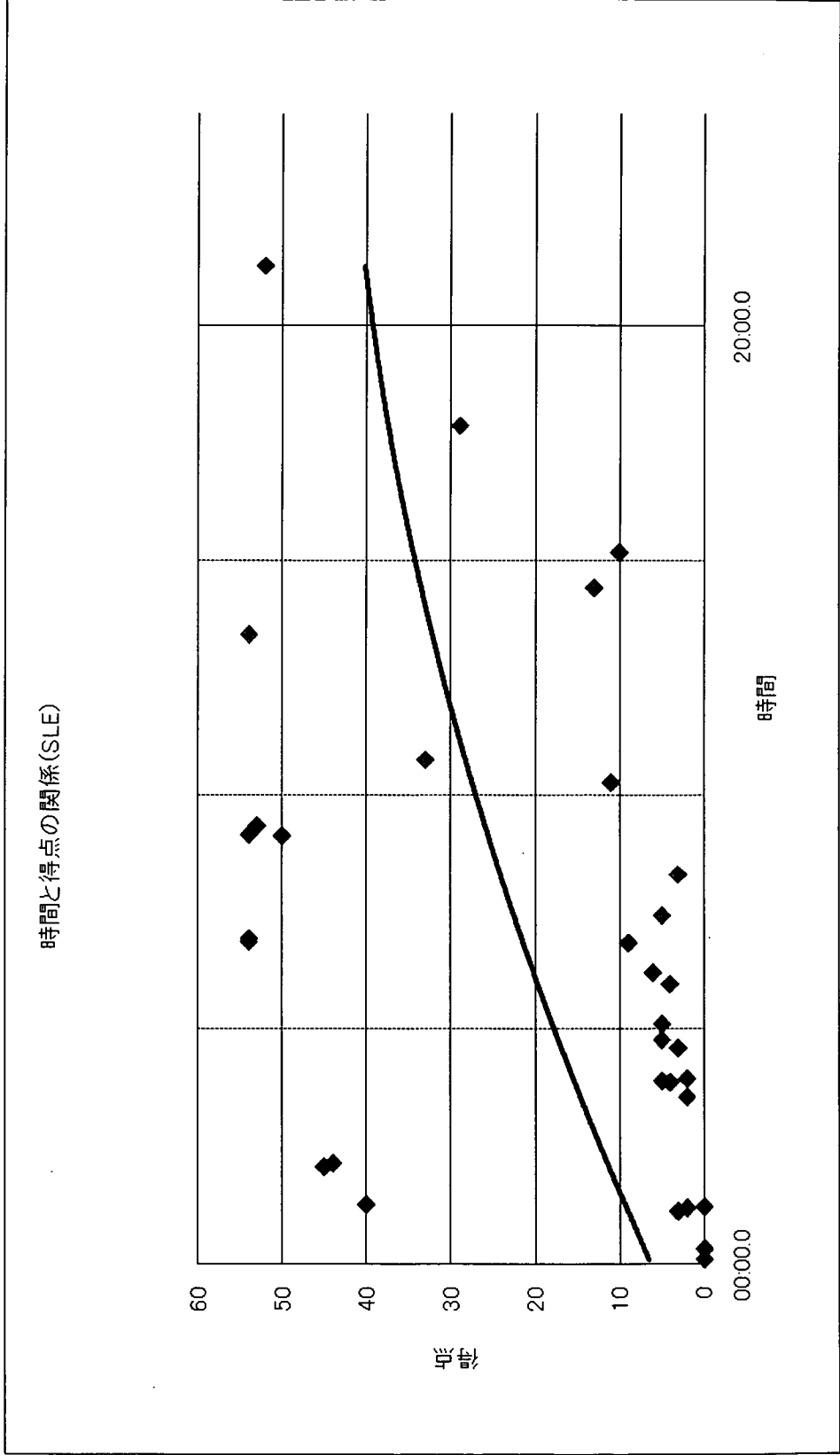


表2 得点と選択数の関係

	数	得点		問診数		検査数		
		医師	学生	医師	学生	医師	学生	
インフルエンザ	24	9	39.7±31.9	66.7±14.2	22.4±12.8	12.4±4.7	30.7±17.7	4.8±6.8
SLE	22	3	16.2±16.1	52.0±4.4	20.0±13.0	18.0±8.1	30.3±17	27.0±26.0
腎盂腎炎	21	4	30.2±18.1	44.3±2.8	14.2±9.0	7.75±3.3	28.2±15.5	17.8±3.9
アニサキス	13	7	96.2±13.9	78.6±26.7	5.2±2.1	8.7±4.5	10.2±10.1	10.0±12.0
癒着性イレウス	8	4	51.3±27.5	55±26.5	16.1±7.3	13±4.7	17.8±11.0	3.8±4.3

平均±SD