200835014A

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

試験問題プール制の推進等国家試験の改善に係る研究

平成20年度　総括・分担研究報告書

主任研究者　相川　直樹

平成21(2009)年3月

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

試験問題プール制の推進等国家試験の改善に係る研究

平成 20 年度　総括・分担研究報告書

主任研究者　相川　直樹

平成 21（2009）年 3 月

平成 20 年度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)
研究報告書

研究課題名
試験問題プール制の推進等国家試験の改善に係る研究

課題番号
H 18 -医 療- 一般- 018

研究実施期間
平成 20 年 4 月 1 日から平成 21 年 3 月 31 日まで(3 年計画の 3 年目)

主任研究者
相川　直樹
(慶應義塾大学医学部救急医学　教授)
〒160-8582 東京都新宿区信濃町 35
Tel: 03-3353-1368
Fax: 03-3226-9877

分担研究者:
畑尾　正彦 (日本赤十字看護大学　教授)
伴　信太郎 (名古屋大学医学部附属病院総合診療医学　教授)

研究協力者:
鈴木　則宏 (慶應義塾大学医学部内科学　教授)

目　　次

Ⅰ．総括研究報告

試験問題プール制の推進等国家試験の改善に係る研究（総括）―――――――― 1

相川直樹・畑尾正彦・伴信太郎・鈴木則宏

II．分担研究報告

1．試験問題プール制の推進に係る研究――――――――――――――――――――――― 4

相川直樹・鈴木則宏・鈴木昌

（相川分担資料 1）

2．OSCE の実施に関する研究――――――――――――――――――――――――――― 35

畑尾正彦・伴信太郎

（畑尾分担資料）

平成２０年度　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）

総括研究報告書

試験問題プール制の推進等国家試験の改善に係る研究

主任研究者　相川　直樹
慶應義塾大学医学部・救急医学教授

研究要旨

本研究班では、「医師国家試験改善検討委員会報告書（平成 15 年 4 月）」で指摘された検討課題を 3 年計画で総合的に検討し、医師国家試験の更なる改善に資することとし、さらに4年ごとに検討される「医師国家試験改善検討委員会」の基礎資料とすることとした。平成 18 年度の研究成果をもとに、主任研究者である相川直樹が部会長を務める「医師国家試験改善検討委員会」の報告書を 19 年 3 月に提出したが、本研究の最終年度である平成 20 年度の研究では、主任研究者相川直樹の「試験問題プール制の推進」のための良質な公募問題の収集の研究に関しては、前年度に作成された公募問題作成時に使用するチェックリストの効果を検討した。その結果、チェックリストを試用した大学から提出された 20 年度の公募問題の質が、チェックリストのなかった 19 年度の公募問題の質と比べて著明に向上したことが示された。このことから、今後の公募問題依頼時には、本研究班で作成したチェックリストを使用することを提言する。
「OSCE の実施に関する研究では」国家試験としての OSCE の実施に係るマニュアル「医師国家試験 OSCE 実施概要」を策定するため、平成 18 年度～20 年度の本分担研究の研究協力者として参画した 15 名の大学医学部教授・病院長等による会議を開催し、国家試験としての OSCE の実施に関する前提条件、事務局、実施日程、課題と評価表、評価者、標準模擬患者、試験会場、要員、事後評価、成績確定等の実際について検討した。会議での検討事項を集約し、その後に各研究協力者が分担した項目の詳細を検討して「医師国家試験 OSCE 実施概要」を策定し、別刷を本研究報告書の抜粋として関係者に配布した。また、厚生労働省、文部科学省、全国の大学医学部、主な臨床研修病院、日本医師会等関係機関に案内状を送付し、シンポジウム「医師国家試験 OSCE 実施に向けて」を開催した。全国から 68 名の参加があり、医師国家 OSCE の課題と評価表、評価者、標準模擬患者、実施組織、評価の信頼性、資格試験としての OSCE の国際比較など 7 演題が発表されたあと意見交換を行い、国家試験 OSCE 実施に向けた問題点を整理した。
以上、公募問題作成時のチェックリスト使用による良質な公募問題の収集と、「医師国家試験 OSCE 実施概要」刊行により、今後の医師国家試験を改善する方策が提言された。

A．研究目的

医師国家試験は資格試験としての一定の質を担保するため定期的に改善を行ってきているが、平成14年7月に再開された「医師国家試験改善検討委員会」において平成17年から適用される新医師国家試験のあり方が提言された。これをまとめた「医師国家試験改善検討委員会報告書（平成15年4月）」では、①臨床実技試験（Objective Structured Clinical Examination 以下「OSCE」と称す。）の客観的な評価手法の確立や　②禁忌肢のあり方に関する検討などの検討課題も指摘されている。
本研究班では、同報告書で指摘された検討課題を総合的に検討し、医師国家試験の更なる改善に資することとし、さらに4年ごとに検討される平成19年版の「医師国家試験改善検討委員会」の基礎資料とすることとした。本研究の初年度である平成18年度の研究成果は、主任研究者（相川直樹）が部会長を務めて平成18年に再開された「医師国家試験改善検討委員会」の基礎資料となり、「医師国家試験改善検討委員会報告書（平成19年3月）」において反映された。
最終年度である本年度の研究では、当初の研究計画を遂行するとともに、「医師国家試験改善検討委員会報告書（平成19年3月）」で提言された国家試験のあり方に関連して、「試験問題プール制の推進」のための良質な公募問題の収集を目的として、前年度に作成したチェックリストの試用による公募問題の質の向上の効果を検討することとした。さらに、医師国家試験改善検討部会報告書の指摘に応え、医師国家試験における実技試験の必要性と可能性について、医学教育関係者の認識を深めるために、医師国家試験OSCE実施のためのマニュアルを作成することを目的として研究を行い、公募問題の質の改善と、医師国家試験OSCE実施のマニュアル整備により、総合的に医師国家試験の更なる改善の方策を提言することを目的とした。

B．研究方法

本研究の主要部分は、「試験問題プール制の推進に係る研究」（担当：主任研究者・相川直樹）と「OSCEの実施に関する研究」（担当：分担研究者・畑尾正彦、伴信太郎）から構成される。それぞれの研究方法については、各分担研究報告書に示した。

C．研究結果と考察

それぞれの分担研究の結果と考察や効果は、各分担研究報告書で詳細に示した。
最終年度である平成20年度の研究結果を総括すると以下のとおりとなる。
「試験問題プール制の推進のための良質な公募問題の収集」に関しては、チェックリストがなかった19年度の公募問題と、チェックリストを試用して作成された20年度の公募問題の質を比較した結果、そのまま採用し得る公募問題は、19年度の65題（9%）から20年度の222題（28%）と大幅に増加し、簡単な修正で出題可能となる問題を含めると、20年度では3分の2以上の問題が、大幅な修正をしなくても国家試験に出題可能な問題となった。本研究班で作成したチェックリストを公募問題作成時に使用することによって、収集された公募問題の質が大幅に向上するものと判断された。これをもとに今後の医師国家試験問題公募時には、本研究で作成したチェックリストを使用することを提言する。
「OSCEの実施に関する研究」では、「医師国家試験OSCE実施概要」を作成した。受験者数を年間で8000～9000名と想定し、テストの内容としては、Case-based の臨床能力と個別課題(共用試験OSCEよりもハイレベル)の15～20課題とし、8ステーションのローテーション方式で、それぞれ

15 分のステーション時間を用いて時間配分方式（指定型と自由設定型の混在）とし、実質テスト時間は一人当たり 120 分間の OSCE 実施に関して、諸課題（事務局と実施日程、課題と評価表、評価者、標準模擬患者 SP（Standardized Patient）、試験会場、要員、当日のスケジュール、事後評価、成績確定、追試験・再試験、など）について具体的概要を示すことができた。
また、公開シンポジウム「医師国家試験 OSCE の実施に向けて」には、全国から 68 名の参加があり、医師国家 OSCE の課題と評価表、評価者、標準模擬患者、実施組織、評価の信頼性、資格試験としての OSCE の国際比較など 7 演題が発表されたあと意見交換を行い、国家試験 OSCE 実施に向けた問題点を整理した。医師国家試験 OSCE の実施に反対する意見はなかった。
以上、公募問題作成時のチェックリスト使用による良質な公募問題の収集と、「医師国家試験 OSCE 実施概要」刊行により、今後の医師国家試験を改善する方策が提言された。

D．健康危険情報

なし。

厚生労働科学研究費補助金（医療技術評価研究事業）
分担研究報告書

試験問題プール制の推進に係る研究

主任研究者　相川　直樹

慶應義塾大学医学部・救急医学教授

研究協力者　　鈴木則宏（慶應義塾大学医学部・内科学教授）
分担研究協力者　鈴木　昌（慶應義塾大学医学部・救急医学助教）

研究要旨

「試験問題プール制の推進等国家試験の改善に係る研究」の総合的研究課題の中で、主任研究者の分担研究項目としての「試験問題プール制の推進」については、最終年度である平成 20 年度においては、前年度に作成したチェックリストを公募問題作成時に使用することによって、公募問題の質が改善するか否かを検討した。問題公募依頼時に大学を 2:1 の割合で無作為に分け、53 大学にチェックリストを配布して試用を依頼、対照とした 27 大学にはチェックリストを配布しないこととし、チェックリストを試用した大学のうち、19 年度と 20 年度にそれぞれ 9 題以上の問題を提出した 20 大学の公募問題を対象として、主任研究者が問題の質とタクソノミーならびにチェックリスト項目抵触の有無を評価した。
チェックリストがなかった 19 年度の公募問題と、チェックリストを試用して作成された 20 年度の公募問題の質を比較した結果、そのまま採用し得る公募問題は、19 年度の 65 題（9%）から 20 年度の 222 題（28%）と大幅に増加し、簡単な修正で出題可能となる問題を含めると、20 年度では 3 分の 2 以上の問題が、大幅な修正をしなくても国家試験に出題可能な問題となった。
チェックリストのそれぞれの項目については、両年度とも「選択肢は論理的順序に配列されているか」、「医師国家試験出題基準の範囲内の内容か」、「学力の低い受験者でも、なんなく除外できる誤答肢（ナンセンス肢）でないか」の項目に抵触する問題が多かったが、各項目に抵触した問題数は 19 年度と比べて 20 年度の公募問題で大幅に減少した。公募問題の質とタクソノミーの関係を検討した結果では、評価のランクとタクソノミーとには明らかな関係は認められなかったが、20 年度の公募問題では高ランクの問題が増加したために、良問とされるタクソノミー3 や 2 の問題の多くが、そのまま出題可能あるいは修正により出題可能となる問題となった。
以上の結果から、本研究班で作成したチェックリストを公募問題作成時に使用することによって、収集された公募問題の質が向上するものと判断された。

A．研究目的

医師国家試験は資格試験としての一定の質を担保するため定期的に改善を行ってきているが、平成14年7月に再開された「医師国家試験改善検討委員会」において平成17年から適用される新医師国家試験のあり方が提言された。これをまとめた「医師国家試験改善検討委員会報告書（平成15年4月）」では、①臨床実技試験（Objective Structured Clinical Examination 以下「OSCE」と称す。）の客観的な評価手法の確立や　②禁忌肢のあり方に関する検討などの検討課題も指摘されている。
本研究班では、同報告書で指摘された検討課題を総合的に検討し、医師国家試験の更なる改善に資することとし、さらに4年ごとに検討される平成19年版の「医師国家試験改善検討委員会」の基礎資料とすることとした。本研究の初年度である平成18年度の研究成果は、主任研究者（相川直樹）が部会長を務めた「医師国家試験改善検討委員会」の基礎資料となり、「医師国家試験改善検討委員会報告書（平成19年3月）」において反映された。
これを受けて、本研究の第2年度である平成19年度では、当初の研究計画を遂行するとともに、「医師国家試験改善検討委員会報告書（平成19年3月）」で提言された医師国家試験のあり方に則って、更なる医師国家試験の改善の方策を検討した。
以上の総合的研究課題のなかで、主任研究者の分担研究項目として、本研究の最終年度である平成20年度では、平成18・19年度に引き続き、「試験問題プール制の推進」について研究を行った。試験問題プール制にあたっては、平成17年の医師国家試験から、従来は回収されていた試験問題が開示され、広く公開されるようになった。このことにより、公開されている既出問題をプールして試験問題として出題した場合の適否が問題となった。これに対応して、全国に試験問題作成を公募して収集した試験問題（以下、「公募問題」と称す。）の多くが、そのまま試験問題とするには種々の問題があることが指摘されている。このような背景をもとに、良質な公募問題を収集する方策を提言することを目的とした。
特に20年度においては、平成19年度に本研究で作成したチェックリストを公募問題作成時に使用することにより、公募問題の質が改善するか否かを検討した。

B．研究方法

平成18年度に作成した18項目からなる「Web公募システムチェックリスト」を推敲して、平成19年度には13項目からなるチェックリスト改訂版（内容・テーマに関して6項目、説明文・設問文・選択肢に関して7項目）を作成した（資料1）。

資料1

チェックリスト改訂版

A. 内容・テーマについて：

□A-1　医師として具有すべき知識が評価できる問題か。医師として第一歩を踏み出すのに必要な能力を問う問題か。

□A-2　医師国家試験出題基準の範囲内の内容か。

□A-3　非常に稀な疾患・病態ではないか。

□A-4　単一のテーマを扱った問題か。

□A-5　学説により意見が分かれることはないか。

□A-6　当たり前のことを問う内容（いわゆるナンセンス問題）ではないか。

B. 説明文・設問文・選択肢（正解肢・誤答肢）について：

□B-1　説明文・設問文に不用意なヒントが含まれていないか。

□B-2　選択肢は論理的順序に配列されているか。

□B-3　一つの選択肢に二つ以上の内容が含まれていないか。

□B-4　学力の低い受験者でも、なんなく除外できる誤答肢（ナンセンス肢）でないか。

□B-5　二律背反の関係にある選択肢のペアはないか（特にX type で注意）。

□B-6　否定形の設問に否定形の選択肢（二重否定）となっていないか。

□B-7　選択肢に「必ず」、「常に」、「すべて」などの限定句が入っていないか。

1．チェックリストの試用

このチェックリストを20年度の試験問題公募時に試用した。試験問題の公募対象は、医科大学・大学医学部（以下、「大学」と称す。）、日本医師会ならびに管理型臨床研修病院であるが、今回の検討では、施設の規模が比較的均一である全国 80 大学とした。問題公募依頼時に、大学を 2:1 の割合で無作為に分け、53 大学にチェックリストを配布して試用を依頼、対照とした 27 大学には、チェックリストを配布しないこととした。

2．検討対象公募問題

大学からの公募問題は、19 年度（チェックリストなし）では 1474 題、20 年度（一部チェックリスト試用）では 1568 題が提出された（図 1）。

今回の検討では、チェックリストを試用した 53 大学のうち、19 年度（チェックリストなし）と 20 年度（チェックリストあり）にそれぞれ 9 題以上の問題を提出した 20 大学の公募問題（総数： 19 年度 761 題、20 年度 787 題、各大学からの応募数の中央値：19 年度 38 題、20 年度 36 題）を対象とした（図 1 の枠内）。

3．評価方法

公募問題の機密性のため、評価者は主任研究者 1 名とした。評価は、公募問題のブラシュアップ時に用いている、公募問題の質の評価(ランク付けの基準)（表１－１）の A～E に対応して、5～1 とした（表１－２）。

また、評価「4」のうち、説明文、設問文や選択肢の内容でなく、記載法や表現上の修正だけで国家試験問題として出題可能となる公募問題を「4*」とし、選択肢が論理的順序に配列しなおせば出題可能となる公募問題を「4S」とした。なお、今回の検討では、視覚素材に関する評価は行わなかった。

同時に、それぞれの公募問題について、タクソノミー（３～１）を評価し、さらに、チェックリスト項目に対する抵触の有無を調べた。

4．以上の評価後に、① 先ず、ブラッシュアップ時の評価結果（A～E）が得られている 19 年度公募問題について、A～E 評価と本研究による 5～1 評価との相違、② 本研究による評価法を用いて、19 年度と 20 年度の公募問題の質の比較、③本研究による評価（ランク 4 以上）とタクソノミーとの関係、④ チェックリスト項目に抵触した問題、について検討した。

（倫理面への配慮）

本研究は、人に対する臨床研究あるいは動物を対象とする実験研究ではないため、倫理的な問題は生じない。

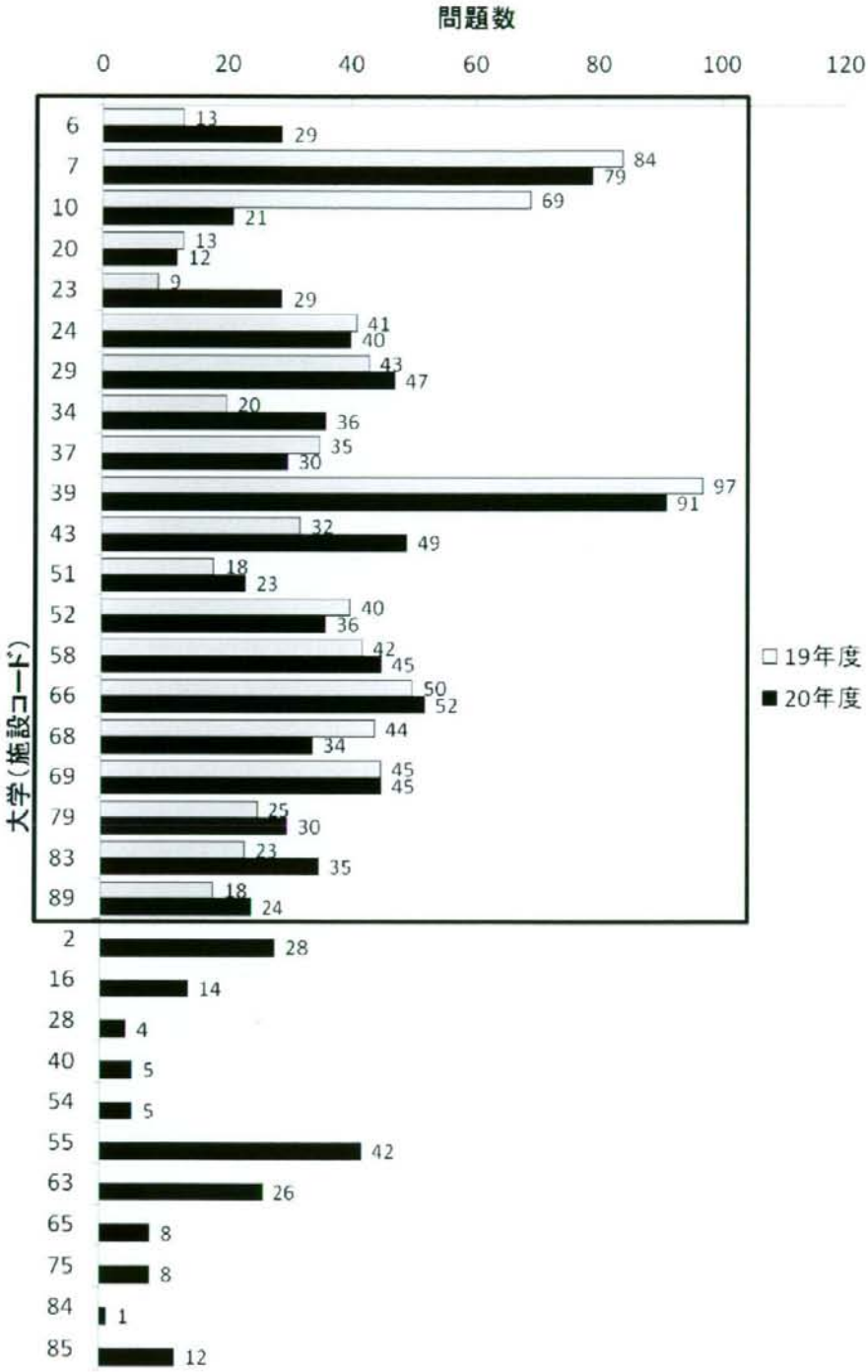
問題数
0
20
40
60
80
100
120
大学（施設コード）
6
7
10
20
23
24
29
34
37
39
43
51
52
58
66
68
69
79
83
89
2
16
28
40
54
55
63
65
75
84
85
13
29
84
79
69
21
13
12
9
29
41
40
43
47
20
36
35
30
97
91
32
49
18
23
40
36
42
45
50
52
44
34
45
45
25
30
23
35
18
24
28
14
4
5
5
42
26
8
8
1
12
□19年度
■20年度

図1．検討対象公募問題

表１．19年度の公募問題評価基準（A～E）と本研究で行った評価（5～1）との関係

表１－１．１９年度の公募問題の評価（A～E）とその基準

カテゴリー	問題の具体的評価
Ａランク （無修正で採用）	内容、表現、視覚素材すべてが適切であり、国家試験に無修正で出題できる。
Ｂランク （一部修正で採用）	ブラッシュアップ委員によって表現等を一部修正していただいた上で、国家試験に出題できる。 （例：問題のねらいは適切だが誤答肢の表現が曖昧なため修正 等）
Ｃランク （アイデア採用）	内容や視覚素材など、部分的に採用できるが、国家試験として出題するための修正には多大な時間を要する。 （例：内容が専門的すぎる、正解肢以外にも正解がある　等）
Ｄランク （不採用）	内容や視覚素材が適切でないため、国家試験としての出題には適さない。
Ｅランク （評価不能）	専門領域がご担当の先生と異なっているため評価が困難である。

表１－２．　19 年度と 20 年度の公募問題の一部に関して本研究で行った評価（5～1）とその基準

カテゴリー	問題の具体的評価
５ランク （内容、表現について無修正で採用）	内容と表現がすべてが適切であり、国家試験に無修正で出題できる。但し、視覚素材については、検討を要するものがある。
４ランク （4*，4Sを含む） （一部修正で採用）	内容と表現等を一部修正すれば、国家試験に出題できる。（例：問題のねらいは適切だが誤答肢の表現が曖昧なため修正 等）。 以上の４のうち、 4*:内容は適切。表現の修正のみで出題可能。 4S: 選択肢の配列を論理的順序に修正すれば出題可能（チェックリストB-２のみが該当） 但し、視覚素材については、検討を要するものがある。
３ランク （アイデア採用）	内容や視覚素材など、部分的に採用できるが、国家試験として出題するための修正には多大な時間を要する。 （例：内容が専門的すぎる、正解肢以外にも正解がある等）
２ランク （不採用）	内容や視覚素材が適切でないため、国家試験としての出題には適さない。
１ランク（評価不能）	評価が困難であるもの。

註１：20 年度の公募問題の一部とは、20 年度の問題公募に際して、チェックリストを使用した施設から提出された公募問題を言う。

C．研究結果

1．ブラッシュアップ時の評価（A～E）と本研究による評価（5～1）との相違

ブラッシュアップ時の評価は、平成19年度の公募問題に対して行われているが、20年度の公募問題に対しては評価行われていない。そこで、ブラッシュアップ時の評価（A～E）がなされている19年度の公募問題のうち、Eランク（評価不能）を除いた664題におけるA～D評価と、本研究による評価との相違を検討し、その結果を表2に示した。
両評価の一致率は67.0%で、一致率の指標であるκ係数は0.54（P＜0.001）であった。
一致率80%以上を良好な一致率、60～79%を中等度の一致率、60%未満を不良な一致率とすれば、ブラッシュアップ時の評価と本研究による評価の一致率は中等度と考えられる。
なお、本研究による評価がブラッシュアップ時の評価を上回っている、すなわち、ブラッシュアップ時より本研究で高く評価された問題は13.4%、逆にブラッシュアップ時の評価が本研究による評価を上回っていた問題は19.6%であった。
但し、本研究による評価では、視覚素材の適否が評価されていないために、視覚素材の適否の評価を追加した場合には本研究による評価のランクが低下することがある。

表2　ブラッシュアップ時の評価と本研究評価の相違（問題数を示す）

評価ランク		ブラッシュアップ時の評価				計
		A	B	C	D	
本研究による評価	5	37	24	0	0	61
	4*, 4S, 4	58	162	16	7	243
	3	5	50	149	42	246
	2	0	4	13	97	114
計		100	240	178	146	664

一致率=67.0%、κ=0.54（κ係数のP<0.001）

ブラッシュアップ時の評価が本研究による評価を上回る：19.6%

本研究による評価がブラッシュアップ時の評価を上回る：13.4%

２．本研究による評価結果

検討対象とした平成19年度と20年度の公募問題の本研究による評価結果を図２と表３に示した。19年度では、ランク５（そのまま採用、但し視覚素材の適否は未検討）は、19年度の761題中65題（9%）に対して、20年度では787題中222題（28%）と大幅に増加した。ランク4*（表現上の修正のみで出題可能、但し視覚素材の適否は未検討）も19年度の120題（16%）から20年度では218題（28%）へと著明に増加した。一方、ランク２（不採用）の問題は、19年度の124題（16%）から20年度では21題（3%）へと著明に減少した（図2）。

ランク５とランク4* ならびにランク4S（選択肢の配列を論理的順序に修正すれば出題可能、但し視覚素材の適否は未検討）の総計は、19年度の305題（40.1%）から20年度の541題（68.7%）となった。（表3）

すなわち、チェックリストがなかった時に作成された19年度の公募問題と、チェックリストを試用して作成された20年度の公募問題の質（ランク）を比較した結果、チェックリストを試用して作成された20年度の公募問題の質が著明に改善した。特に、そのまま出題可能なランク５や、簡単な修正で出題可能となるランク4* とランク4Sの合計が20年度では3分の2以上となった。

これら19年度と20年度のランク別頻度の差は統計的に有意であった（表3）。

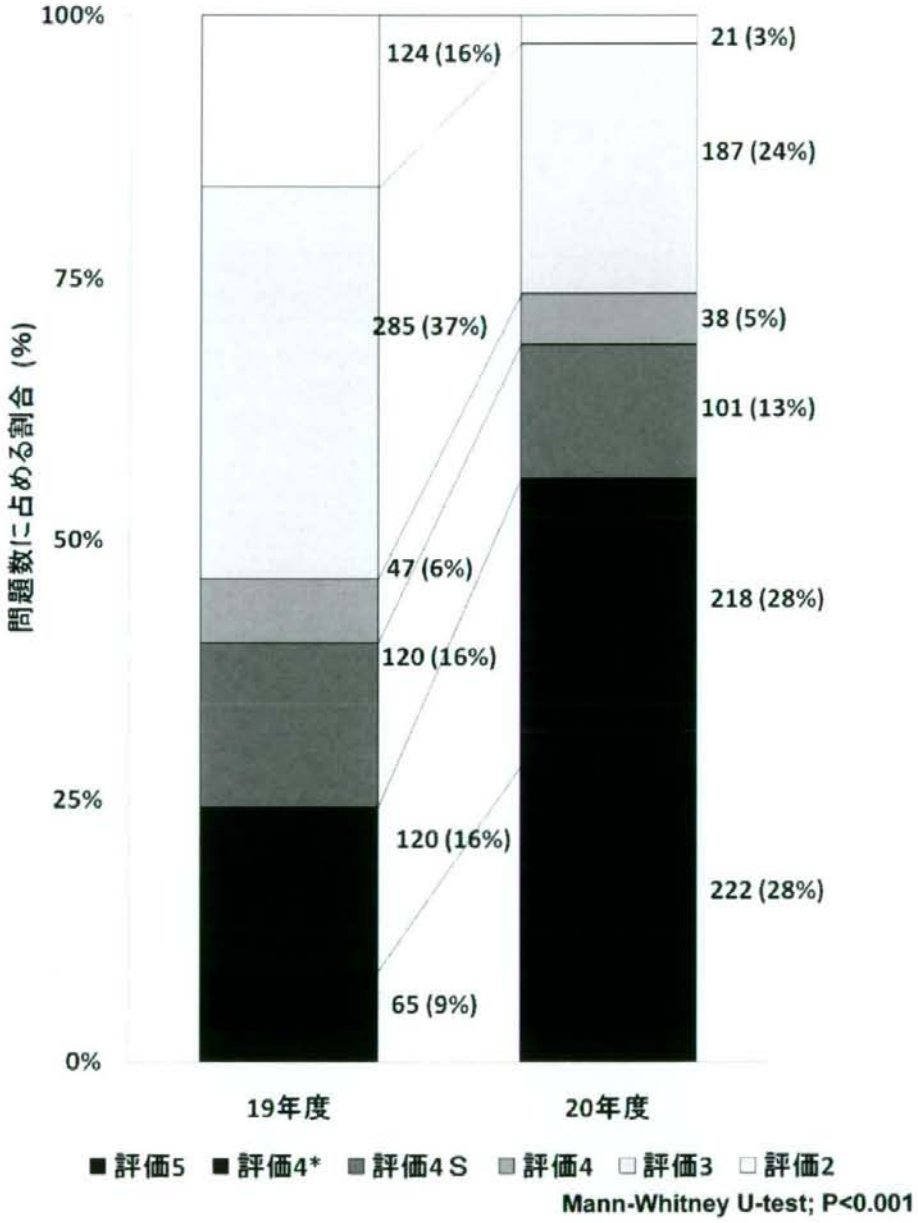

図2. 公募問題の評価結果

表３　19 年度と 20 年度の公募問題の評価結果の比較（ランク 4 以上）

本研究による評価	19 年度 問題数（%）	20 年度 問題数（%）	Odds ratio (95%CI) 19 年度に対する 20 年度の問題数
5	65 (8.5%)	222 (28.2%)	4.27 (3.12−5.67)
4*+5	185 (24.6%)	440 (55.9%)	3.88 (3.12−4.82)
4S+4*+5	305 (40.1%)	541 (68.7%)	3.29 (2.67−4.05)
4+4S+4*+5	352 (46.3%)	579 (73.6%)	3.23 (2.61−4.00)
全問題数	761 (100%)	787 (100%)	−

３．公募問題のタクソノミーと評価結果

平成19年度と20年度の公募問題におけるタクソノミー１～３の占める割合に大きな変動を認めなかった（図３）
両年度の公募問題のうち、本研究による評価のランクが４以上の問題、すなわち国家試験にそのまま出題可能あるいは修正により出題可能となる問題について、評価のランクとタクソノミーの関係を表４に示した。
19年度では、タクソノミーと評価のランクに明らかな関係は認められなかった。
20年度でもタクソノミーと評価のランクに明らかな関係は認められなかったが、19年度に比べてランクの高い問題が増加したため、タクソノミー３に分類される問題85題のうち、ランク５が28題（32.9%）と増加し、ランク５、4*と4Sの合計が60題（70.6%）となった。
また、20年度ではタクソノミー２に分類される問題317題のうちで、4S（選択肢が論理的順序に配列されていない問題）の占める割合が大きくなっていることが示唆され、ランク５、4*と4Sの合計が238題（75.1%）となっている。
以上の結果から、20年度の公募問題では、良問とされるタクソノミー３、２の問題の多くが、国家試験にそのまま出題可能（ランク５）あるいは簡単な修正で出題可能（ランク4*と4S）となっていることが示されている。
ただし、両年度とも視覚素材の適否ついては検討なされていないため、今後さらに視覚素材の適否について検討したうえで、再度解析する必要がある。