

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、公募問題のブラッ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。

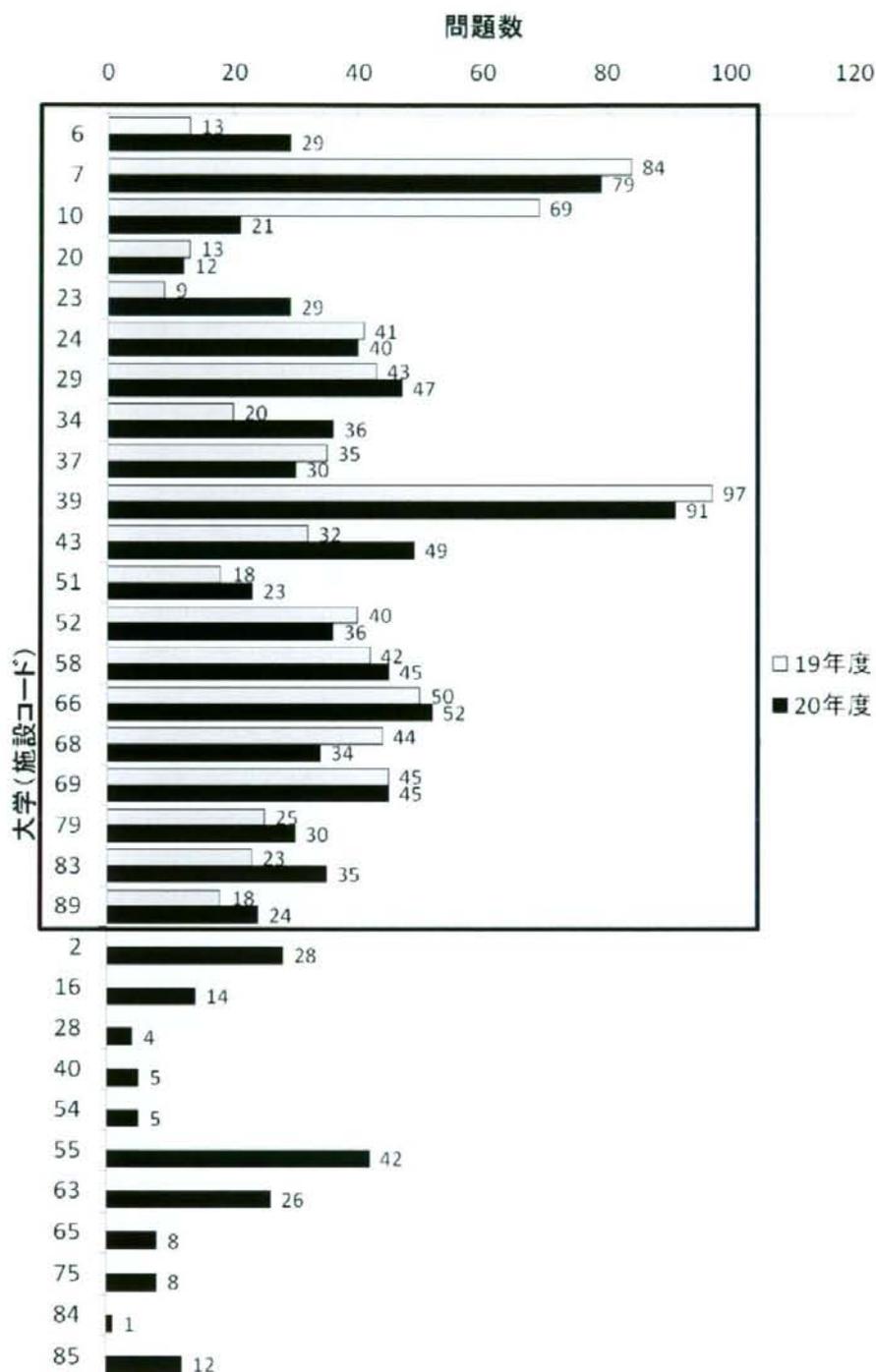


図1. 検討対象公募問題

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)	<p>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。</p>
Bランク (一部修正で採用)	<p>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。</p> <p>(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)</p>
Cランク (アイデア採用)	<p>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。</p> <p>(例：内容が専門的すぎる、正解肢以外にも正解がある 等)</p>
Dランク (不採用)	<p>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。</p>
Eランク (評価不能)	<p>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。</p>

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)	<u>内容と表現がすべてが適切</u> 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)	<u>内容と表現等</u> 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	
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%、 $\kappa=0.54$ (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）。

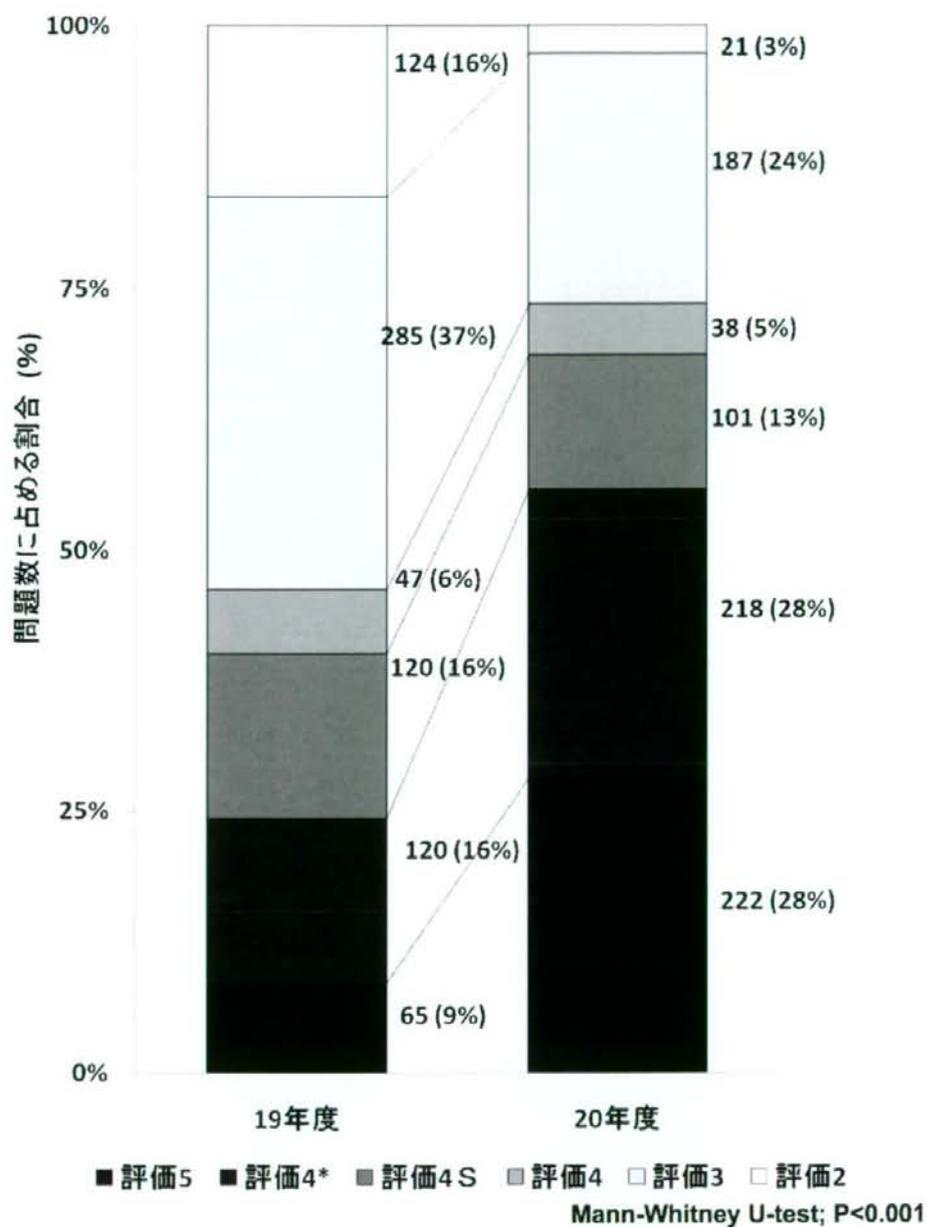


図2. 公募問題の評価結果

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。

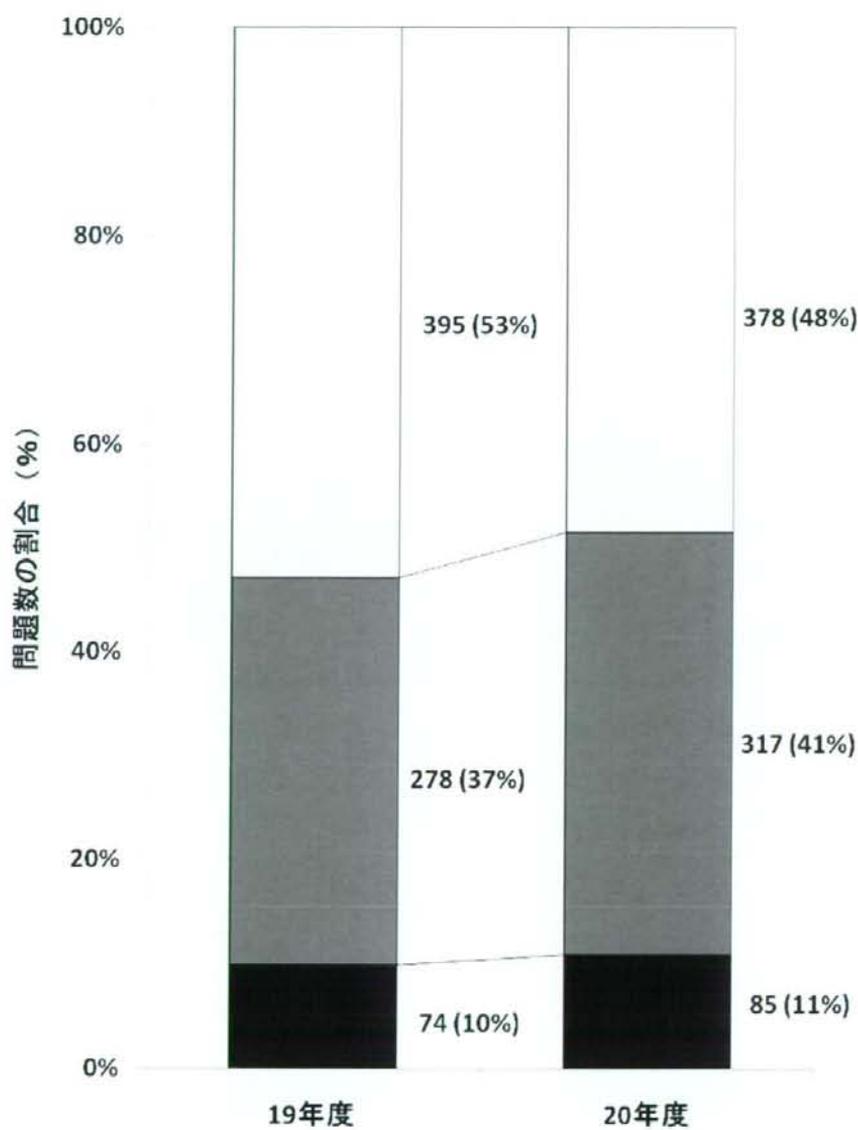
19年度では、タクソノミーと評価のランクに明らかな関係は認められなかった。

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に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。



■タクソミー3 ■タクソミー2 □タクソミー1

図3. 公募問題のタクソミー

Mann-Whitney U-test; P=0.10

表4. 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

年 度	本研究による評価	タクソノミー			Chi-square test
		1 問題数 (%)	2 問題数 (%)	3 問題数 (%)	
19年度	5	38 (9.6)	23 (8.3)	3 (4.1)	P=0.29
	4*+5	89 (22.8)	74 (27.0)	18 (25.0)	P=0.46
	4S+4**+5	157 (39.7)	108 (38.8)	35 (47.3)	P=0.41
	4+4S+4**+5	186 (47.1)	122 (43.9)	39 (52.7)	P=0.38
19年度	全公募問題	395 (100)	278 (100)	74 (100)	
20年度	5	115 (30.4)	79 (24.9)	28 (32.9)	P=0.17
	4*+5	196 (51.9)	191 (60.3)	50 (58.8)	P=0.07
	4S+4**+5	237 (62.7)	238 (75.1)	60 (70.6)	P=0.002
	4+4S+4**+5	263 (69.6)	246 (77.6)	63 (74.1)	P=0.06
20年度	全公募問題	378 (100)	317 (100)	85 (100)	

4. チェックリスト項目に抵触した問題

公募問題のなかで、ランク5と4*に評価された問題は、チェックリストの項目に抵触するものはないが、それ以下のランクの問題では、チェックリストの1項目以上に抵触する問題が多かった。

表5に示したように、1問あたりチェックリスト抵触項目数は両年度の公募問題とも最大4項目で、平均値は、19年度の公募問題（作成時にチェックリストなし）では1問あたり0.82項目であったが、20年度の公募問題（作成時にチェックリスト試用）では0.47と有意に減少した（ $P<0.001$ ）。

表5 問題1問あたりのチェックリスト抵触項目数の比較

	19年度	20年度	Mann-Whitney U-test
中央値	1	0	
平均値	0.82	0.47	P<0.001
(95%CI)	(0.77-0.88)	(0.43-0.47)	
範囲	0-4	0-4	

チェックリストのそれぞれの項目について、抵触する問題数を平成19年度と20年度で比較した結果を表6に示した。

比較検討対象とした19年度の公募問題761題で、最も多かった抵触項目はB2（選択肢は論理的順序に配列されているか）で、294問（38.6%）が抵触、次いでA2（医師国家試験出題基準の範囲内の内容か）が144問（18.9%）、B4（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）が75問（9.9%）であった。

20年度の公募問題787題でも、抵触項目はB2、A2、B4の順に多かったが、それぞれの項目に抵触した問題数は、19年度に比べて大幅に減少した。

両年度の比較で、統計学的にはパワー（ $1-\beta$ ） $>80\%$ で $P<0.05$ であった項目は、B2（odds ratio = 0.40）とA2（odds ratio = 0.48）であった。

特に、「医師国家試験出題基準の範囲内の内容」でない問題が20年度でほぼ半減したことは意義が深い。

表6. チェックリストに抵触する問題数

項目	19年度 問題数 (%)	20年度 問題数 (%)	Odds ratio (95%CI)
A1	25 (3.3%)	14 (1.8%)	0.53 (0.28-1.03)
A2	144 (18.9%)	79 (10.0%)	<u>0.48 (0.36-0.64)</u>
A3	3 (0.4%)	2 (0.3%)	0.64 (0.11-3.86)
A4	8 (1.1%)	4 (0.5%)	0.48 (0.14-1.60)
A5	7 (0.9%)	4 (0.5%)	0.55 (0.16-1.89)
A6	6 (0.8%)	4 (0.5%)	0.64 (0.18-2.29)
B1	10 (1.3%)	4 (0.5%)	0.38 (0.12-1.23)
B2	294 (38.6%)	159 (20.2%)	<u>0.40 (0.32-0.51)</u>
B3	39 (5.1%)	31 (3.9%)	0.76 (0.47-1.23)
B4	75 (9.9%)	60 (7.6%)	0.76 (0.53-1.08)
B5	3 (0.4%)	1 (0.1%)	0.32 (0.03-3.10)
B6	7 (0.9%)	8 (1.0%)	1.11 (0.40-3.07)
B7	6 (0.8%)	3 (0.4%)	0.48 (0.12-1.93)
全問題数	761 (100%)	787 (100%)	

5. 評価（ランク）とチェックリストの意義（図4）

公募問題の構成要素（タクソノミー、チェックリストの有無、および視覚素材の有無）と質の評価（ランク5～1、表現修正要、チェックリストのA1, A2, B2, B3, B4, B5）の関係を最適尺度法を用いたカテゴリーデータの非線形正準相関分析を用いて検討した。なお、チェックリストでは、表6に示されたA1, A2, B2, B3, B4項目を検討対象とし、その他の項目は本検討に含めなかった。

図4は公募問題の構成要素と質の評価との関係を非線形正準相関分析を用いて2つの成分を抽出し、分析に投入した各々の項目に算出される2つの成分負荷量を用いて2次元座標化したグラフである。なお、成分負荷量は、相関係数と同様に、-1から1の間の数値をとり、絶対値の大きいほどその成分の影響を強く受けていることが示される。このグラフでは、次元1を横軸に、次元2を縦軸に示している。チェックリストの有無は縦軸で示した次元2の成分の影響を受けず、横軸で示した次元1において最も強い成分負荷を持つことを示している。したがって、次元1はチェックリストの影響を示していると考えられる。次元1の負の方向が「チェックリストあり」の影響、正の方向が「チェックリストなし」の影響を示す。一方、タクソノミーは横軸で示した次元1の成分の影響は限定的ながら、縦軸で示した次元2において最も強い成分負荷量を持つことが示される。したがって、次元2はタクソノミーの影響を示していると考えられる。次元2の正の方向がタクソノミーのレベルが高い場合の影響、負の方向がタクソノミーのレベルが低い場合の影響を示している。なお、原点付近はこれら2つの成分負荷が限定的であり、その影響が軽微であることを示している。

評価（ランク）はチェックリストがあると高くなるが、タクソノミーの影響は受けていないことが示唆された。また、チェックリストによって改善が期待できる項目はB2（選択肢は論理的順序に配列されているか）とA2（医師国家試験出題基準の範囲内の内容か）と考えられた。B4（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。）とA1（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）はチェックリストの影響は限定的であり、タクソノミーが低い場合に該当することが示唆された。一方、B3（ひとつの選択肢に2つ以上の内容が含まれていないか）はチェックリストもタクソノミーの影響も限定的であった。また、視覚素材の使用はタクソノミーが高いレベルの場合に該当していることが示されたが、視覚素材の適否は今回の質の検討には含まれていない。なお、表現の一部修正を要する「表現修正要」は、ランク2以下（不採用）の問題には該当しない。したがって、このグラフにおいては、ランクが高い場合、すなわち、横軸の次元1で負の方向にある領域に「表現修正要」が位置することは当然であるが、タクソノミーの高い場合の影響も強く受けていることから、タクソノミーの高い問題ではランクが高くても、表現など一部修正を要する問題になることが多いと考えられた。

以上の検討から、問題の構成要素と質の評価との関係には、チェックリストとタクソノミーが強く影響していることが示唆された。タクソノミーはチェックリストでは改善できないが問題の質を向上させることが示された。また、視覚素材の使用はチェックリストの影響やランクに影響はしないものの、タクソノミーを高めることが可能であることが示された。

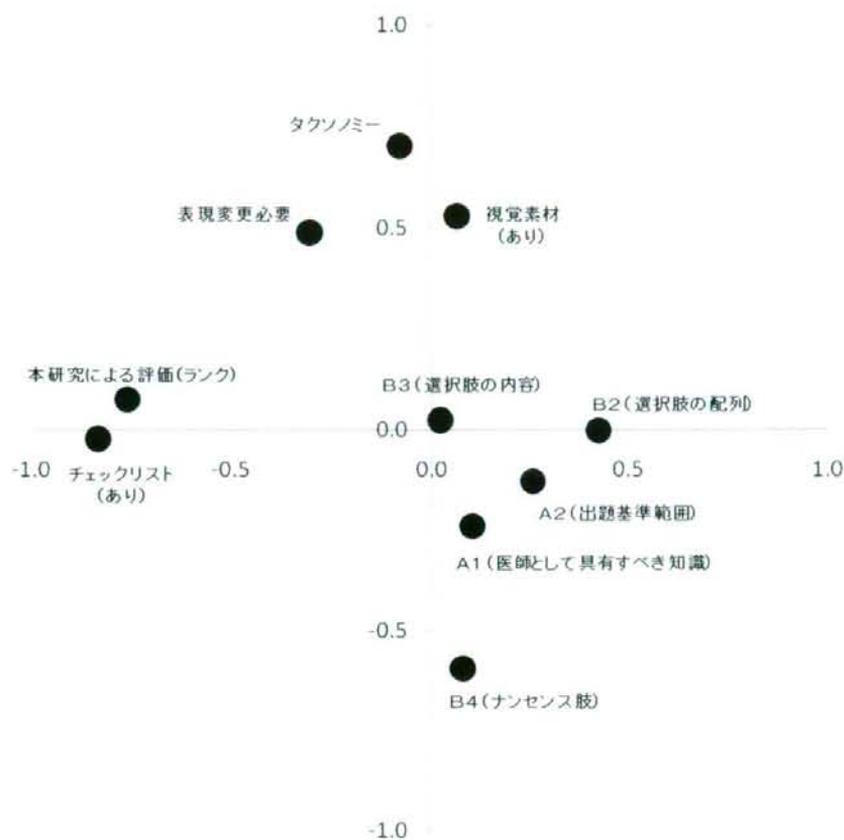


図4. 評価(ランク)とチェックリストの意義

D. 考察

公募問題の質の向上をはかる目的で、本研究の最終年度では公募問題作成時にチェックリストを使った場合の効果について検討した。

平成20年度の問題公募時に、全国80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。

チェックリストの効果の比較方法については、20年度の問題公募を対象として、チェックリストの有無別にブラッシュアップ時の評価で比較する方法があるが、この場合には、大学間のバリエーションが問題となる。

また研究途中で、20年度の問題について21年1月末までにはブラッシュアップによる評価が終了しないことが判明したために、本研究のために、20年度の問題公募の質の評価を、主任研究者が新たに行うことが必要となった。

この主任研究者の評価の妥当性を検証するために、まず、ブラッシュアップ時の評価が得られている19年度の問題公募を対象として、ブラッシュアップ時の評価結果をblindとしたうえで、主任研究者が5~1のランク評価を行った。その結果、主任研究者による評価はブラッシュアップ時の評価と67%で一致していた。なお、将来国家試験問題として出題される公募問題の機密保持のために、今回の研究では、ブラッシュアップ時のように多くの領域の多数の専門家からなる判定者による評価はできなかった。このために、種々の専門領域の視覚素材の適否についての評価も今回はなされていない。

以上のような限界はあるものの、主任研究者による今回の評価は、公募問題の質の絶対評価を目的として用いるのではなく、チェックリストの有無による公募問題の質の比較を目的に用いることから、受け入れられるものと判断した。

そこで、今回の検討では、大学間のバリエーションが問題となる20年度の問題公募を対象としてチェックリストの有無の効果を検討するのではなく、20年度の問題作成時にチェックリストを試用した大学の中から、19年度と20年度の両年度ともに一定数以上の公募問題を提出した20大学の公募問題を対象として、19年度と20年度の問題公募の質の差異を検討することとした。これにより、異なる大学間のバリエーションは排除されるが、19年度と20年度の年度差がバリエーションとなるという限界は残る。

以上の方法による検討結果で、チェックリストを試用して作成された20年度の問題公募の質は著明に向上した。視覚素材の適否については更なる検討が必要なものの、修正なしにそのまま国家試験に出題可能な問題は、19年度の9%から20年度では28%に増加し、簡単な修正で出題可能となる問題（ランク4*と4S）を含めると、19年度の40%から20年度では69%に増加した。このようにチェックリストの試用により公募問題の質が著しく向上した。

随伴した現象として、20年度の問題公募ではチェックリストの項目に抵触する頻度も減少したが、A2項目（医師国家試験出題基準の範囲内の内容か）に抵触した問題がほぼ半減したことも、意義深い。

今後は、20年度の問題公募に対するブラッシュアップ時の評価がなされた時点で、その評価結果を用いて、① 20年度の問題公募でチェックリストを試用した施設と試用しなかった施設との問題の質の差異、② 20年度にチェックリストを試用した施設における20年度と19年度の問題公募の質の差異、の両者の検討から、チェックリストの効果をさらに検証すべきと考えられる。

E. 結 論

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、公募問題の質が向上する。

F. 健康危険情報

なし。

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

総合分担研究報告書

分担研究「OSCEの実施に関する研究」

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

研究要旨：1) 大学医学部で行われたAdvanced OSCEを支援した。2) Advanced OSCEの新たなステーション・課題を開発した。またAdvanced OSCEの課題を公募した。3) スキルスラボ・OSCE実施専用施設に関する全国80大学医学部にアンケート調査を行った。4) 公開シンポジウム「国家試験OSCEトライアルの今までの成果」を開催した。5) Advanced OSCEトライアルの評価データを解析し信頼性を検討した。6) 臨床研修開始時に必要な臨床能力に関するアンケート調査を行った。7) 国家試験OSCEの実施マニュアル「医師国家試験OSCE実施概要」を策定した。8) 公開シンポジウム「医師国家試験OSCE実施に向けて」を開催した。

研究方法：大学医学部教授・病院長等の有識者を研究協力者とする「OSCEの実施に関する研究」班を組織して、Advanced OSCEトライアルの実施を支援し、アンケート調査や公開シンポジウムを通じて、幅広く国家試験OSCEに関する検討を行った。

研究結果：1) 札幌医科大学で行われたAdvanced OSCEにモニターとして参加しフィードバックした。岡山大学で行われたAdvanced OSCEでステーションの運営を支援した。2) 新たに「胃管挿入・抜去」と「目の前が暗くなる」の課題を開発した。課題の公募に対して16課題の応募があった。3) OSCE実施専用施設に関するアンケート調査に76大学（95%）から回答があった。4) 公開シンポジウム「国家試験OSCEトライアルの今までの成果」で評価者や評点等について検討した。5) Advanced OSCE大阪トライアルの評価データを解析し、一般化可能性係数Gを算出した。6) 2年次研修医を対象に、臨床研修開始時に必要な臨床能力に関する自記式アンケートを対象病院ごとに送付し、個別郵送による回収を行った。7) 2008年8月13日に研究班メンバーが一堂に会して「医師国家試験OSCE実施概要」を策定した。8) 公開シンポジウム「医師国家試験OSCE実施に向けて」を開催し、OSCE実施に向けた問題点を整理した。

A. 研究目的

実技試験OSCEを導入して法的にも、社会的にも適正な医師国家試験を実施することによって、医師の資質のさらなる向上に資するために、全国の大学医学部への臨床実習後のOSCE (Advanced OSCE) の普及を支援するとともに、適切なOSCEの課題を収集し、評価者や標準模擬患者の確保を含む実務・運用面の課題に対応したOSCE実施マニュアルを整備することを本研究の目的とする。

B. 研究方法

モデル・コア・カリキュラム策定関係者、共用試験OSCE実施関係者、医師国家試験改善検討部会委員、医師国家試験出題委員等を含む有識者（大学医学部教授・病院長等）を研究協力者とする「OSCEの実施に関する研究」班を組織して、全国規模ならびに大学単位のAdvanced OSCEトライアルを実施あるいは支援し、さらに全国の大学医学部や臨床研修病院へのアンケート調査を行い、また公開シンポジウムを開催して、幅広く国家試験OSCEについて検討した。

C. 研究結果

1) 2006年9月14日(木)に札幌医科大学でAdvanced OSCEが行われたが、「腹痛」、「外来診察」、「神経診察」、「外科手技」の4か所のステーションを4列設定し、ローテーション方式であった。6年生全員(104名)の試験が1日で可能であった。

2006年10月8日(日)に岡山大学で行われたAdvanced OSCEは中国・四国地方では初めてのものであった。

2) 新たに開発した課題「胃管挿入・抜去」と「目の前が暗くなる」は岡山大学のAdvanced OSCEで実施し適切であることが確認された。課題の公募に対して16課題の応募があった。

3) OSCE実施専用施設に関するアンケート調査に76大学(95%)から回答があった。OSCE実施専用施設を有する大学は31であり、10箇所以上のステーションを設定できて、一定期間、その施設を国家試験のために貸し出しできる大学が17であった。

4) 公開シンポジウム「国家試験OSCEトライアルの今までの成果」は2006年11月5日(日)10時30分～16時に東京慈恵会医科大学大学1号館で開催され、評価者や評点等について検討した。卒前臨床実習で臨床実技の学習・修練と関連しなければならないことが確認された。

Ontario大学のQuality Managementの責任者であるDr. Daniel Klassによる「Role of Multiple Station Standardized Patient Exams (MS-SPE) in Licensure」の講演のあと、2006年度に一学年(100名規模)の医学生を対象としてAdvanced OSCEを行った3つの大学の事例が報告された。

5) 評価データの解析

Advanced OSCE大阪トライアルにおける評価データを分散分析した結果は表の通りである。この結果から算出した一般化可能性係数Gは0.817であった。

$$G = \frac{\hat{\sigma}_i^2}{\hat{\sigma}_i^2 + \frac{\hat{\sigma}_{ij}^2}{n} + \frac{\hat{\sigma}_{i(k,j)}^2}{nr}} = 0.817$$

n : 受験者数