

## 総 説

# 看護師等国家試験の CBT (Computer-Based Testing) 化と CBT の解説

西川 浩昭\*

2006 年 3 月 31 日、各府省情報化統括責任者連絡会議において、国家試験業務の業務・システム最適化計画が決定された<sup>1)</sup>。これは現在、国が法令により設けている 280 の資格試験および政府が直接実施する国家公務員採用試験の全てを対象としているが、当面は国が直接実施し、かつ業務の一部をシステム化している 11 府省庁の実施する 32 の国家試験が対象となっている(表 1)。この中には厚生労働省が実施している医師国家試験をはじめとして、看護師、保健師、助産師の各種国家試験が含まれており、看護系の教育機関としても対岸の火事という状況ではなくなっている。

ここで最適化の対象となる業務は、

1. 国民(受験者、合格者等)本位の利便性の高い情報サービスの実現
  - (1) 受験者による、受験案内、受験願書、その他受験に係る手続き及び試験結果に関する情報の迅速な電子的入手の実現
  - (2) 受験願書のオンライン出願による電子申請化の促進
  - (3) 受験費用・登録料等の多様な支払方式の実現
  - (4) 試験結果に対する情報提供及び情報公開
2. 国家試験業務の標準化、共通化及び外部委託化による効率化と汎用化の実現
  - (1) 業務処理方法の標準化、共通化及び外部委託化
3. 各試験で共同利用できる統合管理システムの構築及び運用
  - (1) 柔軟かつ汎用的な統合管理システムの構築及び運用
  - (2) 統合管理システムの 3 つの主要なサブシステム
  - (3) 統合管理システムの利用・運用

表 1 最適化計画対象国家試験一覧

府省名	国家試験名
人事院	国家公務員採用試験
金融庁	公認会計士試験
法務省	司法試験
	司法書士試験
	土地家屋調査士試験
外務省	外務省専門職員採用試験
財務省	通関士試験
	税理士試験
文部科学省	高等学校卒業程度認定試験
厚生労働省	医師国家試験
	歯科医師国家試験
	保健師国家試験
	助産師国家試験
	看護師国家試験
	診療放射線技師国家試験
	臨床検査技師国家試験
	理学療法士国家試験
	作業療法士国家試験
	視能訓練士国家試験
	管理栄養士国家試験
	薬剤師国家試験
農林水産省	普及指導員資格試験
	土地改良換地士資格試験
	獣医師国家試験
経済産業省	計量士国家試験
	弁理士試験
	特定侵害訴訟代理業務試験
国土交通省	測量士試験
	測量士補試験
最高裁判所	裁判所事務官採用Ⅰ・Ⅱ種試験
	家庭裁判所調査員補採用Ⅰ種試験
	裁判所事務官採用Ⅲ種試験

\* 日本赤十字豊田看護大学

- (4) 府省共通システム等との連携
- (5) 先行する試験システム等の再利用

#### 4. システムの安全性・信頼性の確保

- (1) 本業務で保有する受験者等に係る個人情報等のデータの、外部遺漏を防止する
- (2) 構築したシステムの障害発生等による被害の極小化

である。

国家試験業務の標準化の中に、企画・計画業務及び試験開発業務が含まれており、そこでは試験問題作成支援及び試験問題データベースの構築として、「試験問題は、試験問題本文並びに検索及び管理等に必要となる問題の属性を含む試験問題データベースを構築し、試験の品質の向上に資するようにする。これらの試験問題には、実績データ及び個々の問題ごとの分析データも統合的に保存し、新規の問題を作成する際に問題作成者や管理者が参照できるようにする。今後のコンピュータを利用した試験方法（CBT: Computer-Based Testing）への対応、試験問題の機密性及び試験種別ごとの管理を考慮してデータベースを構築する」とされている。こうした点を鑑みると、現在は紙媒体で実施されている看護師や保健師の国家試験も、将来的には CBT で行われる可能性が高い。医学系では既に医学共用試験として臨床実習を行う前の学力判定に利用されているが<sup>2)</sup>、看護系では未だ実施されていない。現代の学生については、コンピュータを利用する試験であっても、多少のとまどいはあるかも知れないが、数回の練習を行えばそれほど大きな問題はないと思われるが、学生の練習用とはいえ、試験問題を作成し、実施する側の教員については大きな負担とストレスがかかることは言うまでもない。ここでは、将来必ずや訪れる看護師等国家試験の CBT 化に備え、CBT の概略と特徴について紹介する。

CBT とは、前述した様に Computer-Based Testing の略称で、コンピュータのディスプレイ上に問題を表示させ、受験生はその問題に対して正しいと思う選択肢をマウスでクリックするか、該当する番号をキーボードから入力することによって回答するテスト形式である<sup>3)</sup>。ディスプレイに質問を表示させ、回答をキーボード等から得る方法は古くから存在しているが、これを多人数に対して体系的なテストとして実用化したのは、1986 年に米国の教育・調査・測定分野の非営利団体である ETS (Educational Testing Service) が大学生の能

力別クラス編成用のテストとしたのが最初とされている<sup>4)</sup>。1990 年代より IT 系企業の認定試験に利用されたことを通して大きく発展し、現在では米国の外国人留学生のための語学テスト (TOEFL Test Of English as a Foreign Language)、大学院入学試験 (GRE Graduate Record Examination)、医師国家試験 (USMLE United States Medical Licensing Examination) や英国の運転免許試験などの公共性の高い試験からベンチャー企業の認定試験まであらゆる分野において普及している<sup>5)</sup>。我が国においては前述した TOEFL や医学部・歯学部との共用試験の他に、漢字能力検定 (漢検) などの試験で用いられている。試験を CBT で行う利点としては、結果が試験の終了と同時に確認することができる、決められた日時や場所で受験する必要がなく、受験者の都合に合わせて受験する場所や日時が自由に選択できる「随時型受験」が可能などなどが挙げられる。ただし、実際の問題として国家試験などの資格試験として行う場合には、受験生間の公平性を保つ必要があるため、同一時刻に受験生を試験会場に集合させて行う必要があり、今後我が国において資格試験として行う場合のハードルとなりうると思われる。

CBT はコンピュータに管理される必要があり、テストを処理するための十分な数のコンピュータと実施用のサーバーが必要となる。実施会場ごとに、試験用のサーバーを設けることは困難であるため、インターネットのサイトを利用して行うことも少なくない。インターネットを利用したテストは特に IBT (Internet-Based Testing) または WBT (Web-Based Testing) と呼ばれるが、これも十分な通信速度と回線の安定性が得られるのであれば可能であり、今後のテスト形式として注目されている<sup>6)</sup>。

なお、前述した CBT の例として、受験生が単一の選択肢を選択する単一選択形式のものを挙げたが、大半の CBT ではこの形式の試験が用いられている。受験生が条件に該当する複数の選択肢を選ぶ複数選択形式や記述式の問題も技術的には出題することが可能であるが、結果が試験終了直後に得られないという理由から、実際にはほとんど行われてはいない。

CBT には、その目的・方法によりいくつかの種類が存在する。代表的なものを紹介する<sup>3)</sup>。

### CFT (Computer Fixed Test)

試験時間を固定し、全受験者が同一の問題を受験する、最も単純な形式であり、紙媒体の試験のコンピュータ版ともいえる方法である。出題の順序については全受験者に対して同じ順序で出題する方法と受験者ごとに無作為に変更する方法がある。

### ATA (Automated Test Assembly)

全受験生について試験時間は固定するが、出題される問題は受験生ごとに、あらかじめ作成し蓄積してある問題プールから無作為に選ばれた問題を受験するため、同一の問題であることは保証されない。一定の難易度の問題のうち、どれだけの数（または割合）の問題に正答したかをもって、受験者の能力がある一定水準以上にあるか否かを判定することを目的とした試験方法であり、我が国の医学・歯学系で現在実施されている共用試験はこの方法が採用されている。

### CAT (Computerized Adaptive Test)

個々の受験生の能力を正確に把握することを目的として実施される方法で、受験者ごとに、ある問題の回答の正誤に基づいて次の問題が問題プールから選ばれるシステムで、出題される問題数、試験時間は各受験生により異なっている。具体的には最初の問題に正解した場合には、2問目は前問より難易度の高い問題を出題し、逆に不正解だった場合には、前問より難易度の低い問題が出題される。この作業を繰り返し、個々の受験生のより詳細で正確な能力を把握する方法である。IBTで受験する様に変更された新しいTOEFLはこの方法で実施されている。

### CCT (Computerized Classification Test)

CATに類似した方法で、CATが個々の受験生の能力を正確に把握することを目的としているのに対し、CCTは各受験生を合・否などの複数のカテゴリーに分類することを目的としている。

看護師等の国家試験に関しては、未だ実施することさえ正式には決定していないが、もしCBTで行われるようになった場合には、試験の公平性という見地から、CFTで実施される可能性が高い。そうしたことを踏まえ、以下ではCFTについて詳述する<sup>7)</sup>。

古くは、Computer-Based Testing または CBT と言え、CFT を指していた。その後様々な新しい出題方法が考案され、CBT はテスト法の総称を意味するようになっていった。そのため、CFT は第一世代のコンピュータテストとされている。問題がコンピュータにより提示され、回答後瞬時に採点され、得点が算出される。問題には、図や写真などの静止画像はもちろんのこと、音声やビデオの画像（動画）などを用いることも可能である。また、CFT では CAT などと異なり、受験生の回答結果によって次に出題する問題を決める必要がないため、受験者が既に回答した問題に戻って、回答を修正できるようにすることも可能である。この融通性は他の出題方法には見られない。さらに CFT では出題する問題は全受験者に対して同一であらかじめ決まっているため、受験者の能力に準拠した問題を選択したり、問題プールから選択したりすることがないため、他の方法に比べ、テスト開発が容易である。他方、全受験生に対する試験問題が同一のため、CBT の本来の随時型受験という方法をとる場合には、安全性を考慮すると、少ない受験生に対して実施するか、試験問題を公開している試験期間を短くし、短いサイクルで出題する問題を変更することが望ましい。このため、CFT ではテスト結果に対して、IRT (Item Response Theory) などのテストデータ用に開発された解析手法を用いるのは対象数の少なさから困難であったが、ある程度の数の受験生を試験会場に集め、同一時刻に実施するのであれば、この点の問題は解消される。

### プレテストと問題プール

十分な数の問題が集まったならば、作成に携わったスタッフや実際に受験する者以外の者によってプレテストを、可能ならば複数回、実施する必要がある。このために、先に述べた医学共用試験では4回のトライアル（試行テスト）を実施している<sup>8)</sup>。この大きな目的は次の項目で解説するが、問題ごとのパラメータを算出し、問題の適・不適を判断することである。ある問題が適当と判断された場合、そうした問題を集めて問題のデータベースである問題プールを作成する。一度使用した問題についても、その際に算出されたパラメータの値とともに、このプールに戻して蓄えておく。こうすることにより、過去に作成した問題を有効に再利用することが可能となるばかり

か、一度用いた問題であってもその後に適当ではないと判断された場合に、問題プールから外していけば、次第に問題が淘汰され、良問のみで構成される問題プールができ上がる。なお、統計数値を用いた数値問題の場合には、変化した数値だけ変更すれば差し支えないとされている。プールする問題数は多いに越したことはないが、あまりに多すぎても使われない問題が増えるだけということもあるため、目安として全領域の合計で1,000～2,000題もあれば十分と考えられる。

CFTに限らず、全てのCBTについて問題プールを作成する必要がある。これは前述した様に、作成された問題を淘汰し、良問のみを残して使用していくためであるが、CFTの場合、CATの様にある問題の出来・不出来によって次に出題する問題を選ぶ必要がなく、全受験生に対して同一の定められた数の問題を出題すればよいいため、それほど大きなプールは必要としない。それでも数百題から千題程度は必要であると考ええる。

## テストの設計に関する問題

### テストの設計

CFTの利点はいつでも自由に回答を中止できること、前の問題に戻って回答の確認ができること、出題される問題を先に見ることが可能な様に設定できることである。しかし、出題される問題を先に見ることができるという点は、この方法の弱点でもあり、このためCFTでは受験者が一度回答を始めたならば、中断することなしに試験を終了させる必要がある。

#### 問題の作成とその管理・評価

CFTではどんな問題でも出題することは可能であるが、実際に用いる試験問題では、2つのパラメータに拘束される。1つは問題の難易度であり、もう1つは問題の識別力である。テストが受験生の能力を測ることを目的としている以上、あまりに難しい問題や易しすぎる問題を出題することは意味がない。これは大半の者が回答できない問題や逆に殆どの者が回答する問題では、受験生の能力を測定するという目的に合致しないためである。それゆえ、一般には難易度の指標である正答率が30%から80%の間になければいけないとされている。また、試験問題としては、成績が良い者が正解し、成績の悪い者が正解できない問題が優れている。これを数値化したものが、識別力である。これは該当している問題

と合計点との点双列相関係数で求められ、0.2未満は不適当な問題とされている<sup>7)</sup>。これらのパラメータはプレテストかまたは過去に実施した試験結果からその都度算出され、個々の問題に関するパラメータとして、問題と一緒にプールされている。

## 受験者の問題

### 回答のチェック

受験生にとって従来までの紙媒体による試験方法と大きく異なる点は、各自の答えを記録しておくものが存在しないという点であると思われる。問題冊子を回収する方法であっても、各問題の答えをメモしておくことができるが、CBTによる試験が実施されれば、そのようなことは不可能となる。実施する側も不正行為や問題流出を避けるため、メモ用紙などを配布することはあり得ない。つまり、CFTの設定として既に回答した問題に戻って、回答を訂正することが可能とした場合でも、チェックミスのための見直しは容易ではない。逆に言えば、最初に回答した際にチェックミスをしないような細心の注意と、全ての問題に回答した後で見直すのではなく、確認は個々の問題を回答した際に行うことが求められる。これは看護師としては、必須の要件であることは言うまでもないが、テクニックに依存する部分も少なくないため、指導する立場にある者がきちんと指導しておく必要がある。

### コンピュータの使用経験による影響

これも、以前から問題とされてきた点であるが、最近の学生については情報処理などの演習で経験しているのみならず、レポート等の作成でもコンピュータを使用しているので、それほど大きな負担にはならないと考えられる。ただし、コンピュータの使用経験は豊富でも、CBTについての経験はほとんどなく、さらにそれが国家試験として実施されるとなると、より大きな緊張を生むことが予想されるので、模擬試験などで十分に経験を積ませることが重要であると思われる。

### 試験時間と試験回数

受験生にとってのもう1つの問題は、試験を受けるという作業の他に、コンピュータのディスプレイを熟視しなければならないことである。VDT作業の連続時間は

1時間程度といわれていることを考慮すると、CBTによる国家試験を実施する場合には、現在の午前に実施されている150問2時間45分は2回か3回に、午後に実施されている90問2時間30分は2回に分けて実施する必要があると生じてくる。人間の緊張の持続時間を考えると2時間45分や2時間半という長時間の試験よりも1時間程度の短時間の試験を数回行う方が良い様に思われるが、どこかのパートで犯したミスの影響が尾を引き、その後の試験結果を左右する可能性も否めない。しかし、これについては過去の実証例がないので判らない。

#### 実施施設のインターフェイス

これは、何もCFTに限ったことではないが、CBTを実施する場合には、実施する機関が、問題や試験実施のためのプログラムなどが収納されているサーバーを受験者の学校、企業などに持参し、実施施設のサーバーに接続して行うのが常である。この場合、実施施設に施設内LANが整備されていれば、ネットワークに接続して試験を実施できるため、実施施設側のシステムに大きな問題は生じない。

#### 等価処理

紙媒体に比べ、CBTによる試験が優れている点は、今実施した試験結果を過去の試験結果と比較して、今回の試験の難易度と個々の受験生の能力を容易に判定できる点である。試験の成績は、問題の難易度と受験生の能力の2つの要因によって決定される。同じ試験問題を使っている以上、今回の受験生の能力を比較することは可能であるが、過去の試験を受験した受験生と比較することは通常は不可能である。同一の問題を使用した場合には、2回目以降に受験した受験生が有利になるのは言うまでもなく、異なった問題では問題の難易度の影響を排除することができない。このため、出題された問題の難易度と受験生の能力値を同時に明らかにする必要がある。この場合に、必要とされる作業が等価処理と呼ばれており、その理論的な裏付けとしてIRT (Item Response Theory 項目反応理論) が存在している<sup>9,10)</sup>。

これは、実施する試験の問題の中に数問(多くとも5%程度)、難易度が明らかで、適当な識別力を有する問題を加えておくことにより、過去の試験で算出されたその問題のパラメータから受験生の能力と他の問題の難易度

を反復推定により算出する方法で、受験者の能力水準とは無関係に、問題の難易度に関する共通な尺度(物差し)を作成し、絶対評価を行うことができる。これにより、今回出題した問題の難易度と過去に出題した問題の難易度を等価させ、今回出題した全問題の難易度を算出することができ、その値から個々の受験生の能力値を算出する。

この方法は、現行の国家試験の様に、1つだけ存在する正答の選択肢を選ばせる、単一回答の場合が最も適しており、複数の正解が存在して部分点(中間得点)が存在したり、結果に採点者の主観が含まれる記述式の問題には適していない。

#### プレテストによる問題選別の例

どんなに優れた問題であっても、1題でその領域の全ての内容を網羅することは不可能であるため、現行の国家試験と同様に1つの領域について数題が出題される。今、A、B、Cの3つの領域からそれぞれ3問ずつ計9問を出題する場合を例としてプレテストにより、実際に使用する問題を選ぶ過程を解説する。

作成された問題が全て、出題に適していることはあり得ないため、実際には出題する問題より多い数の問題が必要となる。今、A、B、Cの各領域において5問ずつ問題が作成されたとする。この全15問を使用してプレテストを実施した結果が表2のようになったとする。先に述べたように、難易度(正答率)が0.3から0.8の間の値で、かつ0.2以上の識別力が必要とされる。この場合、

表2 プレテスト結果

問題	領域	難易度	識別力	
1	A	0.20	0.05	*
2	A	0.50	0.30	
3	A	0.70	0.20	
4	A	0.60	0.45	
5	A	0.80	0.15	*
6	B	0.60	0.50	
7	B	0.90	0.20	*
8	B	0.50	0.40	
9	B	0.20	0.15	*
10	B	0.70	0.35	
11	C	0.50	0.50	
12	C	0.60	0.60	
13	C	0.20	0.10	*
14	C	0.40	0.25	
15	C	0.30	0.15	*

表3 プレテスト結果(2回の合計)

問題	領域	難易度	識別力	
1	A	0.20	0.05	*
2	A	0.50	0.30	
3	A	0.70	0.25	
4	A	0.60	0.45	
5	A	0.80	0.20	
6	B	0.60	0.50	
7	B	0.80	0.20	
8	B	0.50	0.40	
9	B	0.30	0.20	
10	B	0.60	0.35	
11	C	0.50	0.45	
12	C	0.70	0.60	
13	C	0.25	0.20	*
14	C	0.40	0.25	
15	C	0.30	0.15	*

\*印をつけた1、5、7、9、13、15の6問が不適当ということになる。再度のプレテストを実施したところ、表3の様に5、7、9の3問は条件を満足する結果が得られたが、1、13、15の3問は条件を満たさなかったとした場合、1、13、15の3問を除いた12問をプールすることになる。実際に試験を行う場合には、領域AとBについては、それぞれ2～5の4問の中から3問、6～10の5問の中から3問を選んで出題し、領域Cについては11、12、14の3問全てを出題する。この場合、領域Aでは1問、領域Bでは2問交換可能な問題があるが、領域Cでは交換可能な問題は1問もなく、2回目の試験を実施する場合にも11、12、14の3問を再び使用しなくてはならない。この場合には2回目の試験においては、領域AおよびBに比べ、1回目と全く同じ問題が出題された領域Cの成績が高くなることが予想できる。この様に個々の問題について使用される割合が、際だって高いことは好ましくないため、問題プールには十分な数の問題を蓄えておく必要がある。ただし、問題を作成する教員の側から見れば、実際に出題される数以上の問題を作成しなければならない上に、単に問題のパラメータの値が悪いという理由だけで、作成した問題が不適当とされることになるわけで、問題の作成には今以上の労力と覚悟が必要とされることになる。

## おわりに

国家試験をCBT化することの是非は別にして、表1に示した32の国家試験の総出願者数は平成16年度にお

いて約58万人を数え、これらを最適化することにより平成20年度から23年度までの4年間で運用経費が9.1億円削減されることが見込まれ、さらに国家試験の利用者である国民の利便性の向上分として4年間で19.2億円の削減が見込まれている。財政再建のために、政策上経費節約が求められている現状では、最適化計画が実施されることは否めない。

大学において6年制の課程を修了したもののだけが受験する医師国家試験や歯科医師国家試験と異なり、看護師の場合には准看護師を対象とした2年制の課程、専門学校・短期大学の3年制の課程、4年制大学の4年制の課程と様々なコースがあり、各課程において履修する内容も、教育施設が有する設備も異なっているため、これらの問題を完全にクリアして国家試験を最適化しCBTで実施することは、医学ほど容易ではないと思われる。さらに、実施するとなると試験会場や使用する機種、OSなどの問題が山積している。しかし、時代の流れから見てそんなに遠くない将来、CBT化されることは避けられないと思われる。

学生については、コンピュータの使用自体に抵抗はなくても、それが国家試験として実施されることによる緊張感の増加、自分が回答した内容についてメモしておくことができず、ケアレスミスが生じ易いなどの問題が生じる。他方、教員に対しては、試験に使用する以上の数の問題を作成することが必要となり、さらに作成した問題が受験生ができないほど難しいのはもちろんのこと、易しすぎてでき過ぎてもいけないし、他の領域の正答状況と整合性がとれている必要があるため、今まで以上に試験問題作成に労力が必要になることが予想される。現時点においても、これらの問題が予想されており、実施された場合にはこれ以外の問題が生じることもあり得るので、早期の対策と心構えが必要と思われる。

## 文 献

- 1) 情報システム統括責任者(CIO)連絡会議,「国家試験業務・システム最適化計画」[http://www.kantei.go.jp/jp/singi/it2/cio/dai19/19siryou08\\_01.pdf](http://www.kantei.go.jp/jp/singi/it2/cio/dai19/19siryou08_01.pdf), 2006. 3. 31
- 2) 共用試験実施機構: 共用試験とは—我が国の医学教育向上のために—, 医学部学生向けパンフレット, 2003

- 3) Parshall, G. C., Spray, A. J., Kalohn, C. J., et. al.:  
Practical Considerations in Computer-Based Testing, 1-12, Springer, New York, 2001
- 4) Crocker, L. & Algina, J.: Introduction to Classical and Modern Test Theory. Fort Worth: Holt, Rinehart & Winston, 1986
- 5) 池田央, 林規生: Computer Based Testing の現状と開発, 日本行動計量学会第7回セミナー講演論文集, 40-57, 2004
- 6) Parshall, G. C., Spray, A. J., Kalohn, C. J., et. al.:  
Practical Considerations in Computer-Based Testing, 47-69, Springer, New York, 2001
- 7) Parshall, G. C., Spray, A. J., Kalohn, C. J., et. al.:  
Practical Considerations in Computer-Based Testing, 92-105, Springer, New York, 2001
- 8) 仁田善雄, 石田達樹, 前川眞一, 他: CBT 医学系共用試験トライアルの総括報告, 日本テスト学会第4回大会発表論文抄録集, 116-117, 2006
- 9) 池田央: 現代テスト理論, 51-86, 朝倉書店, 東京, 1994
- 10) 豊田秀樹: 項目反応理論 [入門編], 45-129, 朝倉書店, 東京, 2002