

口頭発表D①

## 共用試験 OSCE の項目特性曲線と評価者間変動に関する分析 ーテスト項目の信頼性と妥当性の検討ー

三笠洋明<sup>1,3</sup>, 赤池雅史<sup>1,2,3</sup>, 寺嶋 吉保<sup>2,3</sup>, 谷憲治<sup>1,3</sup>, 高山 哲治<sup>3</sup>, 福井義浩<sup>1,3</sup>

(1: 医学部教育支援センター, 2: 医療教育開発センター, 3: ヘルスバイオサイエンス研究部)

**目的** 共用試験 OSCE は、臨床現場で必要とされる態度と技能を評価し、臨床実習に参加する資格を確認するための、全国の医学部が参加し実施する実技試験である。試験の信頼性と妥当性を高める事を目的として、個々の項目についての評価者間変動と、項目特性曲線による分析を行い、信頼性と、妥当性に問題のある項目について検討を行った。

**対象と方法** 2009 年度の共用試験 OSCE の 7 課題、177 項目において 2 名の評価者が独立して目隠しで評価した試験データ 96 名分を用いた。統計解析は、PASW Ver18.0J を用いた。

**結果および考察** 各項目の 2 名の評価者の一致率の分布を図 1 に示した。

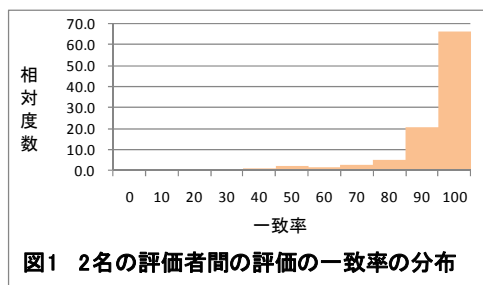


図1 2名の評価者間の評価の一致率の分布

80%以上の一致率を示す項目が全体の 87%を占めている事が示された。

次に、偶然の影響を除いた一致の指標である  $\kappa$  (カッパ) の分布を図2に示した。

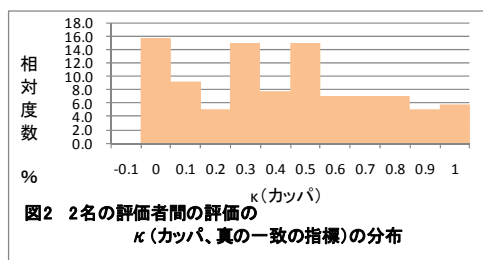


図2 2名の評価者間の評価の  $\kappa$  (カッパ、真の一致の指標) の分布

零付近から完全な一致である1付近まで広く分布しており、統計学的には信頼性に問題を有する項目

が少なくない事が示された。

一致率が 80%を超え、かつ、 $\kappa$  の統計学的な有意性が確認できたものは「信頼性に問題無し」、一致率は 80%を超えるが、 $\kappa$  の有意性が確認できなかった項目は「実用上問題無し」とし、一致率が80%未満の項目は有意性の確認の有無を問わず「何らかの問題あり」と分類した。

表1 一致率と統計学的有意性の確認のクロス表

一致率	有意性の確認		計
	—	確認	
~60%	8	1	9
60%~70%	3	2	5
70%~80%	5	4	9
80%~	80	74	154
計	96	81	177

一致率が 80%未満の 23 (13%) の項目を表2に示した。

表2 評価者間変動の大きい学習項目

課題	分類	項目	一致率	有意の確認
医療面接	オープニング	冒頭で患者さんの訴えを十分聴く	78.1	0
医療面接	医学的情報	症状の程度を聞く	69.8	1
医療面接	医学的情報	症状の経過を聞く	79.2	1
医療面接	医学的情報	症状を増悪、寛解させる因子を聞く	79.2	1
医療面接	医学的情報	症状に随伴する他の症状を聞く	79.2	0
医療面接	良好なコミュニケーション	積極的な傾聴を心がける	52.1	0
医療面接	良好なコミュニケーション	患者さんの気持ちや患者さんのおかれた状況を共感している事を言葉と態度で患者さんに伝える	69.8	0
外科手技	手技のテクニック	手関節の頭側まで洗い、洗い残しが無い事を確認する。	74.0	1
救急	AEDの操作	音声指示に従い胸骨圧迫などを行う。	67.7	0
胸部	診察のテクニック	一側ずつ頸動脈を甲状腺の高さで第2、第3指(または第1指)の指腹を使って軽く触診す	79.2	1
胸部	診察のテクニック	胸部診察 聴診器のイヤピースを外耳道の方向にあわせて装着し、チェストピースを適切に把持する。目的に応じて、膜型、ベル型を使い分ける。	64.6	1
神経	診察のテクニック	神経診察 舌を見たいことを告げ、口を大きく開けてもらい、舌の委縮と繊維拘束性収縮の有無を観察する。	50.0	0
神経	診察のテクニック	課題に関して適切な指示とガイドが出来る	79.2	0
神経	診察のテクニック	課題に関して適切な手技が出来る	60.4	0
頭頸部	診察のテクニック	甲状腺峡部を触診する。輪状軟骨の位置を確認し、効き手の第2指・指腹で甲状腺峡部を軽く触診する	70.8	0
腹部	診察のテクニック	腹部診察 肛門内指診を行う事を患者さんに説明する。	76.0	0
医療面接	概略評価		49.0	1
外科手技	概略評価		43.8	0
救急	概略評価		42.7	0
胸部	概略評価		54.2	0
神経	概略評価		36.5	0
頭頸部	概略評価		38.5	0
腹部	概略評価		57.3	0

医療面接の項目が最も多く、次いで神経診察、胸部診察と続いた。7課題の概略評価の一致率は何れ

も低い、この項目は6段階評価である事が影響していると考えられた。「実用上問題無し」を含め更に信頼性を高めるには、合格基準のみならず、不合格基準の明確化と評価者間での評価基準のすり合わせが必要であると考えられた。また概略評価は、6段階と信頼性には不利であり、一致率は最大でも57.3%であった。しかしながら、学生の処遇に影響するのは合格、境界領域、不合格の3段階である。6段階を3段階評価に変換すると、一致率は7課題とも80%を超える事から、運用上の工夫で改善が期待できると考えられた。

項目の妥当性とは、項目の正答率が測定したい特性(学習到達度)を反映しているかということであり、学習到達度はOSCEの総得点で近似できると仮定した。OSCEの総得点を元に受験生を5群に分け、各成績群の正答率を縦軸にとった、項目特性曲線を全ての項目について作成した。また、成績群と評価結果のクロス表について「線形と線形による関連」を正確確率法により検定を行った。関連が有意であった場合に、識別力が認められる範囲により「模範的」または「一部に識別力」とし、関連が有意でなく全体の正答率が80%を超えるものを「易しい」とした。正答率が低い場合は「識別力無し」全体の正答率が20%を下回った場合は「難しい」と分類した。

表3 グラフ分類と信頼性分類のクロス表

	グラフ分類					合計
	模範的	一部に識別力	易しい	識別力無し	難しい	
信頼性 問題無し	12	14	48	0	0	74
実用上問題無し	1	10	69	0	0	80
何らかの問題あり	7	4	9	2	1	23
合計	20	28	126	2	1	177

妥当性に問題があると考えられるのは「難しい」と「識別力無し」であるが、各々1例と2例のみであった。これら3例の項目特性曲線を図3,4,5に示した。

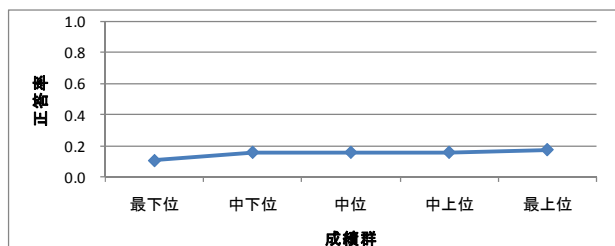


図3 難しいと分類された項目特性曲線  
「肛門内指診を行う事を患者さんに説明する」に関する評価

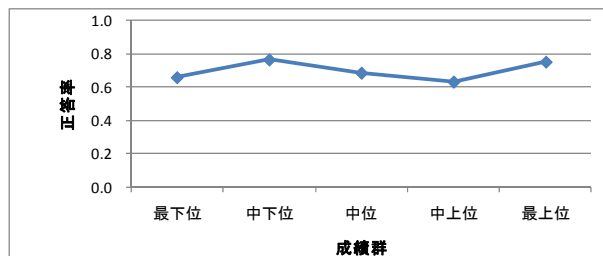


図4 判別力無しと分類された項目特性曲線  
「症状の経過を聞く」に関する評価

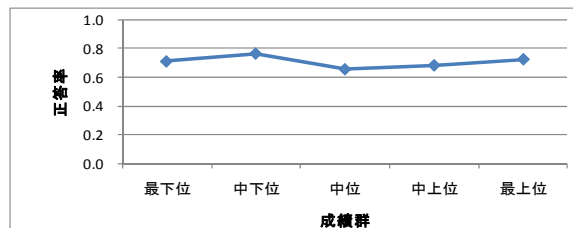


図5 判別力無しと分類された項目特性曲線

「聴診器のイヤピースを外耳道の方にあわせて装着し、チェストピースを適切に把持する。目的に応じて、膜型、ベル型を使い分ける。」に関する評価

一般に妥当性に問題がある項目に関しては、測りたい特性とは直接関係の無い項目である可能性を疑うが、今回の例では「識別力無し」、「難しい」項目は、事前の教育において明示的に示されなかった～教えられなかった可能性が考えられた。

**結論** 共用試験 OSCE の項目の多くは、学生の実施率が高いために、一致率は80%を超える項目が多かった(87%)が、23(13%)の項目は一致率が小さく問題が認められた。特に概略評価(6段階評価)の一致率は最大でも57.3%と低い水準にとどまっていたが、3段階に変換すると実用レベルに達する事から、改善の可能性が認められた。

試験全体の信頼性を更に高めるには、合格基準のみならず不合格基準の明確化、事前の評価者間の打ち合わせ、事前教育の見直しなどにより改善が可能であると考えられた。

項目特性曲線分析の結果、妥当性に疑問がある3項目は事前の教育を適切に行う事で、問題は回避できると考えられた。