

200400960B

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～16年度 総合研究報告書

主任研究者 細田 瑳一

平成17(2005)年4月

目 次

I. 総合研究報告

医師国家試験コンピュータ化に関する研究	-----	1
細田 瑳一		
高林 克日己		

II. 分担研究報告

1. コンピュータ試験問題作成と運用の研究	-----	14
高林克日己		
2. 試験問題のトライアルの実施と検討	-----	25
福島 統		
3. 試験問題のトライアルの実施と検討	-----	27
吉岡俊正		
4. 試験問題のトライアルの実施と検討	-----	28
江口光興		
5. 試験問題のトライアルの実施と検討	-----	29
杉山幸比古		
6. 試験問題のトライアルの実施と検討	-----	31
山内俊雄		
7. 試験問題のトライアルの実施と検討	-----	32
原田研介		

III. 研究成果の刊行に関する一覧表	-----	34
---------------------	-------	----

IV. 研究成果の刊行物・別刷	-----	35
-----------------	-------	----

厚生労働科学研究費補助金（医療技術評価総合研究事業）
（総合）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 瑳一 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
高林 克日己 千葉大学医学部附属病院 医療情報部 教授

研究要旨

【目的】国家試験にコンピュータを利用した試験を導入する蓋然性を検討する。【方法】MCQ（CBT, CAT）、PMP問題作成ツールを独自に開発した。このツールを用いてMCQ問題100題と患者管理問題（PMP）5題、及びそれに対応するMCQ問題10題を作成し、6つの大学の医学生5、6年生を2群に分けて従来のマークシート法（MAR）とコンピュータ試験（COM）で半分ずつを行い、全てをあわせることでCOMとMARの相関を検討した。また受験者のアンケートをとって情報を収集した。また、この研究を進めるに当たって米国NBME、英国General Medical Council、オランダのCITO及びドイツのMEDISを訪問し欧米の実情を調査した。

【成績】COMとMARのMCQに関する結果は近似しており、試験運用上でも大きな問題はなかった。PMPのCOMの結果とMCQ50題の正解率には明らかな相関はみられなかった。またPMPに対応するMCQ問題とも高い相関はなく、PMPでの評価判定は知識のファクターだけではないことが推測された。またPMP問題ではとくに能力の低い学生の識別が明瞭であった。しかし大量の受験生の同時受験などでの運用上の問題が解決されていないことと、学生自身は賛成するものより、反対意見を述べる者がMCQで82:381、PMPで116:402と圧倒的に多かった。決定的に有用な事由がないと、コンピュータ試験を早急に導入することにはさまざまな困難があると考えられるが、試験自身には問題のないことが示されたことから、CATの導入などさまざまな利点を考慮して、欠点を克服した段階で近い将来導入されることは十分考えられる。

【結論】試験方法としては国家試験にコンピュータを導入し、従来のマークシートに変わりCBTのように試験ができること、成績判定の精度を高め、受験生の問題処理時間等の紙の試験にない新しい項目データを収集できること、更にPMPのようにペーパーテストでは検定できない能力の判定ができる可能性を示した。

分担研究者 高林克日己

千葉大学医学部附属病院 医療情報部
教授

吉岡俊正

東京女子医科大学 医学教育学 教授

福井次矢

聖路加国際病院 院長

江口光興

獨協医科大学 小児科学 教授

杉山幸比古

自治医科大学 呼吸器内科学 教授

山内俊雄

埼玉医科大学 精神医学 教授

福島 統

東京慈恵会医科大学 解剖学 教授

原田研介

日本大学医学部 小児科学 教授

研究協力者 神津忠彦

東京女子医科大学 前教授

田村光司

株式会社マイカストレージ（前東京女子医大 循環器内科 助手）

池田 央

立教大学 前教授

植村 研一

弘慈会加藤病院

宮木 浩行

三菱電機インフォメーションシステムズ株式会社

A. 研究目的

わが国の医師国家試験の電子化は、欧米の実情からみてもう少し検討される必要があるだろう。そこで実施する場合の利点欠点を明らかにし、解決法を示して具体的な方向を決定するための問題作成及び試験実施の技術的問題を解決し、同時にコンピュータのMCQ方式と従来のマークシート試験の間に差があるかどうかの検定と、PMP方式の問題と従来のマークシートによるMCQ問題の間に相関があるかどうかを検定すること、および学生の対応、運用についての検討を行うことを目的とする。

B. 研究方法

1 問題作成

平成14年度15年度に作成したCBTに準じたコンピュータでのmultiple choice question(MCQ)のツールを用いて、今回研究協力者から提出された各大学独自の5、6年生の内科系、外科系問題を収集し、その中から同一の分野でレベルの近似した対の問題になるような100題を抽出して、各A、B問題50題とし、それぞれにコンピュータ用問題とペーパー試験用問題とを作成した。PMPは昨年までに作成したPMPのツールを利用して、コンピュータ用に5題を高林が作成し、このうち4題については各PMP問題に対応するようにそれぞれに5問のMCQ問題をペーパー試験用に作成した。

2 試験方法

今回は6大学において主に5学年の学生(一部6学年)を対象に試験を行った。試験はコンピュータによるMCQ試験50題(75分)、PMP試験3題(45分)と、ペーパー試験のMCQ60題(75分)である。各施設ではグループを対の2群(グループI、グループII)に分け、グループIはコンピュータでのMCQのA問題の試験、PMPでは3問を受験し、ペーパー試験ではMCQのB問題、およびPMPで受けなかった2問に対応するペーパー試験10題を受験した。グループIIはこの逆にMCQのB問題の試験、PMPではグループIとは別の2問と共通の1問を受け、ペーパー試験ではA問題、およびPMPで受けなかった2問に対応するペーパー試験10題を受験した。この間、A、Bの受験生は別室でコンピュータ、あるいはペーパー試験を受けるか、あるいは同時に試験を行い、両グループの受験生が試験内容を相互に話し合える時間はないように配慮した。

3 総受験者数

慈恵医大、日大、東京女子医大、埼玉医大、自治医大、独協医大の5年生(2大学では一部6年生を

含む)を対象に、男性264名、女性216名、計580名を同一施設、同一学年の学生を2群に分けて試験を行った。

(倫理面への配慮)

大学名・学生の氏名は用いず、コード番号で表示した。

C. 研究結果

1 MCQにおける検討

MCQ試験における得点を表1および図1にまとめた。平均点では6年生に行ったC大学とF大学の6年生が60点台であったのに対し、もっとも低い大学は30点台と大学群間で乖離した数字となった。コンピュータ試験(COM)とペーパーテスト(MAR)の結果において、2群間には強い相関がみられ、平均点でCOMの方が0.24低い結果が得られた。男性でも女性でもともに同様の結果であり、各大学でみてもCOMの平均点の方が低かったが、6年生に行ったF大学の結果では、COMの平均点の方が高かった(ここでは後に述べる4問の問題を除外した数値で示してある)。

(表1) MCQ問題のコンピュータ試験とペーパーテストの結果

コンピュータ試験 (COM)		ペーパー試験 (MAR)	
全体		全体	
被験者数	504	被験者数	498
平均点	50.88	平均点	51.12
標準偏差	15.01	標準偏差	14.50
最高点	92	最高点	90
最低点	17	最低点	18
A大			
被験者数	23	被験者数	23
平均点	44.20	平均点	44.75
標準偏差	9.01	標準偏差	7.86
最高点	60	最高点	62
最低点	27	最低点	30
B大			
被験者数	112	被験者数	112
平均点	42.06	平均点	42.52
標準偏差	11.58	標準偏差	11.48
最高点	69	最高点	78
最低点	17	最低点	22
C大			
被験者数	104	被験者数	104
平均点	61.78	平均点	62.23

標準偏差	12.42	標準偏差	12.45
最高点	90	最高点	90
最低点	21	最低点	24
D大			
被験者数	89	被験者数	89
平均点	54.99	平均点	54.30
標準偏差	11.19	標準偏差	9.64
最高点	79	最高点	76
最低点	29	最低点	32
E大			
被験者数	92	被験者数	90
平均点	37.86	平均点	39.58
標準偏差	8.85	標準偏差	8.84
最高点	67	最高点	64
最低点	19	最低点	18
F大（5年）			
被験者数	26	被験者数	26
平均点	44.87	平均点	45.58
標準偏差	8.96	標準偏差	10.37
最高点	63	最高点	70
最低点	25	最低点	30
F大（6年）			
被験者数	59	被験者数	58
平均点	66.74	平均点	65.90
標準偏差	11.31	標準偏差	10.87
最高点	92	最高点	88
最低点	40	最低点	36

図1をみるとコンピュータ試験（COM）に比してペーパー試験（MAR）でとくに得点の低い群と高い群が少数であるが存在する。 $r = 0.70$ でほぼよい相関といえる。男女別にみると図でみると男性の方がより高得点に見えるが、これは5年生に女子学生だけの大学が含まれていたためであると考えられる。

次に得点分布を偏差値で比較すると、図4に示すように、A問題、B問題ともにほとんど二つのグラフは重なる（CBT：コンピュータ試験、MAR：ペーパー試験）。

個々の問題についての得点差について検討したのが図5である。ほとんどの問題でコンピュータとペーパーテストの間に正答率に差はないといえる。

このうちA8、A50、B27、B44は紙とコンピュータ問題で問題・解答内容に違いがあり除外することにした。A7は選択問題X2であった。したがって他のデータ、図表ではこれらを除外した結果を示してある。B10のみは両者の正答率に大きな

差が出ているが、その意味を説明できなかった。4問を除外した結果、平均点±標準偏差はコンピュータテスト50.88±15.01に対してペーパーテストは51.12±14.50となり、平均点ではほぼ同じになった。

図6は個々の得点を2群に分けて比較したものである。平均点、最高点、最低点などはほぼ同じであり、両群間に有意差はない。

2. PMP試験（PMP）と従来の紙試験（MAR）との比較

PMPの問題5題のうち3題を施行しているので、全員の3題の平均得点と、全員のそれに対応するMCQ10問の正解率を示す（表2）。

1) MP試験と対応するマークシート試験との比較

3題のPMP問題の得点とMCQの得点との間には強い相関はなかった（相関係数 $r = 0.439$ ）。男女比で見ると女性にやや相関がみられた（図7）。これに対してMCQの得点とPMPに対応するペーパーテストの10問の得点で比較するとこれでは弱い正の相関が見られる。いわゆるペーパーテストでは全体のMCQと同様に知識を問う試験問題になっていることから、PMPと異なった相関を示すものと考えられる（図8）。またPMPのコンピュータ問題とペーパー試験の直接の対比をみると、この散布図（図9）に示すように、両者間には相関がなかった。

2) PMP問題での検討

問題ごとにみていくとそれぞれ別の傾向がある。アニサキスでは回答項目が3項目で得点もその3つに分かれる。特に中間層は少なく、全然できないか、全て答えたか、どちらかがほとんどになる。一方喘息は回答項目（治療法）が多種であり、得点分布もこれによって多数に分かれている。肺梗塞も治療法が多くあるために得点分布も広がるが、途中で大きなギャップがある。イレウスも能力のあるものとなないものの差が明確になっており、かなり能力のあるグループを入れると三群に分かれる（図10、11、12、13）。

3. 運用上の問題

途中で作動しないことが埼玉医大で発生した。またインストールにおいて、LAN上のクライアントに配信する方法では、各施設の設定に影響されて多少の問題が起こった。画像の情報がとくに階段教室などでは周囲の学生に見えてしまい、これがヒントになることが考えられた（図14）。

D. 考察

平成14年度に調査した米国NBMEでは、収録された問題も10,000題を超えており、CATを恒常的に実施できる態勢になっており、PMP問題も100題以上用意されており、常勤者（医学専門家と心理測定士など問題作成の専門家）が数百名いて常に出題問題を評価し改善している。カナダもこれに近い状況である。

平成15年度に調査した欧州各国では、医師国家資格の試験では、コンピュータによる試験よりも臨床技能観察試験が重視される傾向にあり、すぐにコンピュータを用いる試験に移行する状況ではない。しかし、他方でe-learningもすすんでおりVirtual Medical Schoolも18ヶ国の協力で構築されており、その場合、評価もコンピュータを用いて行われることになる。

問題作成に関しては平成14年度から報告している通り、本研究班でもPMPに関する高林教授の研究（三菱電機インフォメーションシステムズ株式会社 宮木氏共用試験）に関連して、慈恵医大福島教授、東京女子医大 神津前教授、吉岡教授の問題作成と収集や、株式会社メディカルストレージ（東京女子医大前助手）田村氏の問題作成ツールの研究などの成果が出ているが、アメリカのUSMLE review 1、2、3を参考にした。

今年度は各大学の問題を集めて試行した。MCQ問題におけるCOMとMARの間には大きな差は見られなかった。個々の問題について、正答率に特徴的な差はなかった。形式によってコンピュータの差異があるわけではなかった。男性の方がややCBTで得点が高い傾向が伺えた。アンケートでは画像などが見やすいとの意見があったのに対して、メモがとりたいたい、目が疲れるなどの意見が出ており、操作性については大きな問題として上げる意見はなかったが、直前の両者を比較した意見ではペーパー試験の方がよいという意見の方が多かった。

PMPでの得点と全体のMCQの得点と比較することで、PMPの独自性と蓋然性を検討した。正の相関は得られたが、とくにPMPでは低い得点として存在する群が認められ、より明確に能力の低い学生を検出するものか、統合的意思決定能力に欠けるもの、あるいはコンピュータを苦手とするものを出しているものと思われた。一方でMCQの結果とある程度の正の相関が見られることから、全く異なった能力を評価しているわけではないと考える。

PMPに問題と対応するMCQの問題は必ずしも内容が一致しておらず、相互の比較をすることはできないと考えられた。ひとつのPMPの中で治療や診断の解答数が多いほど、得点分布がひろがり、喘息などをみても、かなり均一に広がっている。これに対して有効解答が3つしかないアニサキスはとくに

に極端な得点分布になっている。また、全く回答できない例も見られ、ペーパー試験に比べ、得点の低いものの検出力に優れているともいえる。一方で診断名、治療数など解答数により得点分布が大きく異なることから、得点で評価するためにはある程度の解答要求数があったほうがよいのかもしれない。しかし逆に何をもちて評価するのがあいまいのなることを考えると、必ずしも細かいほうがよいのではなく、アニサキスの問題のように3つくらいのカテゴリーに分けるほうが正確な指標であって、これがPMPの限界なのかもしれない。問診の中で適切なものを選んだかどうかの評価まで加えることでより正しい評価を示すものであるといえるが、従来そこまで分析したシステムはなかったし、問題を作るのが大変であろう。しかしせつかくの総合能力の試験であるとするなら問診、所見、検査、診断、治療の総合点で評価すべきものなのかもしれない。問診なども、たとえば決定的なポイントになるものだけにつけるなどの工夫をすればよいのかもしれない。アンケート結果ではMCQよりも、PMPの方が好ましい結果も得られているが、操作性の問題で評判がある。特に検索方式であるが、これらは改善の余地があると考えられる。

運用上の問題としてインストールの方法のトラブルがあったと思うが、これだけの数を正確に同時に動かすことが求められるとなると、試験のリスクが大きい。末端の故障が起こる確率から100台あたり数台以上の予備機の準備は必要である。

PMPの画像に含まれる問題ではこれが周囲に見える状況になると大きなヒントを与えることになりかねない。このためには隔離した環境、あるいは広いスペースが必要になる。

E. 結論

MCQ問題をコンピュータ試験で行うことに大きな問題はなく、またその得点は紙のテスト結果とほぼ一致した。しかしながらメモ機能などを希望する声があり、また同時に多数の受験者を扱うことの問題は残っている。PMP問題の結果はMCQの結果と強い相関を示さず、PMPでは統合能力、意思決定能力など別の能力を見ていると考えられた。またとくに能力の低いものを明確に判別しえた。

F. 今後

現在ペーパーテストをコンピュータ化することの緊急性はないが、今後多くの試験がコンピュータ化していくことは容易に想像される。メモ機能などいくつかの技術的な改良とともに、大量の問題作成方法、および同時に実施する方式をとるのか、別の日

時で試験を行う方法をとるのかの検討が必要なところ
にきている。これにより準備は大きく異なる。

あるいは医師免許の更新制が開始されるのであれば、
このようなところから開始するという方法もあるの
かもしれない。

G. 研究発表

1. 論文発表

2. 学会発表

高林克日己 細田瑳一 福島統ほか 医師国家試験
へのコンピュータ試験の導入の検討 第37回日本医
学教育学会大会

G. 知的所有権の取得状況

1. 特許取得

なし

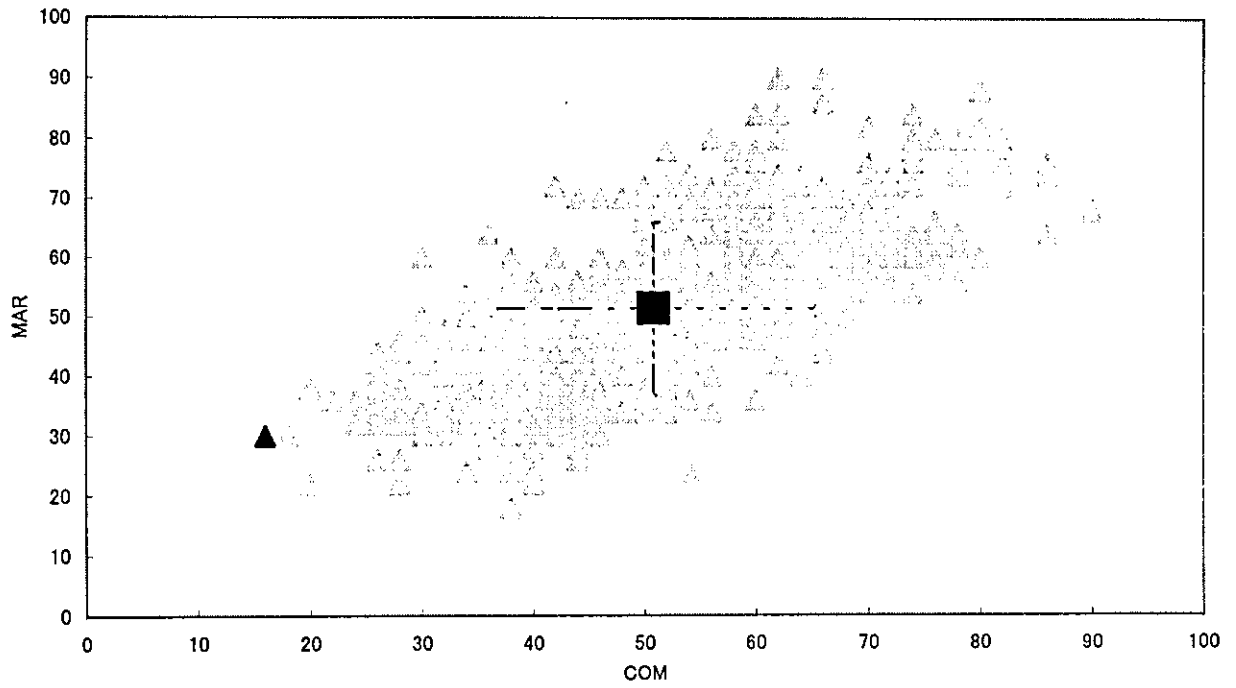
2. 実用新案登録

なし

3. その他

なし

(図1) MCQにおけるCOMとMARの得点比
R=0.70



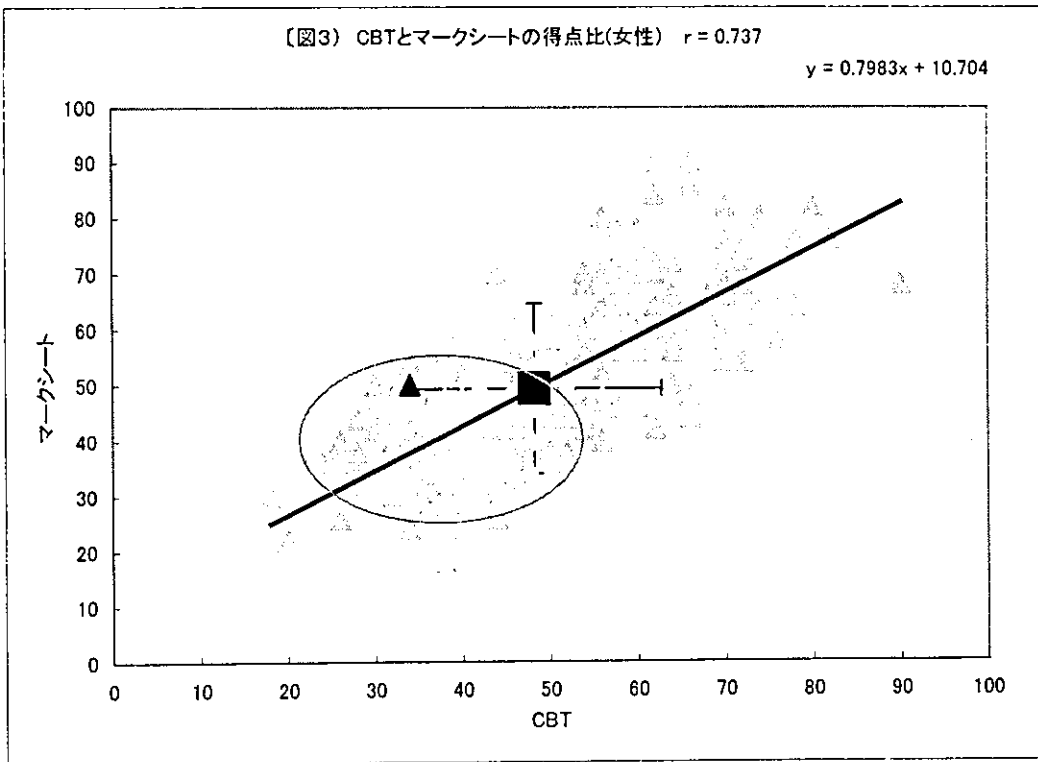
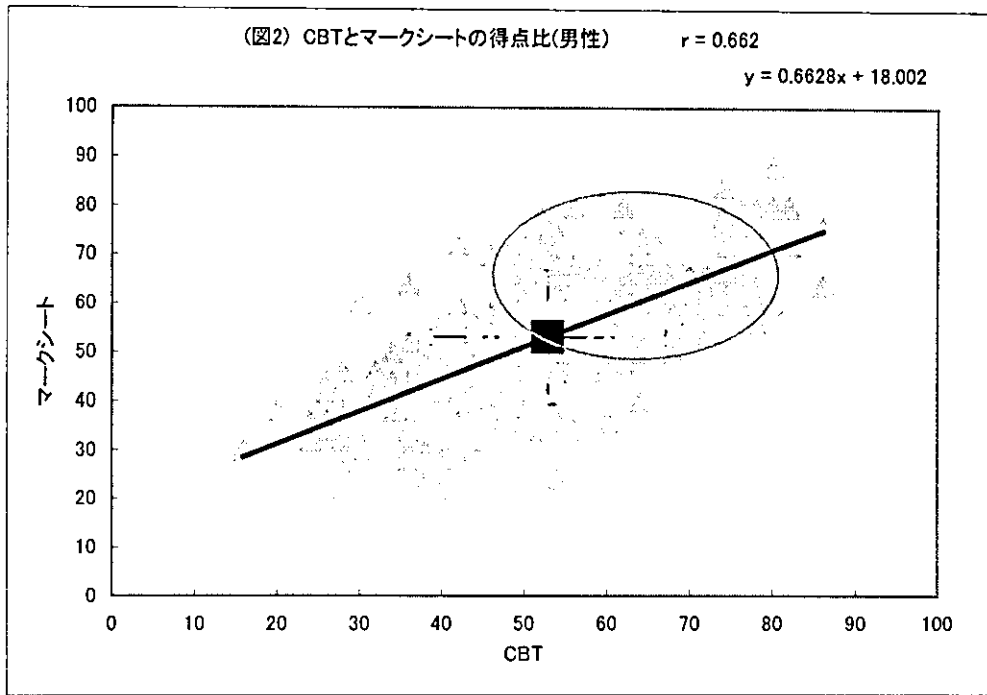
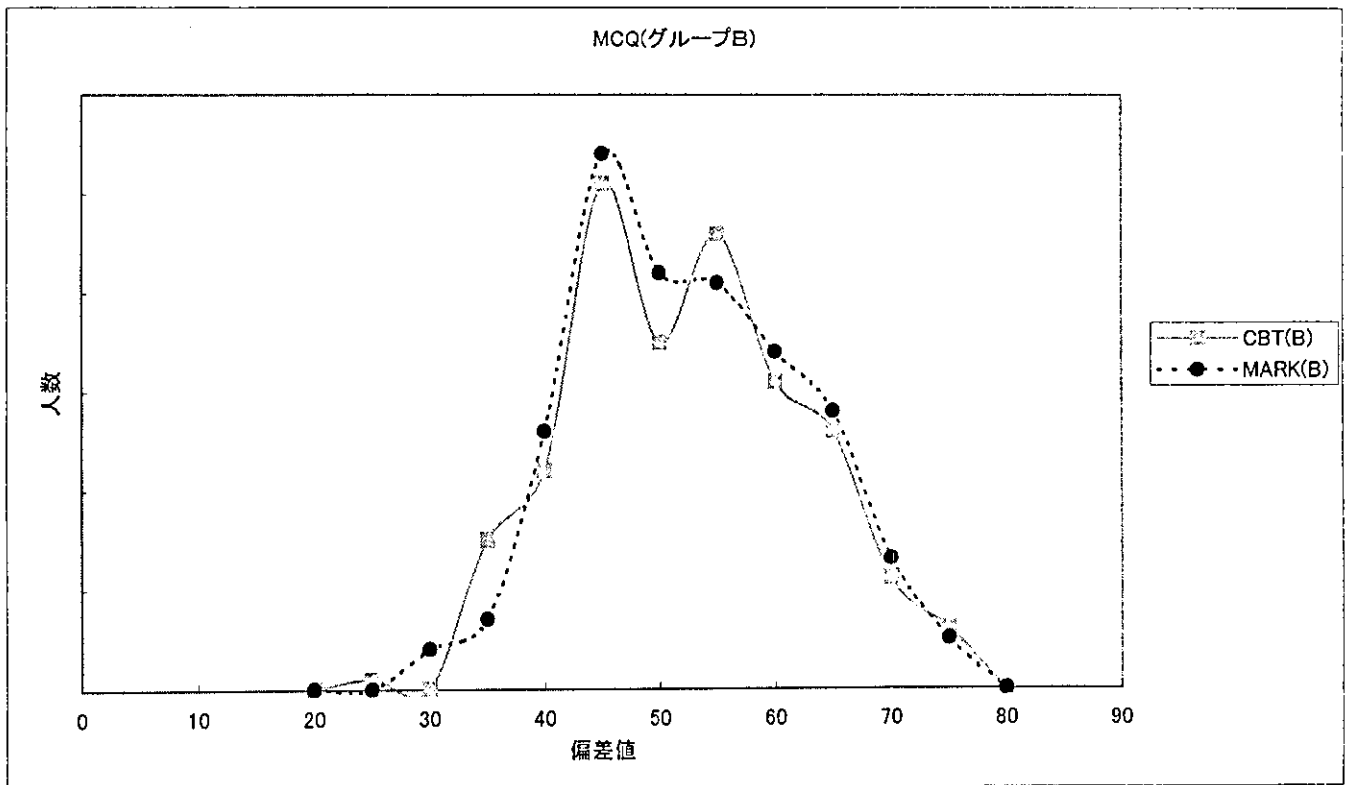
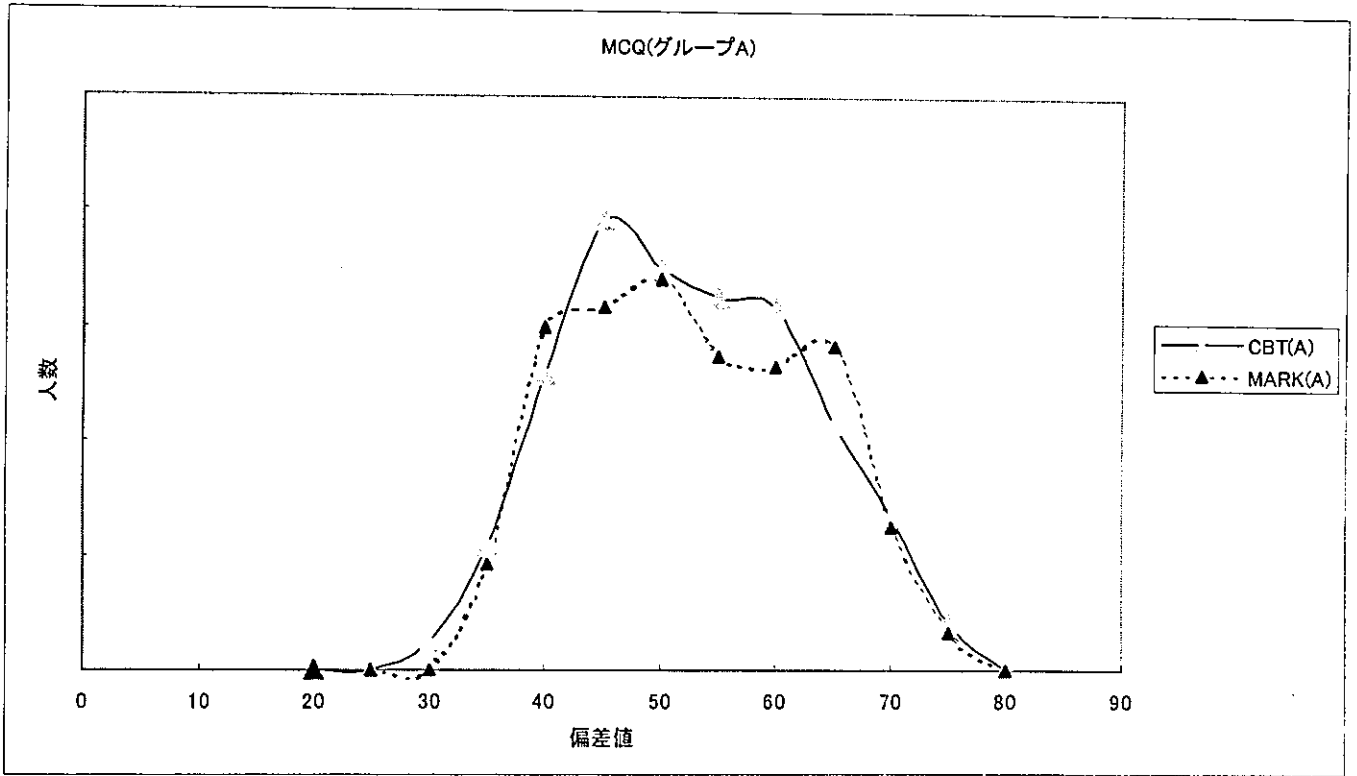
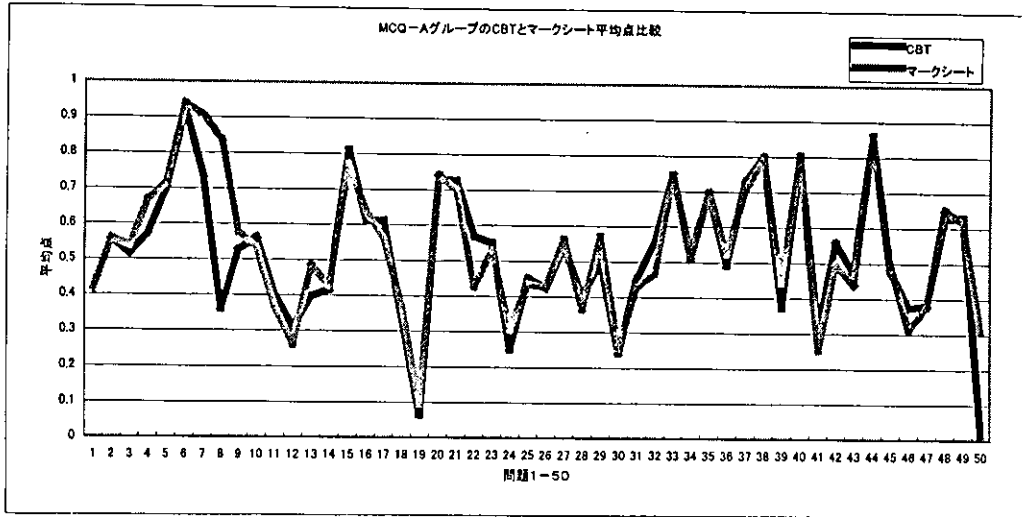
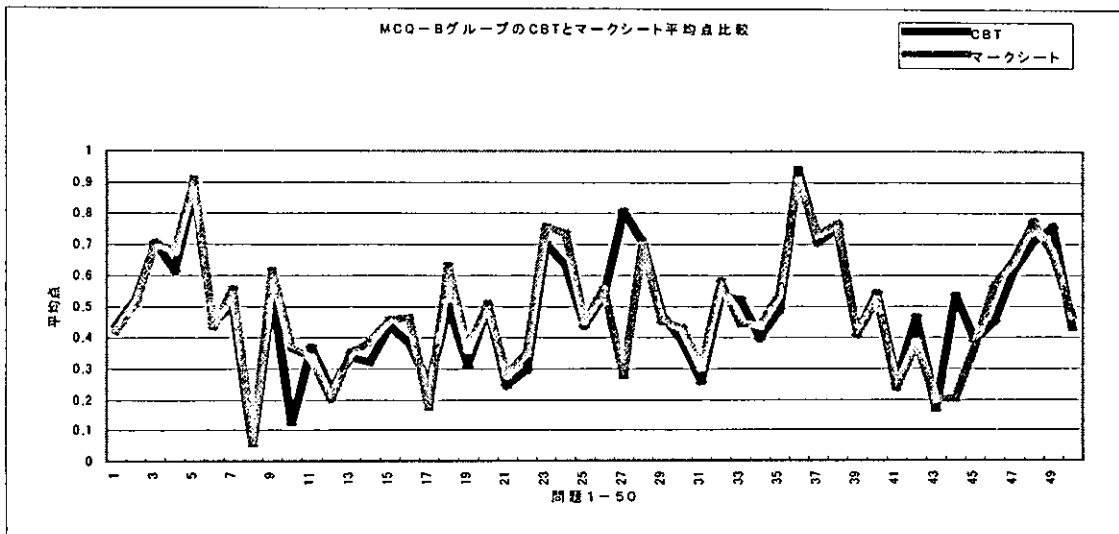


図4 COM (CBT) とMARにおける各問題の正答率

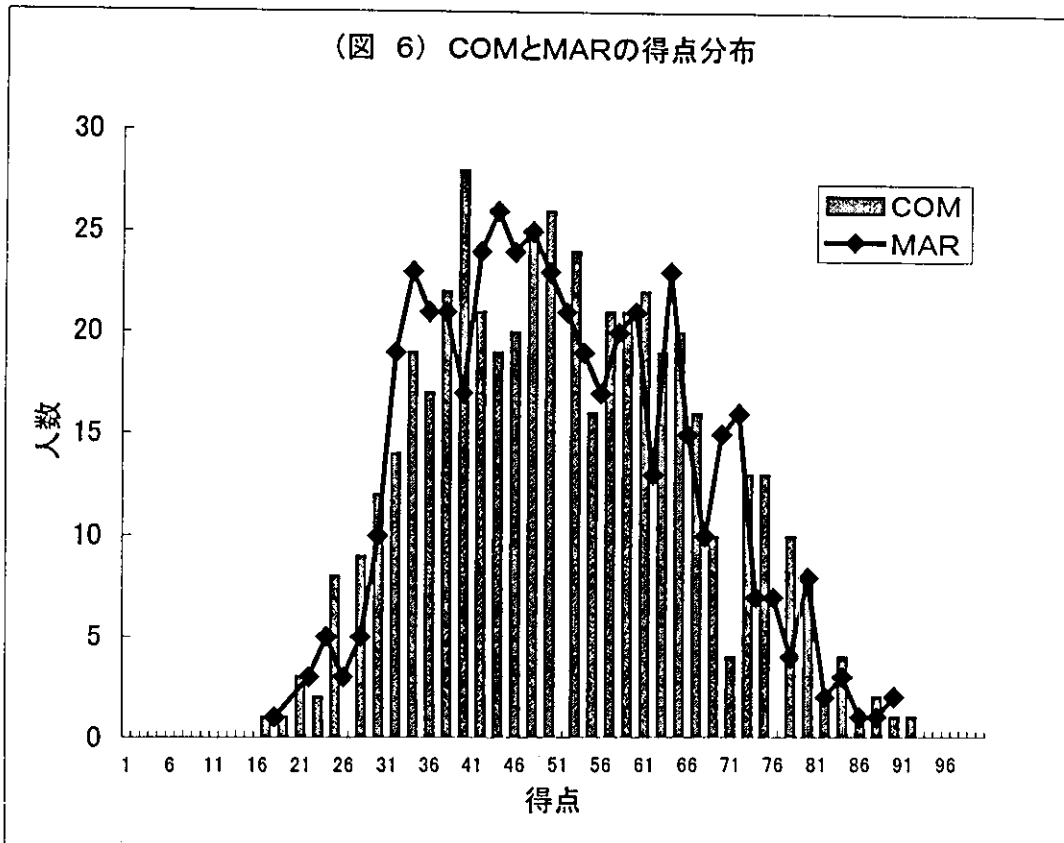




(図 5 A)



(図 5 B)



(表 2) PMP問題の実施結果

C B T

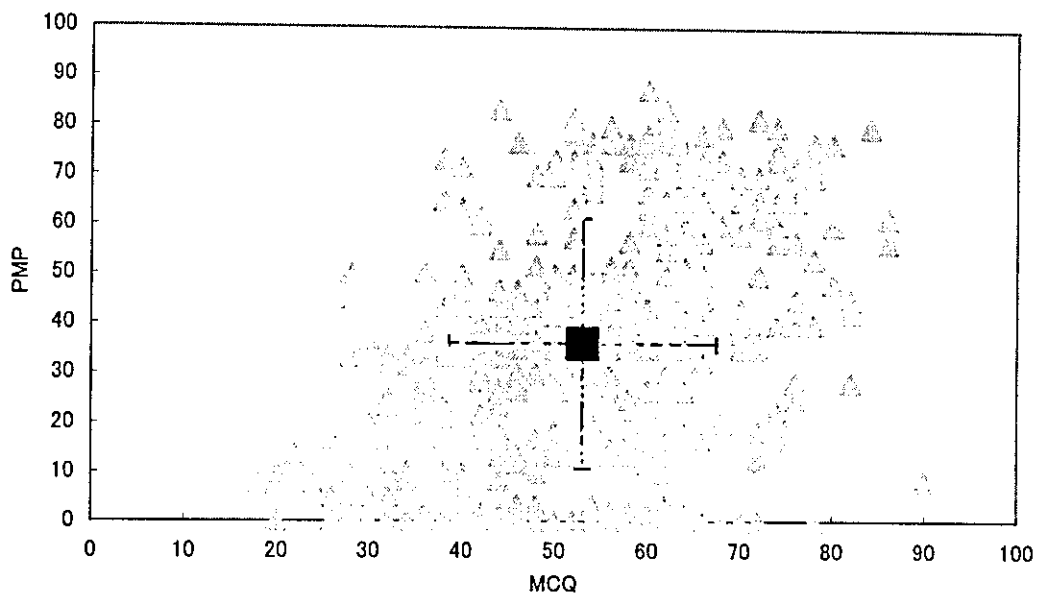
全体	アニサキス	喘息	肺梗塞	イレウス	過換気
被験者数 505	被験者数 244	被験者数 255	被験者数 240	被験者数 231	被験者数 48
	平均点 61.27	平均点 33.82	平均点 29.40	平均点 29.87	平均点 31.
平均点	平均点 45.43	平均点 27.49	平均点 20.74	平均点 31.41	平均点 37.
標準偏差	標準偏差 100	標準偏差 95	標準偏差 85	標準偏差 100	標準偏差 10
最高点	最低点 0	最低点 0	最低点 0	最低点 0	最低点 0

MAR

全体	アニサキス	喘息	肺梗塞	イレウス
被験者数	被験者数 254	被験者数 244	被験者数 254	被験者数 244
平均点	平均点 60.39	平均点 51.39	平均点 74.96	平均点 54.10
標準偏差	標準偏差 19.76	標準偏差 22.14	標準偏差 22.42	標準偏差 20.88
最高点	最高点 100	最高点 100	最高点 100	最高点 100
最低点	最低点 0	最低点 0	最低点 0	最低点 0

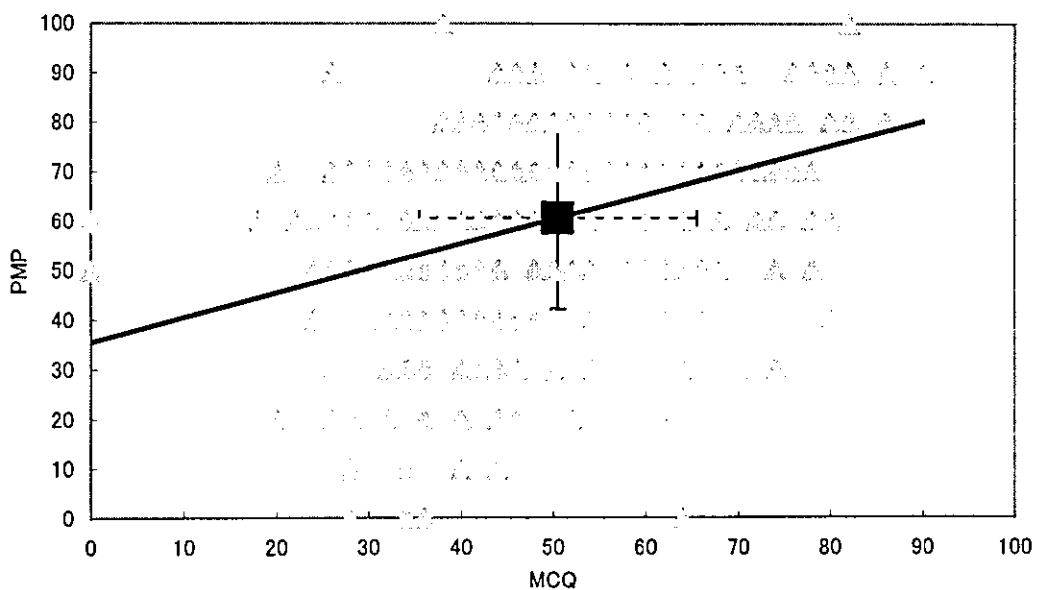
(図 7) コンピュータによるMCQ得点とPMP得点の分布

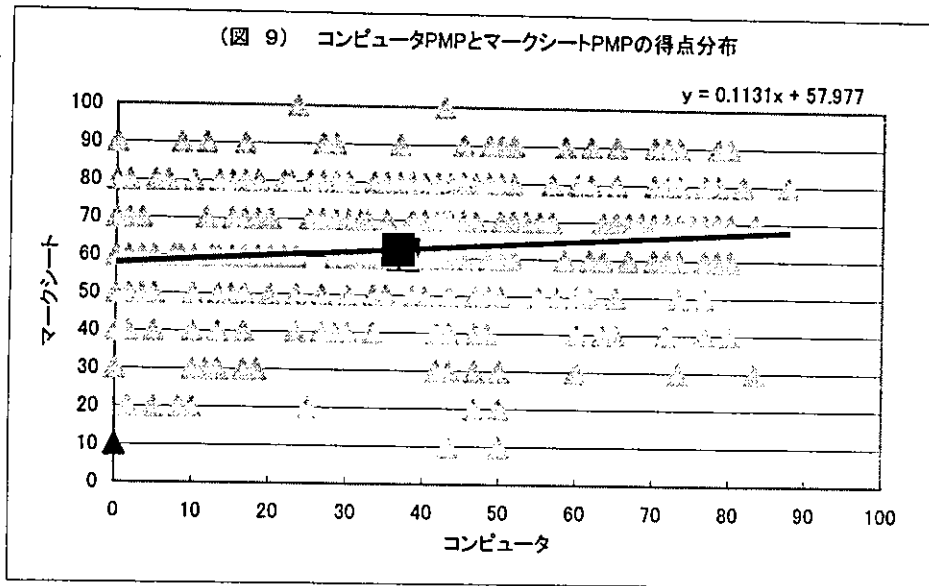
R=0.438845



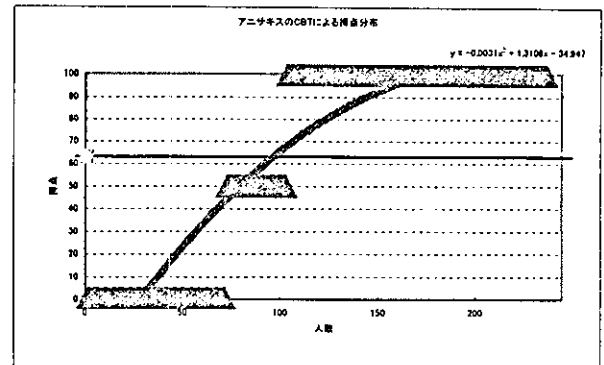
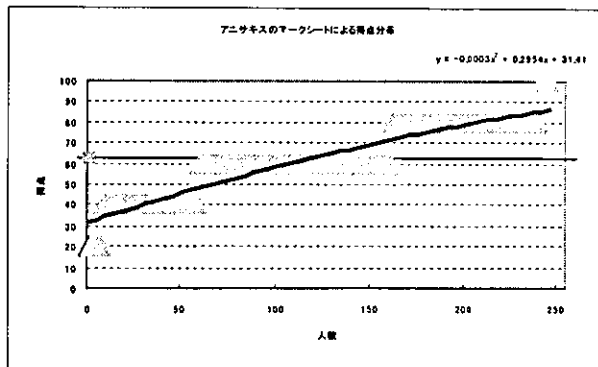
(図 8) コンピュータによるMCQ得点とマークシートPMP得点の分布

$y = 0.4949x + 35.49$

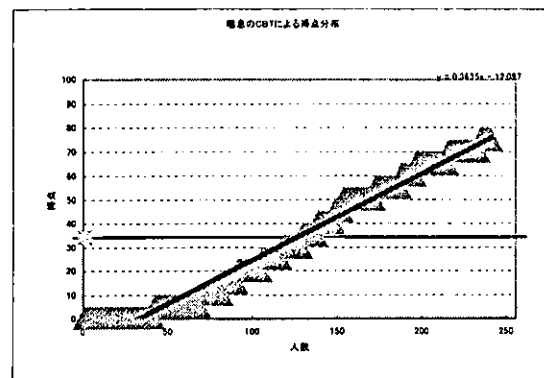
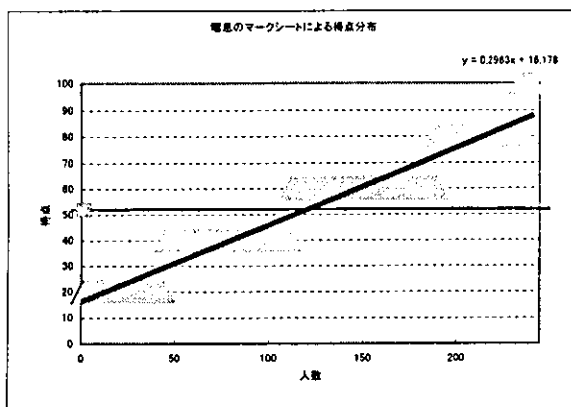




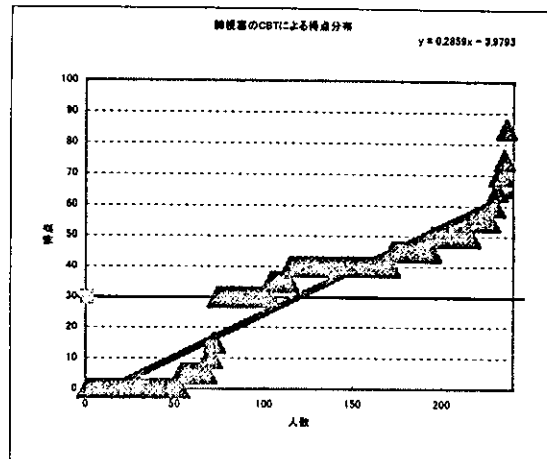
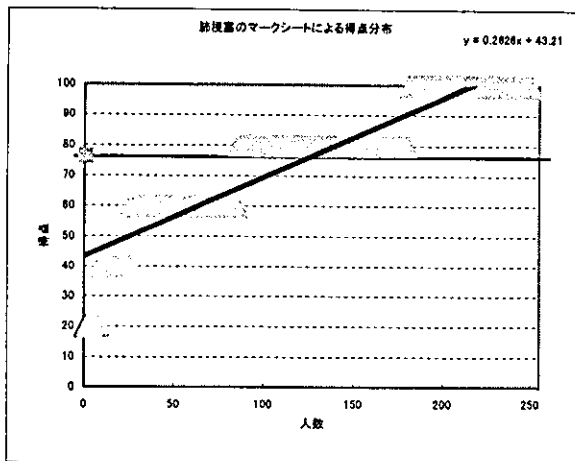
(図 10) アニサキスの問題の得点比較 (左がマークシート、右がPMP)



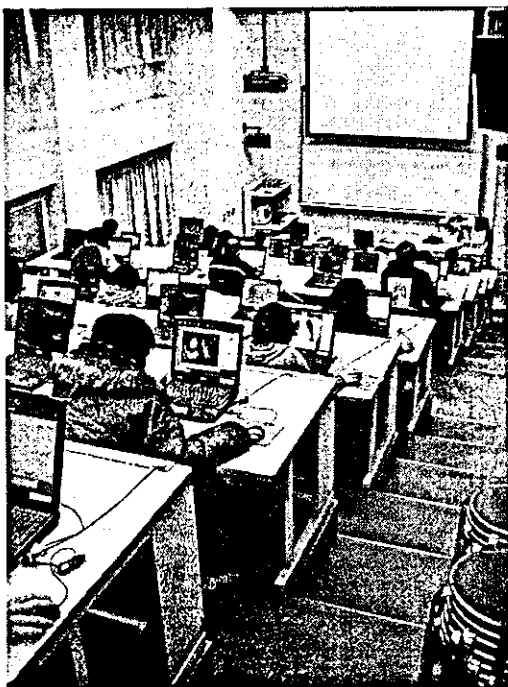
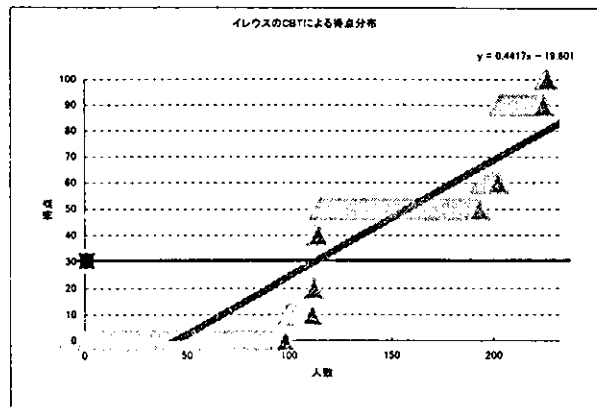
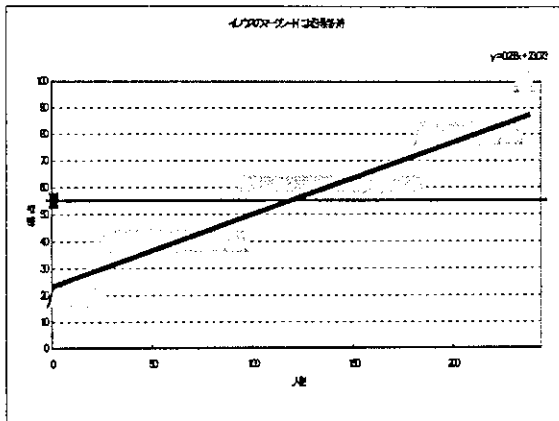
(図 11) 気管支喘息の問題の得点比較 (左がマークシート、右がPMP)



(図 1 2) 肺梗塞の問題の得点比較 (左がマークシート、右がPMP)



(図 1 3) イレウスの問題の得点比較 (左がマークシート、右がPMP)



(図 1 4) 階段教室の試験風景

試験問題のコンピュータ化と諸外国の実情視察

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

研究要旨

MCQ 問題のコンピュータ化、及び PMP 問題の作成を行い、試験の運用に供した。

A. 研究目的

CAT (Computer Adapted Test) および通常の MCQ 試験のコンピュータ化、及び PMP (患者管理問題) の形式で実際の試験に耐えうる問題をコンピュータ上で作成する。

医師国家試験にコンピュータを利用した試験を導入する妥当性を検討する。

B. 研究方法

1) CAT の実施

平成14年度は13年度に入力した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、アンケート調査を含めてその問題点を検討した。試験問題は800問のうち100題が前問までの正答率によってランダムに選ばれる方式である。

2) PMP 作成

平成15年度は各分野専門家9名にPMPによる試験問題の作成を依頼し、その作成にどの程度の能力、時間が必要かを検討した。以前作成したシミュレーションのツールを改変し、Windowsの各バージョンに対応できるようにし、また入力方法を容易にした。マックOSには対応していない。項目として国試出題基準項目を問診、所見、検査、治療、診断の基準項目に用いた。この上で中堅内科医をはじめ9名に入力を依頼し、問題

を作成するとともに、その作成の難易度などを検討した。

2) MCQ、PMPの本格的運用

平成16年度は福島統分担研究者が作成した50問の試験問題2セット、計100問をコンピュータ化して、画像などとセットにし、Windowsマシンにインストール可能とした。同様に一昨年作成したツールを介してPMP問題を作成し、これらMCQ、PMPとアンケートをまとめてインストールできるように設計した。試験中には受験者の解答のほか、試験時間、とくにPMPでは各セクションに関する解答時間を収集し、これらをまとめて、ハードディスク上に一意の名前のファイル作成するようにした。受験者にはデータベースや画像は受験中に侵入できたり閲覧できないように設計した。実際のファイルの展開を示す。図1はMCQの問題の1例である。画像をクリックすることで大きくより鮮明な画像を表示できる。どの問題を終了したかは一覧でみることもできた。どれでも番号を選ぶことでその問題に移ることができる。1選択、でも複数選択でも入力が可能である。また試験終了時にはすぐその場で評価点が表示される。

PMP問題ははじめのオープニングシーンの情報〔図3〕以降は全て受験者が選択するもので、その結果は「カルテ」に貯まっていく〔図4〕。必要に応じて画像が展開することもある。最終的には診断と治療が行なわれ、これらの点数で評価が行なわれる。終了時には選択結果と各セクションの所要時間が表示される〔図5〕。こ

れらを5問作成した。アンケートもこれと同様にコンピュータから行なった

②受験者のアンケートをとって情報を収集する。学生アンケートは全員の学生に試験終了時にコンピュータへの入力として行い、複数回答を許した。

C. 結果

1) ①CATの施行とアンケート

CATは学生の希望の時間で、2時間以内で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（表1・2、図6・7）

2) PMP問題の試験的作成

作成された問題のほとんどが内科系の各分野である。問題の内容については以下のようなものである。

問題1	総胆管結石と胆管炎
問題2	C型肝炎
問題3	気管支喘息
問題4	ウエルニッケ脳症
問題5	糖尿病ケトアシドーシス
問題6	腸重積
問題7	急性虫垂炎
問題8	肥大型心筋症
問題9	前立腺癌
問題10	延髄外側症候群
問題11	抗リン脂質抗体症候群

PMP問題作成上の問題点

i) 問題項目の欠落

国試基準から集めた問題項目では、特に問診項目が欠如していて、追加項目を加えることが必要であった。このような項目の多く（「胸痛の性状」、「しめ鯖を食べた」など）は国試の出題基準にはないので、作成者が判断して追加しなければならない。また国試出題基準ではプライマリケア的な発想が少ないため、common diseaseを考えるにあたり、別に考慮しなければならない。おそらくこれらの問診に関する用語、項目はPMPを作成する限り、今後も多くの追加が必要になると思われる。所見に関する項目につ

いても難解な用語が含まれている一方で、基本的な用語の欠落が多数見ついている。

検査項目においては国試レベルの設問としては用意ができていたと思われる。また治療においても項目はほぼ十分であるが、それぞれの問題作成にはやはり固有の設問項目を追加する必要があった。

ii) 問題作成の難易さ

シミュレーションの意味と問題作成の説明のため、それぞれの依頼医師に約1時間の説明時間を要した。概要の理解はおおむね容易であった。

iii) 問題作成時間

一問の作成の所要時間として、短い作成者で30分、長い作成者で3時間、おおむねが1～2時間が必要との回答であった。この作成時間だけを見る限りではそれほどの作業時間を要するようには考えられなかったが、ここではその後の検討、再修正の時間は入っていない。

iv) 問題完成度のチェック

問題のチェックは通常のMCQとは異なり、多量の時間を要する。個々の項目の値の妥当性を検討する他に、複数項目選択によって起こってくる矛盾などを考慮すると、チェックすべき選択の組み合わせ数は爆発するので、このような検討には限界がある。国家試験などの資格試験に耐える問題とするには、多くの試験官に前もって入念にチェックしてもらう必要がある。また採点法においてもどこにポイントを置くのかは難しかった。

3) MCQ, PMPの実際の運用

この内容に関しては総括報告を参照されたい。

D. 考察

1) CAT

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。しかしこ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。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ると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。CATは米国でも看護師試験には導入されたが、医師には適用されて

いないし、学生の慣れを考えると、まずコンピュータ化した上で段階的に進めるべきかもしれない。

2) PMPシミュレーション問題

PMPの作成には多くのコツがあり、かなり習熟しないと良問ができないのと、以下のようない点に注意しないと、折角の問題が崩れてしまうことになりかねない。米国でみても一問作成するのに、多数の人間が関与して作っており、本格的にこれを施行するには、多数の教育者の理解がまず必要であると考えられる。今回の結果をみても、多くの専門医は問題の作成ができないわけではないので、まずはこれを広める教育者の養成が必要である。

i) 項目の値のありかた

その問題のある項目の特定値から判断して(たとえば患者の年齢や職業などから)、その問題の解を安易に見つけ出してしまうこともありうる。したがってこれらの特定値は適当に乱数化して値を示すほうがよいかもしれないという意見があった。これは一般の検査値においても同じで、問題数が少ないと簡単な項目の解を求めることで問題全体が解けてしまう可能性がある。このため特定値も一般値もいくつかの解を持つか、ある範囲内で乱数化することができることが望ましい。

ii) 選択数の制限など

現在のシステムでは同一セクションで選択が複数でき、制限がない。すなわち禁忌肢を選択しない限り、多くの答えを見て回答を選んだ方が得策ということになる。このために問題選択数の制限を設定するのも一法ではあるが、そうするとこのことが一般の臨床での診療論理思考とは異なった方法を要求することになりかねない。

iii) 項目の選択方法

このPMPでは前もって準備されている選択項目はないが、ある項目を選択するには検索機能で探すことになる。このため白血球数を選択するために、白でも白血でも検索できるが、こうなると他の準備されている用語も同時に表示される。極端に言えば一字で選択する方法をとれば多数の選択肢を得ることができ、この中から適当な用語を選択できるという要領で答えを探すことができってしまう。一方でその単語に完全一致を要求すると、作成時に同義語など定義が複雑になりかねない。これらの矛盾をどのように解決すべきかも今後の課題である。

iv) 客観的評価法

このシミュレーションの問題の評価を何によって行うかも重要な要素である。米国NBMEの場合は時間経過、コストまで計算に入れている。しかしファクターが多くなるほど評価は複雑で難

しくなる。さらに現在の6年制の医学部を出た段階でそこまで要求できるであろうか。今回の評価法は適切な項目が選ばれているか、最終的な治療と診断が正しいか否かで判断するものになっている。しかしそうすると正解、不正解での落差が大きくなる。この部分をどのように判断するかは以前からのPMPの課題になっている。項目数の制限をつけずに評価するとなると、全ての項目に点数をつけることになるが、これは煩雑であるだけでなく、複雑な選択に対して正しく評価できなくなったり、矛盾が生じる可能性がある。あえて単純な評価方法にした方がよいというのが現時点において多い意見であった。

v) 作成された問題の評価

上記にも示したように、問題の蓋然性、完成度のチェックは多大な時間を要し、特に国家試験に用いるとなると、その検定には相当量の時間が必要になるであろう。一方で試験問題数としてかなりの量を作成しないと、どの問題かわかってしまう。このために費やす時間と労力はMCQなどの比ではない。一方で一度完成された問題ができれば、その後は少しの改変で長期間利用できる可能性がある。

vi) Sequential タイプと枝分かれタイプ

これは以前からいわれてきたことであるが、PMPにおける意思決定によるその後の展開が枝分かれするのは学習においては有用であるものの、評価採点をするとなると採点方式が複雑になり、矛盾の起こらないように作成するには困難がともなう。今回の方式であると全てsequentialになっているが、当面はこの方式による検討を進めるべきと考える。

3) MCQ, PMPの実際の運用

この考察については総括報告を参照されたい

E. 結論

- 1) 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段階的に行なわれるべきであろう。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。
- 2) PMPの作成については問題点の指摘は受けたが、作成に関しての困難さは問題作成の依頼ができないレベルでないことは確認された。問題の蓋然性の検討が容易ではないのと、どのように試験で評価するかの最終決定はできていないが、シミュレーションとしての試験への応用は可能であり、受験者側の問題点を解析することで最終的な利用段階も考えられる。解

答結果の回収やデータベース構築などはMCQ、CATと同様で少数であれば難しいことではない。

- 3) MCQとPMPについて6大学での実施での問題作成と運用は、ここまでの懸念以上のものではなかったが、アンケート結果は早急な改変を望まないものが多かった。

F. 研究発表

1. 論文発表

なし

2. 学会発表 (予定)

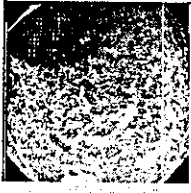
- 1 高林克日己, 鈴木隆弘, 池田央, 福井次矢, 植村研一, 細田瑳一 (2003) 医師国家試験へのComputer Adapted Test導入の試み. 第35回医学教育学会. (医学教育34, Suppl 54, 2003)

G. 知的財産権の出願・登録状況

なし

Multi Choice Question

22歳の男性。1か月前から血性下痢を認めていたが、昨日から腹部全体の痛みと38.5℃の発熱が出現し、鮮血の混じる下痢が1日20回以上となったため来院した。緊急下部内視鏡検査所見を別に示す。この疾患について正しいのはどれか。



設問1 設問2 設問3 設問4 設問5
 設問6 設問7 設問8 設問9 設問10
 設問11 設問12 設問13 設問14 設問15
 設問16 設問17 設問18 設問19 設問20
 設問21 設問22 設問23 設問24 設問25
 設問26 設問27 設問28 設問29 設問30
 設問31 設問32 設問33 設問34 設問35
 設問36 設問37 設問38 設問39 設問40
 設問41 設問42 設問43 設問44 設問45
 設問46 設問47 設問48 設問49 設問50

1: 炎症は大腸壁全層に及ぶ。
 2: 合併症として穿孔が多い。
 3: 10年以内に癌化するものが多い。
 4: サラソスルファピリジンが有効である。
 5: Clostridium difficileの菌毒素が関与する。

1/50 C 1 C 2 C 3 C 4 C 5 答える 前へ 次へ 提出

図1 MCQの画面

結果

試験結果

0.0 %

了解

図2 試験終了時の表示