

# 東邦大学学術リポジトリ

Toho University Academic Repository

タイトル	Inter rater reliability of grade evaluation of Post Clinical Clerkship (Post CC) OSCE based on kappa coefficient and agreement rates
別タイトル	診療参加型臨床実習後(Post CC)OSCE の成績評価における検者間信頼性の検討
作成者(著者)	高山, 充
公開者	東邦大学
発行日	2022.03.31
掲載情報	東邦大学大学院医学研究科 博士論文 内容の要旨及び審査結果の要旨.
資料種別	学位論文
内容記述	主査: 佐藤二美 / タイトル: Inter rater reliability of grade evaluation of Post Clinical Clerkship (Post CC) OSCE based on kappa coefficient and agreement rates / 著者: Mitsuru Takayama, Akiko Nakada, Naoki Hiroi / 掲載誌: Toho Journal of Medicine / 巻号・発行年等: /
著者版フラグ	none
報告番号	32661甲第1042号
学位記番号	甲第721号
学位授与年月日	2022.03.31
学位授与機関	東邦大学
DOI	10.14994/tohojmed.2021_015
その他資源識別子	<a href="https://mylibrary.toho.u.ac.jp/webopac/TD05160149">https://mylibrary.toho.u.ac.jp/webopac/TD05160149</a>
メタデータのURL	<a href="https://mylibrary.toho.u.ac.jp/webopac/TD37728776">https://mylibrary.toho.u.ac.jp/webopac/TD37728776</a>

# 博士學位論文

論文内容の要旨

および

論文審査の結果の要旨

東邦大学

高山 充より学位申請のため提出した論文の要旨

学位番号甲第721号

学位申請者 : 高 山 充  
                  たか          やま          みつる  
                  高          山          充

学位論文 : Inter-rater reliability of grade evaluation of Post-Clinical Clerkship (Post-CC) OSCE based on kappa coefficient and agreement rates

(診療参加型臨床実習後(Post-CC) OSCE の成績評価における  
検者間信頼性の検討)

著 者 : Mitsuru Takayama, Akiko Nakada, Naoki Hiroi

公 表 誌 : Toho Journal of Medicine

論文内容の要旨 :

1. 目的

2020年に全医学部で実施されたPost-CC OSCEは技能・態度修得を総合的に評価するためのhigh-stakes testであり、本試験における高い信頼性の確保が急務だが、成績評価の信頼性を明らかにした報告は少ない。今回、2020年度Post-CC OSCEの評価教員による成績評価の差異を明らかにした。

2. 方法

1) 研究対象

2020年度にX大学医学部6年生119名に対して行われたPost-CC OSCEの、評価者36名による評価点数を分析対象とした。この試験は全3課題、各課題に6系列、1ステーション内には評価者2名を配して行われ、学生1名を評価者2名が同時に評価した。Post-CC OSCEの評価はルーブリックとなっており、評価観点は、A患者への配慮/コミュニケーション、B医療面接、C診断仮説に基づいた身体診察、D症例プレゼンテーション、E臨床推論、および概略評価の6項目である。評価尺度は6段階(1~6)であり、評価尺度がそのまま配点となっている。

2) 統計分析

ルーブリックを用いた成績評価を従属変数、成績評価に関連する因子(系列、試験実施時間帯、評価者の職位)を独立変数とした。ステーション内の2評価者間の評価点数を評価観点ごとに比較する際には $\kappa$ 係数、Spearmanの順位相関係数、評価者間一致率を用いた。評価者間一致率は6段階評価の一致率に加え合格ラインである4点以上と4点未満の2段階での一致率を算出し

た。3 課題の概略評価点数の平均値を系列ごとに比較するため、学生の合否に直結する概略評価点数の 3 課題の平均値のみを代表値とし、Kruskal-Wallis 検定と  $\chi^2$  検定を用いた。評価者の職位による評価点数の違いを評価観点ごとに比較する際には Kruskal-Wallis 検定、試験実施時間帯による評価点数の違いを評価観点ごとに比較する際には Mann-Whitney の U 検定を用いた。

### 3. 結果

#### 1) 概要

学生・評価者全員が研究への参加に同意したため、本調査では評価点数全てを分析対象とした。119 名の学生が 3 課題を受験したため計 357 回の試験が行われ、それを 2 名の評価者が同時に評価したため計 714 回の評価が行われた。評価者の職位は教授 12 名 (33.3%)、准教授 13 名 (36.1%)、講師 11 名 (30.6%) であり、性別は男性 32 名 (88.9%)、女性 4 名 (11.1%) であった。受験学生は留年経験なし 94 名 (79.0%)、留年経験あり 25 名 (21.0%) であった。各評価観点における平均値は観点 A で  $4.58 \pm 0.74$  点、観点 B で  $4.44 \pm 0.76$  点、観点 C で  $3.8 \pm 0.85$  点、観点 D で  $4.3 \pm 0.76$  点、観点 E で  $4.1 \pm 0.84$  点、概略評価で  $4.04 \pm 0.84$  点であった。

#### 2) ステーション内の 2 評価者間の差異

行われた 357 回の試験を対象として、ステーション内の 2 評価者間の差異を評価観点ごとに評価した。一致度  $\kappa$  係数で最も低かったのは「観点 A : 配慮/コミュニケーション」( $\kappa=0.18$ , Poor)、最も高かったのは「観点 E : 臨床推論」と「概略評価」( $\kappa=0.27$ , Fair) であった。Spearman の順位相関係数で最も低かったのは「観点 A : 配慮/コミュニケーション」( $r=0.256$ ,  $p<0.001$ )、最も高かったのは「観点 E : 臨床推論」( $r=0.424$ ,  $p<0.001$ ) であった。6 段階リッカート評価の一致率で最も低かったのは「観点 C : 身体診察」の 41.7% (121/357)、最も高かったのは「概略評価」の 49.9% (128/357) であった。しかし 6 段階リッカートを合格ボーダーの 4 点以上と 4 点未満の 2 群に分けて一致率を算出したところ、その一致率は最も高い「観点 A : 配慮/コミュニケーション」では 95.5% (341/357) であったが、最も低い「観点 C : 身体診察」では 65.3% (233/357) となり、多くのステーション内で 2 評価者の合否判定が一致しなかった。

#### 3) 系列による成績評価の差異

学生 119 名が受験した 3 課題の概略評価点の平均値を系列ごとに算出した。点数が最も高い系列で  $4.27 \pm 0.41$  点、低い系列で  $3.87 \pm 0.33$  点であり、系列間に有意な差はみられなかった ( $p=0.115$ )。しかし概略評価点を合格ボーダーの 4 点以上と 4 点未満の 2 群に分けたところ、合格率は最も高い系列で 90.0% (18/20)、低い系列で 36.8% (7/19) であり、系列間の合格率には有意な差が見られた ( $p=0.023$ )。

#### 4) 試験実施時間帯による差異

留年経験のない学生 94 名の成績評価を試験実施時間帯によって前半群と後半群に 2 分し平均値を算出した。6 評価観点の全てにおいて、前半群と後半群の間に有意な差は見られなかった ( $p=0.207 \sim 0.955$ )。

#### 5) 職位による差異

全試験の各評価観点の成績評価を評価者の職位別に算出した。観点 B ( $p=0.011$ )、C ( $p=0.001$ )、E ( $p=0.005$ )、概略評価 ( $p<0.001$ ) に有意な差が見られた。概略評価を含む 3 つの評価観点で准教授の評価が低い傾向が見られた。

### 4. 結語

1) 評価者の属性による評価点数の差異は認められなかったが、ステーション内 2 評価者間において、概略評価を含む 6 の評価観点の  $\kappa$  係数は poor~Fair、一致率は 41~49% と小さかった。特に「観点 C : 身体診察」と「観点 E : 臨床推論」の評価観点においては、合格ラインである 4 点をまたいでの評価の不一致が多かった。

2) 学生の合否に直結するような評価の不一致が見られており、これは学生にとって大きな不利益となる。試験の実施方法や評価方法をさらに精練するとともに、試験後に統計的な評価の確認を行い、必要時には評価の見直しができるシステムを構築することが重要である。

1. 学位審査の要旨および担当者

学位番号甲第 721 号	氏 名	高 山 充
学位審査担当者	主 査	佐 藤 二 美
	副 査	中 村 陽 一
	副 査	堀 裕 一
	副 査	西 脇 祐 司
	副 査	三 上 哲 夫

学位論文の審査結果の要旨 :

Post-CC OSCE は 2020 年度から全国医学部に正式導入され、医師国家試験への実技試験の導入に向けた試みとも位置付けられるものである。しかしその合否は実施大学で判定していることから、評価の信頼性をいかに担保するかが重要である。本研究では、2020 年度に実施された Post-CC OSCE3 課題 6 系列において、1 ステーションに配置された 2 名の評価担当者の評価点数の分析を行った。計 36 名の評価者について、評価観点 6 項目 (A 患者への配慮/コミュニケーション、B 医療面接、C 診断仮説に基づいた身体診察、D 症例プレゼンテーション、E 臨床推論、および概略評価) の評価点数 (6 段階評価) について、ステーション内の 2 評価者間、6 系列間、実施時間帯、評価者の職位による差異を解析した。ステーション内の 2 評価者間では、一致率は「身体診察」で最も低く、「概略評価」で最も高かった。6 段階評価を合格ボーダーである 4 点以上と 4 点未満の 2 群に分けた一致率では、最も高い「患者への配慮」では 95.5%であったが、「身体診察」は 65.3%で最も低く、多くのステーション内で 2 評価者の合否判定が一致しなかった。6 系列間では、評価の平均点では有意差がなかったが、概略評価点を合格ボーダーの 4 点以上と 4 点未満の 2 群に分けたところ、合格率は最も高い系列で 90.0%、低い系列では 36.8%で、系列間の合格率に有意差があった。実施時間帯による有意差はなく、評価者の職位では、准教授の評価が低かった。評価の信頼性の点で、特に合格ラインをまたいでの評価の不一致は学生の不利益に通じるもので、評価の一致率を上げる方策の導入、必要時の評価実施後の再確認に関するシステム構築の必要性が示唆された。

学位審査会は、2022年2月9日17時から、主査佐藤、副査中村、堀 (書面審査)、西脇、三上の参加のもと、実施された。審査委員からは、成績を扱う研究であるため、特に学生への研究同意はどのような形で取り、どのような理解に基づいてなされたかや、身体診察で評価者間の差が大きかった理由は何か、特定のペアで評価不一致があったのではないかと、解析結果とその解釈に関する質問のみならず、評価の信頼性・妥当性の違い、今後の研究予定など、医学教育学の知識や将来の方向性を含めた多岐にわたる質問がなされた。申請者はこれらの全ての質問に対し、的確に返答した。以上より、本論文は、Post-CC OSCE 評価の信頼性の担保に向けての有意義な研究であり、審査委員全員一致で学位に値する論文と認め、医学研究科委員会に報告することとして、学位審査会を終了した。