

医学教育 2010, 41(3): 169~173

総 説

The Objective Structured Teaching Evaluation (OSTE)

浜田 久之*

要旨:

- 1) 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たが、指導医に対する評価の必要性も重要視されてきている。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、その概要を述べる。
 - 2) 1990年代に、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) が開発された。
 - 3) OSTEは、学生へ対するOSCEのように個人へ対する総括評価や資格試験として用いられることは少なく、FDとしてまたはFDの教育効果を測定するひとつの手段として用いられている。
 - 4) 今後日本でも指導能力の向上を目的としたFDや指導医講習会等で、OSTEを導入するための研究が期待される。
- キーワード: 指導医, 指導技術, 客観評価, 臨床研修制度

The Objective Structured Teaching Evaluation

Hisayuki HAMADA*

- 1) With the implementation of mandatory clinical residency in Japan, resident evaluations have become an important part of clinical education. Recently, however, a greater emphasis has been placed on the evaluation of physician instructors. In the West, research examining the teaching skills of clinical physicians has been ongoing since the 1970s and is reviewed in this paper.
- 2) The Objective Structured Teaching Evaluation (OSTE), which uses standardized students, multiple stations, video recordings, and scoring by observers, was developed in the 1990s.
- 3) Unlike the Objective Structured Clinical Examination, which is an evaluation tool for medical students and residents, the OSTE is rarely used to evaluate individual performances or used as a part of certification exams. Instead, the OSTE serves as a tool for faculty development or as an outcome measure for the effectiveness of faculty development education initiatives.
- 4) If available in Japan, the OSTE would be an excellent resource for improving the teaching skills of physicians and would be a useful tool for training sessions for physician instructors. More research is needed to facilitate the introduction of the OSTE to Japan in the future.

Key words: physician instructor, teaching skills, objective evaluation, clinical training systems

* 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]

受付: 2010年1月20日, 受理: 2010年3月5日

臨床医の指導能力を客観的に評価しようとする研究は、1970年代にすでに始まっている。1970年にBamford JC Jrらは、臨床医の指導能力を向上させるためのFaculty development（以下FD）において、学生に協力してもらい臨床教育現場を再現したり、ビデオカメラで病棟回診の指導の様子を撮影したり、さらに同僚による指導技術の評価などを試みている¹⁾。彼らは詳細な方法や結果は述べていないが、1990年代に開発されることとなるObjective Structured Teaching Evaluation：客観的指導能力評価（以下、OSTE）の多くの要素がすでに含まれていた。本稿では、欧米における臨床医の指導能力を評価する研究からOSTEが開発された経緯に関して文献的に考察する。

臨床医の指導能力に関して何を評価するべきかについて、1975年Stritter FTらは、具体的に研究している²⁾。4名の教育者等が77の評価項目をリストアップし、医学部3-4年の265名（全体の83%）が各項目に関して4段階（非常に重要—1点、まったく重要でない—4点）で点数化した結果、16項目（平均1.5点）が重要とされた。得点の高い項目は、知識や技術の実際的な応用を議論する、学生に何が重要と考えているかに気づかせる、明快に説明する、十分に準備をしている、熱心に取り組む姿勢等であった。さらにStritter FTらは77項目を6つに分類（①積極的に学生を参加させる②指導に積極的態度をとる③問題解決を強調する④学生を中心とした教授方法を持つ⑤人間味のある指南をする⑥参考文献と研究を重要視する）して評価指標とした。次にStritter FTらは、上記の評価指標を用いた教育ワークショップを1977年に開催している³⁾。目的は、下記の5つのステップを経て参加者自らの臨床指導に関して批判的な視点を持ち改善することであった。①自分が経験した最も効果のある、または最も効果のない指導方法について自由に意見を言う。②あらかじめ指導評価項目が12項目挙げられており、これらを重要な順番に各自並べ替え、どういう教え方が重要であるかを批判的にグループで議論する。③2つの典型的な臨床指導（各10分のビデオテープ）を見て、前述の研究

からStritter FTらが作った指標をもとに評価をさせる。④小グループに分かれてロールプレイを行う。参加者は、指導者、観察者または評価者の役を演じ、ひとりの学生に対して15-20分の指導をおこなう。⑤最後にワークショップのまとめをおこない、各自がどのような指導方法で今後教えて行くかを述べる。この研究のように、評価項目を定め、それに沿った教育ワークショップが盛んになって行くが、日常のリアルな臨床指導の場を評価しようとする試みもビデオカメラの普及により可能となった。

1977年にカナダのCassie JMらは、臨床指導能力の評価に用いる道具としてビデオカメラの有用性を発表している⁴⁾。彼らは指導医が病棟回診で指導している様子を録画して、言語相互作用指標を開発した専門家に解析させ、指導医へフィードバックしている。1978年にも、イリノイ大学のBazuin CHらが診療所で教える指導医の指導方法の改善のためにビデオカメラを利用する試みを行っている⁵⁾。指導医が診療所で学生を教えているところを1時間半撮影し、後日指導医にビデオを見せながら教え方について研究者らと討論する。月に1度、12回おこなった後、指導医の疑問、期待や役割について論じている。ビデオを使用し教え方を改善しようとするこれらの研究は、介入的であるが指導能力を評価に重点をあてたものではない。

介入試験を通して臨床指導能力の評価をする研究は、1983年Skeff KMらにより行われた⁶⁾。彼らは、臨床指導医の指導能力を向上させるために、64人の臨床指導医を対象に3つの評価方法で介入している。評価①質問票を用いた自己評価。評価②医学生に対して質問票を用いて指導医を評価した後、個別に指導医に結果を知らせる。評価③自分がおこなった回診のビデオ録画を自分で見る。次に、64人の臨床指導医は以下のように4つの群に振り分けられた。第1群：上記の3つの評価方法を用いて90分で指導方法に関して集中的な個別フィードバックを受ける。第2群：②と③を行う群。第3群：②のみを行う群。第4群：指導医は評価されるがフィードバックを受けない。結果は、集中的なフィードバックを受ける

第1群は、他の群より指導医の満足度が高く、介入後の評価得点は他群より高かった。フィードバックにより臨床医の指導方法の改善ができることを証明した。

1990年代には、客観的臨床能力試験 (Objective Structured Clinical Examination (OSCE)) の普及に伴い、同様の形式で指導能力を評価する研究が本格化する。1992年 Simpson DEらは、外来診療教育に関する90分間のFDに、訓練された学生 (標準化学生) を用いた⁷⁾。5-7名の指導医のグループに1人の標準化学生と教育者が振り分けられ、2人の指導医がそれぞれ標準化学生を相手に指導した後、2人の指導医の違いなどをグループで議論した。終了後アンケート調査が行われ、20名の参加者全員が、他人の指導を見ることは勉強になり、標準化学生を用いたことの有用性を認めている。また、22%の参加者が他人から評価を受ける状況に恐れを感じていた。1994年 Lesky LGらは、OSTEを objective structural teaching exercise とし、指導医が指導方法を訓練するために標準化された学生がいる仮の教育現場を複数設定したものと定義している⁸⁾。彼らは、20ステーション (ST) を設定し各STに訓練された標準化学生を配備した。各STには、簡単なシナリオがあり標準化学生はそれに沿って演技を行い、指導医は10分の指導を標準化学生に対して行った。EvaluationではなくExerciseである所以は、指導医に対する評価は行わず、教師中心から学習者中心のアプローチへ移行することに焦点をあてた議論を行うことのみで終了している点にある。

1998年の Gelula MHの研究は、より評価を意識した研究である⁹⁾。指導医が医学生を教える25本のシナリオを作成した。対象を、知識のない学生を指導する場合、優秀な学生を指導する場合のように幾つかに分け、場面も Problem Based Learning, 朝のカンファランス, 実験室等複数設定した。標準化医学生に合わせて、学年や年齢や性別を調整できる多彩なシナリオを用意した。指導医は標準化学生に対して10-15分間指導し、ビデオに撮影される。その後、指導医は、シナリオに関する教育技法等が含まれたFDへ参加した後、5-6人のグループ (標準化研修医も含まれて

いる) で、自分が撮影されたビデオを見てグループメンバーからフィードバックを受ける。この実験に関しては統計的な解析はなされておらず、研究者らが考える有用性を述べている。昇進のための参考資料とするために評価を受けることを望んだ参加者もいたことは、興味深い。

1998年に、Prislin MDらは客観的な評価に重点をおいた8つのSTを用いたOSTEを開発した¹⁰⁾。この研究は、現在普及している学生へ対するOSCEの形式に沿った初めての試みであるとしている。家庭地域科指導医8名を対象に、医学部3年生へ対する指導能力を評価している。評価者は録画したビデオで全パフォーマンスを評価する教員2名、各STのパフォーマンスを評価する教員3名、各STで演技する標準化研修医5名の10名である。評価項目は、クラークシップの教育的目的に関する知識、医学部3年生の医学知識範囲に関する知識、外来診療に適した学習方法と教授方法の知識、学生の臨床能力を適切に評価する能力、学生の個別のニーズに適応する能力、効果的なフィードバックをする技術、困難を抱えた学生を管理する能力、学習者—指導者の肯定的な関係を構築できる態度、適切な職業人としての態度であった。この評価項目に沿ってシナリオが作成された。最初のSTは学生へのオリエンテーション、第2STから第5STは様々な学生のレベル (初心者から上級者まで4段階) に合わせて教える、第6STは学生の臨床能力の評価、第7STはフィードバック技法、第8STは困難な学生への対応がテーマであった。各STにおける指導医のパフォーマンスは15分間であり、評価者ひとりにつき12.5点満点の試験である。評価者は事前に、書面と口頭で訓練を受けておりチェックリストで評価する。評価結果については、評価者群 (標準化学生、各STの教員、全体を評価する教員) の平均点では優位差はなかった。シナリオは概ね現実的であり、OSTEは教員の教え方を上手く評価していると結論付けている。標準化学生、複数のST、ビデオ撮影等を用いて客観的に指導能力を測定しようとした研究であったが、学生へ対するOSCEのように評価に重点をあててのではなく、Faculty developmentが充実したものとす

るため OSTE を利用すべきと結んでいる。

2000 年前後より OSTE 研究の流れは、評価の信頼性や妥当性を統計学的に検証するものとなっている。判定者の主観的な意識が入り込むことが避けられない観察スコアに関しては、その信頼性、妥当性を議論することは避けられない。1998 年に Litzelman DK らは、スタンフォード大学で使用されている臨床指導医を評価する指標の妥当性に関して統計学的に検討している¹¹⁾。2002 年に Morrison EH らは、OSTE に関する信頼性、妥当性を検討している¹²⁾。信頼性に関しては、OSTE で用いたチェックリストをクローンバックの α 係数、評価者間の相関係数を算出した。妥当性に関しては、3 つの点で証明している。第一に、教育トレーニング受ける介入群とコントロール群に対して OSTE を施行し、介入前、介入後、コントロールを比較すると、介入後が高得点であった。次に、別の指標（スタンフォード大学の FD で用いられる指導医を評価する 26 項目の指標）を用いた指導医評価が OSTE の結果と相関していた。最後に、教育経験が豊富な群が乏しい群より OSTE で良い点数を取ったことを妥当とした。2005 年 Mark Quirk らは、OSTE を設計する場合の ST の数、評価者の属性や数、スコアの種類などを統計学的に信頼性や妥当性を論じている¹³⁾。これらの研究は OSTE 成立の科学的根拠を追求するものであるが、近年は教育効果を測定するツールとして OSTE を利用している。2003 年、Morrison EH らは研修医を指導者として育成するプログラム（Residents-as-teachers program: 以下 RAT プログラム）を評価する目的で、OSTE を利用している¹⁴⁾。23 名の家庭医の研修医を無作為に 13 名の介入群と 10 名のコントロール群に分けた。介入群は 6 か月にわたり合計 13 時間の指導技術の訓練を受けた。介入の前後で全員が 8ST の OSTE を受け、その結果により RAT プログラムの効果を判定している。2007 年、Gaba ND らも同様に RAT プログラムの中で OSTE を行っており¹⁵⁾、OSTE による教育効果を測定した研究は少なくない。

以上のように欧米における臨床医の指導能力に関する研究の歴史は長い。指導能力を客観的に評

価しようとする OSTE は、学生に対する OSCE のように個人に対する総括評価や資格試験として用いられることは少なく、FD としてまたは FD の教育効果を測定するひとつの手段として用いられている。日本においても、2004 年に医師の臨床研修が義務化され指導医の役割についても大きな変化が起こっている。臨床研修の義務化で研修医が評価されることも義務化されたが、指導医に対する評価の必要性も重要となってきている¹⁶⁾。厚生省が示した新医師臨床研修制度における指導ガイドライン¹⁷⁾の中でも、指導医は評価されるべきであることを示している。今後日本でも指導能力の向上を目的とした FD や指導医講習会等で、OSTE を積極的に導入するための研究が盛んになることを期待する。

告 示

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 Prof. Helen P Batty, University of Toronto and Dr. Samuel Lapalme-Remis, McGill University and BA.Naomi H for providing information related to this study and reviewing of the manuscript.

文 献

- 1) Barnford JC Jr, Sall S, Rubin SH, et al. A project to improve faculty performance and enhance student learning. *J Med Educ* 1970 Sep; 45: 709-10.
- 2) Stritter FT, Hain JD, Grimes DA. Clinical teaching reexamined. *J Med Educ* 1975 Sep; 50: 876-82.
- 3) Stritter FT, Hain JH. A workshop in clinical teaching. *J Med Educ* 1977 Feb; 52: 155-7.
- 4) Related Articles, Cassie JM, Collins GF, et al. The use of videotapes to improve clinical teaching. *J Med Educ* 1977 Apr; 52: 353-4.
- 5) Bazuin CH, Yonke AM. Improvement of teaching skills in a clinical setting. *J Med Educ* 1978 May; 53: 377-82.
- 6) Skeff KM. Evaluation of a method for improving the teaching performance of attending physicians. *Am J Med* 1983 Sep; 75: 465-70.
- 7) Simpson DE, Lawrence SL, Krogull SR. Using

- standardized ambulatory teaching situations for faculty development. *Teach Learn Med* 1992; 4: 58-61.
- 8) Lesky LG, Wilkerson L. Using "standardized students" to teach a learner-centered approach to ambulatory precepting. *Acad Med* 1994 Dec; 69: 955-7.
 - 9) Gelula MH. Using standardized medical students to improve junior faculty teaching. *Acad Med* 1998 May; 73: 611-2.
 - 10) Prislun MD, Fitzpatrick C, Giglio M, Lie D, et al. Initial experience with a multi-station objective structured teaching skills evaluation. *Acad Med* 1998 Oct; 73: 1116-8.
 - 11) Litzelman DK, Stratos GA, Marriott DJ, et al. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Acad Med* 1998 Jun; 73: 688-95.
 - 12) Morrison EH, Boker JR, Hollingshead J, et al. Reliability and validity of an objective structured teaching examination for generalist resident teachers. *Acad Med* 2002 Oct; 77 (Suppl): S29-32.
 - 13) Quirk M, Mazor K, Haley HL, et al. Reliability and validity of checklists and global ratings by standardized students, trained raters, and faculty raters in an objective structured teaching environment. *Teach Learn Med* 2005 Summer; 17: 202-9.
 - 14) 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.
 - 15) Gaba ND, Blatt B, Macri CJ, et al. Improving teaching skills in obstetrics and gynecology residents: evaluation of a residents-as-teachers program. *Am J Obstet Gynecol* 2007 Jan; 196: 87. e1-7.
 - 16) 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 2003 p89 東京.
 - 17) 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 Jan. 2010)