

医師国家試験等改善に関する研究

主任研究者	細田	瑛一（日本心臓血圧研究振興会）
分担研究者	畑尾	正彦（日本赤十字武蔵野短期大学）
	齋藤	寿一（自治医科大学）

厚生科学研究費補助金（医療技術評価総合研究事業）
総括研究報告

医師国家試験等改善に関する研究

主任研究者 細田 瑛一 日本心臓血圧研究振興会附属榊原記念病院院長

研究要旨 skills analysis 形式の利点を明らかにしこの形式を用いた問題を作成して研究協力者の施設で、模擬試験を試行し国家試験既出問題と比較検討した結果、学生はこの形式に馴れていないので、正解率が低い傾向があるものの、適正な問題を作成すれば従来の国家試験問題で評価し難い領域について評価できる可能性が認められた。

A.研究目的

Skills analysis 法が医師国家試験出題形式の一つとして利用できるかどうかを検討し適切な例題を作成する。

B.研究方法

昨年度まとめた skills analysis 法の利点を考慮して例題を作成し、既出の国家試験問題と同数出題し、5 学年生の学生に試行して結果を分析する。試験当日、欠席者があった為 6 大学医学部の 5 学年生合計 474 名に対して問題 A、問題 B を試行した。各問、各大学毎に正解率と識別指数を算出し、A-B 問題間の相関、各出題領域毎の集計を行った。

C.研究結果

474 名分の総計では、A 問題（国試既出問題）の正解率の平均は 54.8%（最高 58/60・最低 5/60）であり、正解率の平均は各大学毎に最高 77.6%、最低 28.1%とかなりの差があった。一方 B 問題（skills analysis 形式）の正解率の平均は 51.9%（最高 44/60・最低 16/60）であり、大学毎の平均は最高 54.2%、最低 48.5%であり、最高点が A 問題より低く、最低点が高く、受験生毎、大学毎の差は A 問題に比べて小さかった。両問題の合計では全体の平均正解率が 53.4%（最高は 66%・最低は 38.3%）であった。両問題の領域毎の正解率には相関は認められないが、両問題の評価する項目が異なる可能性を増した。

識別指数については、A 問題は既出問題でその中から良問を選んだこともあり、全て 0.14 以上の信頼性の高い問題であった。B 問

題は、受験生が初めて遭遇した形式であり、識別指数がマイナスとなった問題が 14 題と多く、その中正解率 20%以下の問題が 4 題含まれ、識別指数プラスの問題の中にも正解率 20%以下の問題が 7 題あり、その中 3 問の識別指数は 0.05 以下であった。

skills analysis の利点の一つに理由づけを問うことができる点があげられているが、この理由づけの条件を解答肢に含ませると二つのことを問う形式がとられることになり解答を誤る原因にもなっている。また、複雑な図表の読解を問う問題 5 問では正答率 18~72%で識別指数 0.03~0.06 と低くなっており、reading を問う良問と考えて出題した意図が生かなかった。正答率が適正で識別指数の高い問題の中には、形式的に従来のものと類似しているものが 5 題含まれていた。しかし形式が従来のものであっても、領域が稀な疾患であると正解率が低く識別指数は比較的高い。論理や判断を問う新しい形式の問題では比較的易しい問題でも正解率、識別指数共に低い問題がみられた。真偽形式で出題された 15 題はいずれも正解率が高く、2 題を除いて 70%以上で、その中 7 題が 90%以上の正解率であった。真偽形式（true false）問題の作問出題に当ってはこの点に十分配慮すべきであるが、問題の主題は受験者にもよく理解され、正確な評価法として優れている。識別指数 0.10 以上正解率 30%以上の適正な問題が 14 問あった。

D. 考察

skills analysis 形式の利点を用いた良問を作成することは容易ではない。臨床診療の現場を再現する医療面接からの問題抽出、学術論文や複雑な図表の読解、それらを根拠とした重みづけや診断治療の進め方の判断など設問に用いられるテーマは多いが、この結果からみると適正な問題は 25% 程度であった。設問や選択肢はなるべく簡明にするように心がけ、従来の形式に判断基準などを入れた問題や難易度の高い場合の真偽形式の導入など今後考慮すべき点も明らかとなった。その中で、従来の問題選別の基準で適正と考えられる問題が 14 問認められた点は評価できる。

E. 結論

skills analysis 法の利点を利用して、医師国家試験の例題を作成し、14 題は正答率、識別指数の適正な問題であった。

F. 研究発表

未定であるが、詳細に関して現在報告書を作成中である。

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

客観的臨床能力試験に関する研究

分担研究者 畑尾 正彦 日本赤十字武蔵野短期大学教授

研究要旨 医師の診療実技を適正に評価する試験である客観的臨床能力試験（OSCE）を全国規模で実施するために、実技ステーションで提示する課題と医療面接のシナリオを新たに作成した。OSCEを総括的評価とする場合の成績の表示法および合否判定基準の設定方法として、得点法、「0」評定A方式および「0」評定B方式を策定した。

分担研究者 畑尾 正彦
日本赤十字武蔵野短期大学
教授

A. 研究目的

客観的臨床能力試験（OSCE）を医師国家試験として全国規模で実施するためには、年度によって変化のある課題を提示する必要がある。また総括的評価として、適正な合否判定をしなければならない。より多くのステーション課題を作成し、適正な結果の表示方法と合否判定基準設定方法を策定する目的で、本研究を行った。

B. 研究方法

1. OSCEの実技ステーションで提示する課題の作成

1) 研究協力者の大学で実施されたOSCEで新しく作成・採用された課題を調査した。2) 臨床能力教育ワーキンググループ（日本医学教育学会）に研究協力を依頼し、新たな課題を作成した。

2. 適正な成績の表示方法と合否判定基準設定方法の策定

研究協力者の大学のOSCEで採用している成績表示方法および合格基準設定方法を調査し、適正な方法を策定した。

C. 研究結果

1. 新たな課題の作成

1) 大学の1回のOSCEに採用されている課題数は4～14課題であった。全体では16種類の課題が採用されており、そのうち新しいものは4種類であった。

2) 臨床能力教育ワーキンググループで検査・処置などの実技および「疾患を想定した身体診察法」（従来は基本的診察手技が中心）などの課題が新たに作成され、さらに「医療面接」のステーションで採用する新しい「シナリオ」も作成された。

2. 成績表示方法と合否判定基準

1) 成績表示方法 ①得点による表示法（得点法） 研究協力者のすべての大学での成績表示は、評価表の評定尺度にマークされた各評定項目の評定点を合計した得点で表示するものであった。②「0」評定項目数による表示法（「0」評定法）

評価表の各評価項目のうち、最低評定「0」であった項目の数で成績を表示する方法を採用している大学はなかった。

2) 合否判定基準設定方法 ①得点法における合否判定基準 i) 各課題ごとの合否判定は、研究協力者の大学では60%を分割点とするものがほとんどであった。

ii) OSCE全体としての合否判定は、5課題中3課題、6課題中5課題、あるいは5課題のすべての合格をもってOSCE全体としての合格とするなど、さまざまであった。②「0」評定法における合否判定基準 i) 各課題ごとの合否判定を、「0」と評定された項目の数が、全項目数の40%以下であれば合格とする方式(A方式)を採用している研究協力者の大学はなかった。ii) 各課題の全体としての評定を、評価表の全評定項目数をほぼ均等に分けた4段階(4, 3, 2, 1)の評定尺度で表示した場合に、「0」と評定された数がどの段階に当るかにより、相乗法でOSCE全体としての合否判定を行う方式(B方式)を採用している研究協力者の大学はなかった。

D. 考察

1. OSCEの実技ステーションの課題 研究協力者の大学のOSCEで新たに作成・採用された課題は、「画像診断」「手洗生検」などであった。分担研究者および研究協力者が新たに作成した数種類の課題および「医療面接」のシナリオも6種類が新たに作成された。今後も新たな課題とシナリオが継続的に作成されることが望まれる。

2. 適正な成績の表示方法と合否判定基準設定方法の策定

成績の表示方法：評価表の各項目の評定点を合計して表示する得点法は、受験者に課題の修得度を分かりやすくフィードバックできるので、形成的評価では有用な表示法である。「0」評定法は、上手

さの程度を表わすものではなく、課題の各項目でできなかったものがいくつあるかを示すので、総括的評価に適していると考えられる。

合否判定基準設定方法：OSCEの成績を得点法で表示する場合に合否の分割点を60%とする大学が多いことは、日本の多くの試験が60%を合否の分割点としている現状から、これに合理的根拠はないとしても無理もないと考えられる。

OSCEを構成する実技ステーションの課題のいくつが合格の場合に、そのOSCE全体としての合格と判断するかは今後の検討課題である。

「できない」ものを合格と判定してはならないが、必ずしも「上手にできる」ことを求めない卒前教育や医師国家試験では「できない」項目数で成績を表示する「0」評定法は有用であると考えられる。

ある課題の評価表で「1」以上と評定された項目の数が、全評価項目の60%以上であればその課題の合格と判定する「0」評定A方式は、得点法と同様にOSCE全体としての合否判定をどうするかを検討課題が残る。その点「0」評定法B方式は、OSCEを構成する各課題ごとの評定の数値の積で判定をするので、OSCE全体としての合否判定についての課題は残らない。総括的評価に適した方式の一つと考えられる。

実際のOSCEで、得点法、「0」評定A方式あるいは「0」評定B方式を適用した場合にどのような結果になるかについて、今後検討を重ねる必要がある。

E. 結論

実技ステーションで提示する課題と医療面接のシナリオを新たに作成した。

成績表示法および合否判定基準設定方法として得点法、「0」評定A方式、「0」評定B方式を策定した。

F. 研究発表 未定

6. 研究結果の概要

<p>客観的臨床能力試験(O S C E)の実技ステーションで提示する課題を、研究協力者の大学で新たに作成・採用された課題の他に、分担研究者および研究協力者である臨床能力教育ワーキンググループ(日本医学教育学会)のメンバーが新たに作成した。また実技ステーションの一つである医療面接のシナリオが、臨床能力教育ワーキンググループで新たに作成された。</p>
<p>これらの研究結果はO S C Eを医師国家試験として全国規模で実施する場合に、提示する課題が年度により変化することが求められることに対応している。</p>
<p>国家試験では、適正な合否判定が求められる。O S C Eを総括的評価とする場合の成績の表示法および合否判定基準の設定方法として、得点法、「0」評定A方式および「0」評定B方式を策定した。</p>
<p>第2年度には初年度の研究結果を研究協力者の大学の実際のO S C Eに適用して実用性を確認する。</p>

7. 研究により得られた成果の今後の活用・提供

<p>本研究は客観的臨床能力試験(O S C E)を医師国家試験として全国規模で実施するためには、年度によって変化のある課題を提示する必要性に対応して、より多くのステーション課題を準備することと、総括的評価として、適正な合否判定を行うために、O S C Eの結果の表示方法と合否判定基準設定方法を策定することである。</p>
<p>新たに作成されたステーション課題を、研究協力者の大学の事情が許す範囲で、実際のO S C Eに適用し、課題として適切であるかどうか、修正してより適切な課題にブラッシュアップして、将来の医師国家試験の課題、またはその叩き台として活用することができる。</p>
<p>策定された3種類の成績表示法と合否判定基準設定方法を、研究協力者の大学の実際のO S C Eの結果にシミュレーションで適用し、最も適正な方法を探って、将来の医師国家試験にO S C Eが行われた場合の合否判定方式の叩き台を提供することができる。</p>

8. 研究の実施経過

1. OSCEの実技ステーションで提示する課題の作成
1) 研究協力者の大学で実施されたOSCEで新しく作成・採用された課題を調査した。 大学の1回のOSCEに採用されている課題数は4~14課題であった。全体では16種類の課題が採用されており、そのうち新しいものは4種類であった。
2) 臨床能力教育ワーキンググループ(日本医学教育学会)に研究協力を依頼し、新たな課題を作成した。 臨床能力教育ワーキンググループで検査・処置などの実技および「疾患を想定した身体診察法」(従来は基本的診察手技が中心)などの課題が新たに作成され、さらに「医療面接」のステーションで採用する新しい「シナリオ」も作成された。
2. 適正な成績の表示方法と合否判定基準設定方法の策定
研究協力者の大学のOSCEで採用している成績表示方法および合格基準設定法を調査し、適正な方法を策定した。
1) 成績表示方法
①得点による表示法(得点法) 研究協力者のすべての大学での成績表示は、評価表の評定尺度にマークされた各評定項目の評定点を合計した得点で表示するものであった。
②「0」評定項目数による表示法(「0」評定法) 評価表の各評価項目のうち、最低評定「0」であった項目の数で成績を表示する方法を採用している大学はなかった。
2) 合否判定基準設定方法
①得点法における合否判定基準
i) 各課題ごとの合否判定は、研究協力者の大学では60%を分割点とするものがほとんどであった。
ii) OSCE全体としての合否判定は、5課題中3課題、6課題中5課題、あるいは5課題のすべての合格をもってOSCE全体としての合格とするなど、さまざまであった。
②「0」評定法における合否判定基準
i) A方式:各課題ごとの合否判定を、「0」と評定された項目の数が、全項目数の40%以下であれば合格とする方式を採用している研究協力者の大学はなかった。
ii) B方式:各課題の全体としての評定を、評価表の全評定項目数をほぼ均等に分けた4段階(4,3,2,1)の評定尺度で表示した場合に、「0」と評定された数がどの段階に当るかにより、相乗法でOSCE全体としての合否判定を行う方式(B方式)を採用している研究協力者の大学はなかった。
「0」評定法B方式は、OSCEを構成する各課題ごとの評定の数値の積でOSCE全体としての合否判定をする方式であり、医師国家試験などの総括的評価に適した方式の一つと考えられる。

医師国家試験改善に関する研究：良質な試験問題の作成とプール制導入に関する研究

分担研究者	自治医科大学	齊藤 寿一
研究協力者	岩手医科大学	村井 和夫
	自治医科大学	横山 英明
	独協医科大学	吉田謙一郎
	日本大学医学部	堀江 孝至
	日本医科大学	荒牧 琢己
	順天堂大学医学部	檀原 高
	東京女子医科大学	二瓶 宏
	杏林大学医学部	伊藤 泰雄
	香川医科大学	高原 二郎
	琉球大学医学部	齋藤 厚

目的：

医師国家試験問題は、従来から80名程度の試験問題作成委員が毎年問題を新規に作成し、それを数回の委員会にかけて、選択と修正を経て実際の試験に出題する320題を決定している。しかしながら今後、より精細な評価を行なうためには一回の試験に使用する試験問題の数を増やすことが望ましいと指摘されており、毎年の新規作成で十分に質の高い問題を確保することが困難となることも予想される。その対策としては良質の問題を多数プールしておくことが望ましい。問題をプールするには第一に毎回の試験で出題された問題を志願者が持ち帰って、問題が一般に公表されてきた従来の方式を改め、反復して出題できる様にするのが重要である。また第二には問題の源として、各医科大学から多数の良質な問題の提出を求め、その中から適切な問題を選出して修正・改良してプール問題とすることが有効であると考えられる。

今回の研究では各医科大学から提出された試験問題で医師国家試験を実施するためのフィールド調査として、模擬的な作業を行ない「医大出題方式による問題プール制度」に検討を加えることを目的とした。

方法：

予め研究の趣旨を説明し協力を了承した全国の10医科大学についてそれぞれ試験問題検討委員1名の推薦を求め、表記した10の協力大学と研究協力者を決定した。その大学に対し、各大学の学内試験で使用された試験問題を以下の基準で提出する様に求めた。

- ①学内で6学年学生に対して出題された過去の試験問題から提出する。
- ②出題領域としては現行の必修問題に準じ、医師となるために必須の事項を問う問題とする。
- ③学内試験における正解率は80%以上で、かつ識別指数が0.1以上であること

と。

上記の基準に合致した問題を各大学10題、対照としてこれらの基準から外れた問題2題の計12題の提出を求めた。

これらの集めて参考120題の問題を作成した(参考資料)。平成10年11月16日に厚生省において会議を持ち、9委員に対し、医師国家試験の「必修問題」として適切であるか否かを全120題について判定を求め次の4つの基準で数値化した。

0：適切である。

1：どちらかと言えば不適切である。

2：不適切である。

3：極めて不適切である。

各問題について9委員から提出された数値を集計し、その問題の不適切度とした。

この検討から、特定の医科大学における学内試験での正解率と識別指数から判断した問題の良否が、協力者によって判定出来るか否かを検討した。

結果：

提出された問題について、指摘された問題の不適切度と提出元大学における正解率の分布を図1に示す。判定された不適切度の高い問題は正解率が85%以上または30%以下の問題に多くみられた。また、不適切度と提出元大学における識別指数の分布を図2に示す。図上明らかな様に、識別指数についてはこれが0.2以上と識別指数について適切な範囲にある問題も不適切と判断されやすく、問題から識別指数を予測することは困難であることが明らかになった。

これらを加味して、正解率40%以上でかつ識別指数が0.1以上である適切な問題と、正解率が40%未満または識別指数が0.1未満である不適切問題の2群について、それぞれ2名以上の協力者が不適切であると判断した場合の数を集計した(表1)。この集計について、不適切と判断される頻度が正解率が40%未満また識別指数が0.1未満である不適切問題で、不適切と判断される頻度が高いか否かを検討した。委員9名中8名以上が適切と判断した問題数と2名以上が不適切と判断した問題数の比率は、正解率が40%以上でかつ識別指数が0.1以上である適切な問題では、2.27で正解率40%未満、または識別指数0.1未満の不適切問題では1.17で、前者に高い傾向がみられた。しかしながら、 χ^2 検定では推計学的な有意差はみられなかった。

考案：

フィールド試験の予備段階として今回は特定医科大学で6学年学生実施された試験の結果が協力者の判断とどの程度一致するかを検討した。その結果、問題の難易度についてはある程度、第三者である委員によって判定選別することが可能であろうと思われた。しかしながら個々の問題の識別力も踏まえた適否の予測は必ずしも容易ではないことも明らかになった。今回の判断は、出題者が不在の状

態で下され、出題の意図説明や協力者による質疑検討がなされなかったこと、協力者は全問題について判断しそれぞれの専門性は度外視して判断を求めたことが通常の問題作成委員会とは異なった状況となった。また今回使用した正解率や識別指数が特定の医科大学の学生での試験で得られた数値であって、必ずしも医師国家試験受験者全体の動向を反映していない可能性があること、などの諸点を考慮する必要があるだろう。この点についての詳細な検討は、平成11年度に予定されている参加10大学でのフィールド試験の結果の解析に委ねられる。

今回の基礎調査から、各医科大学から提出された多数のプール用問題をブラッシュアップする作業は問題の作成に匹敵する、ブラッシュアップ委員の専門性や作題者による出題意図の説明・検討が必要であり機械的にこれを取り進めることは困難であることを示唆している。その意味で現在、米国で良質問題プールのために行なわれている様に実際の医師国家試験の一部の問題に採点の対象としない調査対象となる問題を混ぜて試験を行ない、問題そのものの適否を評価する方式の妥当性が、わが国においても検討されるべきであると思われる。

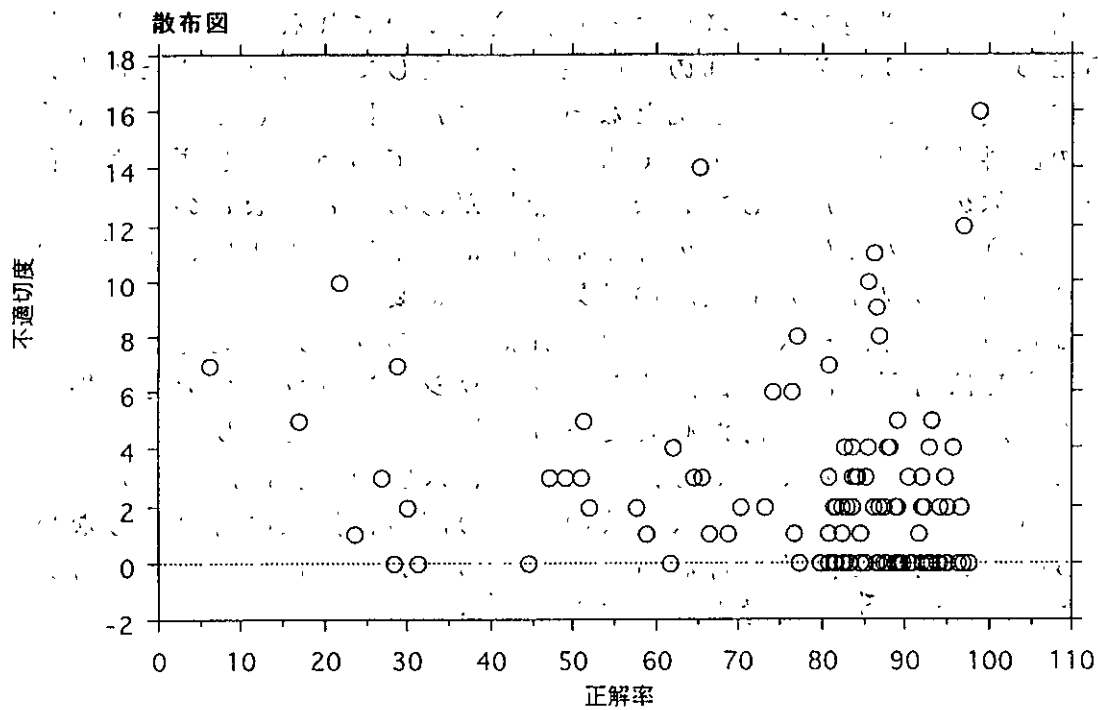


図1 出題医科大学に学内試験における正解率と委員判定による不適切度の相関

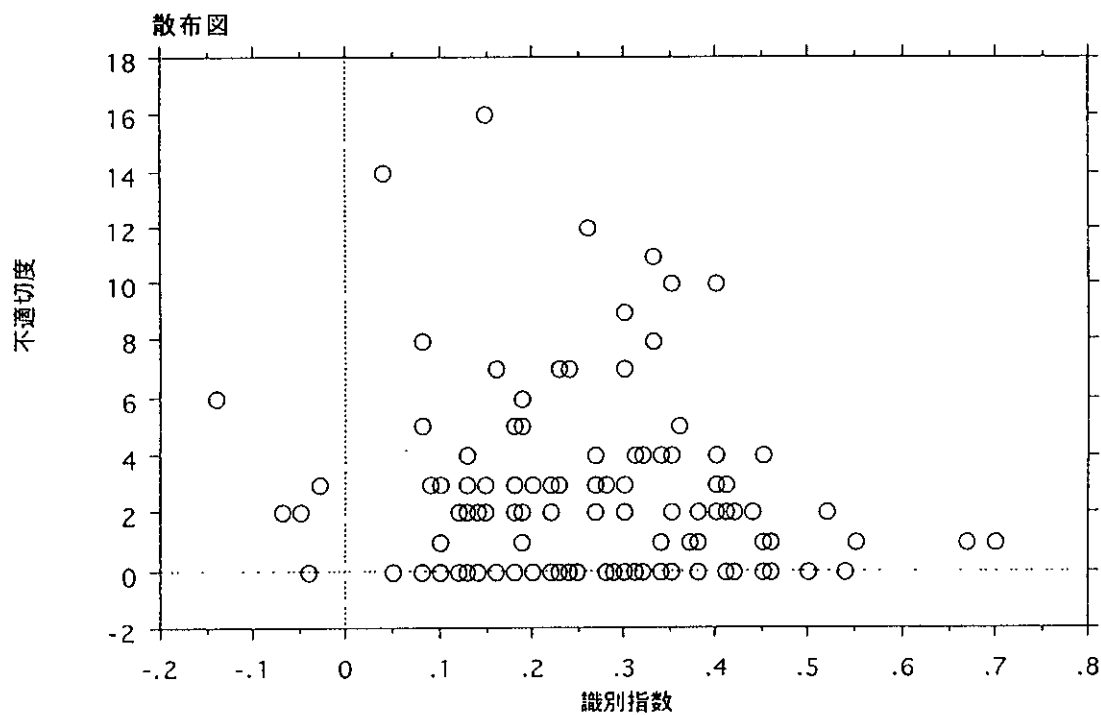


図2 出題医科大学に学内試験における識別指数と委員判定による不適切度の相関

表1 出題元大学における正解率・識別指数と判定委員による判定

	出題元大学の正解率 40%以上で識別指 数0.1以上の問題	出題元大学の正解率 40%未満または識 別指数0.1未満の 問題	合 計
不適切の 判定2名 未満	73	10	83
2名以上 が不適切 と判定	28	9	37
合 計	101	19	120