

医学教育 2010, 41(5): 325~335

原著—探索的研究

客観的指導能力評価 (Objective Structured Teaching Evaluation : OSTE) の実施と分析

浜田 久之^{*1} シェリー・リー^{*2} アバス・ガバンラスル^{*2}
近藤 久義^{*3} 江崎 宏典^{*4} 大谷 尚^{*5}
ヘレン P バティアー^{*2}

要旨:

日本では、近年臨床研修指導医に対する評価の必要性も高まっている。欧米では、標準化学生、複数のステーション、ビデオによる記録、観察スコアによる複数の評価者等を基本とした Objective Structured Teaching Evaluation (OSTE: 客観的指導能力評価) が開発され指導医の評価方法のひとつになっている。今回我々は、日本での OSTE を実施し分析を行った。

論点

- 1) 5つのステーション、標準化研修医、ビデオによる記録、7名の評価者による OSTE に 10名の臨床研修指導医が参加した。
- 2) チェックリストと5段階スケールによる指導医の指導力評価の他に、チェックリストの信頼性と妥当性を評価した。指導医の背景因子による分析もおこなった。
- 3) 「指導医講習会参加歴有り」、「教育歴5年以上」、「非内科医」の因子が、チェックリスト得点とスケール得点で有意に高かった。評価者間での個人差はなかった。
- 4) チェックリストに関する一般化可能性係数は 0.81、信頼度指数は 0.83 であった。チェックリスト得点とスケール得点との相関係数は 0.8 であった。
- 5) 参加者に関するバイアス等があったが、OSTE による客観的な指導能力評価は可能と考えられる。今後、さらなる研究が必要である。

キーワード: 指導医, 指導技術, 客観評価, 臨床研修制度

Enforcement and analysis of the Objective Structured Teaching Evaluation

Hisayuki HAMADA^{*1} Shirley LEE^{*2} Abbas GHAVAM-RASSOUL^{*2}
Hisayoshi KONDO^{*3} Hironori EZAKI^{*4} Takashi OTANI^{*5}
Helen P BATTY^{*2}

*1 長崎大学病院医師育成キャリア支援室, Career Development Center for Medical Doctors, Nagasaki University Hospital [〒852-8501 長崎市坂本1丁目7番1号]

*2 トロント大学医学部地域家庭医療科/FDセンター, Family & Community Medicine/Centre for Faculty Development of Medicine, University of Toronto

*3 長崎大学大学院医歯薬学総合研究科附属原爆後障害医療研究施設, Nagasaki University Graduate School of Biomedical Sciences Atomic Bomb Disease Institute

*4 国立病院機構長崎医療センター/教育センター, National Organization Hospital Nagasaki Medical Center

*5 名古屋大学大学院教育発達科学研究科, Information Science in Education Graduate School of Education & Human Development Nagoya University

受付: 2010年4月19日, 受理: 2010年8月6日

Abstracts

In Japan, awareness has increased in recent years of the importance of evaluating clinical educators. In Europe and North America, the Objective Structured Teaching Evaluation (OSTE), which employs standardized students, multiple stations, video recording, and scoring by multiple observers, is used to evaluate clinical educators. We report on the implementation of an OSTE in Japan.

- 1) Ten clinician-educator physicians participated in the OSTE, which comprised 5 stations and included standardized residents. The stations were video-recorded, and the educators were assessed by 7 different evaluators.
- 2) The educators were evaluated with a checklist and a 5-point scale. We assessed the reliability and validity of the checklist and analyzed the background characteristics of the clinician educators.
- 3) The factors most closely associated with high ratings on the checklist and the 5-point scale were: having a history of attendance at a seminar for clinician-educators, having greater than 5 years experience as an educator, and not being an internist. There was no interobserver variability among the evaluators.
- 4) The generalizability of the checklist was 0.81, and its reliability index was 0.83. The correlation coefficient between the total scale score and the checklist score was 0.8.
- 5) Although biases by participants were identified, our project suggests that the OSTE could be used in Japan to objectively evaluate the teaching skills of clinician-educators. Further research on the OSTE in Japan is warranted.

Key words: teacher's physicians, teaching skills, objective evaluation, clinical training systems

はじめに

日本では、臨床研修の義務化により研修医に対する評価が義務化され、指導医に対する評価の必要性も重要視されてきている¹⁾。欧米ではすでに1970年代より臨床医の指導能力を評価しようとする研究は始まっており、1990年代には、標準化学生、複数のステーション、ビデオによる記録、観察スコアによる複数の評価者等を基本とした Objective Structured Teaching Evaluation (OSTE: 客観的指導能力評価) が開発された。我々は、欧米における OSTE 研究の歴史的な流れと Faculty development 等で応用されている現状を報告し、日本においても OSTE 研究の必要性を述べた²⁾。今回我々は、欧米の文献や実施例を参考に²⁾、日本で OSTE を設計し実施したので報告する。

目的

本研究の目的は、独自に設計した OSTE を施行し、その信頼性及び妥当性の有無を統計学的に検討することである。本研究計画は厚生労働省が示す臨床研究に関する倫理指針に則り、計画書を作成し国立病院機構長崎医療センター倫理委員会の審査を受け受理された。

方法

1) 設計手順

本 OSTE は、Objective Structured Clinical Examination (OSCE: 客観的臨床能力試験)³⁾ 作成と同様の手法で、3段階の過程を経て設計された。最初に海外の文献分析と実施例を参考に全体の概要をカナダ人研究者らと作成した(表1)。次に、臨床研修指導医(以下、指導医)の指導能力に関する評価項目、シナリオ、チェックリスト(CL)、実施概要(ステーション(ST)数、パフォーマンス時間等)の詳細を日本人研究者が中心となり作成した。最後に、本研究に関与してない研修医及び指導医数名、標準化研修医に意見を求め数回修正した。

2) 設計

指導医の指導能力に対する評価指標は、指導医ガイドライン¹⁾、厚生省の指針に沿った指導医講習会^{4,5)}、Cleveland Clinic Foundation (CCF)⁶⁾、Stanford Faculty Development Program (SFDP)⁷⁾ による評価指標等を参考に5つの指標を作成した(研修医が学習しやすい雰囲気作りをする、研修医を指導するための教育的な知識を持っている、研修医を指導するためのフィードバック技術を持っている、医学的知識・技術を教えることができる、研修医に対してロールモデルとなるような態度を示す努力をする)。

ST数は5つ、指導医が研修医に指導する(パ

表1 Objective Structured Teaching Evaluation (OSTE) 概要

目的	臨床研修指導医の指導能力を客観的に評価する OSTE を施行し, その信頼性, 妥当性を検討する.		
準備	指導医に対する評価指標を海外の指標等を参考に作成し, 臨床教育現場を再現した5つのシナリオと48項目のチェックリスト (CL) を作成.		
実施概要	5つのステーション (ST) を指導医が OSCE と同様の方法で回る. 各 ST で指導医は3分間の準備時間を与えられ, 標準化研修医に対して7分間の指導を行い, ビデオテープにより記録される.		
シナリオ	第1ST: 指導医の診療科研修についてのオリエンテーション 第2ST: 新入院患者チャートでの研修医のプレゼンテーション 第3ST: 病棟でのオーダーを忘れた研修医への指導 第4ST: 静脈確保 (ルート確保) の教え方 第5ST: うつ状態が疑われる研修医への対応 (メンタルフォロー)		
参加者	臨床研修指導医10名, 演技訓練を受けた標準化研修医6名, 評価者7名, ビデオ撮影者7名, 運営スタッフ3名, 模擬患者1名		
評価する項目	評価する人	評価時期	評価方法
シナリオの現実性	標準化研修医, 指導医	各 ST の指導医のパフォーマンス終了直後	各 ST 毎に4段階 Likert Scale
	評価者	ビデオ録画観察中 ^(*)	
研修医の演技	標準化研修医	各 ST の指導医のパフォーマンス終了直後	各 ST 毎に4段階 Likert Scale
	指導医		
	評価者	ビデオ録画観察中	
OSTE における指導医の指導力	標準化研修医	各 ST の指導医のパフォーマンス終了直後	各 ST に毎に CL 及び5段階 Likert Scale
	評価者	ビデオ録画観察中	
本研究に関して	参加者全員	自由記載	記述
OSTE に関する感想	指導医	1週間後	聞き取り

*評価者は, ビデオ録画された指導医の全 ST のパフォーマンスを観察する.

パフォーマンス) 時間は7分間, 移動及び次の ST の準備時間は3分間とした. 5つのシナリオ作成は, 上記の5つの評価指標に沿って行われ, 場面の設定は, 日常的に遭遇する初期研修医と指導医の教育現場を再現するように努めた (表2).

チェックリストの作成にあたっては, 各々のリストが, 上記の5つの評価指標, CCF, SFDP のどの項目に対応するかを吟味し, 48項目作成された (表3).

本 OSTE にて統計的に評価する項目は, シナリオの現実性, 研修医の演技, OSTE における指導医の指導力の3項目とした (表1). シナリオの現実性, 研修医の演技については, 標準化研修医, 指導医, 評価者により評価された. 指導医の指導力評価については, 評価者は全指導医の全 ST のパフォーマンス

を評価するが, 標準化研修医は各自の ST のみを評価した. アンケートにて評価する項目は, 本研究の運営に関する意見と OSTE に参加した感想とした.

3) 参加者

被評価者 (指導医) は, 長崎医療センターで実際に研修医を指導している医師30名に依頼し, 同意が得られた10名とした. 性別, 診療科等の11項目の背景因子を調査した (表4-1). 評価者については, 本研究者より3名 (評価者1, 5, 6), 外部より4名を選び, 7項目の背景因子を調査した (表4-2). 評価者は事前にチェックリストを熟読し, 評価前に30分間の評価方法の説明を受けた. 標準化研修医は, 長崎医療センターの研修医40名の中から10名を任

表2 OSTE シナリオ概要

(指導医は、STの中で指示書を読み3分間どのように指導するか準備をした後、標準化研修医へ対して7分間指導する)

ST	内 容
1: オリエンテーション	研修医 A は指導医の診療科を3カ月間研修する予定である。指導医がオリエンテーションをするために研修医 A を呼び出し、部屋へ研修医 A が入ってくる場面から指導医のパフォーマンスが始まる。研修医 A の当該診療科に対するモチベーションは標準的であり、7分間の中で2つの質問をするように設定した。質問1は「学会発表をしたい」、質問2は「研修期間中の休日に友達の結婚式に出席したい」という希望である。
2: 新入院患者チャート	研修医 B が、貧血を主訴に入院した新患についてカンファランスで3分間プレゼンする(ビデオ録画されたプレゼンがパソコン上で再生される)。指導医のパフォーマンスは、プレゼン終了後から始まる。研修医 B は標準的な能力であり、指導医の質問には短く端的に答えるように設定した。さらに、研修医 B と同僚の研修医を1名配備した。同僚研修医から指導医への質問を2つ用意した。質問1は「この患者を入院させる必要があったのか?」、質問2は「この患者は輸血が必要か?」という点である。
3: オーダー忘れた研修医への指導	研修医 C はやる気はあるが仕事の効率が悪く、現在患者を7名担当し能力の限界に近い。さらに過去の研修事例から「採血は毎日するべきではない」と思い込んでいるとした。次のセリフ後よりパフォーマンスは開始される。研修医「おはようございます。いま、ちょっとよろしいですか? 昨日、胃潰瘍で入院した長崎さんですが、昨晩はバイタルも安定していて、吐血もなく特に問題ありませんでした。とても元気で、お腹がすいたといっていますが、食事の方は……どうすればいいのでしょうか?」指導医「本日のヘモグロビンのデータは?」研修医「あ〜、申し訳ありません、忘れていました〜」この後の会話は制約しない。
4: 静脈確保の教え方	研修医 D は不器用であるが、静脈確保について他の指導医の手技を5-6回見学し、さらに研修医同士で5回練習をしている。本日、初めてのルート確保臨むという設定である。処置室で点滴セットを研修医 D が準備し終えた段階で、指導医が登場しパフォーマンスは開始される。隣には病室があり模擬患者がベッドに臥床している(腕についてはシミュレーター)。研修医 D には、ルート確保を1回失敗するように設定している。
5: メンタルフォロー	指導医は、自分が指導をしている研修医 E が、元気がなく落ち込んでいるという情報を同僚研修医と病棟師長より得て、面談を行う。研修医 E は「自分の思うように仕事ができない。自分はミスが多い。同僚の研修医と比べて自分は能力が低いのではないか。」と思い込み自信をなくしている。また他の指導医に学会発表するようにいわれ、少し焦っている。さらに1か月前に担当した患者さんが急変して亡くなり、それを引きづっている。時々眠れないことがある。ここ1か月土日を休んでない。パフォーマンスは、研修医が指導医の部屋へ入ってきた時から始まる。7分間面談を行い研修医が外へ出た後に指導医は、ビデオカメラに向かって、今後指導医としてどのような行動をとるのかを約20秒間〜60秒述べる。

意に抽出し、同意が得られた6名(各ST1名、第2STのみ2名)に依頼した。標準化研修医は、演技について個別訓練(60分×2回)をおこない、全員で予行演習(1回)を行った。演技については、演劇の経験のある医師が指導した。模擬患者1名(第4ST)、ビデオ撮影者7名(各ST1名、機材調整係2名)、運営スタッフ3名(タイムキーパー1名、指導医の誘導者1名、統括者1名)を配備した。

4) 評価方法及び解析方法

(シナリオの現実性)

4段階 Likert Scale (「非常に現実的」4点、「ほぼ

現実的」3点、「あまり現実的でない」2点、「非現実的」1点)で行い、判定者間、ST間の差を分散分析により検討した。

(研修医の演技)

上記と同様に4段階 Likert Scale で行った。

(指導医の指導能力)

チェックリストと5段階 Likert Scale の2種類の評価方法を用いた。チェックリストに関しては、各質問内容に対して評価者が「できた」と判断した場合を1点(第3STのみ1.25点)、「できなかった」と判断した場合を0点とし、各STごとの合計点を用

表3. 各ステーション (ST) のチェックリスト概要

	1ST オリエンテーション	2ST 新患チャート	3ST 'オーダー忘れ'への対応	4ST ルート確保	5ST メンタルフォロー
1	自己紹介 ^(*)	教育的態度	教育的態度	患者への配慮	呼び出しの理由の説明
2	ラポール形成	丁寧な言葉づかい	丁寧な言葉づかい	研修医の技量の確認	非評価的な面談である旨の説明
3	オリエンテーションの説明	フィードバック技術の使用	フィードバック技術の使用	手技のポイントの説明	悩みを聞きだす雰囲気作り
4	ニーズアセスメント	研修医の尊厳への配慮	研修医の尊厳への配慮	適切な見守り方	話を中断しない
5	双方向的な会話	意見の聴取	忘却の理由についての質問	手技中のアドバイス	悩みを解決する姿勢
6	研修医からの質問の受諾	誤った知識の訂正	誤った知識の訂正	明確なアドバイス	アドバイスの具体性
7	到達目標の提示	プレゼンテーション指導	明確な指示	失敗を叱責	休養に関する質問
8	研修医の役割提示	EBMに基づくコメント	まとめのフィードバック	手技の後のアドバイス	体調に関する質問
9	研修の評価方法提示	明確な指示		フィードバック技術の使用	今後の適切な方針
10	スケジュールの説明	まとめのフィードバック		理論的なフィードバック	プライバシーへの配慮

*チェックリスト例：指導医は、研修医へ対して自己紹介をしたか？ Yes No

い、評価者間により差があるか否かを分散分析により検討した。次に、各STにおけるチェックリスト評価点数の合計（チェックリスト評価総点数）が、判定者（評価者と標準研修医）または指導医で差があるか否かを、分散分析およびTukeyの多重比較により検討した。また、チェックリスト評価総点数に影響をおよぼす要因として、指導医の属性では診療科（内科か内科以外）、年齢（40歳未満か40歳以上）、医師経験年数（15年未満か15年以上）、研修医教育歴（5年未満か5年以上）、年間指導研修医数（5人未満か5人以上）、指導医講習会受講の有無、指導に対する自信の有無、指導上の悩みの頻度（時々か頻繁）を、専門評価者の属性では職業（医師か医師以外）、性別（男か女）、OSCE講習会受講の有無、指導医講習会受講の有無を考え、評価者間または指導医間の個人差を考慮して分散分析により検討した。

さらに、各STで指導医の指導能力に対する5段階Likert Scale（「非常に良い指導の仕方である。他の

指導医の模範となる指導方法である」5点、「平均以上の指導の仕方である」4点、「平均的な指導の仕方である」3点、「指導の仕方に問題が多い」2点、「指導の仕方に問題が多く、研修医の研修を阻害する」1点）による評価を行い、チェックリストと同様の解析を行うと同時に、両者の総点数の間の相関についても検討した。

チェックリストの信頼性に関しては、一般化可能性理論 (Generalizability theory) を用いて検討した。本研究では、測定対象は指導医の指導能力の測定（チェックリスト評価得点の総点数）を規定する要因として、信頼度指数 (Dependability Coefficient: ϕ -指数) と一般化可能性係数 (Generalizability Coefficient: G-係数) を求めた。

結果

OSTE実施日の1週間前に、標準化研修医及び運営スタッフによる入念な予行演習が行われ、実施日

表4-1 被評価者(指導医)背景一覧

指導医 番号	性別	診療科	年齢	臨床 年数	研修医 への指 導経験	年間指 導研修 医数	指導医 講習会 受講	研修医指導 が好きか	研修医指導に 自信があるか	研修医指導 で悩んだ経 験はあるか	事前筆記 テスト *(10点満点)
1	男性	外科	40代	20年 以上	10-15 年	10-15 人	○参加	どちらかと いえば好き	ある	頻繁	9
2	男性	病理	50代	20年 以上	5-10 年	1-5人	○	どちらかと いえば好き	ある	時々	6
3	男性	内科	30代	10-15 年	5年以 下	15-20 人	×	どちらかと いえば好き	ある	時々	5
4	男性	内科	30代	10-15 年	5-10 年	20人 以上	○	どちらかと いえば好き	ある	頻繁	4
5	男性	内科	20代	5年 以下	5年 以下	15-20 人	×	好き	ある	時々	5
6	男性	小児科	30代	5-10 年	5-10 年	5-10 人	○	どちらかと いえば好き	ない	時々	3
7	男性	内科	30代	5-10 年	5年 以下	1-5人	○	あまり好き でない	ない	時々	8
8	男性	内科	40代	15-20 年	5年 以下	0人	×	どちらかと いえば好き	ない	頻繁	10
9	男性	内科	40代	10-15 年	5年 以下	1-5人	○	どちらかと いえば好き	ない	頻繁	9
10	男性	救命	40代	20年 以上	15-20 年	20人 以上	○	好き	ある	時々	8

*厚労省の指針に沿った指導医講習会時に用いられるプレテストをOSTE前に施行

表4-2 評価者背景

	年齢	性別	医者/医者以外	研修指導歴	年間指導研修医数	OSCE 評価者 研修受講歴	指導医講習会 受講歴
評価者1	50代	男性	医師	20年以上	20名	あり	あり
評価者2	40代	女性	教育専任看護師	0	0	なし	あり
評価者3	40代	男性	医師	10-20年	30名	なし	あり
評価者4	30代	女性	医師	5-10年	30名	あり	あり
評価者5	50代	男性	教育学部教員	0	0	なし	なし
評価者6	40代	男性	医師	10-20年	10名	あり	なし
評価者7	20代	女性	臨床研修担当事務職	なし	0	なし	なし

は予定通り約2時間で終了した。1カ月後に、評価者は、ビデオ録画された全指導医の全STのパフォーマンスを観察し評価した。

(シナリオの現実性)

指導医による評価の平均値は3.34、評価者は3.17

で両者間に有意な差はなかった。標準化研修医は3.00であり、指導医や専門評価者に比べてやや評価が低かった。STごとの指導医と専門評価者を合わせた平均得点は、第1STから順に3.60, 3.21, 3.29, 2.94, 3.29で、第4ブースの評価がやや低かったが、ST間

表 5-1 チェックリスト (50 点満点) による評価

	標準化 研修医	評価者								全平均
		*R-1	R-2	R-3	R-4	R-5	R-6	R-7	平均	
指導医-1	39.00	31.50	29.25	33.25	34.25	26.75	33.75	41.75	32.93	33.69
指導医-2	35.75	36.75	38.00	32.50	34.75	38.75	37.75	30.50	35.57	35.59
指導医-3	22.00	18.25	20.00	12.25	23.50	19.75	20.00	10.50	17.75	18.28
指導医-4	43.00	39.00	33.50	37.50	37.75	38.75	33.75	30.00	35.75	36.66
指導医-5	37.00	36.75	25.00	27.25	32.50	35.75	32.75	32.00	31.71	32.38
指導医-6	37.75	36.50	28.00	38.00	35.75	36.75	38.75	40.50	36.32	36.50
指導医-7	27.00	26.25	29.25	27.25	29.50	29.25	30.50	24.75	28.11	27.97
指導医-8	33.50	31.00	29.00	25.75	33.75	29.25	32.25	27.50	29.79	30.25
指導医-9	36.75	34.75	31.50	31.75	34.75	35.75	35.75	32.50	33.82	34.19
指導医-10	40.50	37.50	38.25	32.25	37.50	42.25	39.50	37.00	37.75	38.09
平均	35.23	32.83	30.18	29.78	33.40	33.30	33.48	30.70	31.95	32.36

*R 評価者

表 5-2 5 段階スケール (25 点満点) による評価

	標準化 研修医	全平均								全平均
		*R-1	R-2	R-3	R-4	R-5	R-6	R-7	平均	
指導医-1	19	17	16	20	16	15	17	19	17.1	17.4
指導医-2	17	18	20	20	16	19	18	17	18.3	18.1
指導医-3	15	16	14	13	13	14	11	11	13.1	13.4
指導医-4	17	22	18	23	18	18	16	19	19.1	18.9
指導医-5	20	18	14	16	15	18	17	18	16.6	17.0
指導医-6	20	20	15	20	18	20	19	19	18.7	18.9
指導医-7	15	16	17	15	17	16	16	13	15.7	15.6
指導医-8	20	16	14	12	14	15	16	15	14.6	15.3
指導医-9	19	20	19	19	18	21	19	19	19.3	19.3
指導医-10	20	22	19	18	18	22	20	19	19.7	19.8
平均	18.2	18.5	16.6	17.6	16.3	17.8	16.9	16.9	17.2	17.4

*R 評価者

に有意な差はなかった。総平均は 3.27 であった。

(研修医の演技)

指導医の評価得点の平均値は 3.54, 評価者は 3.46 で両者間に有意な差はなかった。標準化研修医全体は 2.75 であり, 指導医や専門評価者に比べてやや評価が低かった。ST ごとの指導医と専門評価者を合わせた平均得点は, 第 1ST から順に 3.46, 3.52, 3.47, 3.37, 3.71 で ST 間に有意な差はなかった。総平均は 3.50 であった。

(指導医の指導能力)

指導医に対するチェックリスト評価得点および 5 段階スケールによる得点の判定者 (評価者と標準化研修医) および指導医別の平均値 (表 5-1 および表 5-2) で, 最も高く判定した者は 35.23, 最も低く判定した者は 29.78 であったが, 分散分析の結果は有意ではなく, 判定者間 (研修医と評価者も含めて) の個人差はなかった。指導医ごとの全判定者のチェックリスト評価得点の平均値は, 最も高い指導医が

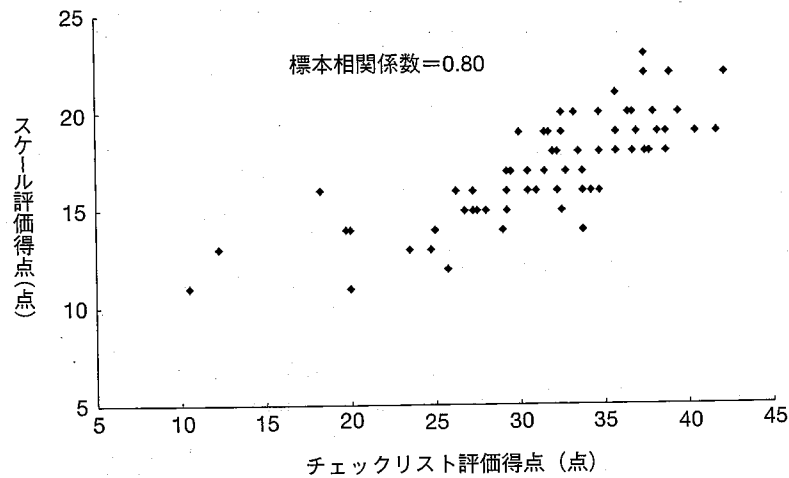


図1 チェックリスト評価得点とスケール評価得点の相関

38.09であったのに対して最も低い指導医は18.28と大きくばらついており、分散分析の結果も1%水準で有意であった。この結果に基づき、指導医ごとのチェックリスト評価得点の平均値に関する多重比較を実施したところ、最下位の指導医と他の全ての指導医の差および上位3名の指導医と下位5名の指導医における差は統計的に有意であった(表5-1)。一方、スケール評価得点の判定者ごとの平均値は最も高い判定者で18.5、最も低い判定者で16.3であったが、分散分析の結果は有意ではなく、チェックリストの総点数と同様に判定者間(研修医と専門評価者も含めて)の個人差はなかった。指導医ごとの平均値は、最も高い指導医が19.8であったのに対して最も低い指導医は13.4とやはり大きくばらついており、分散分析の結果も1%水準で有意であり、多重比較を行ったところ、総点数の上位4名の指導医と下位3名の指導医における差が統計的に有意であった。

指導医に対するチェックリスト評価得点とスケールによる評価得点は相関があり(図1)、統計的にも(標本相関係数0.80)有意であった。

指導医の属性の中で属性の違いによるチェックリスト評価得点で統計的に有意であったのは、診療科、年齢、医師経験年数、研修医教育歴、指導医講習会の受講の有無の5項目であった(表6-1上段)。一方、スケール評価得点で有意であったのは、診療科、研修医教育歴、指導医講習会の受講の有無の3項目であった(表6-1下段)。

専門評価者の属性の違いによるチェックリスト得

点で統計的に有意であったのは、OSCE講習会受講の有無のみであった(表6-2上段)。一方、スケール評価得点で有意であったのは、性別のみであった(表6-2下段)。

チェックリストに関する一般化可能性係数は0.81、信頼度指数は0.83であった。

考 察

日本の医学教育研究において評価指標等を定めて指導医を評価する研究は少ないが、田中ら⁸⁾は研修医からの聞き取り調査にて指導医の評価指標を示し、福井ら⁹⁾は日本における指導医評価の問題点や海外で頻用されている指標(CCFの15項目、SFDPの25指標)を紹介し分析している。これらの指標は、基本的には成人教育理論に基づいて作成されており大きな差はない。我々(日本人、カナダ人)が作成した評価指標はチェックリストを作成するための基礎となるもので、日本の臨床研修指導医の置かれている立場を考慮し上記の指標を大きく網羅するようにした。チェックリストの48項目の作成にあたっては、CCF指標、SFDP指標、本研究指標を網羅するように作成された。我々のチェックリスト48項目中15項目がCCF指標の「学習者に適した雰囲気をつくる」とSFDP指標の「学習するための雰囲気」と対応している。STの設定が、各指導医が異なる5名の研修医を指導することになっているために、雰囲気作りが重視されるチェックリストとなった。さらに5項目がCCF、SFDPの「フィードバック」に対応している。短時間で指導力を評価する

表 6-1
指導医の属性とチェックリスト評価得点 (上段)
指導医の属性とスケール評価得点 (下段)

診療科	年齢		医師経験年数		研修医教育歴		年間研修医指導数		指導医講習会受講		指導に対する自信		指導上の悩み		
	非内科	40歳未満	40歳以上	15年未満	15年以上	5年未満	5年以上	5人未満	5人以上	なし	あり	なし	あり	時々	頻繁
29.5	35.6	29.9	34.0	30.6	34.0	28.2	35.7	31.8	32.0	26.4	34.3	32.0	31.9	33.1	31.2
±	±	±	±	±	±	±	±	±	±	±	±	±	±	±	±
6.7	4.0	7.7	4.1	7.2	4.5	6.5	3.8	3.9	7.8	7.4	4.3	4.2	7.7	3.8	7.7
P<0.01		P<0.01		P<0.05		P<0.01		N.S.		P<0.01		N.S.		N.S.	
16.4	18.5	16.7	17.8	17.1	17.4	15.9	18.6	17.0	17.4	14.8	18.3	17.1	17.3	17.0	17.5
±	±	±	±	±	±	±	±	±	±	±	±	±	±	±	±
2.8	1.9	2.8	2.4	2.8	2.4	2.5	2.0	2.3	2.8	2.1	2.1	2.4	2.8	2.7	2.5
P<0.01		N.S.		N.S.		P<0.01		N.S.		P<0.01		N.S.		N.S.	

N.S.: not significant

表 6-2
評価者属性とチェックリスト評価得点 (上段)
評価者属性とスケール評価得点 (下段)

職種		性別		OSCE 講習会受講		指導医講習会受講	
医師	医師以外	男	女	なし	あり	なし	あり
32.4±6.0	31.4±7.1	32.3±6.5	31.4±6.5	31.0±7.1	33.2±5.3	32.5±7.1	31.5±6.0
N.S.		N.S.		P<0.01		N.S.	
17.3±2.7	17.1±2.6	17.7±2.8	16.6±2.3	17.2±2.8	17.2±2.4	17.2±2.7	17.3±2.6
N.S.		P<0.01		N.S.		N.S.	

N.S.: not significant

OSTE においては、フィードバックの有無や方法が重視される¹⁰⁾。唯一網羅できなかった項目は CCF の「適切な医療資源の利用方法や費用について考慮するときの原則を教える」という項目である。今後、他職種、他診療科、他病院へ紹介する場面や薬剤等のコストについて教える場面設定をしたシナリオを作成すれば網羅することができるであろう。設計段階での最も大きな問題点は、参加者の偏りであった。自由意思で参加することが原則であるので、評価を受けることを厭わない教育熱心な指導医が必然的に多くなり均一集団となった可能性がある。さらに、同一施設の標準化研修医、指導医、評価者が多くバイアスがかかっていることは否定できない。

結果に関して考察する。シナリオの現実性は、統計学的には ST 間で有意差はなかったが、第 4ST では得点が低かった。原因として、手技に関する ST では医師以外の評価者にとっては評価が難しいこと

や、シミュレーターを用いるため現実性に乏しかったことが考えられる。研修医の演技に対する評価でも第 4ST が低かった。不器用な研修医を演じ点滴手技を一度失敗するシナリオとしていたが、演じた標準化研修医によると時間的にすべてを演じるのは困難であったと述べている。手技の指導に関する設定は、7 分間という時間では困難である可能性がある。また、全体的にシナリオの現実性に対する評価は標準化研修医の演技に対する評価よりも低い。統計学的な証明はできないが、シナリオの如何にかかわらず、標準化研修医の演技次第で OSTE の出来不出来の印象が決まる可能性がある。標準化研修医の演技に関してどれくらい訓練すれば良いか明確な根拠はない。海外の文献²⁾では、合計数時間～10 時間前後であり、OSCE の模擬患者の育成プログラムより短いようである (岡山 SP 研究会¹¹⁾では、6 カ月以上にわたり 1 回 2.5 時間×10 回)。OSCE の模擬患者は、

一般の人を患者役として様々なシナリオに合わせて養成する機会が多いが、OSTEの標準化研修医（あるいは医学生）は医学、医療行為や研修内容を知っていることが前提条件とされるので、医療者でありかつ研修を経験した研修医や若手医師が適任である。海外の文献でも医療者以外を標準化研修医としている論文は少なく²⁾、シナリオのコンセプトを理解し演技を厭わない研修医や若手医師を養成する方法が最良と思われる。

指導医の指導能力に関するチェックリストでの評価得点では、上位3名と下位5名の指導医、総括評価得点では上位4名と下位3名の指導医は統計学的に有意差があった。分散分析の結果、判定者間（研修医と評価者も含めて）の個人差はなかった。これにより本OSTEでの指導医の指導力に対する高低の判断の客観性はある程度存在するものと考えられる。さらに客観性を高めるためには、チェックリストの判定基準を厳密に決め、評価者に対して講習を数回行う必要がある。

指導医の属性と評価との関連については、「非内科系医師」、「年齢40歳以上」、「医師経験年数」、「教育歴が5年以上」、「指導医講習会受講歴」がチェックリストで高得点を得た。「非内科系医師」、「教育歴が5年以上」、「指導医講習会受講歴」がスケールで高得点を得た。「非内科系」医師が両評価方法で有意差を持って高得点であった理由は明らかでないが、内科系と非内科系において、研修医と接する機会にほとんど差がないこと（本施設では、内科、外科、小児科、救命科が必修）とシナリオに特定の科が不利になる要素がなかったことが考えられる。海外の文献の多くは、OSTEは診療科別に行われておりシナリオの内容も診療科の特色を出している。本OSTEはどの診療科でも起こりうることを想定としてシナリオを作成したが、特定の診療科を対象としたOSTEも次の段階で考慮すべきであろう。「年齢」、「医師経験年数」がチェックリスト得点では有意差があったがスケール得点と関係しなかった。「教育歴が5年以上」が両評価で有意差がでていることより、研修医教育にあたる医師は、早くから教育経験を積むことが望まれる。新臨床研修制度では屋根瓦方式の研修体制が推奨され⁹⁾、若い医師が研修医を教えることが多くなっている。海外では、研修医の頃より教育者

として育てる、Residents as teachers (RAT) プログラム¹²⁾がありOSTEを用いて研修医を訓練している。日本でもこのような試みや研究が必要であろう。「指導医講習会受講」については、両評価方法で有意差があった。指導医講習会はプログラムプランニングを中心に構成されているが、「良い指導医、悪い指導医」を演じるロールプレイングのセッション等への参加経験などが直接的に本OSTEへ影響し、受講者に有利に働いたものと考えられる。また、指導医講習会で使用されるプレテストは医学教育の基礎知識を問うものであるが、この点数とOSTEの得点には相関がなかった。さらに、「指導に対する自信」や「指導上の悩み」という属性に関しても有意差がないことは、指導力は講習会等で指導技術を習得し、実際の指導の中で後天的に獲得できるものではないかと推測される。

スケール得点（25点満点）とチェックリスト総得点（45点満点）の関係は、相関係数0.8であり統計学的に相関が認められた。チェックリストの評価と他の方法での評価が相関していることはチェックリストの妥当性の裏付けのひとつとなっている。妥当性をさらに高めるためには、参加した指導医へ介入試験（指導技法を教える講習会等を受講させる）を行い、再度OSTEを施行して介入前後の得点を比較することが望まれる¹³⁾。チェックリストに関する信頼性は、OSCEの望ましいG係数の値は0.8以上であり¹⁴⁾、本研究は一般化係数0.83、信頼指数0.81で信頼性は保たれていると判断される。G係数を高めるためには評価者よりブース数を増やすことが効率的であることは明らかであるが（図2）、大幅な信頼性の上昇は期待できず、現状の5ブース、7人評価者で十分と判断する。

指導医に対するアンケート調査では、「評価されることにストレスは感じたものの自分の指導の仕方をふり返ることができた」「OSTEは指導医の資格試験として受け入れられる可能性は少ないが、FDのひとつとして受け入れられる」との声が多かった。今後、診療科別、指導医の教育経験年数別、研修区分別（外来、入院）、対象別（医学生、研修医）等の多様なOSTEの研究開発が日本においても必要とされるであろう。

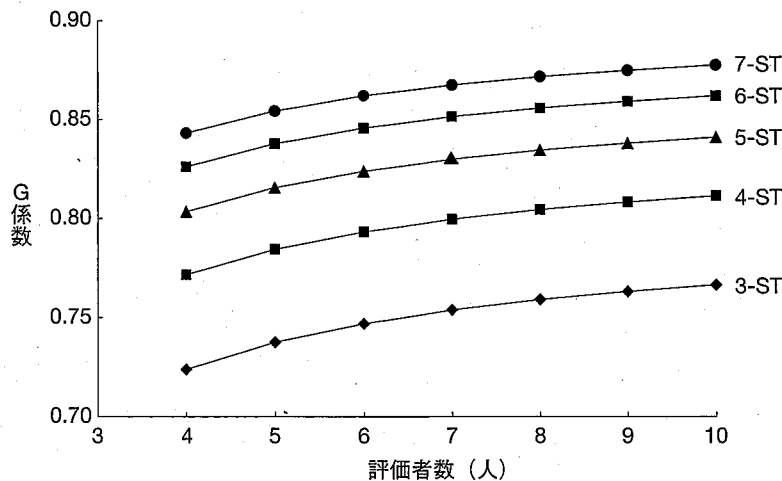


図2 ステーション (ST) 数と評価者数の違いによる G 係数の変化

告知

本研究は財団法人ファイザーヘルスリサーチ振興財団助成金 (平成 19 年度) による国際共同研究の一環として行われた。Thank you to Dr. Samuel Lapalme-Remis, McGill University Faculty of Medicine and BA. Naomi H for providing information related to this study and reviewing of the manuscript.

文献

- 1) 新医師臨床研修制度における指導ガイドライン <http://www.niph.go.jp/soshiki/jinzai/kenshu-gl/index.html> (accessed 10 Apr. 2010)
- 2) 浜田久之 The Objective Structured Teaching Evaluation (OSTE) 医学教育 2010; 41: 169-73.
- 3) 伴信太郎. 客観的臨床能力試験—臨床能力の新しい評価法—. 医学教育 1995; 26: 157-63.
- 4) 平成 18 年度国立病院機構長崎医療センター病院群. 臨床研修指導医講習会の記録, 2007.
- 5) 平成 19 年度国立病院機構長崎医療センター病院群. 臨床研修指導医講習会の記録, 2008
- 6) Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center. *Acad Med* 2000 Feb; 75: 161-6.
- 7) Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Acad Med* 1998 Jun; 73: 688-95.
- 8) 田中雄二郎, 丸野文昭. 卒後初期臨床研修医による指導医評価. 医学教育 2001; 32: 187-91.
- 9) 福井次矢監修. 研修指導医ガイドブック第 2 版. インターメディカ, 東京, 2003.
- 10) Prislin MD, Fitzpatrick C, Giglio M, Lie D, Radecki S. Initial experience with a multi-station objective structured teaching skills evaluation. *Acad Med* 1998 Oct; 73: 1116-8.
- 11) 岐阜大学医学教育開発研究センター編集. 新しい医学教育の流れ'09 冬, 三恵社, 愛知, 2009, p.42.
- 12) Morrison EH, Rucker L, Boker et al. A pilot randomized, controlled trial of a longitudinal residents-as-teachers curriculum. *Acad Med* 2003 Jul; 78: 722-9.
- 13) Morrison EH, Boker JR, Hollingshead J, Prislin MD, Hitchcock MA, Litzelman DK. Reliability and validity of an objective structured teaching examination for generalist resident teachers. *Acad Med* 2002 Oct; 77 (10 Suppl): 29-32.
- 14) 社団法人医療系大学間共用試験実施機構. 認定評価者用更新資料 2007 年度版, 東京, 2006, p.22.