

## 医学教育シリーズ

# 良質な試験問題の作成法とその評価の仕方

平野 豊<sup>1</sup> 松尾 理<sup>2</sup>

<sup>1</sup>近畿大学医学部附属病院中央臨床検査部・循環器内科学教室

<sup>2</sup>近畿大学医学部医学教育研究センター・第2生理学教室

## 抄 録

客観試験 (objective examination) とは、出題時にすでに正答が決定されており、採点者が異なっても同一の採点結果が得られるという利点がある。そのためには、「問題文はある共通の約束の枠内で作成し、簡潔で日本語としての曖昧さを残さない明瞭な文章」としなければならない。多数の受験者の誰が読んでも、誤解を招かない多義性のない文章にする必要がある。さらに問題作成には十分な時間をかけると共に、作成後もその問題について充分吟味をする必要がある。ここに問題評価の重要性が秘められている。

望ましい試験の条件としては、健全な自己学習を促進させ、問題解決能力を向上させるものが望ましい。いわゆる山かけや一夜漬けをできる限り排除できるような問題作成を心掛けるべきである。知識の記憶を重視した出題に偏重し、学生の応用力や問題解決力を問わない出題が多いと、実際の医師国家試験や共用試験 CBT で役に立たないことがある。したがって難問や奇問を排し、暗記のみでは正解に至らないような解釈や問題解決に重点をおいた出題を心がけるべきである。またその試験問題によって、良く学習し理解している学生とそれ以外の学生を高率に識別できなければならない。したがって問題作成後に、その問題が良い問題であったかどうかを評価する必要がある。問題に対する正解率だけでなく、識別指数や最低合格水準 (MPL; minimum pass level) あるいは修正イーベル法等を用いて、問題自体の評価を確実に実行する事によって、はじめて問題作成の作業が完結するのである。この点を全ての問題作成者は十分に認識する必要がある。また試験を実施して充分吟味する事によって、次の問題作成時にさらに良い問題作成が可能になる。

**Key words:** 試験問題作成, 試験問題評価

## はじめに

大学教員として学生の進級試験や卒業試験を作成する事が多いが、試験問題作成には多くの労力と時間を要する。実際には「試験問題作成のための多くの決まり事」を知った上で作問しなければならない。しかし、試験問題作成について詳細に述べられたものは少ない。今回、試験問題作成に携わる方々のために、試験問題作成とその評価について詳細なコツあるいは注意点について述べる。

### 客観試験 (objective examination)

出題時にすでに正答が決定している試験である。採点者が異なっても同一の採点結果が得られるという利点がある。客観試験によって評価できる学習領域 (taxonomy) は、知識、態度、技能の3領域のうち、知識の領域の想起、解釈および問題解決能力である。

## MCQ (Multiple Choice Question)

MCQ は客観試験の1形式であり、多肢の解答肢が用意されている。選択肢は単純に真偽 (いわゆる○, ×) を問うのではなく、選択肢間に「いろいろ違った段階の正しさのもの」があって、その違いを識別する能力、判断力を問うようなものが望ましい (one-best-answer)。

### 共用試験と国家試験

医師になるために受けなければならない客観試験としては、共用試験と国家試験がある。

共用試験とは「臨床実習開始前の学生評価のための共用試験システム」における CBT (Computer Based Testing) と OSCE (Objective Structured Clinical Examination) のことである。実施機関は全国の医科大学・医学部が創設した医療系大学間共用試験実施評価機構である。学生が診療参加型臨床

実習 (clinical clerkship) を行う前の4年次に受ける関門である。

診療参加型臨床実習とは医学生が指導医のもとで、診療チームの一員として実際の診療に従事する。そのため、一定の医行為が許容される。医学生が診療への責任が生じる医行為を行うためには、適切な評価試験を実施して、それに合格した学生には、指導医のもとで一定範囲内の医行為が許されることとなる。したがって、それ相応の知識が備わっている必要があると共に、医療面接、診療録記載、プレゼンテーション技法、診察技法、簡単な手技、および臨床推論が出来るようにならなければならない。そのための評価がCBTとOSCEである。前者が知識と問題解決能力を評価し、後方で技能と態度を評価する。共用試験は、2005年12月から正式に実施された。合格レベルは実施大学が決定することになっている。

CBTの問題は、プール問題から出題され、新規問題は試行問題とされる。難易度調整は項目反応理論<sup>1</sup>でおこない、コンピューターを用いてランダム出題で実施される。また試験当日には公平に実施されていることを担保するために共用試験実施評価機構から各校に試験モニター員の派遣が行われる。さらに、共用試験実施評価機構は参加大学に対してCBTの新作問題作成を依頼し、得られた問題については全問題をブラッシュアップしている。ブラッシュアップされた問題がプール問題となる。全問題の解答データを踏まえ、問題の再検証(問題評価)も実施している。受験する学生には一人ずつ異なった問題が画面上に出るが、問題全体の難易度は同一になるように問題の組み合わせを作っている。

### 総合試験あるいは卒業試験問題の作成

基本的には医学教育モデル・コア・カリキュラム—教育内容ガイドライン—平成19年度改訂版<sup>2</sup>、あるいは新しく改訂された平成21年度版医師国家試験出題基準<sup>3</sup>に準じる必要がある。

### 医師国家試験について

医師国家試験の問題の種類としては、必須問題、医学総論、医学各論に分けられる。また問題の形式は一般問題と臨床問題がある。問題数は必須問題(一般問題と臨床問題含む100題)、一般問題200題、臨床問題(200題)の合計500題で行われている。

### 必修問題の内容

必修問題の内容としては、プライマリケアを主体に簡単な機器を利用して、医療面接や通常的身體診

察で行える内容(診療所や一般病院での外来入院診療を想定)、多科にまたがるような基本的な問題内容(例えば全身性疾患と皮膚病変、眼病変)、あるいは医の社会的問題や倫理的問題等が考えられる。詳細な項目はガイドラインにリストアップされている。必修問題の形式としては、必ずA型で作問されている。

### 禁忌肢問題

禁忌肢問題としては、患者の死亡や臓器の不可逆的な機能廃絶に直結する事項、あるいは医の社会的問題および倫理的問題が問われる。

### 試験問題の形式

医師国家試験ではA型:単純択一形式(One-Best-Response);(a b c d e)の中から1つ選べと、X型:多真偽形式(五肢複択形式)X2:定数2肢(a b c d e)の中から2つ選べ、が実際の試験問題として用いられてきた。しかしながら、平成21年度からは、正解肢数を指定しないで解答させる問題(Xn型)、選択肢が5肢以上あり、その中から正解肢を1つ選択させる問題、さらに実際に計算させる問題等の新しい型の問題が何問か出題されることになっている。

一方、CBT試験では形式タイプA、タイプL、タイプQ問題が出題されている。

#### タイプL問題

タイプL問題は、Theme;その問題で扱うテーマ、Options;選択肢(8~15肢)、Lead in;設問文でなる。この形式では選択肢数が8~15肢と多いので、偶然正解に当たる可能性は激減する。さらに設問文によっては単純想起だけでなく、解釈力、応用力、あるいは問題解決能力も測定できる。

#### タイプQ問題

タイプQ問題は臨床問題4連問である。第一問が医療面接、第二問が身体所見、第三問が検査、第四問が病態生理を問う。学生が診療参加型臨床実習において要求される能力として臨床推論ができるかどうかを問うものである。

### 一般問題作成の留意事項

#### (1) 選択肢が名詞の場合

「質問文」+「選択肢(名詞)」の問題である(図1)。質問文は「〜で正しいのはどれか。」とする。選択肢は簡潔な名詞であるため主に知識の有無を問う問題として用いられる。

#### (2) 選択肢が文章の場合

「質問文」+「選択肢(文章)」の問題である(図2)。

医療面接で良好な医師患者関係を築く上で有用なのはどれか。

- (a) 干渉
- (b) 説得
- (c) 共感
- (d) 教育
- (e) 誘導

図1 例題；選択肢が名詞（A type）

診療録の記載について正しいのはどれか。2つ選べ。

- (a) 空欄を作って記載する。
- (b) できるだけ略語を用いる。
- (c) できるだけ外国語を用いる。
- (d) 読みやすい字ではっきり記載する。
- (e) 治療行為や検査の直後に記載する。

図2 例題；選択肢が文章（X2 type）

質問文は「～について正しいのはどれか。」とする。選択肢の文章は、可能な限り同じ長さとして一行以内にまとめるようにする。

### 試験問題作成の心構え

問題の文章は、共通の約束の枠内で作成し、簡潔で日本語としての曖昧さを残さない明瞭な文章とする。多数の受験者の誰が読んでも、誤解を招かない多義性のない文章とする。カルテ用語や学会発表用語はできるだけ避ける。作題には十分な時間をかけて多面的に何回も吟味し、細心の注意を払う必要がある。

### 問題作成時の一般的注意点

- (1) 人名は原語によること。（例：Down 症候群）ただし CBT ではカタカナ名を括弧内に併記する。Down（ダウン）症候群。
- (2) 薬品名は「～薬」という記載にし、日本医薬品集（赤本）に準拠する。市販名（商品名）は避けるか、一般名の後に記載する。
- (3) 年齢別の呼称については、4 週未満：新生児、4 週から 1 歳未満：乳児、1～12 歳：男児、女児、13～18 歳：男子、女子、19 歳以上：男性、女性、というように細かく決められている。
- (4) 「必ず」、「常に」、「すべて」等の限定語は使わない。
- (5) 「～ことがある。」という表現は否定できないので使わない。
- (6) 設問文は原則として肯定文にする。
- (7) 設問が否定形の場合は、選択肢を否定形にしない（二重否定を避ける）。
- (8) 原則的に選択肢はすべて対等の重みを持ち、同一範疇の事象にまとめる。
- (9) 一つの選択肢に 2 つ以上の内容を含ませない。

例 1：短文～長文の順に配列（治療薬、診断など）

- (a) ST合剤
- (b) 抗結核薬
- (c) 抗真菌薬
- (d) 副腎皮質ステロイド薬
- (e) 非ステロイド性抗炎症薬

例 2：解剖学的高位～低位順に配列

- (a) 第1頸髄
- (b) 第3頸髄
- (c) 第5頸髄
- (d) 第7頸髄
- (e) 第1胸髄

図3 選択肢の作成

口腔内所見と疾患の組合せで正しいのはどれか。

- (a) 巨舌 ————— 強皮症
- (b) 口唇の毛細血管拡張 ——— Behçet病
- (c) 口腔内潰瘍 ————— CREST症候群
- (d) う歯の増加 ————— Sjögren症候群
- (e) 再発性アフタ性口内炎 ——— 全身性エリテマトーデス

図4 組み合わせ問題

- (10) 各選択肢の長さは大体等しくする（長い肢は正答肢となることが多いため）。
- (11) 否定句を用いる場合はアンダーラインを引く（誤っているのはどれか）。
- (12) 各選択肢は独立した内容であれば、短いものから長いものへ順に、原語であれば ABC 順に、検査であれば侵襲度順に、解剖学的高位～低位順などできるだけ「論理的順序」とする（図3）。

### 組み合わせ問題（図4）

説明文は、「AとBの組合せで正しい（誤っている）のはどれか。」とする。また選択肢については、A、B各群は同一範疇の内容に統一し、症候や身体所見は最初に配置する方が良い。

### 臨床問題作成の時の注意点

- (1) 先頭に「年齢と性」を記載する（例：64歳の男性）。
- (2) 来院理由は「〇〇を主訴に来院した。」、あるいは「〇〇のために来院した。」と表記する。
- (3) 現病歴の記載は時系列順に簡潔な言葉で記載する。
- (4) 既往歴・家族歴の記載する事項がなければ「特記すべきことはない」と書かずに省く。
- (5) 現症の記載  
「身長」「体重」「体温」「呼吸数」「脈拍」「血圧」の順に記載する。必要がないものは省略してよい。
- (6) 検査所見の記載  
「尿検査」「血液検査」「血清生化学検査」の順に記載し、画像検査等は最後に記載する。  
1) 尿所見；肉眼所見、尿量、比重、浸透圧、pH、

定性試験結果(蛋白, 糖, ウロビリノゲン, ビリルビン, 尿潜血 etc), 定量検査(蛋白, 糖 etc)の順に記載する。その後ろに尿沈渣検鏡, 細菌検査を記載する。

2) 血液所見はまず血沈, 赤血球, Hb, Ht, 網赤血球, 白血球(分画を含む), 血小板の順に記載する。次に止血機能検査(出血時間, 全血凝固時間, PT, PTT…), 造血能/溶血に関する検査(TIBC, UIBC…), 血液型/輸血関連検査の順に記載する。

3) 血清生化学検査所見は血糖, 蛋白(総蛋白, Alb, 蛋白分画, 免疫グロブリン…), 含有窒素, 脂質, 生体色素(ビリルビン), 酵素[AST (GOT), ALT (GPT), LDH, ALP…], 電解質, ホルモンの順に記載する。

4) 免疫学検査については, 感染症抗体, 自己抗体, 補体, 免疫蛋白, 腫瘍マーカーの順に記載する。

5) 画像検査については, 通常, エックス線(単純, 造影), CT 検査(単純, 造影), MRI 等の順に記載する。

### 問題の評価

多肢選択式問題(MCQ)形式による出題の利点としては, 一定時間内に多数の問題が出題できるので信頼性が高い。学生のいわゆる「山かけ」を排除でき, しかも客観的な採点ができることが挙げられる。しかし, 過去に出題された問題ばかりを出題すると, 学生は記憶に頼る勉強をするので, 過去に出題された問題の出題は避けて, 暗記型ではなく問題解決型の出題をすることが必要である。過去に出題された問題の吟味を充分行って, 誤答肢を上手に改変すれば, 学生の弱点を補強でき, また実習や講義での力点とする事ができる。

### 良い試験の条件

良い試験問題の条件としては, いわゆる「山かけ」や「一夜漬け」を排除して, 健全な自己学習を促進させ, 問題解決能力を向上させることが必要である。そして試験問題による評価の妥当性・客観性・信頼性が高く, 良く学習し理解している学生とそれ以外の学生とを識別することができなければならない。知識重視の出題に偏重し, 学生の応用力や問題解決力を問わない出題が多いと, 良い成績で卒業しても実際の国家試験に合格出来ないことになる。難問, 奇問を排し, 暗記のみでは正解に至らないような解釈や問題解決に重点をおいた出題に心がけるべきである。厚生労働省からの国家試験問題作成の依頼もそのような問題を求めている。問題作成方法としてはまず試験計画書を先に作り, どの範囲から何を問

う問題を何題出題するかを決定してから, 問題を作るのがよい。

### 試験問題の良質化へ

試験を実際に行って, 得られた結果を分析する必要がある。実際には正解率だけでは, その問題は学生にとって簡単なのかあるいは難しいのかは解らない。

#### 1. 識別係数(Discriminating Index)<sup>4</sup>

良く出来る学生(良く勉強している)とそうでない学生(勉強していない)を識別する。-1から+1の間に分布し, 0の場合, 良い学生=悪い学生であり, その識別はできない。1では, 良い学生の全てが正解し, 悪い学生の全てが間違っていたことになる。さらに0から-1では, 悪い学生の正解率の方が高いことになり, 良い問題とは言えない。そのため個々の問題において, 0.20以上になるよう努力する。そのため選択肢に誤答肢をうまく組み込むことによって, 受験者を惑わせて, 良不良を区別する事が必要である(distracters)。誤答肢のいずれも最低2%以上の受験生がが, 必ず引っかけられるようにすることが良問への第一歩である。

例えば, 識別指数が0.02で正解率92%であれば distracter が効いていないことになる。一方, 識別指数0.04で正解率22%であれば, 選択肢全体に分布し

#### 難易度コード

2: 正解もしくは最低の学生が, 正解と区別出来ない選択肢(大いに同情出来る)

1: 最低の学生が, 間違いに気付くかもしれないし, 気付かないかもしれない(多少同情出来る)

0: これを選ぶ学生は許せないという選択肢(全く同情出来ない)

注) ソフトでは正解 ②を表示できない。そのため限りなく正解に近い 2 と区別できないという難点があることを承知しておいて欲しい

図5 MPL (Minimal Pass Level; 最低合格水準)を求めるときの難易度コード

- 5つの選択肢に難易度コードをつける  
a): 1 b): 2 c): 2 d): 1 e): 0とする

- 最低の学生が正解b)を当てる確率を計算する  
=最低合格指数: Minimal Pass Index (MPI)

- $MPI = 2 / (1 + 2 + 2 + 1 + 0) = 0.33$

- これを各問題毎に計算し, 総和した結果がMPLである。

図6 MPI (Minimal Pass Index; 最低合格指数)とMPL

ているので設問に問題ありとなる。一方、識別指数が-0.15で正解率25%であればほとんどが当てずっぽうで正解したことが疑われる。

**MPL ; Minimum Pass Level (最低合格水準)**<sup>4,5</sup>

イリノイ大学医学部教育開発センターによって開発された。許容可能な最低能力の学生が正答を選ぶ可能性を推定するという理論に基づいている。卒業試験や医師国家試験のような資格認定試験では、最低限度の医学的知識と臨床能力を有しているか否かを判定することが要求される。

5つの選択肢がある時20%得点出来る確率がある。したがってMPLを20-100%の間に設定する。

具体的な方法としては、まず選択枝をよく吟味して評価をする事が必要である。その後、選択枝それぞれに難易度をつける。難易度コードからMPI ; Minimal Pass Index (最低合格指数)をつけてMPLを計算する具体的な方法を図5、6に示した。

**修正イーベル法**<sup>6</sup>

問題の難易度と必要度から合格水準を設定する。難易度を選択枝ではなく、問題全体として判定し、必要度は到達目標やその後の学習などとの関連からどれほど重要かを判定する。MPLはMCQにしか使えないが、修正イーベル法は論述問題にも適用出

来るので有用性が高い。また難易度と必要度を複数の人が判定すれば、試験問題の質が評価されていることになる。

近畿大学総合教育委員会が使用している問題入力ソフトでは、試験問題の入力画面(図7)に、出題分野、出題科、出題者、連問の有無、画像の有無、問題種別、カテゴリー、項目、MCQの分類、求められる学力、問題文、設問、答え、Key Word、作文のねらいおよびMPI等を入力できるようになっている。各科で作成された試験問題は、後日行われる総合教育委員会で問題の吟味が行なわれるが、さらにその時に吟味画面で問題の必要度と難易度を入力している。このソフトでは、そのことによってMPLや修正イーベルが簡単に計算され表示されるようになっている(図8)。試験問題を吟味しチェックすることにより、いわゆる難問、奇問を減らすことができる。また、必要度が高く、難易度のほどよい問題を作る目安ができるので、さらに試験問題の質の向上が期待出来る。

**ま と め**

試験問題作成の実際について述べた。問題作成には作成のための決まりがあるので、まずそれを理解してから作成する。絶えず過去問の結果を吟味し、類似性をチェックし、弱点を発見するような誤答肢

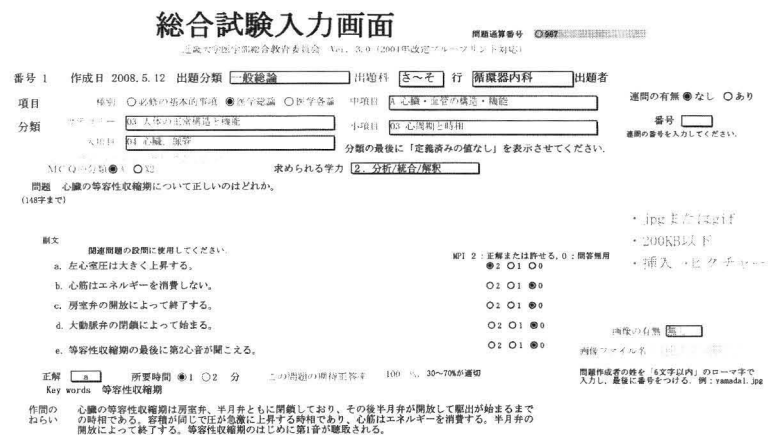


図7 総合試験問題入力画面(総合教育委員会)

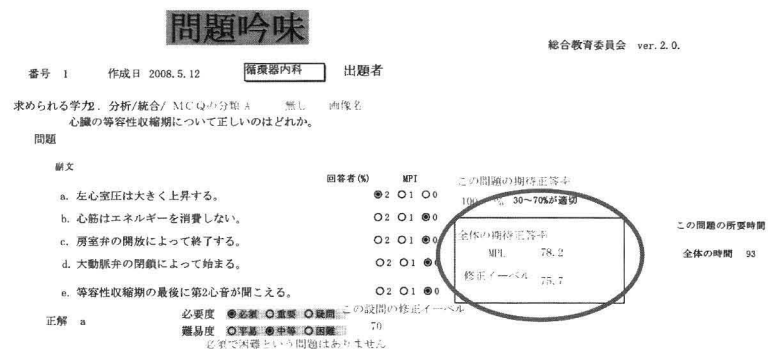


図8 試験問題吟味画面(総合教育委員会)

を作成し、問題の質を高めていく努力が必要である。そして試験問題作成後の試験問題の評価が良問作成に最も重要なのである。

#### 文 献

1. 高橋正視編。項目反応理論入門～新しい絶対評価～。東京：アイデア出版局，2002
2. 医学教育モデル・コア・カリキュラム—教育内容ガイドライン—平成19年度改訂版
3. 平成21年度版医師国家試験出題基準・ブループリント。2008年医道審議会医師分科会医師国家試験改善検討部会
4. 日本医学教育学会監修，医学教育マニュアル4。評価と試験。篠原出版，東京，1982，24-27
5. 日本医学教育学会監修，医学教育マニュアル1。医学教育の原理と進め方。篠原出版，東京，1982，92-94
6. 牛場大蔵，畑尾正彦，堀原一，山下文雄，石井哲夫，植村研一，岡島道夫，尾島昭次，酒井紀，高久史磨，田中勸，真島英信(1985)試験の合格水準の理論と実際。医学教育；16：175-182