Preface

As physicians' clinical skills cannot be measured solely by written examination, a National objective structured clinical examination (OSCE) has been considered necessary as a part of medical licensure examination. By the end of 2008, about 20 medical centers/ hospitals in Taiwan announced that OSCE has been a regular clinical examination for their trainees. However, there is little consensus about how to implement a high-stake, large scale OSCE.

The Medical Council of Canada (MCC) has administered an OSCE for the license to practice medicine since 1992. The high stake, large scale OSCE is to test physicians' skills of history taking, physical examination, and communication. The examination results with psychometric evidence indicate that a full-scale national administration of an OSCE model for licensure is feasible in Canada.

Dr. Sydney Marla Smee is currently the Manager of MCCQE Part II, Evaluation Bureau Medical Council of Canada (1990-Present). She is an internationally recognized expert in implementing a high stake OSCE. The workshop in Taiwan lead by Dr. Smee is to facilitate the establishment of a Taiwanese model of high stake OSCE.

Contents

- **Agenda -- 3**
- **Speaker's Curriculum Vitae -------------------------- 5**
- **Instructor Assistant's Curriculum Vitae ------------ 10**
- **Moderator's Curriculum Vitae ----------------------- 12**
- **PowerPoint -- 13**
- **Case Writing Guidelines ------------------------------ 33**
- **Example of MCCQE Part II Case Physical Examination -- 43**
- **Example of MCCQE Part II Case History Station 48**
- **The Blank Case Document ---------------------------- 52**
- **Related Articles**
  - **The Use of Standardized Patient Assessments for Certification and Licensure Decisions ------------------- 56**
  - **The Comparison of Physician Examiners and Trained Assessors in a High-Stakes OSCE Setting ---- 64**
  - **Physician Scores on a National Clinical Skills Examination as Predictors of Complaints to Medical Regulatory Authorities ------------------------------------ 68**
- **Participants -- 78**

Day1 Agenda

Case writing for High Stakes OSCE: The Medical Council of Canada's Approach

Date: January 9, 2010 (Saturday)
Place: 臺北醫學大學 醫學綜合大樓前棟 4 樓-**誠樸廳**
台北市信義區 110 吳興街 250 號
Instructor: Dr. Sydney M. Smee
Instructor Assistant: Dr. Charity T.C. Tsai

Time	Activity	Moderator
08:30-08:45	**Registration**	
08:45-09:00	**Opening Remarks** President Wen-Ta Chiu **邱文達校長** Professor Chi-Wan Lai **賴其萬教授** Superintendent Chung-Ye Hong **洪傳岳院長**	Dean Chii-Ruey Tzeng **曾啟瑞院長**
09:00-10:15	**Presentation:** Writing cases for high stakes OSCE Part One	
10:15-10:30	**Break**	
10:30-12:00	**Presentation:** Writing cases continues…	
12:00-13:00	**LUNCH**	
13:00-14:45	**Presentation:** Writing cases for high stakes OSCE Part Two	
14:45-15:00	**Break**	
15:00-16:00	**Presentation:** Continue work on cases and checklists	
16:00-17:00	**Questions and Discussion:** Creating Clones	
17:00	Adjourn	

Day2 Agenda

Case writing for High Stakes OSCE: The Medical Council of Canada's Approach

Date: January 10, 2010 (Sunday)
Place: 臺北醫學大學 醫學綜合大樓前棟 4 樓-**誠樸廳**
台北市信義區 110 吳興街 250 號
Instructor: Dr. Sydney M. Smee
Instructor Assistant: Dr. Charity T.C. Tsai

Time	Activity
08:30-09:00	**Check-in:** Questions/Answers
09:00-09:30	**Presentation:** Cases and Examiners
09:30-10:15	**Individual work:** Individual Work on Cases
10:15-10:30	**Break**
10:30-12:00	**Individual work:** Work on Cases…
12:00-13:00	**LUNCH**
13:00-14:30	**Group activity:** Group review of work to date
14:30-14:45	**Break**
14:45-16:00	**Make edits/corrections**: Update all work
16:00-17:00	**Wrap-up - Review principles - Feedback - Next Steps**
17:00	**Closing Remarks 閉幕致詞** Superintendent Chi-Hsiung Wu **吳志雄院長**

Speaker's Curriculum Vitae

Dr. Sydney M. Smee

Higher Education

2007 *Doctor of Philosophy*
University of Ottawa
Major: Education
Minor: Measurement and Evaluation

1994 *Master of Education*
Ontario Institute of Studies in Education
University of Toronto
Major: Adult Education

1982 *Bachelor of Arts*
McMaster University
Major: Political Science

Professional Positions

1990-Present ***Manager, MCCQE Part II***
Evaluation Bureau, Medical Council of Canada

1987 - 1990 ***Coordinator, Volunteer Services***
Casey House Hospice, Toronto

1986-1988 ***Consultant, Standardized Patient Program Development***
University of Massachusetts, University of Toronto, McMaster University

1984 - 1985 ***Coordinator, Patient Instructor Program***
University of Massachusetts Medical School, Worcester, MA

Editorial Activities

Ad Hoc Reviewer:

Advances in Health Sciences: 2001, 2007

Medical Education: 2001 – 2002, 2005, 2009.

Publications

Boulet, J.R., Smee, S.M., Dillon, G.F., and Gimpel, J.R. (2009).The use of standardized patient assessments for certification and licensure decisions. Simulation in Health Care 4:1 Spring.

Smee, SM. (2008). High Stakes OSCE scoring: Station-specific rating scales versus checklists. Paper presented at the 13th Ottawa Conference on Medical Education: Melbourne, Australia.

Smee, SM. (2008). Impact of judgmental weights for OSCE checklist items on station pass marks. Paper presented at the 13th Ottawa Conference on Medical Education: Melbourne, Australia.

Boursicot, KA, Smee, SM, & Paterson, J. (2008). Ten years of monitoring test security in graduation level OSCEs. Paper presented at the 13th Ottawa Conference on Medical Education: Melbourne, Australia.

Wood, TJ & Smee, S. (2008). Does editing an OSCE station after an examination improve its performance on subsequent examinations? Paper presented at the 2008 annual meeting of the Association of Medical Educators of Europe (AMEE): Prague, Czech Republic.

Wood, TJ, Smee, SM, Bartman, I, & Blackmore, DE. (2008) Do two different processes for limiting false positive errors add to the quality of the pass/fail decision on a high stakes examination? Paper presented to the annual meeting on Research in Medical Education (RIME): San Antonio, USA.

Tamblyn, R, Abrahamowicz, M, Dauphinee, D, Wenghofer, E, Jacques, A, Klass, D, Smee, S, Blackmore, D, Winslade, N, Girard, N, Du Berger, R, Bartman, I, Buckeridge, D, & Hanley, J. (2007). Physician Scores on a National Clinical Skills Examination as Predictors of Complaints to Medical Regulatory Authorities. Journal of the American Medical Association, 298, 993-1001.

Birtwhistle, R, Bartman, I & Smee, S. (2006). Effect of SP gender on candidate performance in an OSCE station in a high stakes examination. Paper presented at the 12th Ottawa Conference on Medical Education, New York.

Humphrey-Murto, S, Smee, SM, Touchie, C, Wood, TJ, & Blackmore, DE. (2005). A comparison of physician examiners and trained assessors in a high-stakes OSCE setting. Academic Medicine, 80, s59-s62.

Boursicot, KAM & Smee, SM. (2004). Setting standards for a finals Objective

Structured Clinical Examination (OSCE): Comparing the borderline group method with an Angoff approach. Paper presented at the 10th Ottawa International Conference on Medical Education, Ottawa, Canada.

Smee, SM. (2003). ABC of learning and teaching in medicine: Skill-based assessment. British Medical Journal, 326, 703-706.

Smee, SM, Dauphinee, WD, Blackmore, DE, Rothman, AI, Reznick, R, & Des Marchais, J. (2003). A sequenced OSCE for licensure: Administrative issues, results and myths. Advances in Health Sciences Education: Theory and Practice, 8, 223-236.

Birtwhistle, R, Blackmore, DE, Smee, SM, & Wood, T. (2002). Does specialty play a role when physicians are used as examiners in a nationally administered OSCE? Paper presented at the 9th Ottawa International Conference on Medical Education Capetown, South Africa.

Blackmore, DE & Smee, SM. (2002). Weighted vs. unweighted OSCE checklists. In Paper presented at the 9th Ottawa International Conference on Medical Education Capetown, South Africa.

Smee, SM & Blackmore, DE. (2002). Setting standards for an objective structured clinical examination: The borderline group method gains ground on Angoff. Medical Education, 35, 1009-1010.

Smee, SM & Blackmore, DE. (2002). Authors' reply: Setting standards for an objective structured clinical examination: The borderline group method gains ground on Angoff. Medical Education, 36, 388-389.

Smee, SM & Blackmore, DE. (2001). Commentary - Setting standards for an objective structured clinical examination: The borderline group method gains ground on Angoff. Medical Education, 35, 1009-1010.

Dauphinee, WD, Boulais, AP, Smee, SM, Rothman, AI, Reznick, R, & Blackmore, DE. (2000). Examination results of the Licentiate of the Medical Council of Canada: Trends, Issues and Future Considerations. In D. E. Melnick (Ed.), Proceedings of the Eighth International Ottawa Conference - Evolving Assessment: Protecting the Human Dimension (pp. 92-98). Philadelphia: National Board of Medical Examiners.

Dauphinee, WD, Blackmore, DE, Smee, SM, Rothman, AI, Des Marchais, J, & Reznick, RK. (2000). Adaptive testing: A report on the results and myths arising from the use of a sequenced OSCE for national licensure. In D. E. Melnick (Ed.),

Proceedings of the Eighth Ottawa International Conference - Evolving Assessment: Protecting the Human Dimension (pp. 241-246). Philadelphia: National Board of Medical Examiners.

Smee, SM, Blackmore, DE, Rothman, AI, Reznick, RK, & Dauphinee, WD. (2000). Pioneering a sequenced OSCE for the Medical Council of Canada: An administrative overview. In D. E. Melnick (Ed.), Proceedings of the Eighth International Ottawa Conference - Evolving Assessment: Protecting the Human Dimension (pp. 234-240). Philadelphia: National Board of Medical Examiners.

Poldre, P, Smee, SM, Reznick, RK, Blackmore, DE, Birtwhistle, R, Blouin, D, Chalmers, A, Galway, B, Hodges, B, MacFadyen, J, & Spady, D. (1999). The experience of thousands: The post-examination OSCE station review process of the Medical Council of Canada. In D. E. Melnick (Ed.), Evolving assessment: Protecting the human dimension (CD-ROM) Philadelphia: National Board of Medical Examiners.

Dauphinee, WD, Blackmore, DE, Smee, SM, Rothman, A. I, & Reznick, RK. (1997). Using the judgments of physician examiners in setting the standards for a national multi-center high stakes OSCE. Advances in Health Sciences Education: Theory and Practice, 2, 201-211.

Dauphinee, WD, Blackmore, DE, Smee, SM, Rothman, AI, & Reznick, RK. (1997). Optimizing the input of physician examiners in setting standards for a large scale OSCE: Experience with Part II of the Qualifying Examination of the Medical Council of Canada. In A. J. J. A. Scherpbier, C. P. M. van der Vleuten, J. J. Rethans, & A. F. W. van der Steeg (Eds.), Advances in Medical Education: Proceedings of the Seventh Ottawa International Conference on Medical Education (pp. 656-658). Dordrecht: Kluwer Academic Publishers.

Reznick, RK, Blackmore, DE, Dauphinee, WD, Smee, SM, & Rothman, AI. (1997). An OSCE for licensure: The Canadian experience. In A. J. J. A. Scherpbier, C. P. M. van der Vleuten, J. J. Rethans, & A. F. W. van der Steeg (Eds.), Advances in Medical Education (pp. 458-461). Dordrecht: Kluwer Academic Publishers.

Smee, SM. & Sumawong, V. (1997). Advancing the use of standardized patients: A workshop for the consortium of Thai medical schools. In AJJA. Scherpbier, CPM van der Vleuten, JJ Rethans, & AFW van der Steeg (Eds.), Advances in Medical Education (pp. 714-716). Dordrecht: Kluwer Academic Publishers.

Smee, SM. & Blackmore, DE. (1997). Preparing physician examiners for a high stakes, multi-site OSCE. In AJJA Scherpbier, CPM. van der Vleuten, JJ Rethans,

& AFW van der Steeg (Eds.), Advances in Medical Education (pp. 462-469). Dordrecht: Kluwer Academic Publishers.

Reznick, RK, Blackmore, DE, Dauphinee, WD, Rothman, AI, & Smee, SM. (1996). Large-scale high-stakes testing with an OSCE: Report from the Medical Council of Canada. Academic Medicine, S71, 19-21.

Smee, SM. (1994). Medical Education Clinic: Using SPs for teaching and evaluation. Adult Education Quarterly, 6, 9-10.

Reznick, RK, Blackmore, DE, Cohen, R, Baumber, JS, Rothman, AI, Smee, SM, Chalmers, A, Poldre, P, Birtwhistle, R, Walsh, P, Spady, D, & Bérard, MJ. (1993). An objective structured clinical examination for the licentiate of the Medical Council of Canada: From research to reality. Academic Medicine, S68, 4-6.

Reznick, RK, Smee, SM, Baumber, JS, Cohen, R, Rothman, AI, Blackmore, DE, & Bérard, M. (1993). Guidelines for estimating the real cost of an objective structured clinical examination. Academic Medicine, 68, 513-517.

Reznick, RK, Smee, SM, Rothman, AI, Chalmers, A, Swanson, DB, Dufresne, L, Lacombe, G, Baumber, J, Poldre, P, Levasseur, L, Cohen, R, Mendez, J, & Bérard, M. (1992). An objective structured clinical examination for the licentiate: Report of the pilot project of the Medical Council of Canada. Academic Medicine, 48, 487-494.

Stillman, PL, Swanson, DB, Smee, SM, Stillman, AE, & Ebert, TH. (1986). Assessing the clinical skills of residents with standardized patients. Annals of Internal Medicine, 105, 762-771.

Instructor Assistant's Curriculum Vitae

Charity TC Tsai, MD, PhD

China Medical University
University of Calgary, MSc in Medical Education
University of Calgary, PhD in Medical Education
Present Academic and Administrative approintment:

Director, Department of Pediatrics, Taipei Medical University WanFang Hospital
Vice Director, Faculty of Medicine, Taipei Medical University
Vice Director, Education and Research, Taipei Medical University WanFang Hospital

Publication (in recent 5 years)

1. Tsuen-Chiuan Tsai, Using children as standardized patients for assessing clinical competence in pediatrics. Arch Dis Child 89 (12): 1117-1120, Dec. 2004
2. Tsuen-Chiuan Tsai, M.D., Peter H. Harasym, Ph.D., Cheri Nijssen-Jordan, and Greg Powell, Learning gains derived from a high fidelity simulation in emergency department J Formos Med Assoc 105 (1):94-98, 2006
3. JD Tsai, FU Huang, CC Lin, TC Tsai, HC Lee, and JC Sheu. Intermittent hydronephrosis secondary to ureteropelvic junction obstruction: clinical and imaging features. Pediatrics. 2006 Jan;117(1):139-46.
4. Tsai TC, Harasym PH. Challenges of pediatric residency education in Taiwan. Acta Pediatrica Sinica. 47(1):3-6, 2006
5. Tsuen-Chiuan Tsai . Psychosocial effects on caregivers for children in Taiwan on chronic peritoneal dialysis. Kidney Int. 2006 Dec;70(11):1983-7
6. Tsuen-Chiuan Tsai . University of Washington 家醫科及臨床技能中心參訪. J Med Education. Jan. 10(1): 86-88, 2006
7. Tsuen-Chiuan Tsai,[1] Pei-Jung Chang,[2] Shin-Yuan Fang, [2] Chyi-Her Lin[3.] A Mannequin-based Simulation on Teaching Emergent Crisis Care. J Med Education, 10(2): 115-125, 2006
8. Sheu JC, Koh CC, Chang PY, Wang NL, Tsai JD, Tsai TC. Ureteropelvic junction obstruction in children: 10 years' experience in one institution. Pediatr Surg Int. 2006 Jun;22(6):519-23
9. Tsai YC, Tsai TC, Tsaf JD, Huang FY, Lin CC, Sheu JC. Clinical analysis of chronic peritoneal dialysis related peritonitis in children. Pediatr Neonatol. 2006 Mar-Apr;47(2):72-6
10. Tsuen-Chiuan Tsai, Chyi-Her Lin, Chung-Lin Chen, Co-Chi Chao, Taung-Lieh Yeh, Jing-Jane Tsai, Yin-Fan Chang. Analysis of OSCE results: experience in National Cheng Kung University Medical College. J Med Education 10

(4):313-23, 2006

11. Huang DTN, Tsai TC, Huang FY, Tsai JW, Chiu NC, Lin CC. Clinical differentiation of acute pyelonephritis from lower urinary tract infection in children. Journal of microbiology, immunology and infection. J Microbiol Immunol Infect. 2007;40:513-517
12. Tsai TC. Resistance to educational change: management and communication. Pediatr Neonatol.48:3-6, 2007
13. Tsai TC, Lin CH, Harasym PH, Violato C. Students' perception on medical professionalism: the psychometric perspective. Med Teach. 2007 Mar;29(2-3):128-34.
14. Peter H. Harasym, Tsuen-Chiuan Tsai, and Payman Hemmati. Current trends in developing medical students' critical thinking abilities. Kaohsiung J Med Sci July 2008. 24 (7) 341-354
15. Tsuen-Chiuan Tsai. The Use of Medical Cognition in Medical Curriculum Reform in Taiwan. Pediatr Neonatol 2008;49(3):53−57
16. 蔡淳娟、邱文達、王先震、連吉時、粟發滿、郭雲鼎、徐明義.The use of portfolio in internship clinical education. J of Med Edu 12(1): 8-19.2008.
17. Lee MD, Lin CC, Huang FY, Tsai TC, Huang CT, Tsai JD. Screening young children with a first febrile urinary tract infection for high-grade vesicoureteral reflux with renal ultrasound scanning and technetium-99m-labeled dimercaptosuccinic acid scanning. J Pediatr. 2009 Jun;154(6):797-802.
18. 顏如娟,蔡淳娟,郭耿南,張般瑞,陳泰宏.台灣醫師人力需求之探討.投稿台灣公共衛生雜誌 2009/07
19. Tsuen-Chiuan Tsai, Peter H. Harasym, Sylvain Coderre, Kevin McLaughlin, & Tyrone Donnon. Assessing ethical problem solving by reasoning rather than decision making. Med Edu 2009: 43: 1188–1197
20. Ju-Chuan Yen,Tsuen-Chiuan Tsai, Min-Huei Hsu, Kung-Jiang Chang, Du-Jian Tsai, Wei-Hua Lee. The attitudes toward disclosure of medical errors: the perspectives of Taiwanese with different occupational backgrounds. Submit to The American Journal of Bioethics (UAJB-2009-0218) 2009/04/14
21. Tsuen-Chiuan Tsai, Peter H. Harasym. An Ethical Reasoning Model: Contributions to Medical Education. (Submitted to Med Educ in 2009/09)
22. 蔡淳娟,林其和,劉克明. 台灣各界對醫學系學制變革可行性的看法. 投稿醫學教育雜誌. 2009/09

Moderator’s Curriculum Vitae

Professor Wen-Ta Chiu (邱文達校長)

Present Positions

Professor and President, Taipei Medical University

Professor Chi-Wan Lai (賴其萬教授)

Present Positions

Executive Secretary, Medical Education Committee, Ministry of Education

CEO, Taiwan Medical Accreditation Council

Professor Chung-Ye Hong (洪傳岳院長)

Present Positions

Professor and Superintendent , WanFang Hospital, Taipei Medical University

---

Professor Chii- Ruey Tzeng (曾啟瑞院長)

Present Positions

Dean, College of Medicine, Taipei Medical University

Professor and Chairman, Department of Obstetrics and Gynecology, School of Medicine, Taipei Medical University

---

Professor Chi- Hsiung Wu（吳志雄院長）

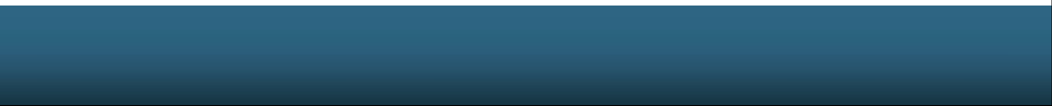
Present Positions

Professor and Superintendent, Shuang Ho Hospital, Taipei Medical University

Case writing for High Stakes OSCE: The Medical Council of Canada's Approach.

Writing Cases for High Stakes Objective Structured Clinical Examinations (OSCE) Part One

- A Medical Council of Canada Approach
- Dr. Sydney Smee
- Manager, MCCQE Part II
- Evaluation Bureau



What is the MCCQE Part II?

- Performance-based assessment of clinical skills
- Multi-site, administered twice per day
- Timed circuit of 14 stations
- Patient-based
- Physician-scored
- Prerequisite for licensure in Canada since 1993

Presented at Taipei Medical University, WanFang Hospital for the Taiwan Association for Medical Education January 2010 2

Why do we have the Part II?

- Requested by the Medical Licensing Authorities because they were facing:
  - Increasing number of complaints, often based on a physician's communication skills.
  - Need to be publicly accountable; e.g., reports that not all trainees were being assessed in a clinical setting.
  - Obligation to audit the training of all medical graduates seeking licensure in Canada.

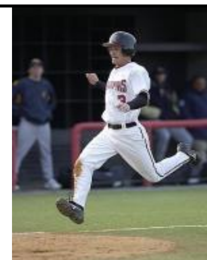
Content of Part II

- Multidisciplinary, patient-based cases
  - Some have a written component based *directly* on the patient problem
- Common or acute presenting problems
  - Some problems include legal and ethical issues
- Assesses skills:
  - History taking
  - Physical examination skills
  - Counseling / patient education skills
  - Patient management ability

Part II - Fall 2009

- 12 case OSCE with 2 pilot cases
  - 5+5 minute couplets (patient + written components)
  - 10-minute (patient + (sometimes) oral component
- OSCE was run twice per day and ran over two days (Saturday and Sunday)
- 16 university-based sites
- 2,644 test takers assessed in two days across four time zones and two languages
- Most common site model runs two tracks for one day and requires the following:
  - 44 clinic rooms - 50 to 60 standardized patients
  - 44 physician examiners - 16 to 20 staff people

Goals for Today and Tomorrow



- **Present case writing and review process**
- **Provide criteria for case review**
- **Lead you through the process once**
- **Have everyone complete one good case**
  - **More if possible….**

Characteristics of Successful Cases

- Feasible
  - "OSCE-able"
  - Task-time congruence
- Valid
  - Authentic problem
  - Task assesses what you want to assess
- Complete
  - Candidate Instructions
  - Examiner Instructions
  - Mark sheet
  - Simulation Instructions
  - Props specified

Feasible = "OSCE-able"

- Patient affect and symptoms can be presented with sufficient realism
  - May use simulated patient or a manikin
  - May use make-up and props (e.g., taping an IV line to the simulated patient, bruising, pallor)
  - Pain, aphasia, perspiration, sadness, irritability, gait problems and much more can be simulated
  - Jaundiced sclera and swollen knee and others – cannot be simulated

Feasible = Task–Time Agreement

- Task is appropriate for time allotment
  - Completing a focused physical examination for abdominal pain may be do-able in five minutes
  - Providing a bad prognosis should not be do-able in five minutes
  - Some tasks may be framed by instructions to make them appropriate for a short station
- Task can be done within time allotment
  - Pilot test cases!

Valid

- Authentic patient problem
  - Simulated patient problem and clinical situation are sufficiently realistic to elicit required performance
    - Always balance realism with feasibility
    - Draw from actual cases as much as possible
- Patient problem matches examination specifications – the OSCE blueprint
  - Patient problem is relevant to objectives of the assessment
  - Task is geared to level of test takers

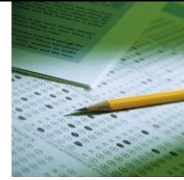
Complete

- Cases have multiple users
  - Standardized patient (SP) learning the role
  - Trainer preparing the standardized patient
  - Administrator setting up equipment & signage
  - Examiner scoring the case
  - Test takers taking the case
- Cases have multiple components
  - Ensure that all users have what they require
  - Ensure components are in agreement – no contradictions or overlap

Writing in Groups is Good

- Working in a group allows the facilitator to prepare authors for the task all at once
- Authors provide input and advice to each other which improves case quality
- Working in a group encourages creativity and learning
- Writing good cases is hard work but it does not have to be boring
- Facilitator can focus clinical experts on assessment issues

Cases are Test Items



- Clinicians want to write about the patient
- Test developers write the test item first
- Basic steps:
  1. **Articulate the *purpose* for the case**
  2. **Develop the *instructions* to the test takers**
  3. **Create the *scoring instrument(s)***
  4. **Develop the patient presentation**

Part of an MCC Objective

HYPERCALCEMIA

- **RATIONALE**
- Hypercalcemia may be associated with an excess of calcium in both extracellular fluid and bone (e.g., increased intestinal absorption), or with a localised or generalised deficit of calcium in bone (e.g., increased bone resorption). This differentiation by physicians is important for both diagnostic and management reasons.
- **CAUSAL CONDITIONS**
  - Increased intestinal absorption
  - Increased intake (e.g., milk-alkali syndrome)
  - Vitamin D mediated (e.g., granulomatous diseases)
  - Increased bone resorption
  - Malignancy
  - Primary/Secondary/Tertiary hyperparathyroidism
  - Hyperthyroidism
  - Immobilization
  - Paget disease
  - Diminished excretion (familial hypocalciuric hypercalcemia, thiazides)
- **OBJECTIVE**
- Through efficient, focused, data gathering, differentiate hypercalcemia caused by increased intake from that of excess bone resorption.

http://www.mcc.ca/Objectives_online/

1. Case Purpose

- State the purpose of the case, for example:

  To elicit a history from a patient that demonstrates an understanding of hypercalcemia and its etiologies.

- Purpose must relate to the OSCE objectives
- The purpose determines what to include or exclude as the case is written

2. Instructions to Test Takers

- Instructions are the test question
- Include the following:
  - Patient's name and age
  - Setting
    - Is the patient in the emergency department, a clinic, a hospital ward?
- Add relevant background information if necessary
- Specify the task
- (Specify the time limit)

Writing Instructions: Be Clear

- If you are assessing the test taker's ability to do a specific task, say so; for example:
  - Conduct a mini-mental status examination
  - Include relevant information to frame the task
- If you are assessing the test taker's judgment about what should be done (as well as how to do it), say so; for example:
  - Take a focused history and make an appropriate assessment of this patient

Writing Instructions: Be Careful

- Do not ask test takers to pretend they have met this patient before; instead tell them:
  - This patient is being seen in the emergency department, a walk-in clinic or the patient was previously seen by your colleague (who is inconveniently away), the patient is new to your practice , the patient is just visiting in this city…
- Do not ask test takers to pretend it is a different time of day or season – make the case work in real time (more about this later)

Writing Instructions: Be Creative

- By providing more or less information you can make a case fit the OSCE framework and time limits better
- Examples:
  - The nurse just arrived…. (or comes in after)
  - The child is being weighed – speak with the parent
  - Provide relevant history – then ask the test taker to do the appropriate history
  - Provide lab results, imaging, etc. and ask them to interpret for the patient (previously seen by your colleague, at the hospital etc.)

3. Scoring Instruments

- Checklists are useful when assessing
  - Thoroughness or key elements
  - Student (beginner) levels of ability
  - Procedural tasks
  - Limited time for training markers and/or for marking
- Rating scales are useful when assessing
  - Behaviours
    - Emphasis is on "done more or less" rather than on "done / not done"
  - Higher levels of expertise
    - Emphasis is on judgment rather than thoroughness
- To weight or not to weight…

How many items?

- Five-minute stations
  - 8-25 checklist items
  - 1-2 rating scale items
- Ten-minute stations
  - 15-50 checklist items
  - 3-7 rating scale items
- Ensure items reflect the clinical problem
  - Scoring items should reflect expected actions and behaviours, not the patient's responses

Presented at Taipei Medical University, WanFang Hospital for the Taiwan Association for Medical Education January 2010 21

Sample Case

PURPOSE:

To elicit a history from a patient with hypercalcemia, being aware of the manifestations of hypercalcemia and its etiologies.

Test takers are eligible for entry to independent practice (with a minimum of 18 months post graduate clinical training.

INSTRUCTIONS:

Susan Newton, 58 years old, presents to your office because of abnormal blood tests found on an insurance physical done by your colleague. You received notification that her serum calcium is elevated at 3.3 mmol/L (N=2.18-2.58 mmol/L). This result has been repeated and verified.

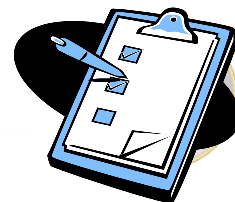
In the next five minutes, obtain a focused and relevant history.

Write a LEGIBLE checklist on paper 22

Are these good items?

- Asks about alleviating and aggravating factors
- Asks about previous medical history
- Inspects abdomen
- Patient has shortness of breath
- Takes blood pressure
- Asks about alcohol use
- Assesses alcohol use with CAGE questions or similar

Checking a Checklist



- Items must be separate
- Items begin with a verb
- Items must be observable
- Item specificity matches patient problem
- Number of items relates to length of station
- Weights are assigned to items (if using weights)
- Verify that findings can be simulated
- Ensure completeness
  - Responses for every checklist item are in the case
- Checklist items do not conflict with instructions

Time to start writing…..

What is the purpose of your case?

(15 minutes)

Form small groups

1. Three to four people per group
2. Have each person explain the purpose of their case to the group
3. Provide suggestions and feedback to each other based on the OSCE objectives and framework
4. Take about 5 minutes per case

Write instructions for your case

- Include:
  - Patient's name and age.
  - The setting.
  - Information about the patient that the test taker should have before starting.
  - The task.

Be Clear, Be Careful, Be Creative.

(Consult - Take about 20 minutes.)

Review Process

- Ongoing review improves cases
- Early review improves cases
- Perspectives form other clinicians (of different disciplines) helps an author refine their work
- Feedback from a test developer helps an author refine their work

Re-form into small groups

1. Each person presents the instructions for their station
2. Provide feedback to each other:
    - Is the task clear?
    - Is there enough information? Too much?
    - Discuss what should be covered in the checklist for each case
3. Take about 5-8 minutes per case

Write the checklist

- On your own, with your colleagues….
- Remember:
    - Items must be separate
    - Items begin with a verb
    - Items must be observable
    - Item specificity matches patient problem
    - Number of items relates to length of station
    - Examiner directions included as needed
    - Verify that findings can be simulated
    - Checklist items do not conflict with instructions

Written Components – What?

- Short answer write-in questions related to a specific patient presentation in the OSCE
  - What is most likely diagnosis?
  - What are three most likely differential diagnoses?
  - Interpret the x-ray, ECG, etc.
  - What were three key features of the history that led to your primary diagnoses?
  - What four initial investigations would you order for this patient?
  - Write your admission orders for this patient.
  - Write the prescription for this patient.
  - What advice would you give this patient?
  - Who, if anyone, should be informed about…. ?

Written Components: Why? Why Not?

- Why?
  - Assess test taker's synthesis of patient information
  - Aids in assessing other aspects of the patient presentation (like ethical and legal issues)
  - Collect more data from test takers without needing an examiner or SP present
  - Increases capacity of an examination site with limited increase in examination day costs
- Why Not?
  - Difficult to write good questions
    - Must be specific to the OSCE patient (otherwise put in a knowledge-based test)
    - No questions can overlap with or cue answers for other questions
  - Difficult to write good answer keys
  - Require post-OSCE scoring

Writing the rest of your case

- There are many users:
  - Administrators
  - Standardized Patients
  - Examiners
  - Test Takers

Administrators

- Edit cases to produce mark sheets
- Produce training materials for SPs and Examiners
- Plan the examination set-up

- Case writer's task:
  - Ensure case material is clear
    - Use few acronyms
    - Make sure descriptions are complete
    - Specify equipment and room requirements

Standardized Patients (SPs)

- Give concrete descriptions of patient's behaviour, affect and symptoms
- Give patient's reason for visit and their belief about the problem
  - This information guides the SP when they need to respond to unexpected questions or information from the test takers
- Use patient-based language; for example:
  - Belly pain - not abdominal pain
  - Headaches come and go – not intermittent HA
- Give relevant information only; for example:
  - Deskbound job – not accountant for the WanFang Hospital
- Include critical and pertinent negatives
- Identify any information that can be volunteered (non essential)

More about SP information

- Link case information to the checklist
  - Patient information should include responses to every item in the checklist, as well as responses to likely actions from test takers
- Only use physical findings and affects that can be simulated
- Suggest how findings may be simulated
  - Be aware that simulating a finding is different than describing it
  - Describe impact of symptoms on activities of daily living, work, mobility, etc.
- Give specific details for pain levels, ROM, location, etc.

WHAT CAN BE SIMULATED

Adapted from *An Overview of the Uses of Standardized Patients for Teaching and Evaluation Clinical Skills*, HS Barrows, ACADEMIC MEDICINE, June 1993; 6:445.

- Tenderness
- Restricted ROM
- Aphasia
- Kernig's sign
- Babinski reflex
- Lid lag
- Brudzinski sign
- Muscle weakness
- Parkinsonism
- Cheyne-Stokes respirations
- Perspiration
- Confusion
- Photosensitivity
- Costovertebral-angle tenderness

MORE THAT CAN BE SIMULATED

Adapted from *An Overview of the Uses of Standardized Patients for Teaching and Evaluation Clinical Skills*, HS Barrows, ACADEMIC MEDICINE, June 1993; 6:445.

- Decerebrate fit
- Rebound tenderness
- Rigidity
- Facial paralysis
- Seizures
- Gait abnormalities
- Bruises
- Pneumothorax
- Shortness of breath
- Hearing Loss
- Tremor
- Hypomania
- Visual loss

Examiners and Test Takers

- Purpose of Station, Instructions and Checklist (plus rating scale, if included) are all used by the examiner
- Test takers use the Instructions

New challenges

- How to assess cultural, legal, ethical and organizational objectives?
- How to assess professionalism?
- How to assess ability to work in interdisciplinary teams?
- Are there other ways to score?
  - Global ratings?
  - Key Feature checklists?

MEDICAL COUNCIL OF CANADA / LE CONSEIL MÉDICAL DU CANADA

OVERVIEW FOR CASE AUTHORS
QUALIFYING EXAMINATION PART II (MCCQE PART II)

The Examination

The Qualifying Examination Part II is a prerequisite to obtaining the Licentiate of the Medical Council of Canada (LMCC). The LMCC, in turn, is essential to acquiring a license for independent medical practice in all Canadian provinces and is essential to portability of licensure throughout Canada. The purpose of the MCCQE Part II is to assess candidates for the knowledge, skills, and attitudes expected of all physicians entering independent practice. The prerequisites for the MCCQE Part II are that candidates have successfully completed at least one year of postgraduate training and have passed the Medical Council of Canada Qualifying Examination Part I.

The MCCQE Part II is a criterion-referenced Objective Structured Clinical Examination (OSCE). Examination objectives developed by the MCC are used to identify the domains of cognitive and clinical skills that are evaluated on this national examination and serve as the basis for developing the content of the OSCE cases.

The examination's objective is to assess the candidate's knowledge and skills in the following domains: physical examination, problem solving, data gathering and interpretation, investigation, crisis management, patient education and communication; as well as their approach to legal and ethical issues.

The Stations

The MCCQE Part II is composed of a series of clinical cases that are presented in one of two station formats. In the first type, the candidate has a five-minute encounter with a Standardized Patient (SP) and then has five minutes to complete a Post Encounter Probe (PEP), which consists of written questions related to the patient encounter. In the second type, candidates have a ten-minute encounter with a SP or a 9-minute patient encounter, followed by 1 or 2 oral questions from the examiner. Stations are categorized as History, Physical Examination, History plus Physical Examination, Communication, Management or CLEO (Consideration of Legal, Ethical and Organizational issues). Management stations usually include a combination of history taking, physical examination and initial decisions that must be made by the candidate, based on the findings.

The Patients

Standardized patients are selected individuals trained to reproduce in a standardized fashion a patient problem, including physical findings and affect, according to the scenario that has been developed.

The Scoring Instruments

During the patient encounter, a physician examiner observes the interaction and scores each candidate according to an Examiner's Checklist, Patient Interaction Rating Scale and a generic global rating scale. The PEPs are usually in one of two formats: short answer write-in responses or extended-match items where candidates must select answers from a list. Another option is completion of important documents, for example writing a prescription.

CASE DEVELOPMENT PROCESS

Step 1: The OSCE Test Committee identifies content areas to be developed.

Step 2: Authors prepare scenarios from a set of guidelines.

Step 3: Authors write and review cases with colleagues and MCCQE Part II staff.

Step 4: Cases are formatted at MCC and then subjected to multiple levels of review.

Step 5: Cases accepted by the Central Examination Committee for inclusion in a blueprint.

Step 6: Center pre-exam review of cases.

Step 7: Post-exam review by the test committee.

MCCQE PART II GUIDELINES FOR AUTHORS

WRITING OSCE CASES

(PLEASE READ THE GUIDELINES BEFORE COMING TO THE CASE WORKSHOP)

PRELIMINARY CONSIDERATIONS

Consider adapting an actual patient case as this often makes case writing easier. You are more likely to provide relevant and consistent information, even when you have to strip out considerable amounts of extraneous information. Remember that first and foremost you are developing a test question.

Being able to clearly state what aspect of clinical skills the case should assess is the critical first step. From that statement comes the stem or test question, which is the information about the presenting problem that candidate's will receive before seeing the patient and must include a clear statement of their task. Based on the stem, you develop the scoring instruments. Given the presenting problem, the task and the time limit, what actions should a candidate carry out to demonstrate their clinical competency? This list of actions can then be refined into a checklist for scoring purposes. Only after this list is well drafted should you work on the patient's history and/or findings.

A critical step to finishing a case is to ensure that the scoring instrument(s) and the patient presentation are linked. Have your provided patient responses to the actions expected of candidates? Have you made suggestions for how the SPs should respond to unexpected but perhaps likely actions? Given the patient's presentation, should you add other items to the scoring instrument(s)?

Lastly, if you are developing a case that will rely on an x-ray, ECG or some other prop, please **find an appropriate one and bring it with you.** Finding props to match developed cases is a painful process and frequently results in a case being abandoned. In short, the recommended order for developing a station is as follows:

1. Draft the Case Information, which includes a statement of purpose.
2. Develop the stem or Candidate Instructions.
3. Draft the scoring instrument (and select rating scale items where appropriate).
4. Outline the Patient Information.
5. Review of checklist and ensure that it is linked to patient information.
6. If applicable, develop the Post Encounter Probe.

MCCQE PART II GUIDELINES FOR AUTHORS

STATION INFORMATION

STATION TYPE	State whether the station is a 5+5 couplet station OR ten minute station (maybe 9+1)
	PATIENT NAME **Occupation:** Specifying the occupation/activities of the patient aids in standardizing the presentation; e.g., university student, salesperson active in sports, accountant, unemployed librarian, homemaker, etc.
SP CHARACTERISTICS	Gender, Age, Physical features Patient Demographics : (In addition to what was provided in the candidate's instructions) Race (if clinically relevant):
PROPS	Describe the prop(s) necessary for the station; e.g. X-ray (s), EKG, photo or lab results. Note that if the station requires a prop; **please bring it with you** to the workshop.
	Medical equipment: Describe medical equipment necessary for the station: exam table , x-ray viewer, IV pole, emergency kit, cardiac dots (to be taped), BP cuff, nasal prongs, stickers for IV identification
TRAINING RESOURCES	If you used a text, journal article or any diagrams, please provide references for MCC staff and SP trainers.
DOMAIN / SPECIALTY	Domain/Indicate one: history, physical, communication, counseling or management History/Indicate one: medicine, surgery, psychiatry, pediatrics, ob-gyn OR other (specify)
PROBLEM / DIAGNOSIS	Provide a brief statement of the presenting problem and state the diagnosis.
*** PURPOSE OF STATION ***	With an actual patient case in mind, write a statement that will focus the examiner on what is expected of the candidate for the specific patient problem you are creating. There may be more than one purpose.
OBJECTIVE(S)	
*** SCORING GUIDELINES ***	Guidelines provide brief criteria for scoring specific items for physician examiners (specify Checklist items #.) Only necessary when an item needs further definition to promote multiple physicians scoring the item reliably and are best written after the peer review.

CANDIDATE'S INSTRUCTIONS

Include patient's age, gender, setting of the encounter, chief complaint, vital signs and/or lab. results if appropriate. State the candidate's task; e.g. take a focused history. For a physical examination include as synopsis of the history that would have been obtained prior to the examination.

E.g., Joseph Trans, 12 years old, has been brought to your <u>office</u> with a history of right hip pain which occasionally radiates to the knee.

IN THE NEXT 5 MINUTES, CONDUCT A FOCUSED AND RELEVANT PHYSICAL EXAMINATION.
As you proceed. EXPLAIN WHAT YOU ARE DOING AND DESCRIBE YOUR FINDINGS

At the next station, you will be asked to answer questions about this patient.

MCCQE PART II GUIDELINES FOR AUTHORS

EXAMINER'S CHECKLIST

1.	Checklist item	SCORES
	-checklist sub-items if appropriate	Assign a relative weight
	-checklist sub-items if appropriate	to each item by specifying
	-checklist sub-items if appropriate	how many marks it is worth

1. Each item should be as specific as the clinical context demands.
2. For 5+5 couplets, 8 to 25 checklist items are recommended.
3. Each item should be written as an action verb; e.g. elicits, specifically asks about, examines, etc.
4. Oral question(s) may be administered in the last minute of a 10 minute station.
   4.1. The question(s) may be asked by the Examiner, the SP or the Nurse (if present).
   4.2. Questions must be related to the patient scenario
   4.3. Questions may relate to clinical problem-solving (e.g., diagnosis) or to CLEO objectives (e.g., "don't tell my insurer.)
5. Items should be discreet (avoid “and / or” statements). Each assessable/observable item should be a separate item
6. Key words indicating the SP’s answer or response must be given for every item, even if the response is simply “yes”, or “Normal”.

PHYSICAL EXAMINATION STATIONS: Items may direct examiner to report to candidates (e.g. “BP is normal) if initiating the action is sufficient for a candidate to receive credit.

HISTORY-TAKING STATIONS: Include a verb; e.g., elicits, means latitude in how candidate gathers the information from the patient; asks, indicates candidate must be specific; e.g., “Are you taking Aspirin®?” versus “Are you taking any medication?”

COUNSELING / PATIENT EDUCATION STATIONS: Items *may* include information to be gathered from the patient and *must* include education and/or advice, and/or support to be given to the patient. Instructions to patients may include questions that the SP will ask all candidates; i.e., questions that cue candidates to certain items.

MANAGEMENT STATIONS: May be acute care (e.g. ER setting) or may require decisions about patient management; for example, resolving problems related to multiple medications. Items *may* include history, physical examination and / or communication actions but *must* include some items regarding decisions, orders, treatment initiatives, etc.

RATING SCALES

Patient Interaction Rating Scales: (see attached list)

Select up to seven items based on the station task and time limit.

Example:

PATIENT / PHYSICIAN INTERACTION
1-RELATIONSHIP TO THE PATIENT
2-ENSURING PHYSICAL COMFORT
3-ORGANIZATION OF EXAMINATION
RATIO INTERACTION/CHECKLIST = 10%

PATIENT INFORMATION

SP STARTING POSITION: Specify where the encounter is taking place: office, emergency, health clinic, other E.g., ER, lying on his back on a stretcher/examining table

Clothing: E.g., wearing sweat pants and socks. No shirt, no shoes.

Make-up: E.g., abrasion to chest, facial pallor and clamminess

PATIENT'S OPENING STATEMENT: This is the brief opening statement that the standardized patient makes. Please use basic, common language. E.g., "I've got this pain in leg and I'm worried."

PATIENT BEHAVIOR, AFFECT AND MANNERISMS: Briefly describing the patient's posture, movement (e.g., restless), and their concerns &/or perception of the problem (e.g., serious, not serious) and their expected response to the candidates (e.g., good eye contact, challenging, attentive). E.g., Tense and anxious.

QUESTIONS PATIENT MUST ASK: "Must ask questions" are prompts SPs must ask all candidates and do cue candidates to the checklist. This ensures that all candidates are assessed on their response to the issue. E.g., "Should I keep taking my mom's Tylenol 3 at night?"

HISTORY OF PRESENT PROBLEM

Information must be from the patient's perspective and include enough information so that the SP can reliably and realistically answer a wide range of questions regarding their problem.

Include (as appropriate): patient issues, onset/duration, progression, frequency, location, radiation, quality and intensity, alleviating/aggravating factors, precipitating incident, prior events, associated symptoms, current drugs and other considerations.

Relevant past medical history:

Focus only on **pertinent positive** information that contributes to the purpose of the station. An unnecessarily detailed history complicates training and makes it more difficult to standardize SP portrayals.

Include (as appropriate) past illnesses, past medications, hospitalizations, allergies, accidents/injuries and/or other considerations.

Relevant social history

As above, but do ensure that **relevant** information about patient is included. (Balance between too much and too little is everything!) Include (as relevant): living environment (where, with whom), sexual history, drug, alcohol and smoking habits, and/or other considerations relevant to the purpose of the station.

Relevant family history

Unless otherwise specified, SPs are instructed to give benign, "boring" family histories. DO provide any pertinent **positive information** about parents, siblings or other family, but only as necessary. Simple information is helpful; e.g., parents alive and well, or uncle died of MI at age 46, or no family history of diabetes, or niece had febrile seizures as in infant, etc.

Critical review of systems

Only provide information NOT included in the above sections and only provide information critical to this patient's presentation. Generally, only pertinent positives are relevant, although if this patient's presentation should include an uncommon or less frequent pertinent negative, please specify.

Review the following systems to consider if you need to add anything to the patient's history or the checklist: HEENT, SKIN, HEMATOPOIETIC / LYMPHATIC, RESPIRATORY, CARDIOVASCULAR, GASTROINTESTINAL, GENITOURINARY, MENSTRUAL / REPRODUCTIVE ENDOCRINE, MUSCULOSKELETAL, NEUROLOGIC, PSYCHIATRIC, SLEEP, ENERGY, APPETITE, & SENSE OF WELL BEING

QUESTIONS PATIENT MAY ASK: "May ask questions" are prompts SPs may use, which do NOT cue the candidate to the checklist. E.g., "Is it serious?"

REVIEW OF CHECKLIST (TO LINK WITH PATIENT INFORMATION)

The checklist must be coordinated with the SP script; i.e., **key words indicating the SPs answer** or response are given on the checklist for every item, even if it is simply "Normal

PHYSICAL FINDINGS THAT MAY BE SIMULATED FOR MCCQE PART II **

Babinski reflex	Muscle weakness
Brudzinski sign	Parkinsonism

Bruises	Perspiration
Coma/unresponsiveness	Petechiae
Confusion	Photosensitivity
Costovertebral-angle tenderness	Pneumothorax
Decerebrate fit	Rebound tenderness
Facial paralysis	Rigidity
Gait abnormalities	Seizures
Hearing Loss	Shortness of breath
Hypomania	Tremor
Lid lag	Visual loss

**Adapted from An Overview of the Uses of Standardized Patients for Teaching and Evaluation Clinical Skills, HS Barrows, ACADEMIC MEDICINE, June 1993.

NOTE: This list is not exhaustive; there may be other symptoms that can be used. Far more symptoms than are listed here can indeed be simulated. However, there are some symptoms that have not been used to date because of concerns about reliable presentation given the strenuous nature of the MCCQE Part II (e.g., hyper/hypoactive reflexes). Authors wanting more information are encouraged to call the MCCQE Part II staff: Sydney Smee, Manager, sydsmee@mcc.ca or Lise Martineau, Test Development Officer, lmartineau@mcc.ca , or call (613) 521-6012.

POST ENCOUNTER PROBE

PEPs are a written 5-minute station that follows a 5-minute patient encounter. General principles are:

1. The PEP must be inextricably linked to the specific patient problem that the candidate has just seen, i.e., generic questions are better to general knowledge tests and should not be included in a PEP.
2. The candidate has only 5 minutes to complete the PEP -- ensure that the candidate can complete the allotted tasks in this time frame.
3. If questions rely on a prop (e.g., an x-ray), the **author should supply this for review at the workshop.**
4. Be aware of a PEP's structure. Does an x-ray inadvertently provide the answer to more than one question? Does the answer/scoring key supply all the likely answers, including those that are worth zero, e.g., must the answer be "colon cancer" for 2 marks or will "cancer" suffice for 2 marks or is "cancer" only worth 1 mark?
5. Write-in PEPs has been the norm for a long time. Consider alternate formats – a task that could be completed prior to seeing the patient (the reverse PEP) or use extended match questions .

	On physical examination baby Kayla weighs 4000 grams (8 lb.) and appears tired; she is otherwise unremarkable.	**Assign a relative weight to each item by specifying how many marks it is worth**
Q1.	**List three conditions that may account for this child's problem.**	
A1.	Cow's milk protein intolerance	4
	Excess, or inappropriately mixed, formula	4
	Infectious gastro-enteritis	
	Lactose intolerance	2
	Gastrointestinal malformation (e.g. malrotation)	1
	Pyloric stenosis	0
	Urinary tract infection	0
	Maximum	

Include good answers, acceptable answers and wrong answers

MCCQE PART II GUIDELINES FOR AUTHORS

EXTENDED MATCH PEP

Like any other PEP, extended match PEPs are based on the preceding patient encounter (AC Station).

The PEP should challenge the candidate's ability to synthesize information gained from the patient; for example, by asking for a diagnosis, a differential diagnoses, a treatment plan, or by asking for key facts gathered from the patient. Questions may also challenge the candidate's understanding of related legal-ethical issues. Each question must specify how many options should be selected.

Stem

For any given PEP a stem can be provided that gives the candidate additional information, for example lab results or the results of initial management. If the task in the AC station was to conduct a physical examination, then the stem may include relevant history data and / or add relevant laboratory data. Vice versa, if the task in the AC station was to take a history, then the stem may provide the candidate with the relevant physical findings and / or add relevant laboratory data.

Questions / Items - Examples

- **select the most likely diagnosis** *from a list of diagnoses organized around the presenting problem and the differential diagnoses for it*
- **select the [author's choice] that is most likely to be** [abnormal **/**defective / deficient/ nonfunctioning] in this patient from the following *list of hormones, list of pathologic processes, etc.*
- **select the result/finding/contributing factor that you would expect given your primary diagnosis** from the *list of laboratory results, list of physical signs, list of risk factors….*
- **select the most likely cause** from the *list of causes, including underlying mechanisms of the disease, list of medications that might cause side effects*
- **select the [e.g., drug] that should be administered.**
- **select the most appropriate next step in patient care.** *e.g., list of pharmacologic therapies, list of laboratory studies, could include a mixed set of treatments and additional studies*

Guidelines for creating answer lists

- no reason to make answers tricky – use single words or very short phrases
- homogeneous content (e.g., all diagnoses, all management options)
- candidate should be able to generate an answer(s) and then find that answer in the list
- only ONE Best answer
    - Or more - the number of best answers should match the number given in the stem
- include a few acceptable answers
- add at least four reasonable distractors
- alphabetical list unless there is a logical order

Scoring

Experience indicates that partial credit scoring is preferable where possible.

If you use all-or-nothing scoring, the items may be extremely difficult, and it is best to require examinees to select only one or two options, rather than more.

Reviewing your work

- select 3 to 4 potential diagnoses (including likely but wrong diagnoses).
- provide answers to all questions keeping in mind the rationale behind each diagnosis.
- add to your answer lists any missing answer that would be logical given each diagnoses.

MCCQE PART II GUIDELINES FOR AUTHORS

RATING SCALES – PATIENT / PHYSICIAN INTERACTION

Selected items from the following patient interaction rating scales are part of the scoring key for most of the stations or patient cases included in the MCC Qualifying Examination Part II. Up to a total of seven items may be selected, depending on the task that candidates are asked to complete.

The OSCE Test Committee weights individual items on their relative importance to each other for a specific station. The combined items are also weighted relative to the task-specific checklist for that station. The weight for the combined items ranges from 10% to 50% of the total score for any one station.

Initiation of interview

0	1	2	3	4	5
Lack of introduction	Minimal acknowledgement of patient	Borderline unsatisfactory, Acknowledges patient introduces self	Borderline satisfactory, Acknowledges patient introduces self	Acknowledges patient, moderately at ease & attentive	Attentive to patient, introduces self, at ease, personable

Listening skills

0	1	2	3	4	5
Interrupts inappropriately, ignores patient's answers	Impatient	Borderline unsatisfactory Somewhat attentive	Borderline satisfactory Somewhat attentive	Attentive to patient's answers	Consistently attentive to answers & concerns

Questioning skills

0	1	2	3	4	5
Awkward, exclusive use of leading or closed ended questions, jargon	Somewhat awkward, inappropriate terms, minimal use of open-ended questions	Borderline unsatisfactory, moderately at ease, appropriate language, uses different types of questions	Borderline satisfactory, moderately at ease, appropriate language, uses different types of questions	At ease, clear questions, appropriate use of open and closed ended questions	Confident and skilful questioning

Organization of interview

0	1	2	3	4	5
Scattered, shot-gun approach	Minimally organized	Borderline unsatisfactory, flow is somewhat logical	Borderline satisfactory, logical flow	Logical flow with sense of purpose	Purposeful, integrated handling of encounter

Organization of physical examination

0	1	2	3	4	5
Scattered, patient moved unnecessarily	Minimally organized	Borderline unsatisfactory, flow is somewhat logical	Borderline satisfactory, logical flow	Logical flow with sense of purpose	Purposeful, integrated handling of examination

Demonstration of technical skills (NOT CLEO)

0	1	2	3	4	5
No skill Manoeuvres cannot provide reliable / useful information	Manoeuvres too rushed or clumsy, unlikely to provide reliable/ useful information	Borderline Unsatisfactory: Some skill but likelihood of reliable / useful findings minimal	Borderline Satisfactory: Some skill, some reliable / useful findings likely	Consistent skill, manoeuvres likely to provide reliable / useful information	Consistent skill, manoeuvres performed will elicit reliable / useful information

MCCQE PART II GUIDELINES FOR AUTHORS

RATING SCALES – CLEO ITEMS

CLEO: Considerations of Legal, Ethical and Organizational aspects of the practice of medicine.

Rapport with person (CLEO)

0	1	2	3	4	5
Condescending, offensive, judgemental	Minimal courtesies only	Borderline unsatisfactory	Borderline satisfactory	Polite and interested	Warm, polite, empathic

Information giving (CLEO)

0	1	2	3	4	5
No attempt or inappropriate attempt to give information; e.g., not truthful	Awkward and / or incomplete attempts to give information	Borderline unsatisfactory, somewhat at ease, attempts to give information	Borderline satisfactory, somewhat at ease, attempts to give information	Gives information easily, somewhat attentive to patient's understanding	Confident and skilful at giving information, attentive to patient's understanding, truthful

Professional behaviour (CLEO)

0	1	2	3	4	5
Offensive or aggressive; frank exhibition of "unprofessional conduct"	Negative attitude to patient	Borderline unsatisfactory, does not truly instil confidence	Borderline satisfactory, manner inoffensive but does not necessarily instil confidence	Attempts professional manner with some success	Overall demeanour of a professional, caring, listens, communicates effectively

Ethical conduct (CLEO)

0	1	2	3	4	5
Markedly inappropriate or awkward handling of ethical issues	No consideration to ethical issues	Borderline unsatisfactory, minimal consideration of ethical issues	Borderline satisfactory, minimal and appropriate consideration of ethical issues	Responds to ethical issues satisfactorily	Considers and responds to ethical issues with care and effectiveness

Compliance optimization (CLEO) [Did the candidate do everything possible to optimize the patient's compliance?]

0	1	2	3	4	5
Candidate's approach may negatively affect compliance	Patient's compliance unlikely to be optimized	Weak attempt to encourage patient's compliance	Candidate's approach may positively affect patient's compliance	Candidate's approach encourages patient's compliance	Candidate's approach highly likely to optimize patient's compliance

Relationship to the patient (CLEO)

0	1	2	3	4	5
Introduction absent or inappropriate, no consent, awkward, uses jargon, no interaction or acknowledgment of patient	Minimal interaction and/or minimal acknowledgment of patient	Borderline unsatisfactory in approach to patient	Borderline satisfactory in approach to patient	Moderately clear and understandable, acknowledges patient, moderately at ease with patient	Clear, concise instructions, elicits consent to physical examination, at ease with patient

Attention given to patient's physical comfort (CLEO)

0	1	2	3	4	5
Inattentive to patient's comfort or dignity; e.g., no draping and/or causes pain unnecessarily	Causes some unnecessary discomfort or embarrassment	Borderline unsatisfactory in attending to patient's comfort and needs	Borderline satisfactory in attending to patient's comfort and needs	Mostly attentive to patient's comfort and dignity	Consistently attentive to patient's comfort and dignity

Lastly, if you are developing a case that will rely on an x-ray, ECG or some other prop, please find an appropriate one and bring it with you. Finding props to match developed cases is a painful process and frequently results in a case being abandoned. In short, the recommended order for developing a station is as follows:

1. Provide your information (see Author Information below).
2. Draft the Case Information, which includes a statement of purpose.
3. Develop the stem or Candidate Instructions.
4. Draft the scoring instrument (and select rating scale items where appropriate)
5. Outline the Patient Information
6. Review of checklist and ensure that it is linked to patient information.
7. If applicable, develop the Post Encounter Probe.

AUTHOR INFORMATION

NAME:

SPECIALTY:

AREA OF PRACTICE:

ADDRESS:

PHONE (Work):

FAX:

E-MAIL:

SAMPLE

CONFIDENTIAL

EXAMPLE OF MCCQE PART II CASE
PHYSICAL EXAMINATION
5+5 / COUPLET

	SP INFORMATION
	Marie Perone / Student
SP CHARACTERISTICS	F-16
ROTATION	Shared between 2 SPs
TRAINING	1 training session + DRY RUN
SP STARTING POSITION	In emergency department; lying on back with knees slightly flexed.
PROPS	**Make-Up:** flushed face **Clothing:** hospital gown, underwear and socks on **Furniture:** stretcher **Secure props:** x-ray **Medical equipment:** x-ray viewer
TRAINING RESOURCES	N/A

	STATION INFORMATION
TYPE	5 + 5 / couplet
DOMAIN / SPECIALTY	Physical Examination / Pediatrics
PROBLEM / DIAGNOSIS	Adolescent with abdominal pain, likely cause is appendicitis.
PURPOSE OF STATION	To perform a physical examination of the abdomen and identify positive findings.
OBJECTIVES	MCC – 003A Acute abdominal pain
SCORING GUIDELINES	***Item #6**:* SP will not permit deep palpation because it is too painful. Credit should be given to candidates who attempt to palpate deeply.

CANDIDATE'S INSTRUCTIONS

Marie Perone, 16 years old, has come to the Emergency Room with a 16 hour history of abdominal pain.

IN THE NEXT FIVE MINUTES, CONDUCT A FOCUSED AND RELEVANT PHYSICAL EXAMINATION.

As you proceed, EXPLAIN WHAT YOU ARE DOING AND DESCRIBE YOUR FINDING

At the next station you will be asked to answer questions about this patient.

SAMPLE

SAMPLE

CONFIDENTIAL

PATIENT INFORMATION

PATIENT'S OPENING STATEMENT: "My stomach really hurts and it's getting worse."

PATIENT BEHAVIOUR, AFFECT AND MANNERISMS: Patient is uncomfortable, and a little scared.

HISTORY OF PRESENT PROBLEM

Marie Perone felt uncomfortable last night (16 hours ago) because of a dull, crampy ache in the centre of her stomach. It has been constant since then although she was able to sleep last night. She did get up once and threw up a little bit. Since then the pain has become sharp and it has moved lower down to the right lower quadrant. She has never had this kind of pain before, it is a level 6. She has noticed that the pain is worse if she coughs or moves quickly. She is feeling kind of feverish and did throw up after trying to eat this morning. She does not feel like eating now.

SIMULATION

Marie is lying on back with knees slightly flexed, it is her position of choice. If asked to sit up, she moves carefully and sits slightly bent over.

The pain is worse if she coughs, moves quickly, jumps up and down, takes a deep breath, and with Valsalva manoeuvre. Sucking in her belly hurts, blowing it out does not hurt, she can do both manoeuvres. If either the bed or she herself is shaken, it hurts her right side.

Light palpation - minimal tenderness in RLQ

Deep palpation - causes pain, there is guarding

Rebound - Pain is felt in area of inflamed appendix when this test is done, there is no referred pain.

Patient demonstrates positive right sided psoas sign: When right knee raised against resistance there is pain on the right side in area of appendix. Same result if patient is on her left side and her right leg is extended behind her.

Patient demonstrates positive obturator sign: Pain low on the right (closer to midline) with right sided flexion and internal rotation of hip.

How to answer questions: Since candidates only have 5 minutes, **SPs should only answer questions related to the physical exam** (e.g., "is it painful when I press here?"). SPs should NOT respond to questions regarding the PROBLEM, the DIAGNOSIS or the MEDICATION, or to HISTORY questions, unless it is to repeat what is already given in the Candidate's Instructions. It is always better deflect questions during a physical examination.

If asked history questions (e.g., "How did it start?"), SPs will deflect questions. He should not give history information but answer with "I hope this physical will help to find what's wrong" **OR** "Could we go over your questions when you've finished examining me? "

Questions patient MAY ask: SP should not ask question since this is a physical examination

PERTINENT NEGATIVES: The absence or presence of certain symptoms allows the candidate to make a diagnosis. Only the symptoms **present** are mentioned in the script. Therefore when the candidate takes a history and asks about certain symptoms which are not in the case, the answer should be "no", "I'm not sure", etc.

Negative findings that are important in this case: (SPs don't have to memorize the list.).

ADD IN IF APPLICABLE

SAMPLE

SAMPLE

CONFIDENTIAL

EXAMINER CHECKLIST

ITEM		SCORE
1.	**Drapes patient appropriately** [i.e., breast and pubic area covered]	2
2.	**Inspects abdomen** [ADD FINDINGS]	1
3.	**Auscultates abdomen** [ADD FINDINGS]	1
4.	**Percusses abdomen** [ADD FINDINGS]	1
5.	**Lightly palpates each quadrant** [ADD FINDINGS]	2
6.	**Deeply palpates each quadrant** [EXAMINER: see scoring guidelines]	2
7.	**Palpates for liver enlargement** [ADD FINDINGS]	1
8.	**Palpates for spleen enlargement** [ADD FINDINGS]	1
9.	**Attempts to elicit signs of peritoneal irritation** [any of guarding, percussion tenderness or rebound] (Rovsing's sign, psoas sign]	2
10.	**Palpates costovertebral angles for tenderness** (no findings)	2
11.	**Examines inguinal region** (no findings)	2
12.	**Asks for results of rectal examination** [EXAMINER: report: mild tenderness on right side]	2
13.	**Asks for results of genital examination** [EXAMINER: report: no discharge, no masses, normal cervix; mild tenderness on bimanual palpation on right side.]	2

SAMPLE

SAMPLE

CONFIDENTIAL

POST ENCOUNTER PROBE

	Marie Perone's vital signs are as follows: **BP: 120/80. Temp: 38.2°C. Pulse: 100/minute** **Her laboratory results are:** **White blood count: 13,500** **Hemoglobin 125** **Urinalysis showing no RBC's, 2-3 WBC per high-power field, and no bacteria.**		
Q1.	**Examine the x-ray of this patient and describe the abnormalities if any, and their location. If normal, state so.**		
A1.	X-ray of abdomen shows FECALITH in right lower quadrant	3	
	ANSWERS NEEDED -WRONG OR ACCEPTABLE		
	Maximum	**3**	
Q2.	**What is your working diagnosis?**		
A2.	Acute appendicitis	3	
	ANSWERS NEEDED -WRONG OR ACCEPTABLE		
	Maximum	**3**	
Q3.	**List the 5 most likely conditions to be included in the differential diagnosis**		
A3.	Crohn disease	1	
	Ectopic pregnancy	1	
	Meckel diverticulitis	1	
	Mesenteric adenitis	1	
	Mittelschmerz	1	
	Pelvic Inflammatory Disease	1	
	Ruptured ovarian cyst	1	
	Maximum	**5**	
Q4.	**Outline your initial management of this patient.**		
A4.	Surgical consultation	4	
	Admit patient for observation with frequent re examinations	2	
	Send patient home with return visit in 48 hours	0	
	Maximum	**4**	

SAMPLE

SAMPLE

CONFIDENTIAL

COMMENTS

PATIENT HISTORY INFORMATION:
REMOVED FROM THIS CASE SINCE IT IS A PHYSICAL EXAMINATION STATION
Her menstrual periods started at 12, have been regular for 2 years. Her last period started 11 days ago and finished 6 days ago. She is seeing a boyfriend and has just started having sex, but only a few times and not this week.

She has been taking Ortho Novum for 5 months. She got the prescription from a friend's doctor, and her parents do not know about the pills.

Her health is excellent, there is no family history of relevance. She has never had anaesthetic or been to hospital prior to this episode.

SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE

CONFIDENTIAL

EXAMPLE OF MCCQE PART II CASE
HISTORY STATION
5+5 / COUPLET

	SP INFORMATION
	Luc Leger / Sales adminsitrator
SP CHARACTERISTICS	M-59, (somewhat overweight, perhaps 40-60 pounds)
ROTATION	One SP
TRAINING	1 training session + DRY RUN
SP STARTING POSITION	Office setting
PROPS	**Make up:** jaundiced **Clothing:** casual **Furniture:** room with three chairs and room with chair and table **Medical equipment:** N/A
TRAINING RESOURCES	N/A

	STATION INFORMATION
DOMAIN/SPECIALTY	History/Surgery
PROBLEM/DIAGNOSIS	Jaundice
PURPOSE OF STATION	To elicit a history from a patient with jaundice to establish differential diagnosis.
OBJECTIVES	MCC - 049 Jaundice
SCORING GUIDELINES	N/A

SAMPLE

CANDIDATE'S INSTRUCTIONS

Luc Leger, 59 years old, has come to your office complaining that is has jaundice.

IN THE NEXT 5 MINUTES, OBTAIN A FOCUSED AND RELEVANT HISTORY.

At the next station, you will be asked to answer questions about this patient.

SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE

CONFIDENTIAL

PATIENT INFORMATION

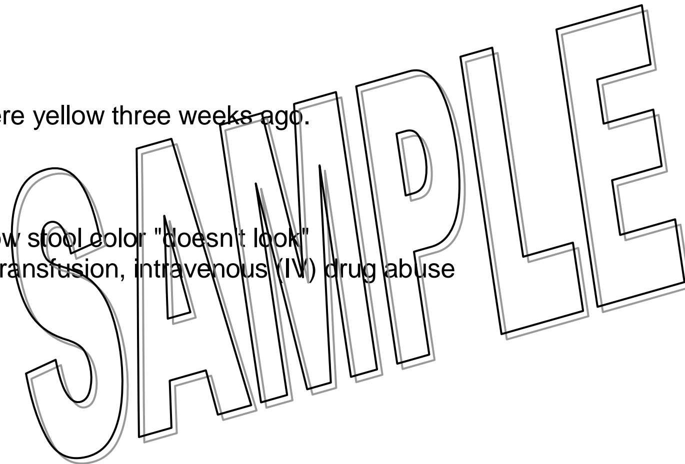
PATIENT'S OPENING STATEMENT: "I think that I have jaundice."

PATIENT BEHAVIOUR, AFFECT AND MANNERISMS: The patient is very concerned about the jaundice. He wants to get to the bottom of this and is very eager to answer all questions.

QUESTIONS PATIENT MAY ASK: Do I have cancer? The patient asked the doctor two or three times during the interview.

HISTORY OF PRESENT PROBLEM

Chief complaint: My niece told me my eyes were yellow three weeks ago.
Onset: 3 weeks ago
Progression: Unchanged
Frequency: Continuous
Associated symptoms: Dark urine; doesn't know stool color "doesn't look"
Other: No recent exposure to hepatitis, blood transfusion, intravenous (IV) drug abuse



RELEVANT PAST MEDICAL HISTORY
Heathy

RELEVANT SOCIAL HISTORY
Habits: No drugs, no alcohol

CRITICAL REVIEW OF SYSTEMS
HEENTS: Yellow sclera
SKIN: Yellow skin pruritis
GI: Mild nausea, mild anorexia no pain, no vomiting, no food intolerance
GENITO-URINARY: Tea colored urine
SLEEP, ENERGY, APPETITE, SENSE OF WELL BEING ETC: Decreased energy, decreased appetite, weight loss of a few pounds

PERTINENT NEGATIVES

The absence or presence of certain symptoms allows the candidate to make a diagnosis. Only the symptoms **present** are mentioned in the script. Therefore when the candidate takes a history and asks about certain symptoms which are not in the case, the answer should be "no", "I'm not sure", etc.

Negative findings that are important in this case: (SPs don't have to memorize the list.). SPECIFY IF APPLICABLE

SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE

CONFIDENTIAL

EXAMINER'S CHECKLIST

ITEM		SCORE
1.	**Establishes onset/duration**	2
2.	**Establishes progression**	2
3.	**Elicits associated symptoms**	
	-dark urine	2
	-pain	2
	-color of stool	2
	-fever	2
4.	**Asks about risk factors**	
	-previous exposure to hepatitis	2
	-recent blood transfusion	2
	-intravenous drug use	2
	-foreign travel	2
5.	**Specifically asks about alcohol use**	2
6.	**Review of systems**	
	-skin	2
	-gastrointestinal	2
	-general	2
	-weight loss	2
	-change in appetite	2

SAMPLE

SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE SAMPLE

CONFIDENTIAL

POST ENCOUNTER PROBE

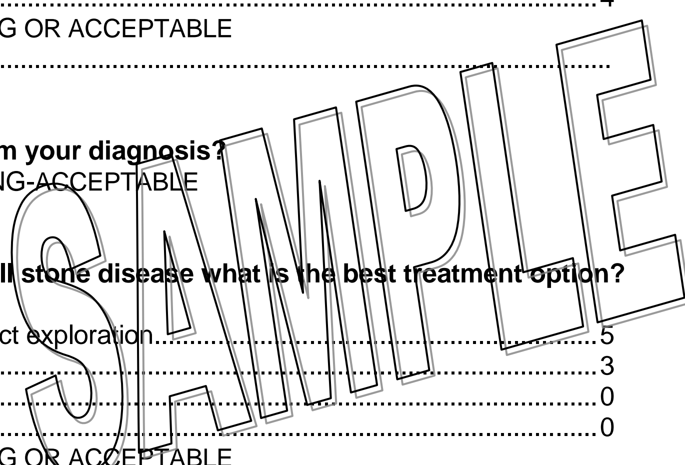
ITEM **SCORE**

Q1. **Based on the history that you have obtainedand given an unremarkable abdominal examination, with no tenderness and no masses, what are the TWO most likely diagnoses?**

A1. Tumor of either bile duct (ampulla, perianpullary) or pancreas 4
Choledocholithiasis 4
MORE ANSWERS NEEDED - WRONG OR ACCEPTABLE
Maximum...

Q2. **What test would you order to confirm your diagnosis?**
A2. ANSWERS NEEDED – GOOD-WRONG-ACCEPTABLE

Q3. **If all the symptoms are related to gall stone disease what is the best treatment option? (be specific)**
A3. Cholecystectomy with common bile duct exploration..5
ERCP with stone extraction 3
Cholecystectomy 0
Cholecystostomy 0
MORE ANSWERS NEEDED - WRONG OR ACCEPTABLE
Maximum...**4**



CASE WRITER:

STATION INFORMATION (Length in time?)	
PATIENT OCCUPATION: (Specifying the type of occupation may help standardize the case)	
SP CHARACTERISTICS (e.g., no abdominal scars, required Body Mass Index)	
NUMBER OF SPs (SP Trainer Field)	
TRAINING SESSIONS (SP Trainer Field)	
PROPS (Indicate any special equipment required for your case / e.g., nasal prongs, make-up)	
CASE PROPS PROVIDED (e.g., x-ray, ECG)	
TRAINING RESOURCES (SP Trainer Field)	
	STATION INFORMATION
DOMAIN / DISCIPLINE (e.g., Management / Medicine) Domain (History, Physical, Management, Hx+ Phys) Discipline (Medicine, Surgery, Psychiatry, Pediatrics, Ob-gyn)	
PROBLEM / DIAGNOSIS	
PURPOSE OF STATION	
OBJECTIVES	
SCORING GUIDELINES for EXAMINER	

*** CANDIDATE'S INSTRUCTIONS ***

- MUST include the following information
  - Patient's name and age
  - Where (e.g., office, walk-in clinic, ER)
  - Time allowed (e.g., IN THE NEXT TEN MINUTES)
  - Task (e.g., take an history, physical, counselling)

- MAY include other RELATED information (e.g., vital signs, lab results or as synopsis of the history that would have been obtained prior to a physical examination)

*** EXAMINER'S CHECKLIST ***

I#	Checklist items	SCORE

*** POST ENCOUNTER PROBE * (WRITE-IN OR EXTENDED MATCH PEP)**

Q1.		SCORES
A1.		

PATIENT INFORMATION

WHAT CANDIDATES KNOW: Leave blank
WHAT CANDIDATE SHOULD BE DOING: Leave blank

SP STARTING POSITION:

MAKE-UP:

PATIENT'S CLOTHING:

PATIENT'S OPENING STATEMENT:

PATIENT BEHAVIOR, AFFECT AND MANNERISMS:

HISTORY OF PRESENT PROBLEM
Include patient's belief about the problem, a timeline, and remember to use patient-based language

SIMULATION (e.g., pain location, pain with specific maneuvers, ROM, pain intensity, splinting)

The Use of Standardized Patient Assessments for Certification and Licensure Decisions

John R. Boulet, PhD;
Sydney M. Smee, PhD;
Gerard F. Dillon, PhD;
John R. Gimpel, DO

Although standardized patients have been employed for formative assessment for over 40 years, their use in high-stakes medical licensure examinations has been a relatively recent phenomenon. As part of the medical licensure process in the United States and Canada, the clinical skills of medical students, medical school graduates, and residents are evaluated in a simulated clinical environment. All of the evaluations attempt to provide the public with some assurance that the person who achieves a passing score has the knowledge and/or requisite skills to provide safe and effective medical services. Although the various standardized patient-based licensure examinations differ somewhat in terms of purpose, content, and scope, they share many commonalities. More important, given the extensive research that was conducted to support these testing initiatives, combined with their success in promoting educational activities and in identifying individuals with clinical skills deficiencies, they provide a framework for validating new simulation modalities and extending simulation-based assessment into other areas.
(*Sim Healthcare* 4:35–42, 2009)

Key Words: Licensure, Certification, Simulation, Standardized patient, Simulated patient, OSCE

There are many types of simulations that are currently being used to assess healthcare professionals.[1–4] In both Canada and the United States (US), many of these simulation modalities, including multiple choice questions, part-task trainers, and computer-based case simulations, have been used as part of the examination process used to certify and license physicians.[1,5,6] These simulation-based examinations, which can vary somewhat in terms of purpose and focus, all attempt to provide the public with some assurance that the person who achieves a passing score has the knowledge and/or requisite skills to provide safe and effective medical services, either independently or under supervision. Here, as with any simulation-based assessment, the structure, content, fidelity, and difficulty of the modeled exercises, combined with the scores, will determine what inferences one can make about the individual test taker.

From a simulation perspective, the use of standardized patients (SPs) for certification and licensure decisions has been a relatively recent phenomenon.[7] Historically, SP-based assessments were implemented as part of formative evaluation activities.[8–10] Individuals were trained to portray specific patient conditions, allowing medical students to practice their clinical skills and receive immediate feedback concerning strengths and weaknesses. In the 1980s, with an increased emphasis on evaluating what medical trainees could do, as opposed to what they knew, various organizations started research programs aimed at determining how assessments employing SPs could be structured to make valid skills-based proficiency decisions. Over the next two decades, the end result of these research activities was the implementation of a number of high-stakes assessments all aimed at measuring abilities in key clinical skills domains. Although these research efforts required extensive resources, they were successful in identifying the specific conditions and structures that are needed to produce defensible scores and decisions for multistation, performance-based, simulation activities.[11–17]

The introduction of SP-based certification and licensure examinations in medicine was a monumental achievement. Although other high-stakes simulation-based assessments have been developed and used in other professions, the logistical, economical, and psychometric challenges associated with national multistation clinical skills assessments were staggering.[18,19] Organizations that built these assessments all had to address concerns regarding test content (eg, types of scenarios to model), test administration models (eg, fixed versus temporary sites; number, timing and sequencing stations), measurement rubrics (eg, holistic or analytic), eligibility requirements, scoring models (eg, compensatory or conjunctive), and the establishment of defensible standards, just to name a few. Nevertheless, even with these hurdles, and despite numerous objections concerning the need to measure

From the Foundation for Advancement of International Medical Education and Research (J.R.B.), Philadelphia, PA; Medical Council of Canada (S.M.S.), Ottawa, ON, Canada; National Board of Medical Examiners (G.F.D.), Philadelphia, PA; and National Board of Osteopathic Medical Examiners (J.R.G.), Conshohocken, PA.

Reprints: John R. Boulet, PhD, Foundation for Advancement of International Medical Education and Research, 3624 Market Street, Philadelphia, PA 19104 (e-mail: jboulet@faimer.org).

The authors have indicated that they have no conflict of interest to disclose.

Copyright © 2009 Society for Simulation in Healthcare

DOI: 10.1097/SIH.0b013e318182fc6c

Copyright © Society for Simulation in Healthcare. Unauthorized reproduction of this article is prohibited

clinical skills as part of certification/licensure process,[20] each of these organizations was able to produce a high-quality simulation-based assessment that was appropriate for their particular needs. In doing so, many lessons were learned, the most important being that simulation-based summative assessment of clinical skills was viable, even with large examinee populations, differing testing purposes, and varying examination administration protocols.

PURPOSE

The purpose of this article was to describe and contrast the Clinical Skills Assessment (CSA) programs that are employed in Canada and the US as part of the certification and licensure process for physicians. These assessments include the Medical Council of Canada (MCC) Qualifying Examination Part II (MCCQE Part II),[21] the United States Medical Licensing Examination (USMLE) Step 2 Clinical Skills (USMLE Step 2 CS),[22] and the National Board of Osteopathic Medical Examiners (NBOME) Comprehensive Osteopathic Medical Licensing Examination Level 2-Performance Evaluation (COMLEX-USA Level 2-PE).[23] To better understand the USMLE Step 2 CS, a brief overview of the Educational Commission for Foreign Medical Graduates (ECFMG) CSA is also provided.[24] The CSA was used to assess the clinical skills of international medical graduates (IMGs) before the introduction of USMLE Step 2 CS. Following this overview, a brief synthesis of the similarities and differences in the assessments and assessment programs is provided. With these distinctions in mind, and knowing the success and scope of the individual testing programs, it is possible to envision where summative simulation-based assessment activities could be enhanced, applied in other areas, and used for the evaluation of nonphysician healthcare professionals.

ASSESSMENT OF CLINICAL SKILLS

In general terms, clinical skills refer to information gathering and communication skills, applied during the patient encounter, that help to establish an accurate diagnosis and support high-quality treatment. Within the medical education and practice community, these skills have long been recognized as essential to patient care. Several organizations, including those responsible for the accreditation of undergraduate and graduate medical education (GME) programs, have included clinical skills among the competencies deemed important to the education and assessment of practicing physicians.[25–27] As a result, it is not surprising that considerable efforts have been made to develop, and subsequently defend, testing methods than can be used to reliably and validly measure these skills.

STANDARDIZED PATIENTS

SPs, often referred to as simulated patients or programmed patients, are people who have been trained to accurately portray the role of a patient with a specific medical condition or conditions. The term "standardized" refers to the fact that the person is specifically trained to model the "real" patient's condition, including symptoms and emotional states, and to do so consistently over time. Examinees who interview the same SP with the same presenting complaint will receive, on questioning, the same patient history. The physical findings relevant to the case, either real or simulated, need to be stable and, for a given modeled scenario, they must not vary from one SP to another.

LARGE-SCALE SP EXAMINATIONS

Medical Council of Canada Qualifying Examination Part II

Since 1912, the MCC has been setting an examination that is a prerequisite for medical licensure in Canada; the Licentiate of the MCC is granted to those who successfully complete it. In 1992, the MCC added the Qualifying Examination Part II (MCCQE Part II) to the assessment sequence. Initially the MCCQE Part II was a 20-station Objective Structured Clinical Examination (OSCE).[7,28] Although the use of OSCEs is now commonplace throughout the world, implementing a national summative, performance-based, assessment based on a series of SP encounters was, at the time, unprecedented. The impetus for implementing the MCCQE Part II came largely from the licensing authorities. In the late 1980s, because of the number and nature of related complaints that they received each year, members of these authorities began calling for an assessment of clinical and communication skills. The existing paper-and-pencil test of medical knowledge and problem solving (MCC Qualifying Examination Part I—MCCQE Part I) was not sufficient to address the emergent belief that candidates for medical licensure should be assessed more broadly.

To qualify for the MCCQE Part II, candidates must have completed successfully 12 months of postgraduate clinical training and passed the MCCQE Part I, currently a computer-adaptive test of knowledge and clinical decision-making. The number of candidates who qualify for the MCCQE Part II continues to grow. In 1992, 401 candidates took the examination. In 2007, 3481 candidates completed this assessment, a more than eightfold increase.

As the measurement qualities of the MCCQE Part II became better understood, the number of stations was reduced from 20 to 14, and is now set at 12. This reduction in station length could be attributed to evolving test development processes, allowing for a more efficient and appropriate targeting of test content to examinee ability. Each station is based on a clinical problem presented by a SP; scoring is completed by physicians who observe from within the room. Checklists and rating scales are used to generate the station scores. At this time, the MCCQE Part II is comprised of eight 10-minute encounters with a SP and six couplet stations that include a 5-minute encounter with a SP followed by a 5-minute written component (Two of the stations in the assessment, including one of the couplets, are used for pilot testing purposes). Four domains are assessed based on common presenting problems: history-taking skills, physical examination skills, patient management, and doctor-patient interactions. Patient safety issues and professionalism are also evaluated.

Each scored station, while potentially measuring slightly different skill sets, counts equally in terms of generating a total score. Although station scores are compensatory, mean-

Copyright © Society for Simulation in Healthcare. Unauthorized reproduction of this article is prohibited

ing poor performance in one station may be compensated by superior performance in another, the overall pass/fail decision is based on a conjunctive standard; candidates must pass both by total score (the sum of their station scores) and by the number of stations passed.

Results from the MCCQE Part II are reported as a standard score (mean = 500, standard deviation = 100). The examination is criterion-referenced, with the individual station pass marks set using the borderline group method.[29] Candidates receive a bar graph indicating their performance in each of four domains relative to the mean score for their testing cohort. The four domains are data gathering (from history taking and physical examination tasks), patient interaction (from rating scale items across stations), problem-solving and decision-making (based on certain stations; eg, acute care of trauma and the written work from the couplet stations), and legal, ethical, and organizational issues (which comprises a minimum of 10% of the total score). More extensive feedback is provided to those candidates who are unsuccessful; specifically, they are told which stations they failed and are provided with a more extensive description of the four domains.

To balance accessibility and costs, a multisite, fixed test form model with two administrations per year is employed. In the spring, one test form is administered twice over 1 day at 10 university sites across Canada. At most sites, the examination runs in two or more parallel tracks. In the fall, there are two test forms, one for each of 2 days of testing, and the examination runs at 16 sites. In spring, over 500 SPs are trained to simulate the patient problems. Twice that many are recruited for the fall. Ensuring that the SPs present their problems consistently and with sufficient fidelity for valid testing is critical. Each site has its own trainers who recruit and prepare the SPs according to the protocols developed centrally. Training videos, meetings with MCC staff, consultation with supervising physicians, along with telephone support are all part of a process aimed at ensuring the SPs are ready for the examination.

Like all large-scale testing programs, there have been some administrative challenges. Developing feasible, psychometrically sound cases (simulated scenarios) is an ongoing task and takes considerable time and effort. Because the MCCQE Part II is a national examination, the scoring instruments and the supporting materials for SP training are developed centrally by a multidisciplinary test committee. Cases range from those requiring relatively little simulation (eg, history of diarrhea) to those where the SP must accurately simulate specific patient presentations (eg, shortness of breath, decreased consciousness, pain, anxiety).

The MCC is continuously assessing different aspects of the MCCQE Part II. Numerous research studies suggest that both valid and reliable competency decisions are being made.[30–32] Most recently, the predictive validity of the MCCQE Part II was investigated by looking at the relationship between MCCQE Part I and Part II scores and complaint records from two licensing jurisdictions.[33] The authors concluded that poor performance on the MCCQE Part II patient-physician communication component and the clinical decision-making component from the MCCQE Part I were predictors for complaints.

Educational Commission for Foreign Medical Graduates Clinical Skills Assessment

Based on several years of extensive research and consultation with the MCC, the ECFMG CSA was instituted in July 1998.[34,35] This 11 station clinical skills examination was developed to evaluate whether graduates of international medical schools (IMGs) possessed the skills necessary to enter supervised GME programs in the US. Successful completion of this examination became one of the required elements for ECFMG certification. Initially, the assessment was offered at one fixed site in Philadelphia, Pennsylvania. In 2002, in collaboration with the National Board of Medical Examiners, a second testing site was constructed in Atlanta, Georgia. Between 1998 and 2004, 43,624 IMGs were tested (37,930 first-time takers) in a total of 372,674 simulated clinical encounters. During this time, numerous studies were published, several providing evidence to support the validity of the assessment scores.[36–38] Of particular note, research was conducted to show that SP and physician evaluations of clinical skills were comparable.[39] In 2004, administration of the ECFMG CSA ceased. Instead, IMGs were required to take and pass USMLE Step 2 CS (described below), a similar simulation-based assessment that was developed to measure the clinical skills of American allopathic medical students and graduates. The USMLE Step 2 CS examination is part of the USMLE sequence (There are three "Steps" to the USMLE. Step 1 is intended to assess whether the examinee understands and can apply important concepts of the sciences basic to the practice of medicine. Step 2 focuses on the examinee's knowledge, skills, and understanding of clinical science essential for provision of patient care "under supervision"—typically the point that medical school graduates begin their postgraduate education and experience. Step 3 is intended to assess whether the examinee can apply medical knowledge and understanding of biomedical and clinical science essential for the unsupervised, independent practice of medicine.) To qualify for a medical license to practice in the US, graduates of MD-granting schools in the US and graduates of medical schools located outside the US must take and pass all components of USMLE.

United States Medical Licensing Examination Step 2 Clinical Skills

From the time that introduction of the USMLE program was first proposed in the late 1980s, it was the intent of the National Board of Medical Examiners and the Federation of State Medical Boards (the organizations that sponsor USMLE) to include clinical skills among the areas assessed as part of the examination program supporting the US medical licensing system. After many years of development, this goal became a reality in June 2004 when USMLE Step 2 CS was administered for the first time.[40] At this point, the previously existing Step 2 examination, a 1-day, computer-based multiple choice questionnaire test, was renamed the Step 2 Clinical Knowledge examination. The introduction of Step 2 CS in the USMLE sequence was informed by the research of many organizations interested in the assessment of these important

© 2009 Society for Simulation in Healthcare

Copyright © Society for Simulation in Healthcare. Unauthorized reproduction of this article is prohibited

skills and by the operational experiences of organizations that brought this type of format to the arena of large-scale, high-stakes assessment, in particular, the MCC and the ECFMG.[41–43]

The USMLE Step 2 CS examination, which is delivered at each of five regional testing centers (Atlanta, Chicago, Houston, Los Angeles, and Philadelphia), requires test takers to move through a series of 12 simulated encounters (stations), interacting with SPs, individuals who are trained to portray real patients. Examinees are given up to 15 minutes to interact with each SP. During that time they are expected to take a history and to perform a physical examination that is focused on the chief complaint of the patient and on the information that is revealed during the encounter. After the simulated encounter, examinees are given 10 minutes to write a patient note that summarizes and synthesizes their findings, including possible diagnoses. The mix of cases seen by any one examinee is guided by a group of content experts who are charged with overall design and development of Step 2. Based on a test blueprint established by this committee, each test form contains a blend of patient presentations that would not be uncommon for clinical practice in the US. This same committee is involved in the process used to establish passing standards.[44,45] Because the Step 2 CS examination is offered daily across five sites, a variable test form administration model is used. The test form (mix of clinical presentations and SP characteristics) for any given administration, at any site, is individually constructed to meet blueprint specifications. Efforts are made to minimize case and SP exposure for previously failing examinees who are repeating the assessment.

USMLE Step 2 CS examinees are required to pass three subcomponents: the integrated clinical encounter, which includes demonstration of skills in history taking, physical examination, and documentation; communication and interpersonal skills, which includes skills in information gathering/sharing and establishing rapport; and spoken English proficiency, which requires clear communication with the patient. With the exception of the postencounter notes, which are scored by a group of physicians who are specially trained to the specifics of the case, all scoring is done by the SPs who are extensively trained and monitored in their use of a series of checklists and rating scales that were specifically designed for gathering reliable and valid measures of these components. To pass the USMLE Step 2 CS, an examinee must pass all of the three subcomponents (integrated clinical encounter, communication and interpersonal skills, and spoken English proficiency) in a single administration. Failing examinees are provided with feedback outlining relative strengths and weaknesses in the various clinical skills components that are measured.

The USMLE Step 2 CS program has been fully operational for almost 4 years, delivering, scoring, and reporting results year round. More than 120,000 examinations have been administered, representing more than 1.4 million examinee-SP encounters. Because of the complexities of an overall system that handles, at any one time, thousands of examinees, hundreds of SPs, and multiple testing centers, there are substantial quality assurance measures in place[46] and, as a result, for the most part, the examination process has been completed with relatively few problems. Similar to the other USMLE examinations, significant efforts are dedicated to all phases of testing, including content development and validation, examinee scheduling, administration, scoring, equating, standard setting, and score reporting.

Despite the technical and administrative challenges, the implementation of the USMLE Step 2 CS program has been successful. USMLE Step 2 CS identifies examinees with deficiencies in important practice skills who might not otherwise have been identified based on the other examinations in the USMLE sequence.[47] In this way, the examination has made a significant contribution to the medical licensing process in the US and, at the same time, has called special attention, within the education and practice community, to the role of clinical skills in patient care activities. In a recent study that was based on interviews of 25 leaders of medical school CSA programs, respondents noted that the new national examination validated the importance of clinical skills for medical students.[48] Also, of particular note, numerous schools have changed the objectives, content, and emphasis of their preclinical curriculum in response to the implementation of the Step 2 CS.[49]

Comprehensive Osteopathic Medical Licensing Examination Level 2-Performance Evaluation

In 1994, the NBOME started the process of developing a SP-based clinical skills examination for osteopathic physician licensure. After considerable research and several feasibility and pilot studies, the COMLEX-USA Level 2-PE was launched in 2004.[50] Similar to both the MCC and the USMLE, this new assessment complemented the other examinations that are part of the licensure process for osteopathic physicians (COMLEX-USA or Comprehensive Osteopathic Medical Licensing Examination is a series of three osteopathic medical licensing examinations administered by the NBOME. The examinations include Level 1, Level 2-CE, Level 2-PE, and Level 3. COMLEX-USA is the most common pathway by which osteopathic physicians (DOs) apply for licensure, and is accepted in all 50 states and numerous international jurisdictions.) The COMLEX-USA Level 2-PE, which is usually taken in the 4th year of osteopathic medical school, tests the clinical skills of graduating students of osteopathic medical schools in the US. As of 2008, the accreditation body for osteopathic medical schools in the US (Commission on Osteopathic College Accreditation of the American Osteopathic Association) requires that all students pass COMLEX-USA Level 2-PE before graduation, and examinees are not eligible to take the COMLEX-USA Level 3 examination, the final examination in the COMLEX-USA series, unless they have passed COMLEX-USA Level 2-PE. Through the end of the 2007 calendar year, there have been a total of 992 COMLEX-USA Level 2-PE test administrations, involving more than 11,800 examinees.

Based on the COMLEX-USA Level 2-PE assessment design, examinees encounter 12 SPs in a simulated ambulatory clinical medical environment. The assessment takes 7 hours and is administered at a single fixed site (NBOME National Center for Clinical Skills Testing) located in the Philadelphia,

Copyright © Society for Simulation in Healthcare. Unauthorized reproduction of this article is prohibited

Pennsylvania area. For each of the 12 simulated encounters, examinees have 14 minutes to evaluate and treat the SP based on the clinical presentation. Following the 14-minute encounter, the examinee has an additional 9 minutes to complete a written patient note. Content design for the examination, including test form specifications, was informed by analysis of national practitioner databanks and expert consensus.[51] The mix of cases for a given test form is balanced with respect to acute, chronic, and health promotion/disease prevention presentations. To enhance content validity, the mix of SPs is governed by specifications related to patient characteristics, including gender and age. The COMLEX-USA Level 2 PE is administered almost every day, and sometimes both in the morning and in the evening. Consequently, a variable test form administration model is employed.

The COMLEX-USA Level 2-PE assesses skills in four clinical skill areas: doctor-patient communication, interpersonal skills, and professionalism; data gathering, which includes medical history-taking and physical examination; documentation and synthesis of clinical findings (including treatment); and osteopathic principles and osteopathic manipulative treatment (OMT). Doctor-patient communication, interpersonal skills, and professionalism are evaluated by the SPs using behaviorally anchored holistic scales. Data gathering proficiency is derived from case-specific checklist items, documented by the SPs following the clinical encounter. Written notes are evaluated by physician examiners located throughout the US using a holistic rubric. Unique to COMLEX-USA Level 2-PE, osteopathic principles and OMT are evaluated by physician examiners via a distributed video review system. Here, the physician examiners, also located across the US, access assigned clinical encounters through a secure web link and then provide structured performance ratings.

The four skill area scores, summarized over the encounters, are combined into two domains. The Humanistic domain summary score is based solely on the SP ratings of doctor-patient communication, interpersonal skills and professionalism. The Biomedical/Biomechanical domain summary score is a weighted composite of an examinee's data gathering, written patient notes, and OMT scores. For both domains, the generation of a summary score, over encounters, is compensatory, meaning that an examinee can compensate for poor performance in one station with excellent performance in another. However, across the two domains, COMLEX-USA Level 2-PE uses a conjunctive scoring model; examinees must achieve passing scores in both domains to receive a passing score for the examination. Examinations standards were initially set in 2004–2005 and, based on widely accepted testing protocols, updated in 2007. Only candidates who fail the examination are given specific feedback on their skills performance in the two domains and four skills areas.

To ensure that decisions based on the COMLEX-USA Level 2-PE examination scores are fair, an extensive quality assurance program has been implemented. In addition to pilot testing cases prior live usage, double scoring a large percentage of the encounters, investigating the relationships among scores, and regularly checking physician and SP rater stringencies, the performances of failing candidates are systematically reviewed to ensure that the decisions are accurate and can be defended.

The introduction of COMLEX-USA Level 2-PE, although logistically challenging, helps to fulfill the public and licensing authority mandate for enhanced patient safety through the documentation of the clinical skills proficiency of graduates from osteopathic medical schools. As a consequence, it has effectively highlighted the importance of clinical skills training as part of the osteopathic medical school curriculum.[52–54] Moreover, there has been an associated increase in the use of simulation throughout the medical school curriculum. Based on a survey of the deans of the 23 fully accredited Colleges of Osteopathic Medicine and branch campuses, Gimpel et al.[55] concluded that the use of SPs and mechanical simulators at colleges of osteopathic medicine increased substantially from 2001 to 2005.

DISCUSSION

The clinical skills examinations described above (MCCQE Part II, USMLE Step 2 CS, NBOME COMLEX-USA Level 2-PE) share many commonalities. They all use a multistation format where candidates rotate through series of clinical encounters, alternating between patient interviews and some form of postencounter exercise. Here, the development and choice of clinical encounters (stations, cases) is governed by detailed test specifications. Multiple stations are used in an effort to broadly sample the practice domain and to ensure that the scores, and associated pass/fail decisions, are reliable. All of the examinations model typical patient settings and doctor-patient interactions. This high-fidelity simulated environment provides the means to measure fundamental clinical skills, including history taking, physical examination, doctor-patient communication, and interpretation of clinical data. In measuring these skills, some combination of rating scales and checklists is used to produce examinee scores. Given the high-stakes nature of these examinations (access to the medical profession), significant resources are allotted to development and validation of the simulated clinical scenarios. For all three examinations, unscored pilot stations are incorporated into live examinations before their active use in making decisions about clinical skills proficiencies. In this way, data can be gathered to establish the fidelity of the simulation, the appropriateness of the clinical content, and the ability of the resultant scores, both ratings and checklists, to discriminate between those who possess the skills and those who do not. Finally, and likely most important, they all employ highly structured training and quality assurance protocols, both for the SPs and physician evaluators. This helps to ensure that valid inferences (ie, pass/fail decisions) can be made from the available scores and ratings.

Although the assessments share a common structure, there are some important differences that, taken collectively, serve to broaden the potential assessment domain and provide potential test administration frameworks that could be useful to other health professions that wish to evaluate clinical skills. First, the USMLE Step 2 CS and NBOME COMLEX-USA Level 2-PE run at fixed sites, whereas the MCCQE Part II operates periodically on weekends at actual clinics

Copyright © Society for Simulation in Healthcare. Unauthorized reproduction of this article is prohibited

across Canada. Although choice of variable or fixed sites is dependent on candidate volume, political considerations, and economics, quality exams can be offered under either administrative model as long as steps are taken to ensure proper standardization and security. Second, because of the almost daily administration of the COMLEX-USA Level 2-PE and USMLE Step 2 CS exams, test forms are continuously changed and are rarely repeated. For the MCCQE Part II administrations, which take place at the same time across different sites, a fixed form model is appropriate (The actual examination does not take place at exactly the same time across Canadian sites. Examinees at sites in later time zones are sequestered so that examination information cannot be shared.) Third, unlike the MCCQE Part II, which is usually taken in the second year of residency, the US-based examinations (COMLEX-USA Level 2-PE, USMLE Step 2 CS, former ECFMG CSA) are targeted at individuals who are just entering GME programs. As a result, the content of the MCCQE Part II is somewhat more challenging, requiring more advanced management and clinical decision making abilities. Fourth, because of differences in the practice characteristics of allopathic and osteopathic medicine, the clinical content modeled in the various assessments is not identical. For example, on the COMLEX-USA Level 2-PE there are proportionally more encounters involving SPs with musculoskeletal complaints. Moreover, unlike any of the other assessments, the evaluation of osteopathic principles and OMT is a fundamental part of this examination.[56] Given the differing purposes of these assessments, it is not surprising that they diverge somewhat in terms of focus. Modeling clinical encounters that are important to the profession, combined with tailoring the examinations to the expected performance level of examinee, provides a basis for establishing the content and construct validity of the assessments. A similar strategy could easily be used for non SP-based simulation activities, including those employing mannequins or part-task trainers.

Although the skills that are measured in these performance-based assessments are similar, the measurement protocols vary. For both the USMLE Step 2 CS and COMLEX-USA Level 2-PE, a score equating strategy is employed.[42] Because the examination content, and associated SPs, can vary considerably from day to day, it is important to account for potential differences in the difficulty of the test forms administered. Unlike the other assessments, the MCCQE Part II employs physician examiners who sit in the room while the clinical interview takes place. These physicians are trained to score the encounters and also to make summary, holistic, judgments of the adequacy of the performance. These summary measures are then used, in combination with assessment scores, to derive performance standards.[29] In contrast, for both the COMLEX-USA Level 2-PE and USMLE Step 2 CS, where SPs complete history taking and physical examination checklists, separate standard setting exercises are conducted periodically. Interestingly, while all three assessments employ some form of assessment of doctor-patient communication skills, there are no common rubrics or training protocols. For both the COMLEX-USA Level 2-PE and USMLE Step 2 CS, the SPs provide ratings of interpersonal and communication skills; for the MCCQE Part II, the physician in the room evaluates these traits. Finally, although employed somewhat differently, all of the examinations have both compensatory and conjunctive scoring elements. Test-level scores are generated by averaging performance in specific domains over the series of modeled encounters. For the COMLEX-USA Level 2-PE and USMLE Step 2 CS, a candidate's pass fail status is determined by summary performance in multiple areas. For the MCC Part II examination, candidates must also demonstrate an acceptable level of performance across a minimum number of stations.

Overall, based on a fairly limited usage of mock-up settings and simulation modalities, the three SP-based examinations are successful in fulfilling their assessment goals. For the most part, the restricted use of simulation modalities can be attributed to the fundamental purposes of the assessments, the logistics and economics of large scale assessment, technological limitations, and psychometric issues pertaining to scoring. Nevertheless, going forward, one can envision the adoption of other simulation strategies to broaden the assessment domain. For example, if logistical and psychometric issues could be effectively addressed, incorporating paired SP-Part task trainer stations could be an effective way to measure procedural skills and clinical decision making.[57–59] Likewise, although stations involving one SP and one examinee are efficient, at least from a testing perspective, the measurement of communication skills in this context is restricted to the doctor (examinee) and the patient. To evaluate teamwork, and certain facets of professionalism and ethical behavior, it would be appropriate to include other simulated healthcare workers and even standardized family members.[60–63] The MCC has already integrated some stations of this nature into their clinical skills examination; for example, working with a nurse to care for a trauma patient in an acute care setting and advising another healthcare professional over the telephone. Finally, even though some physical findings can be simulated by SPs quite well, many cannot (eg, trauma, breathing difficulties). As a result, for an OSCE that only includes SP-based encounters, it can be difficult to fully evaluate physical examination skills. Here, provided financial and logistical concerns can be addressed, electromechanical mannequins could be employed in some stations.[64]

Although the incorporation SP-based performance assessments as part of licensure and certification has spurred substantial research, there remain several important areas where further investigations are warranted. With respect to scoring, the available checklist and rating scales used for SP-based assessments, although appropriate for measuring basic clinical skills including history taking and physical examination, may not yield valid and reliable measures when employed for acute care situations, especially those modeled with electromechanical mannequins or even part-task trainers. Here, other constructs (eg, timing, sequencing, accuracy) will need to be incorporated within the measurement framework. In terms of content sampling, additional research focusing on the choice and structure of the various forms of simulation exercises is needed. Knowing which types of simulated scenarios provide for the best assessment conditions, and most valid and reliable scores, is essential if one seeks

Copyright © Society for Simulation in Healthcare. Unauthorized reproduction of this article is prohibited

meaningful and generalizable measures of ability. Likewise, if new ability measures are constructed, additional psychometric work will be needed to delimit the score, or scores, that separate those who are proficient from those who are not. Finally, and arguably most important, there is still relatively little published research that shows that performance in the simulated environment translates to real-world patient care. Designing and completing outcome studies that provide support for the validity of the performance measures derived from simulation-based assessments is paramount.

Conducting large scale, high-stakes performance assessments for medical licensure has been extremely successful. Although the MCC, USMLE and NBOME clinical skills exams have somewhat different purposes, administration models, and scoring protocols, they are all effective in providing a fair and equitable assessment of the clinical skills of their test populations. All three assessments are supported by a substantial number of research studies aimed at establishing the validity and generalizability of the test scores. As medical simulation further expands into other areas (eg, specialty board certification, selection of residents, continuing medical education, maintenance of certification), the processes used to develop and administer these examinations, with some modification, can be used as a model for assessment design and delivery. Should simulation-based assessment be adopted more broadly, especially for high-stakes competency decisions, one ought to expect a fairly large consequential educational impact, including an enhanced curricular emphasis on any particular skills that are evaluated as part of new assessment strategies. As other health provider groups seek to evaluate their trainees and make defensible competency decisions, the lessons learned in developing high-stakes, SP-based assessments in medicine will certainly prove to be quite valuable.

REFERENCES

1. Melnick DE, Dillon GF, Swanson DB. Medical licensing examinations in the United States. *J Dent Educ* 2002;66:595–599.
2. Cosby JC Jr. The American Board of Dental Examiners clinical dental licensure examination: a strategy for evidence-based testing. *J Evid Based Dent Pract* 2006;6:130–137.
3. Austin Z, Gregory P, Tabak D. Simulated patients vs. standardized patients in objective structured clinical examinations. *Am J Pharm Educ* 2006;70:119.
4. Vu N, Baroffio A, Huber P, Layat C, Gerbase M, Nendaz M. Assessing clinical competence: a pilot project to evaluate the feasibility of a standardized patient—based practical examination as a component of the Swiss certification process. *Swiss Med Wkly* 2006;136:392–399.
5. Dillon GF, Boulet JR, Hawkins RE, Swanson DB. Simulations in the United States Medical Licensing Examination (USMLE). *Qual Saf Health Care* 2004;13(suppl 1):i41–i45.
6. Hallock JA, Melnick DE, Thompson JN. The step 2 clinical skills examination. *JAMA* 2006;295:1123–1124.
7. Reznick RK, Blackmore D, Dauphinee WD, Rothman AI, Smee S. Large-scale high-stakes testing with an OSCE: report from the Medical Council of Canada. *Acad Med* 1996;71:S19–S21.
8. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. *Acad Med* 1993;68: 443–451.
9. Barrows HS, Abrahamson S. The programmed patient: a technique for appraising student performance in clinical neurology. *J Med Educ* 1964;39:802–805.
10. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979;13: 41–54.
11. Cusimano MD, Cohen R, Tucker W, Murnaghan J, Kodama R, Reznick R. A comparative analysis of the costs of administration of an OSCE (objective structured clinical examination). *Acad Med* 1994;69: 571–576.
12. Reznick RK, Smee S, Baumber JS, et al. Guidelines for estimating the real cost of an objective structured clinical examination. *Acad Med* 1993;68:513–517.
13. Cohen R, Rothman AI, Ross J, Poldre P. Validating an objective structured clinical examination (OSCE) as a method for selecting foreign medical graduates for a pre-internship program. *Acad Med* 1991;66:S67–S69.
14. King AM, Pohl H, Perkowski-Rogers LC. Planning standardized patient programs: case development, patient training, and costs. *Teach Learn Med* 1994;6:6–14.
15. Klass DJ. "High-stakes" testing of medical students using standardized patients. *Teach Learn Med* 1994;6:28–32.
16. Norcini JJ, Stillman PL, Sutnick AI, et al. Scoring and standard setting with standardized patients. *Eval Health Prof* 1993;16:322–332.
17. Ben-David MF, Klass DJ, Boulet J, et al. The performance of foreign medical graduates on the National Board of Medical Examiners (NBME) standardized patient examination prototype: a collaborative study of the NBME and the Educational Commission for Foreign Medical Graduates (ECFMG). *Med Educ* 1999;33:439–446.
18. Friedman M, Mennin SP. Rethinking critical issues in performance assessment. *Acad Med* 1991;66:390–395.
19. Boulet JR, Swanson DB. Psychometric challenges of using simulations for high-stakes assessment. In: Dunn W, ed. *Simulations in Critical Care Education and Beyond*. Des Plains, IL: Society of Critical Care Medicine; 2004:119–130.
20. Wettach GR. A standardized patient enrolled in medical school considers the national clinical skills examination. *Acad Med* 2003;78: 1240–1242.
21. Medical Council of Canada. Medical Council of Canada Qualifying Examination Part II (MCCQE Part II). 2008. Medical Council of Canada.
22. Federation of State Medical Boards, Inc. and National Board of Medical Examiners. United States Medical Licensing Examination: Step 2 clinical skills (cs) content descriptions and general information. 2008. Federation of State Medical Boards, Inc. and National Board of Medical Examiners.
23. National Board of Osteopathic Medical Examiners. Bulletin of information. 2008. National Board of Osteopathic Medical Examiners.
24. Whelan G. High-stakes medical performance testing: the Clinical Skills Assessment program. *JAMA* 2000;283:1748.
25. Heffron MG, Simspon D, Kochar MS. Competency-based physician education, recertification, and licensure. *WMJ* 2007;106:215–218.
26. Accreditation Council for Graduate Medical Education and American Board of Medical Specialties. Toolbox of assessment methods. Version 1. 1. 2000. Accreditation Council for Graduate Medical Education and American Board of Medical Specialties.
27. Frank JR. *The CanMEDS 2005 Physician Competency Framework: Better Standards. Better physicians. Better Care*. Ottawa: The Royal College of Physicians and Surgeons of Canada; 2005.
28. Reznick RK, Blackmore D, Cohen R, et al. An objective structured clinical examination for the licentiate of the Medical Council of Canada: from research to reality. *Acad Med* 1993;68:S4–S6.
29. Smee SM, Blackmore DE. Setting standards for an objective structured clinical examination: the borderline group method gains ground on Angoff. *Med Educ* 2001;35:1009–1010.
30. Wood TJ, Humphrey-Murto SM, Norman GR. Standard setting in a small scale OSCE: a comparison of the Modified Borderline-Group Method and the Borderline Regression Method. *Adv Health Sci Educ Theory Pract* 2006;11:115–122.

© 2009 Society for Simulation in Healthcare

Copyright © Society for Simulation in Healthcare. Unauthorized reproduction of this article is prohibited

31. Smee SM, Dauphinee WD, Blackmore DE, Rothman AI, Reznick RK, Des MJ. A sequenced OSCE for licensure: administrative issues, results and myths. *Adv Health Sci Educ Theory Pract* 2003;8:223–236.

32. Humphrey-Murto S, Smee S, Touchie C, Wood TJ, Blackmore DE. A comparison of physician examiners and trained assessors in a high-stakes OSCE setting. *Acad Med* 2005;80:S59–S62.

33. Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *JAMA* 2007;298:993–1001.

34. Ziv A, Ben David MF, Sutnick AI, Gary NE. Lessons learned from six years of international administrations of the ECFMG's SP-based clinical skills assessment. *Acad Med* 1998;73:84–91.

35. Sutnick AI, Stillman PL, Norcini JJ, et al. ECFMG assessment of clinical competence of graduates of foreign medical schools. Educational Commission for Foreign Medical Graduates. *JAMA* 1993; 270:1041–1045.

36. Whelan GP, Boulet JR, McKinley DW, et al. Scoring standardized patient examinations: lessons learned from the development and administration of the ECFMG Clinical Skills Assessment (CSA). *Med Teach* 2005;27:200–206.

37. Whelan GP, McKinley DW, Boulet JR, Macrae J, Kamholz S. Validation of the doctor-patient communication component of the Educational Commission for Foreign Medical Graduates Clinical Skills Assessment. *Med Educ* 2001;35:757–761.

38. Boulet JR, Rebbecchi TA, Denton EC, McKinley DW, Whelan GP. Assessing the written communication skills of medical school graduates. *Adv Health Sci Educ Theory Pract* 2004;9:47–60.

39. Boulet JR, McKinley DW, Norcini JJ, Whelan GP. Assessing the comparability of standardized patient and physician evaluations of clinical skills. *Adv Health Sci Educ Theory Pract* 2002;7:85–97.

40. Hawkins RE. The introduction of the Clinical Skills Assessment into the United States Medical Licensing Examination (USMLE): a description of USMLE Step 2 Clinical Skills (CS). *J Med Licensure Discip* 2005;91:22–25.

41. De Champlain AF, Swygert K, Swanson DB, Boulet JR. Assessing the underlying structure of the United States Medical Licensing Examination Step 2 test of clinical skills using confirmatory factor analysis. *Acad Med* 2006;81:S17–S20.

42. Swanson DB, Clauser BE, Case SM. Clinical Skills Assessment with standardized patients in high-stakes tests: a framework for thinking about score precision, equating, and security. *Adv Health Sci Educ Theory Pract* 1999;4:67–106.

43. De Champlain AF, Macmillan MK, Margolis MJ, et al. Modeling the effects of security breaches on students' performances on a large-scale standardized patient examination. *Acad Med* 1999;74:S49–S51.

44. Margolis MJ, De Champlain AF, Klass DJ. Setting examination-level standards for a performance-based assessment of physicians' clinical skills. *Acad Med* 1998;73:S114–S116.

45. Boulet JR, De Champlain AF, McKinley DW. Setting defensible performance standards on OSCEs and standardized patient examinations. *Med Teach* 2003;25:245–249.

46. Boulet JR, McKinley DW, Whelan GP, Hambleton RK. Quality assurance methods for performance-based assessments. *Adv Health Sci Educ Theory Pract* 2003;8:27–47.

47. Harik P, Clauser BE, Grabovsky I, Margolis MJ, Dillon GF, Boulet JR. Relationships among subcomponents of the USMLE Step 2 Clinical Skills Examination, the Step 1, and the Step 2 Clinical Knowledge Examinations. *Acad Med* 2006;81:S21–S24.

48. Hauer KE, Hodgson CS, Kerr KM, Teherani A, Irby DM. A national study of medical student clinical skills assessment. *Acad Med* 2005;80: S25–S29.

49. Gilliland WR, La Rochelle J, Hawkins R, et al. Changes in clinical skills education resulting from the introduction of the USMLE step 2 clinical skills (CS) examination. *Med Teach* 2008;30:325–327.

50. Gimpel JR, Boulet JR, Errichetti AM. Evaluating the clinical skills of osteopathic medical students. *J Am Osteopath Assoc* 2003;103: 267–279.

51. Boulet JR, Gimpel JR, Errichetti AM, Meoli FG. Using National Medical Care Survey data to validate examination content on a performance-based clinical skills assessment for osteopathic physicians. *J Am Osteopath Assoc* 2003;103:225–231.

52. Errichetti AM, Gimpel JR, Boulet JR. State of the art in standardized patient programs: a survey of osteopathic medical schools. *J Am Osteopath Assoc* 2002;102:627–631.

53. Gimpel JR, Boulet JR, Weidner AC. Survey on the clinical skills of osteopathic medical students. *J Am Osteopath Assoc* 2006;l106:296–301.

54. Baker HH, Cope MK, Adelman MD, Schuler S, Foster RW, Gimpel JR. Relationships between scores on the COMLEX-USA Level 2-Performance Evaluation and selected school-based performance measures. *J Am Osteopath Assoc* 2006;106:290–295.

55. Gimpel JR, Weidner AC, Boulet JR, Wilson C, Errichetti AM. Standardized patients and mechanical simulators in teaching and assessment at colleges of osteopathic medicine. *J Am Osteopath Assoc* 2007;107:557–561.

56. Boulet JR, Gimpel JR, Dowling DJ, Finley M. Assessing the ability of medical students to perform osteopathic manipulative treatment techniques. *J Am Osteopath Assoc* 2004;104:203–211.

57. Nestel D, Kneebone R, Black S. Simulated patients and the development of procedural and operative skills. *Med Teach* 2006;28: 390–391.

58. Kneebone RL, Nestel D, Vincent C, Darzi A. Complexity, risk and simulation in learning procedural skills. *Med Educ* 2007;41:808–814.

59. Hatala R, Issenberg SB, Kassen B, Cole G, Bacchus CM, Scalese R. Assessing cardiac physical examination skills using simulation technology and real patients: a comparison study. *Med Educ* 2008;42: 628–636.

60. Kobayashi L, Shapiro MJ, Gutman DC, Jay G. Multiple encounter simulation for high-acuity multipatient environment training. *Acad Emerg Med* 2007;14:1141–1148.

61. Morgan PJ, Pittini R, Regehr G, Marrs C, Haley MF. Evaluating teamwork in a simulated obstetric environment. *Anesthesiology* 2007; 106:907–915.

62. Malec JF, Torsher LC, Dunn WF, et al. The Mayo high performance teamwork scale: reliability and validity for evaluating key crew resource management skills. *Simul Healthcare* 2007;2:4–10.

63. Rosen MA, Salas E, Wilson KA, et al. Measuring team performance in simulation-based training: adopting best practices for healthcare. *Simul Healthcare* 2008;3:33–41.

64. Huang GC, Gordon JA, Schwartztein RM. Millenium conference 2005 on medical simulation: a summary report. *Simul Healthcare* 2007;2: 88–95.

Copyright © Society for Simulation in Healthcare. Unauthorized reproductionof this article is prohibited

Moderator: John Boker, PhD
Discussant: Brian Clauser, PhD

A Comparison of Physician Examiners and Trained Assessors in a High-Stakes OSCE Setting

Susan Humphrey-Murto, Sydney Smee, Claire Touchie, Timothy J. Wood, and David E. Blackmore

Abstract

Background
The Medical Council of Canada (MCC) administers an objective structured clinical examination for licensure. Traditionally, physician examiners (PE) have evaluated these examinees. Recruitment of physicians is becoming more difficult. Determining if alternate scorers can be used is of increasing importance.

Method
In 2003, the MCC ran a study using trained assessors (TA) simultaneously with PEs. Four examination centers and three history-taking stations were selected. Health care workers were recruited as the TAs.

Results
A 3 × 2 × 4 mixed analyses of variance indicated no significant difference between scorers ($F_{1,462} = .01$, $p = .94$). There were significant interaction effects, which were, localized to site 1/station 3, site 3/station 2, and site 4/station1. Pass/fail decisions would have misclassified 14.4–25.01% of examinees.

Conclusion
Trained assessors may be a valid alternative to PE for completing checklists in history-taking stations, but their role in completing global ratings is not supported by this study.

Acad Med. 2005;80(10 suppl):S59–S62.

Background

The objective structured clinical examination (OSCE) is widely used to evaluate medical students, select foreign medical graduates for training, and for medical licensure. The OSCE has proven to be a reliable and valid assessment of clinical skills.[1] One area of controversy is who should be observing and rating the encounters. In the Medical Council of Canada Qualifying Examination Part II (MCCQE Part II) physician examiners are used as raters and standard setters, whereas in the Educational Commission for Foreign Medical Graduates (ECFMG) examination standardized patients (SP) are used. One argument for the use of physician examiners is that experienced physicians are essential to judge the ability level of examinees for making high-stakes decisions. However, with higher clinical demands on physician's time and difficulty recruiting physician examiners, the use of nonphysicians is an attractive alternative.

Correspondence: Susan Humphrey-Murto, MD, FRCPC, MEd, Riverside Campus, Box 67-1967 Riverside Drive, Ottawa, Ontario, Canada K1H 7W9; e-mail: ⟨shumphrey@ottawahospital.on.ca⟩.

Although there are studies that support the use of SPs,[2–4] several have identified concerns with their use as examiners. Rothman and Cusimano,[5,6] in two separate studies of the Ontario International Medical Graduate Program OSCE, found poor consistency between physician examiners and SPs in their ratings of interviewing skills and little agreement between them in identifying potentially problematic examinees regarding English proficiency.

One challenge for SPs is that they are commonly scoring by recall. One study that explicitly examined this issue compared physician examiners to SPs in five history-taking stations. Martin and colleagues[7] compared physician examiners, SP observers, and SPs completing checklists from recall. Their findings suggest physicians should be used to rate examinees whenever practical. SP observers were considered better than the SPs who rated from recall.

One source of alternate scorers is medical students, and Van der Vleuten et al.[2] demonstrated that trained medical students were almost as good as trained faculty. One interpretation of this study is that individuals with some medical knowledge may be superior to lay persons. Medical students are not appropriate for use in high-stakes examinations where they will be eventual test-takers.

In Canada, the MCCQE Part II is a requirement for medical licensure. The examination is run twice per year and each administration requires between 400 and 900 physicians. Securing sufficient numbers of physician examiners is becoming more challenging. We therefore wanted to determine if nonphysicians are a viable alternative. We chose individuals with a medically related background, as the literature suggested they might perform better. Physicians and nonphysician raters were compared on checklist scores and global rating scales in a high stake OSCE.

Method

The MCCQE Part II is a 12-station OSCE consisting of seven ten-minute patient encounters and five couplets. The couplets are five-minute patient encounters paired with five-minute written exercises. The patient problems of each station are derived from one of the five major disciplines of medicine

(medicine, obstetrics and gynecology, pediatrics, psychiatry, and surgery). The study was based on the fall 2003 administration and data was collected at four of 15 sites. Only English sites were selected to exclude variance due to language. Sites were chosen from different regions to provide a broader sample of data.

Three five-minute history-taking stations were selected for this study. Two of the cases were from the domain of obstetrics and gynecology and the third was a pediatrics case. The objectives tested were vaginal bleeding, pelvic pain, and eliciting a history regarding a crying and fussing child. There were 24 to 30 checklist items for each of the stations. One station also contained a rating scale item for "questioning skills" and another for "rapport with person."

Scoring procedures have been described elsewhere.[8] In essence, for each station, examiners complete a checklist measuring the observed performance of the examinees' clinical skills and subsequently completed a global rating. The global rating scale is a six-point scale ranging from "inferior" to "excellent," with the two middle categories described as "borderline unsatisfactory" and "borderline satisfactory." Cut-scores for each station were established by the modified borderline group method. With this method, each station cut-score was the mean of the case scores for individuals rated as "borderline."[9] The physician examiner score and cut-score were considered the "gold standard."

Thirty-three nonphysicians were recruited to be trained assessors. They were trained to score one of the three history-taking stations selected for this study. Most (27/33) of the trained assessors had a medical background such as nursing, pharmacy, physiotherapy, occupational therapy, paramedics, or psychology. The other six had no medical background but had been SPs in previous examinations. The training involved a two-hour general training session in which trained assessors were provided with a self-study booklet and then participated in one hour of "dry runs" of their patient problem. This step involved watching four to eight SPs portray the case and completing a checklist along with a physician examiner. Each trained assessor scored only one case.

Each of the 33 trained assessors was paired with two physician examiners, one in the morning session and one in the afternoon session. Each pair viewed the encounter in real time and scored up to 32 examinees. They completed the same checklist and global rating scale but were not allowed to discuss results at any time. Each trained assessor scored up to 64 candidates and a total of 466 examinees completed all three stations.

The data were analyzed using SPSS 13.0 (SPSS Inc., Chicago, IL) to calculate correlations between examiner types and to conduct a $3 \times 2 \times 4$ repeated-measures analysis of variance. For this latter analysis, the three stations and examiner types (trained assessor versus physician examiner) were treated as within subject variables and the examination site (1–4) was treated as the between subject variable.

Results

The main effect for examiner was not significant ($F_{1,462} = .01$, $p = .94$). The mean scores and standard deviations are shown in Table 1. However, the interaction between station and examiner was significant ($F_{2,924} = 17.46$, $p < .001$), as was the three-way interaction among station, examiner, and site ($F_{6,924} = 7.50$ $p < .001$). As shown in Table 2, which displays the means and standard deviations by site for stations and examiners, the significant three-way interaction likely occurred because there was a difference between the scorers at some sites and stations that did not occur at other sites and stations. This observation was confirmed by running post hoc comparisons for each pair of trained assessor and physician examiners as a function of site and station. Table 2 displays the resulting level of significance and effect size measure for these comparisons. To protect against an inflation of the family-wise error rate, a significance level of .02 was used for these comparisons. As shown in the table, there was a significant difference in mean scores between trained assessors and physician examiners at site 1/station 3, site 3/station 2, and site 4/station 1 that did not occur elsewhere.

Despite these differences in mean scores, the correlations between scores assigned by the examiners were relatively high. As shown in Table 2, the correlation between examiner scores range from .49 to .92 indicating a relatively high level of agreement between examiners for each site and station. The high correlation and similar mean scores between pairs of examiners suggest that there were few differences between trained assessors and physician examiners, other than some isolated differences due to the interaction of site and station. In the full examination, all three stations were psychometrically sound with means and standard deviations well within expected norms. Item total score correlations were station 1 = .347, station 2 = .421, and station 3 = .359.

Table 1 displays the cut scores for each station as determined by the global ratings of the trained assessors and physician examiners. Although the cut scores for each station appeared to be similar, the agreement in terms of pass/fail decisions was not high. Examinees were classified in opposing pass/fail categories as follows: station 1, 67/466(14.4%); station 2, 78/466

Table 1
Means, Standard Deviations, and Cut-Scores by Examiner-Type across Stations (n = 466)

Station	PE Mean %	PE SD %	PE Cut-score %	TA Mean %	TA SD%	TA Cut-score %
1	60.8	10.0	54.8	60.3	10.5	54.8
2	57.7	11.0	51.6	55.9	10.5	49.2
3	60.3	11.6	49.4	62.2	11.7	53.9

PE = physician examiner; TA = trained assessor.

Table 2
Means, Standard Deviations, by Examiner-Type across Stations and Geographical Sites

			PE		TA				
Site	Station	N	Mean %	SD %	Mean %	SD %	ES Mean %	sig	*r* Standard Deviation %
1	1	118	62.0	11.6	62.0	11.1	.00	.93	.92
	2	118	58.6	11.5	58.1	12.4	.03	.40	.90
	3	118	57.7	10.8	63.7	11.6	.56	.00	.60
2	1	127	58.2	10.8	58.8	10.9	.06	.10	.93
	2	127	56.3	11.0	56.5	10.0	.02	.78	.69
	3	127	58.2	11.8	59.9	12.1	.14	.09	.58
3	1	160	61.4	10.6	60.5	9.8	.08	.12	.75
	2	160	59.2	10.3	53.7	9.3	.53	.00	.77
	3	160	61.9	12.1	61.5	11.7	.03	.64	.49
4	1	61	62.4	8.5	59.6	9.9	.32	.01	.73
	2	61	55.2	10.8	56.1	9.7	.09	.38	.66
	3	61	65.1	8.7	65.4	9.9	.03	.74	.72

PE = physician examiner; TA = trained assessor; ES = effect size of the comparison between the mean scores for the PE and TA; sig = level of significance of the comparison of mean scores for PE and TA.

(16.74%); station 3, 117/466 (25.01%). Physician examiners failed more examinees in every station compared to the trained assessors (136 versus 103 in station 1, 127 versus 109 in station 2, and 174 versus 99 in station 3).

Discussion

The purpose of this study was to determine if a nonphysician trained to score examinees on a particular case could produce ratings similar to that of a physician. In this study there was very good agreement between the physician examiner and trained assessor checklist scores for history-taking stations that were administered as part of a high-stakes OSCE. There was poor agreement, however, on pass/fail decisions. Up to 25% of candidates were misclassified by the trained assessors. This study confirms the findings of previous research suggesting that trained observers are a viable alternative for scoring checklists. The findings also raise the same concern identified by other studies regarding the ability of nonphysicians to complete global rating scales.

The finding that nonphysicians may have difficulty making judgments regarding the appropriateness of certain lines of questioning should not be surprising. A physician examiner may interpret a certain line of questioning favorably, for example recognizing a candidate who is ruling in or out disease, which the nonphysician would not have the medical knowledge to credit. For the Ontario International Medical Graduate OSCE, Rothman and Cusimano[5,6] reported good consistency between physician examiner and SP ratings of English proficiency, but less agreement in their ratings of interviewing skills and little agreement in identifying problematic candidates. For similar reasons, Colliver et al.[10] recommended caution in the interpretation of scores obtained from a case checklist completed by multiple SPs, especially if scores would be used for pass/fail decisions.

This study differed from some of the other studies because it was based on real-time simultaneous observations by the physician examiner and trained assessor pairs. The qualitative loss that may be associated with viewing videotaped encounters was avoided. A second difference lies in the approach to the recruitment and training of the trained assessors. The trained assessors were required to have a university-level degree and a professional background that would support their role as an examiner in a clinical skills examination. In addition they received three or more hours of training related to medical history taking and the case they would be observing. This is less than the 15 hours given to SPs who score the ECFMG examination,[3] but these SPs are trained to portray a case as well as to score it. Newble and colleagues[11] studied the effect of training in physician examiners. They concluded that training for physicians was not effective and that selection of inherently consistent raters was the critical factor. Van Der Vleuten et al.[2] reported similar results and concluded that training was least effective and least needed for medical faculty. However, they also noted that with only two hours of training, laypersons approached the accuracy of untrained faculty.

Conclusions

In conclusion, the study demonstrated that trained nonphysician assessors may be a valid alternative to physician examiners for scoring checklists in a high-stakes OSCE. As a preliminary study, this is encouraging. The next step is to develop a better understanding of the interaction effect that occurred at two of the four sites.

The ability of trained assessors to make valid global judgments that contribute to pass/fail decisions was not supported by the present study. This challenge to the standard setting methodology will need to be addressed before trained assessors

are incorporated in this high-stakes OSCE.

The authors wish to acknowledge Dr. Richard Birtwhistle for supporting this project, as well as Jodi Harold McIlroy and Ilona Bartmann for their statistical expertise.

References

1 Petrusa E. Clinical Performance Assessment. In: Norman G, van der Vleuten CPM, Newble DI (eds). International Handbook of Research in Medical Education. Dordrecht, The Netherlands: Kluwer, 2002.

2 Van der Vleuten CPM, VanLuyk SJ, Van Ballegooijen AMJ, Swanson DB. Training and experience of examiners. Med Educ. 1989;23: 290–96.

3 De Champlain AF, Macmillan MK, King AM, Klass DJ, Margolis MJ. Assessing the impacts of intra-site and inter-site checklist recording discrepancies on the reliability of scores obtained in a nationally administered standardized patient examination. Acad Med. 1999;74(10 suppl):S52–S54.

4 Colliver JA, Swartz MH, Robbs RS, Lofquist M, Cohen D, Verhulst SJ. The effect of using multiple standardized patients on the inter-case reliability of a large-scale standardized-patient examination administered over an extended testing period. Acad Med. 1998; 73(10 suppl):S81–S83.

5 Rothman AI, Cusimano M. Assessment of English proficiency in international medical graduates by physician examiners and standardized patients. Med Educ. 2001;35: 762–66.

6 Rothman AI, Cusimano M. A comparison of physician examiners' standardized patients', and communication experts' ratings of international medical graduates' English proficiency. Acad Med. 2000;75:1206–11.

7 Martin JA, Reznick RK, Rothman A, Tamblyn RM, Regehr G. Who should rate examinees in an objective structured clinical examination? Acad Med. 1996;71:170–75.

8 Reznick RK, Blackmore DE, Dauphinee WD, Smee SM, Rothman AI. An OSCE for licensure: the Canadian experience. In: Scherpbier AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg, AFW (eds). Advances in Medical Education. Dordrecht, The Netherlands: Kluwer, 1997.

9 Dauphinee WD, Blackmore DE, Smee S, et al. Using the judgments of physician examiners in setting standards for a national multicenter high stakes OSCE. Adv Health Sci Educ. 1997; 2:204.

10 Colliver JA, Robbs RS, Nu VV. Effects of using two or more standardized patients to simulate the same case on case means and case failure rates. Acad Med. 1991;66:616–18.

11 Newble DI, Hoare J, Sheldrake PF. The selection and training of examiners for clinical examinations. Med Educ. 1980;14: 345–49.

Online article and related content
current as of January 3, 2010.

Physician Scores on a National Clinical Skills Examination as Predictors of Complaints to Medical Regulatory Authorities

Robyn Tamblyn; Michal Abrahamowicz; Dale Dauphinee; et al.

JAMA. 2007;298(9):993-1001 (doi:10.1001/jama.298.9.993)

http://jama.ama-assn.org/cgi/content/full/298/9/993

Correction	Contact me if this article is corrected.
Citations	This article has been cited 25 times. Contact me when this article is cited.
Topic collections	Medical Education; Patient-Physician Relationship/ Care; Patient-Physician Communication Contact me when new articles are published in these topic areas.
Related Articles published in the same issue	The Value of Assessing and Addressing Communication Skills Gregory Makoul et al. *JAMA*. 2007;298(9):1057.

Subscribe
http://jama.com/subscribe

Permissions
permissions@ama-assn.org
http://pubs.ama-assn.org/misc/permissions.dtl

Email Alerts
http://jamaarchives.com/alerts

Reprints/E-prints
reprints@ama-assn.org

Downloaded from www.jama.com at Taipei Medical University on January 3, 2010

Physician Scores on a National Clinical Skills Examination as Predictors of Complaints to Medical Regulatory Authorities

Robyn Tamblyn, PhD
Michal Abrahamowicz, PhD
Dale Dauphinee, MD
Elizabeth Wenghofer, PhD
André Jacques, MD
Daniel Klass, MD
Sydney Smee, MSc
David Blackmore, PhD
Nancy Winslade, PharmD
Nadyne Girard, MSc
Roxane Du Berger, MSc
Ilona Bartman, MA
David L. Buckeridge, MD, PhD
James A. Hanley, PhD

Context Poor patient-physician communication increases the risk of patient complaints and malpractice claims. To address this problem, licensure assessment has been reformed in Canada and the United States, including a national standardized assessment of patient-physician communication and clinical history taking and examination skills.

Objective To assess whether patient-physician communication examination scores in the clinical skills examination predicted future complaints in medical practice.

Design, Setting, and Participants Cohort study of all 3424 physicians taking the Medical Council of Canada clinical skills examination between 1993 and 1996 who were licensed to practice in Ontario and/or Quebec. Participants were followed up until 2005, including the first 2 to 12 years of practice.

Main Outcome Measure Patient complaints against study physicians that were filed with medical regulatory authorities in Ontario or Quebec and retained after investigation. Multivariate Poisson regression was used to estimate the relationship between complaint rate and scores on the clinical skills examination and traditional written examination. Scores are based on a standardized mean (SD) of 500 (100).

Results Overall, 1116 complaints were filed for 3424 physicians, and 696 complaints were retained after investigation. Of the physicians, 17.1% had at least 1 retained complaint, of which 81.9% were for communication or quality-of-care problems. Patient-physician communication scores for study physicians ranged from 31 to 723 (mean [SD], 510.9 [91.1]). A 2-SD decrease in communication score was associated with 1.17 more retained complaints per 100 physicians per year (relative risk [RR], 1.38; 95% confidence interval [CI], 1.18-1.61) and 1.20 more communication complaints per 100 practice-years (RR, 1.43; 95% CI, 1.15-1.77). After adjusting for the predictive ability of the clinical decision-making score in the traditional written examination, the patient-physician communication score in the clinical skills examination remained significantly predictive of retained complaints (likelihood ratio test, $P<.001$), with scores in the bottom quartile explaining an additional 9.2% (95% CI, 4.7%-13.1%) of complaints.

Conclusion Scores achieved in patient-physician communication and clinical decision making on a national licensing examination predicted complaints to medical regulatory authorities.

JAMA. 2007;298(9):993-1001 www.jama.com

Decades of research have confirmed that poor skills in patient communication are associated with lower levels of patient satisfaction, higher rates of complaints, an increased risk of malpractice claims, and poorer health outcomes.[1-16] Medical schools have responded by incorporating training in patient communication and clinical skills into the curriculum. However, these skills were not systematically evaluated, nor was a minimum level of proficiency required for medical licensure.[17] To address this problem, licensure reforms were undertaken in North America.[18] The Medical Council of Canada (MCC) (1993),[19] the Educational Commission for Foreign Medical Graduates (1998),[20] and most recently the United States Medical Licensing Examination (USMLE) (2004)[21] have all introduced a clinical skills examination (CSE)—a nationally standardized assessment of patient-physician communication, clinical history taking, and examination skills—as a requirement for licensure. All US and Canadian medical

For editorial comment see p 1057.

CME available online at www.jama.com

Author Affiliations: Departments of Medicine (Drs Tamblyn and Dauphinee) and Epidemiology & Biostatistics (Drs Abrahamowicz, Winslade, Buckeridge, and Hanley, and Mss Girard and Du Berger), McGill University, Montreal, Quebec, Canada; Ontario College of Physicians and Surgeons, Toronto, Ontario, Canada (Drs Wenghofer and Klass); Quebec College of Physicians, Montreal (Dr Jacques); and Medical Council of Canada, Ottawa, Ontario (Dr Blackmore and Mss Smee and Bartman).
Corresponding Author: Robyn Tamblyn, PhD, McGill University, 1140 Pine Ave W, Montreal, QC Canada, H3A 1A3 (robyn.tamblyn@mcgill.ca).

©2007 American Medical Association. All rights reserved.

Downloaded from www.jama.com at Taipei Medical University on January 3, 2010

school graduates must now pass a multiple-case standardized patient assessment, where patient and physician examiners observe and grade clinical and communication skills to predict a candidate's competence to practice.

While mandatory assessment of clinical and communication skills is supported by the general public,[22] concerns have been raised about the cost of the examination and the lack of evidence that a 1-day assessment could predict future practice, particularly as it relates to deficiencies in patient-physician communication.[23-27]

Since instituting its CSE for all Canadian physicians, the MCC has tested more than 25 000 medical graduates using an examination format similar to the USMLE Step 2 clinical skills examination.[19] We investigated the ability of CSEs to predict future complaints in medical practice. We tested the hypothesis that lower scores in patient-physician communication would be associated with a higher rate of patients' complaints about quality of care and communication. We also assessed whether the use of clinical examination scores improved the prediction of complaints beyond results from the traditional written examination.

METHODS

Context

In Canada and the United States, medical regulatory authorities (state medical boards and provincial colleges of physicians and surgeons) use a common framework to govern how physicians are trained, accepted into practice, regulated, disciplined, and removed from practice.[28-32] A principal obligation of state and provincial medical regulatory authorities in both countries is to address and resolve public complaints against physicians. In accordance with a common set of principles and procedures, all complaints that are received in writing are investigated. A triage system is used to collect information from the patient and physician for each complaint, weed out frivolous or vexatious actions, and undertake informal steps to attain early resolution of minor issues. When these informal steps are either unsuccessful or deemed inappropriate, the complaint is managed by a more formal committee or panel process that determines further action. Most complaints are resolved through a graded series of regulatory actions, typically education, cautions, and warnings. For the most serious complaints, and for all complaints involving issues of sexual misconduct, formal disciplinary hearings of a quasi-judicial nature are convened. These hearings can result in a variety of sanctions up to loss of license. When a patient complaint about a physician is made directly to a hospital, the hospital in most state and provincial jurisdictions is required to report problems of professional misconduct to the medical regulatory authority.

Design and Population

The cohorts of physicians who took the MCC clinical skills examination between 1993 and 1996 and were licensed to practice in Ontario and/or Quebec were identified. Nearly two-thirds of the Canadian population and approximately 50% of all physicians reside in these 2 provinces. All complaints filed against these physicians with the medical regulatory authority in either province were retrieved between the date of licensure and March 2005. The MCC identified the 6677 physicians taking the examination during this period and provided the first and last name, sex, medical school, and year of graduation of each candidate to the medical regulatory authority in Ontario and Quebec. These 5 nominal fields were used to link to the registry of licensed physicians in each province. Physicians who matched on all fields were retained. Partial matches were manually inspected and adjudicated. Specialty, postgraduate training location and dates, and license year were obtained from the provincial medical regulatory files as well as from the national training registry of all physicians completing postgraduate medical training in Canada. Of the 6677 physicians, 8.6% could not be linked to Ontario/Quebec medical regulatory files or the national postgraduate training registry. Compared with linked physicians, unlinked physicians were more likely to be older (>45 years, 44.2% vs 11.4%; χ^2 $P<.001$), men (73.6% vs 57.4%; χ^2 $P<.001$), have trained outside Canada (83.4% vs 12.7%; χ^2 $P<.001$), have not yet passed the CSE (15.7% vs 1.8%; χ^2 $P<.001$), and have lower traditional written examination scores (495.4 vs 524.7; *t* test $P<.001$) and CSE scores (436.8 vs 517.8; *t* test $P<.001$).

Physician identity and confidential information were protected by replacing all nominal data with an MCC-generated study number, which was used to link demographic, score, and complaint files for each study physician. The McGill Faculty of Medicine institutional review board provided ethical approval. The provincial privacy commission, the Ontario and Quebec medical regulatory authorities, and MCC approved and oversaw data access, linkage, and anonymization procedures.

Measurement of Complaints

Provincial medical regulatory authorities collect standardized information for each written complaint against a physician. This information includes the names of the patients and physicians involved, and a description of the problem, circumstances, medical interventions, outcome, and the location of the incident. The investigation process includes a review of the letter with the complainant, the physician response, the patient's medical records, information from the hospital if applicable (eg, for surgical complications), and information from witnesses. All evidence is reviewed by physician investigators (Quebec) or a complaints committee (Ontario) who determine the legitimacy of the complaint, the type and seriousness of problem, and the recommended approach for resolution and subsequent action. Complaints are classified by investigators into 1 of 55 (Quebec) or 57 (Ontario) mutually exclu-

 ©2007 American Medical Association. All rights reserved.

Downloaded from www.jama.com at Taipei Medical University on January 3, 2010

sive categories (eg, complication due to medical or surgical error, breach of confidentiality, incomplete medical reports), along with the outcome (retained or not) and the action (warning, counseling/training, license withdrawal, suspension, or restriction).

All complaints recorded for study physicians were retrieved by medical regulatory personnel. Data included the physician study number, date of filing and closure, the classification of problem type, and the outcome (retention decision and action taken). Complaint classification codes from the respective regulatory authorities were grouped into 6 categories based on comparable groupings used by the Ontario and Quebec regulatory authorities: communication and attitude; quality of care; professionalism; office-related problem; physician health-related behavior problem (eg, mental illness); and other (eg, false advertising). Assignment of complaint classification codes was independently verified by medical regulatory investigators who arbitrated disagreements on final assignment.

The primary outcome was the complaint rate: the number of complaints retained as valid by the medical regulatory authority after investigation per year of practice time. Because judgment about the validity of a complaint may vary between provincial regulatory authorities, we conducted a sensitivity analysis including all complaints to assess whether our findings were influenced by retention decisions. The subset of retained complaints that were related to problems in communication and quality of care were assessed as secondary outcomes, as these problems should be more strongly associated with the competencies being assessed by the examination.

The complaint rate for each physician was calculated using as the denominator years in practice, defined as the number of years between the final year of postgraduate training exit date and the end of follow-up (March 2005). To assess the validity of using exit date from postgraduate training as the starting date for practice time, we retrieved for 1161 Quebec physicians a count of the number of years in which the physician billed for patient services to the provincial insurance agency. In comparison with actual billing data between 1993 and 2003, our approach modestly overestimates the number of practice-years (mean [SD] from billing, 4.2 [2.4] years; from training exit year estimate, 4.9 [2.2] years). However, there was very good agreement between the 2 methods (intraclass correlation, 0.67; 95% confidence interval [CI], 0.54-0.75) and no relationship between practice-years and communication score (Pearson $r = -0.06$). Thus, potential errors in measurement of practice-years should not confound the association between complaints and communication score.

Medical Council of Canada Examinations

Traditional Written Examination. This examination tests an individual's competence to enter postgraduate training. It is generally taken at the end of medical school and must be passed to be eligible for licensure. Medical knowledge is assessed using approximately 450 multiple-choice questions to assess knowledge in medicine, surgery, obstetrics-gynecology, psychiatry, pediatrics, and preventive medicine.[33] Clinical decision making is assessed using key feature problems.[34] Examinees are asked to respond to critical aspects of diagnosis or management in 36 to 40 clinical problems using write-in or menu-selection response formats.[34] Unlike multiple-choice questions, key feature questions focus exclusively on the components of a case where physicians are required to make critical decisions where errors could have an effect on patient outcome. Grading is based on the relative quality of the response, rather than a single correct answer, and errors of both omission and commission are considered in scoring. The score is calculated as the weighted sum of the multiple-choice (weight = 0.75) and clinical decision-making skills components (weight = 0.25), where the weights reflect the amount of testing time devoted to each component. A criterion-based passing score is established by a modified Nedelsky method,[33,34] and scores for first-time takers are standardized to a mean (SD) of 500 (100). For the study population, the Cronbach α estimate of the reliability of the written examination varied from 0.90 to 0.92 for the multiple-choice component, and from 0.60 to 0.69 for the clinical decision-making component in different administrations.

Clinical Skills Examination. This examination tests competence in data collection (history, physical examination), patient communication, and problem solving (diagnosis and management) through a 20-case objective structured clinical examination, and can be taken after 1 year of postgraduate training.[19] Most physicians take the examination in the second postgraduate year or the first half of the third postgraduate year (93% of physicians taking the examination between 1993 and 1996). Data collection is assessed in a 5- or 10-minute interaction with a standardized patient, by trained physician observers using case-specific checklists.[19] Patient-centered communication is assessed in 3 to 4 cases, selected to represent situations where communication is required for effective management (eg, discuss refusal of treatment for a terminal illness, counsel an adolescent about birth control). Examples of patient-physician communication that would receive a low score include condescending, offensive, or judgmental behaviors, or ignoring patient responses during the encounter. Problem solving is assessed by postencounter written responses to short-answer questions on diagnosis, investigation, interpretation of test results, and management. Responses are scored by physician examiners using an answer key. The passing score for the overall examination is established using criterion-referenced methods,[19,33-35] and scores for first-time takers are standardized to a mean (SD) of 500 (100). For the study population, the Cronbach α estimate of

©2007 American Medical Association. All rights reserved.

Downloaded from www.jama.com at Taipei Medical University on January 3, 2010

the reliability of the CSE scores ranged from 0.25 to 0.50 for communication, 0.59 to 0.75 for data acquisition, and 0.41 to 0.67 for problem solving in different administrations.

Covariates

Physician characteristics that may be associated with communication ability or complaint rate were measured as potential confounders and effect modifiers.[6,10] They included information on the sex of the physician, international medical graduate status, and specialty, which were retrieved from the MCC master file, postgraduate training registry, and the medical regulatory authorities. Practice province also was considered a potential confounder because differences may exist in health service delivery and the management of complaints between jurisdictions.

Statistical Analysis

Correlations between examination scores were estimated by Pearson product-moment correlation coefficients. Score reliability was assessed using a weighted Cronbach α, where weights were based on the number of candidates taking the examination in each administration. Disattenuated correlations also were calculated to determine the expected correlation if both scores were measured with perfect reliability, using the formula[36]

$$r_{xy} \text{ Disattenuated} = \frac{r_{xy}}{\sqrt{\text{Reliability}_x \times \text{Reliability}_y}}$$

The relationship between the CSE scores and complaint rate was assessed using multivariate Poisson regression (SAS version 9.1, SAS Institute, Cary, North Carolina), adjusting for physician sex, specialty, country of training (Canada or international), and province. A 2-sided test with a *P* value of .05 was used to assess statistical significance. Number of complaints was the dependent variable, and number of years in practice was used to measure person-time for each physician. The predictive ability of each examination score was assessed in a separate model that adjusted for sex, specialty, international medical graduate status, and practice province, using continuous scores as well as score quartiles. To determine if the relationship between examination scores and complaints was modified by characteristics that may be associated with communication scores, including practice jurisdiction, physician sex, and foreign training, we assessed interactions between the examination score and these characteristics and used the likelihood ratio test to determine if the interaction terms improved the model fit.

Licensing examinations aim to assess a required level of proficiency, and thus minimum thresholds of communication ability may exist, below which the complaint rate is high and above which the rate is lower and relatively uniform. To assess whether a linear association provided an appropriate representation of the association between examination score and the complaint rate, we tested the multivariate Poisson models for nonlinearity using generalized additive models (GAM) nonparametric extension of Poisson regression.[37] The adjusted effect of examination score was estimated using smoothing splines with 4 *df* and the statistical significance of the nonlinear effect was tested by nonparametric χ^2 test. All models were estimated separately for primary and secondary outcomes.

To determine if including the CSE communication score improved the prediction of complaints beyond the traditional written examination results, we first estimated the independent relationship between scores achieved in the traditional written examination and complaint rate. The CSE communication score was then added to the model that included the traditional written examination score, and improvement in the prediction of complaints was assessed by the likelihood ratio test. The explanatory power of the CSE communication score in predicting complaints was estimated by the population attributable fraction, the proportion of all complaints that were explained by physicians in the bottom communication score quartile,[38] after adjustment for existing predictors.

Power was estimated using the approach proposed by Signorini[39] for Poisson regression. Based on a type I error of 5%, a baseline complaint rate of 3.1% in the study population, and 3424 physicians followed up for a mean 6.5 years, the study had a power of 95% to detect a relative rate difference of 12% per 2-SD decrease in score.

RESULTS

Among 6677 physicians taking the CSE between 1993 and 1996, 3424 (51.3%) were licensed to practice in Ontario and/or Quebec. At the time of the examination, 71.6% of study physicians were 25 to 30 years of age, 55.5% were men, and 12.3% were international medical graduates. Following the examination, 84% completed postgraduate training in primary care or medical subspecialties, and two-thirds entered practice in Ontario (**TABLE 1**). The mean score of the study population for both the clinical skills and traditional written examinations was approximately one-quarter of an SD above 500. However, the range was considerable—approximately 7 SDs for the CSE and 5 SDs for the traditional written examination. Overall, 230 physicians (6.7%) failed the CSE on the first attempt, and 52 of these physicians never passed the CSE but were licensed to practice during the transition to the new licensure requirements.

Correlations between the clinical skills and traditional written examinations overall scores and subscores varied between $r=0.10$ and $r=0.40$ (**TABLE 2**). The communication score had the lowest correlation with the traditional written examination scores and with other scores on the CSE. Even when corrected for unreliability, the correlation between the communication and traditional written examination scores was low (disattenuated $r=0.23$). Communication ability previously has been shown to be a domain independent from more cognitive abilities that are assessed in traditional written examinations.[40]

Overall, 1116 complaints were filed in a total of 22 585 practice-years (4.9

 ©2007 American Medical Association. All rights reserved.

Downloaded from www.jama.com at Taipei Medical University on January 3, 2010

complaints per 100 practice-years) (TABLE 3). The mean (SD) follow-up time per physician was 6.5 (2.4) years, corresponding to the first 2 to 12 years in practice. Of the 3424 physicians, 21.5% had at least 1 complaint filed, and 17.1% had complaint(s) retained in their file after investigation. The majority (81.9%) of retained complaints were for attitude/communication and quality-of-care problems. Communication problems in management and inappropriate treatment/follow-up were the most common causes of quality-of-care complaints. Among the 696 retained complaints, none led to an immediate loss of license, 71 (10.2%) led to recommendations for additional counseling/training or discipline, and the remainder led to verbal and written warnings.

Lower CSE communication scores were associated with a higher rate of retained complaints, particularly in the lowest quartile of these scores (TABLE 4). The 853 physicians in the bottom communication score quartile had 236 retained complaints filed in their combined total of 5542 practice-years. This yielded an overall rate of 4.26 complaints per 100 practice-years compared with 2.51 per 100 practice-years for physicians in the top communication score quartile (Table 4). In multivariate models that adjusted for other physician characteristics, significantly higher complaint rates also were found for male vs female physicians, surgeons and primary care physicians vs medical subspecialists, and physicians practicing in Ontario vs those practicing in Quebec (Table 4). Even after adjustment for these characteristics, physicians in the lowest communication score quartile had an excess complaint rate of 1.75 per 100 practice-years compared with physicians in the top score quartile (adjusted relative risk [RR], 1.52; 95% CI, 1.30-1.78), and an excess complaint rate of 2.15 per 100 practice-years compared with the upper 3 quartiles (adjusted RR, 1.43; 95% CI, 1.22-1.68). The population attributable fraction indicated that 10.0% (95% CI, 6.0%-13.9%) of all retained complaints were explained by physicians in the bottom communication score quartile.

There was no evidence of significant nonlinearity ($P=.25$ for the GAM nonparametric test). According to the lin-

Table 1. Characteristics of the 3424 Physicians Taking the National Postgraduate Clinical Skills Examination Between 1993 and 1996 Who Were Licensed to Practice in Ontario and/or Quebec, Canada[a]

Physician Characteristics	No. (%)
Sex	
Female	1525 (44.5)
Male	1899 (55.5)
Age at the clinical skills examination, y	
<25	159 (4.6)
25-30	2451 (71.6)
>30	814 (23.8)
Mean (SD)	28.9 (4.7)
Undergraduate medical education	
Quebec/Ontario medical school	2655 (77.5)
Other Canadian medical school	349 (10.2)
International medical graduate	420 (12.3)
Postgraduate specialty program	
Family/general medicine	1393 (40.7)
Medical specialty	1481 (43.3)
Surgical specialty	550 (16.1)
Practice location	
Ontario	2263 (66.1)
Quebec	1009 (29.5)
Both provinces	152 (4.4)
Licensing examination performance[b]	**Mean (SD) [Range]**
Clinical skills examination	
Overall score	525.1 (79.9) [50-749]
Communication subscore	510.9 (91.1) [31-723]
Data acquisition subscore	508.8 (90.7) [19-875]
Problem-solving subscore	541.6 (98.5) [170-864]
Traditional written examination	
Overall score	526.5 (77.6) [338-787]
Multiple-choice questions subscore	524.4 (82.7) [278-793]
Clinical decision-making subscore	525.0 (75.8) [221-739]

[a] Percentages may not total 100% due to rounding.
[b] Scores standardized with a mean (SD) of 500 (100) for all first-time takers from Canadian medical schools in a given examination administration.

Table 2. Correlation Between Overall Scores and Subscores on the Medical Council of Canada Traditional Written and Clinical Skills Examinations[a]

	Traditional Written Examination[b]		
Clinical Skills Examination	Overall Score	Multiple-Choice Questions	Clinical Decision-Making Skills
Overall score	0.40	0.36	0.33
Communication	0.14	0.10	0.17
Data acquisition	0.23	0.21	0.16
Problem-solving	0.38	0.36	0.30

[a] For the traditional written examination scores, weighted Cronbach α in different examination administrations was overall score, 0.92; multiple-choice questions, 0.91; and clinical decision making, 0.64. For clinical skills examination scores, weighted Cronbach α was overall score, 0.77; communication, 0.41; data acquisition, 0.66; and problem solving, 0.54. Weights were based on the number of candidates taking the examination in each administration.[36] Disattenuated correlations (r)[30] between the scores for the traditional written examination and clinical skills examination were overall score, 0.47 and communication, 0.23; between the traditional written examination clinical decision-making score and the clinical skills examination score: overall score, 0.47 and communication, 0.43.
[b] Pearson product-moment correlation coefficients. All correlations were statistically significant ($P<.001$).

©2007 American Medical Association. All rights reserved.

Downloaded from www.jama.com at Taipei Medical University on January 3, 2010

ear model, a 2-SD decrease in communication score was associated with a relative 38% increase in the complaint rate (1.17 more complaints per 100 practice-years) (Table 4). The relationship between communication scores and complaint rate was significantly stronger in Quebec (RR, 1.84; 95% CI, 1.51-2.24) compared with Ontario (RR, 1.34;

Table 3. Frequency of Complaints by Type, Status, and Number of Physicians Among 3424 Physicians Followed Up for the First 2 to 12 Years of Practice and 22 585 Combined Practice-Years in Ontario and Quebec, Canada

	Proportion of Physicians With Complaints (N = 3424)[b]				Complaint Rate by Type (N = 22 585 Practice-Years)			
	≥1 Complaint		≥1 Retained Complaint		All Complaints		Retained Complaints	
Type of Complaint[a]	No.	% (95% CI)	No.	% (95% CI)	No.	Rate/100 Practice-Years (95% CI)	No.	Rate/100 Practice-Years (95% CI)
Attitude/communication	422	12.3 (11.1-13.4)	307	9.0 (8.0-10.0)	548	2.4 (2.2-2.6)	367	1.6 (1.4-1.8)
Communication problem in management of care[c]	356	10.4 (9.4-11.4)	239	7.0 (6.1-7.8)	357	1.6 (1.4-1.8)	240	1.1 (1.0-1.3)
Rude, abusive conduct to patients[c]	94	2.7 (2.2-3.2)	57	1.7 (1.3-2.1)	94	0.4 (0.3-0.5)	57	0.3 (0.2-0.4)
Quality of care	289	8.4 (7.5-9.3)	161	4.7 (4.0-5.4)	385	1.7 (1.5-1.9)	203	0.9 (0.8-1.0)
Inappropriate treatment/follow-up[c]	138	4.0 (3.3-4.7)	81	2.4 (1.9-2.9)	138	0.6 (0.5-0.7)	81	0.4 (0.3-0.5)
Inadequate assessment[c]	54	1.6 (1.2-2.0)	32	0.9 (0.5-1.2)	54	0.2 (0.2-0.3)	32	0.1 (0.1-0.2)
Professionalism	72	2.1 (1.6-2.6)	42	1.2 (0.8-1.6)	79	0.4 (0.3-0.5)	45	0.2 (0.2-0.3)
Office-related	37	1.1 (0.7-1.4)	28	0.8 (0.5-1.1)	39	0.2 (0.2-0.3)	30	0.1 (0.1-0.2)
Physician health problem	5	0.2 (0.1-0.3)	4	0.1 (0-0.2)	8	0.1 (0-0.1)	6	0.03 (0.01-0.07)
Other	50	1.5 (1.1-1.9)	42	1.2 (0.8-1.6)	57	0.3 (0.2-0.4)	45	0.2 (0.2-0.3)
Total	735	21.5 (20.1-22.9)	584	17.1 (15.8-18.4)	1116	4.9 (4.6-5.8)	696	3.1 (2.9-3.3)

Abbreviation: CI, confidence interval.
[a] Examples of professionalism included conflict of interest and advertising. Examples of office-related included inadequate records and office staff problems. Examples of physician health problems included mental health and alcohol-related behavior problems. Other complaints included those that were classified in multiple categories.
[b] 118 Physicians had 2 or more retained complaints. The distribution of the 191 retained complaints against these physicians was similar to the distribution of all 696 retained complaints: communication, 86 (45%); quality of care, 68 (35.6%); professionalism, 24 (12.5%); office-related, 10 (5.2%); physician health problem, 1 (0.5%); and other, 2 (1.0%).
[c] The most prevalent subcategories of complaints within each category.

Table 4. Medical Council of Canada Clinical Skills Examination Communication Score and the Rate of Retained Complaints

	Population		Retained Complaint Rate		Relative Rate of Complaints Adjusted for Physician Characteristic[a]	
	No. of Physicians	Combined No. of Practice-Years	No.	Rate/100 Practice-Years (95% CI)	Relative Rate (95% CI)	*P* Value
Communication score, by quartile[b]						
1	853	5542	236	4.26 (3.75-4.84)	1.52 (1.30-1.78)	<.001
2	847	5444	159	2.92 (2.50-3.41)	1.13 (0.96-1.32)	.29
3	867	5672	152	2.68 (2.29-3.14)	1.06 (0.90-1.24)	.63
4	857	5929	149	2.51 (2.14-2.95)	1 [Reference]	
By continuous score (per 2-SD decline in score)					1.38 (1.18-1.61)	<.001
Physician sex						
Female	1525	10 281	211	2.05 (1.79-2.35)	1 [Reference]	
Male	1899	12 305	485	3.94 (3.60-4.31)	1.64 (1.39-1.94)	<.001
Medical school						
Canadian	3004	19 615	580	2.96 (2.73-3.21)	1 [Reference]	
International	420	2970	116	3.91 (3.26-4.70)	1.11 (0.93-1.34)	.25
Specialty						
Medical specialty	1481	8162	163	2.00 (1.72-2.33)	1 [Reference]	
Family medicine or GP	1393	11 633	394	3.39 (3.07-3.74)	1.79 (1.49-2.16)	<.001
Surgical specialty	550	2790	139	4.98 (4.22-5.88)	2.43 (1.93-3.04)	<.001
Province of practice						
Ontario	2263	15 086	553	3.67 (3.38-3.99)	1 [Reference]	
Quebec	1009	6486	107	1.65 (1.37-1.99)	0.49 (0.40-0.61)	<.001
Both provinces	152	1014	36	3.55 (2.56-4.92)	1.00 (0.71-1.40)	.99

Abbreviations: CI, confidence interval; GP, general practice.
[a] Estimated by multivariate Poisson regression, adjusting for physician sex, specialty, country of training (Canada or international), and province.
[b] Cutoffs for the quartiles were first quartile, <457; second quartile, 457-518; third quartile, 519-575; fourth quartile, >575.

 ©2007 American Medical Association. All rights reserved.

Downloaded from www.jama.com at Taipei Medical University on January 3, 2010

95% CI, 1.25-1.49). Physician sex and international medical graduate status were not significant modifiers of the communication score effect. Sensitivity analysis incorporating all complaints (retained and not retained) showed the same significant increase in the relative rate of complaints with declining communication score (6.55 per 100 practice-years in the lowest quartile compared with 4.78, 4.46, and 4.05 in the third, second, and upper quartile, respectively); however, the risk was smaller for all complaints (RR, 1.30; 95% CI, 1.22-1.39).

Among the CSE scores, only the communication score was significantly associated with complaint rates (TABLE 5). The CSE data acquisition and problem-solving scores had no relationship to complaint rate, including quality-of-care complaints. The CSE communication score was most strongly associated with the risk of communication complaints. The traditional written examination also was significantly associated with complaint rate, with the strongest association being for the clinical decision-making (CDM) score. The association between multiple-choice scores and complaint rate was significant for overall retained complaints but not significant for communication or quality-of-care complaints. Statistically significant nonlinearity was found in the relationship between CDM scores and overall complaint rate ($P = .02$, for 3 *df* GAM test). The complaint rate increased with declining CDM scores between 600 and 450, with no systematic effect beyond this score range.

The CSE communication score, when added to a model that included traditional written examination CDM score, significantly improved the prediction of overall retained complaints and communication complaints, but not complaints about quality of care (Table 5). After adjustment for the traditional written examination CDM score, an additional 9.2% (95% CI, 4.7%-13.1%) of retained complaints and 11.2% (95% CI, 5.8%-16.9%) of communication complaints were explained by physicians in the bottom communication score quartile.

COMMENT

In a longitudinal study of physicians who took the MCC clinical skills examination and entered practice in Ontario and/or Quebec, scores obtained in patient-physician communication were statistically significant predictors of future complaints to medical regulatory authorities. The credibility of the association was strengthened by evidence of a linear relationship between complaint rates and communication scores, a slightly stronger association when the outcome was limited to communication complaints, consistency of the direction and statistical significance of the association in Ontario and Quebec, and the persistence of the association after adjustment for physician sex, specialty, international medical graduation status, and time in practice.

We observed a complaint rate of 0.0491 per physician. This rate is within the range of US state medical boards, where the mean complaint rate for all licensed physicians (including those with no complaints) varied from 0.02 per physician in Wisconsin to 0.20 per physician in Alabama between 2001 and 2003.[41] Similar to others, we found that communication problems were the most common reason for complaints[42]: 49.1% of complaints in our study compared with 55% of com-

Table 5. Scores on the Medical Council of Canada Qualifying Examinations and the Rate of Retained Complaints: Overall and by Type of Complaint

	Relative Rate of Retained Complaints by Examination Score After Adjustment for Physician Characteristics[a]					
	Any Retained Complaint		Communication Complaint		Quality-of-Care Complaint	
Examination Scores	Relative Rate (95% CI)	*P* Value	Relative Rate (95% CI)	*P* Value	Relative Rate (95% CI)	*P* Value
Clinical skills examination						
Overall score	1.19 (1.00-1.42)	.05	1.28 (1.00-1.64)	.05	1.06 (0.76-1.48)	.74
Communication score	1.38 (1.18-1.62)	<.001	1.43 (1.15-1.77)	.001	1.38 (1.03-1.86)	.03
Data acquisition score	0.98 (0.83-1.16)	.85	0.97 (0.78-1.22)	.82	1.00 (0.74-1.35)	.92
Problem-solving score	1.02 (0.88-1.19)	.76	1.13 (0.92-1.41)	.25	1.01 (0.76-1.33)	.97
Traditional written examination						
Overall score	1.39 (1.14-1.70)	.001	1.34 (1.01-1.76)	.04	1.54 (1.06-2.22)	.02
Multiple-choice score	1.25 (1.03-1.50)	.02	1.22 (0.94-1.57)	.14	1.29 (0.92-1.80)	.14
Clinical decision-making score	1.51 (1.25-1.84)	<.001	1.47 (1.13-1.92)	.004	1.77 (1.25-2.56)	.002
Clinical skills examination communication score, adjusted for clinical decision-making score[b]	1.32 (1.13-1.71)	<.001	1.37 (1.10-1.71)	.005	1.30 (0.97-1.75)	.09

Abbreviation: CI, confidence interval.
[a] Estimated by multivariate Poisson regression, adjusting for physician sex, specialty, country of training (Canada, international), and province, using the examination score as a continuous variable. Results presented as the change in relative rate per 2-SD decrease in score. A separate model was used to estimate the association of each score with retained complaints, adjusting for all of the same physician characteristics.
[b] Estimated by multivariate Poisson regression. Model includes communication score, clinical decision-making subscore of traditional written examination, physician sex, specialty, country of training (Canada, international), and province. Improvement in the fit of the model with clinical decision-making score alone and communication plus clinical decision-making score was assessed by likelihood ratio test.

©2007 American Medical Association. All rights reserved.

Downloaded from www.jama.com at Taipei Medical University on January 3, 2010

plaints to 1 US state medical board between 1989 and 2000[43] and 74.7% in an investigation of hospital complaints between 2001 and 2003.[6]

Our results provide some feedback for medical educators and licensing authorities. Our study supports the predictive validity of providing a standardized assessment of communication skills prior to entry into practice. Almost 1 in every 5 physicians had a retained complaint filed with the provincial medical regulatory authorities in the first 2 to 12 years of practice. The risk of complaints was significantly greater among physicians in the lowest quartile of communication scores. This result suggests that direct observation and assessment of patient communication skills may be useful in identifying trainees who are more likely to experience difficulties in practice. Assessment of communication could play a role at different stages in training—to select candidates for medical school admission[44] or to identify trainees who may benefit from more intensive communication skill training, as these skills can be improved with training.[45]

In addition, our results suggest that a minimum passing standard should be established for communication on the CSE, as has been done in the US Step 2 Clinical Skills Examination.[21] To do so, the number of cases in which communication is assessed would need to be increased from the 3 to 4 cases to approximately 10 to 14 to obtain a sufficiently reliable score to make pass-fail decisions.[46] The MCC has already increased the number of cases in which communication is assessed to meet this reliability threshold.

Complaints were mainly associated with 2 subscores—clinical decision making and communication. Clinical decision-making assessment was specifically designed to select problems and test aspects of the decision-making process where physicians were more likely to make errors that would have an effect on patient outcome.[34] This approach to the selection of test material may explain why this component of the examination was predictive of complaints, while the data collection and problem-solving components of the CSE were not. The key features approach to clinical decision-making assessment was first instituted by the MCC in 1992, and to our knowledge this is the first evaluation of its ability to predict future practice outcomes.[47] It may be useful to increase the use of key feature problems in traditional written assessment, as this format appears to be more predictive of quality-of-care complaints than ordinary multiple-choice questions. Selecting case and test elements for the national CSE on the same basis as key feature written problems also may be beneficial. The discriminating ability of data acquisition and problem-solving assessment on the CSE may be improved by selecting aspects of data collection that are critical for a given clinical problem, and where physicians tend to make errors.

Our study had several limitations. The poor-to-moderate reliability of the communication score component of the examination likely led to an underestimation of the strength of the relationship between communication and complaints.[48] The use of practice-years as a denominator for estimating the rate of complaints would not take into account differences between physicians in the frequency of patient contact, the type of patients, and the procedures performed, all of which may be associated with the risk of complaints. However, it seems unlikely that physicians with lower scores in communication would systematically seek out work activities and patient populations that are more likely to generate complaints.[13] On the other hand, higher rates of complaints that we found for surgeons, family physicians, and male physicians, even after adjustment for lower scores in communication, may be related to higher practice volume or differences in work activities or practice populations. As higher complaint and malpractice claim rates also have been found for these physician subgroups in other studies,[1,10] a better understanding of the contributing factors would be important. Finally, we did not have information on language of greatest proficiency for the physician or language in which the test was taken, and could not include these factors in the analyses.

In summary, we found that communication and clinical decision-making ability were important predictors of future complaints to regulatory authorities. Current examinations could be modified to test these attributes more efficiently and at earlier points in the training process. Future research should examine whether remediation of communication problems can reduce complaints, and whether other indicators of the quality of practice could be assessed by a clinical skills examination.

Author Contributions: Dr Tamblyn had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Tamblyn, Abrahamowicz, Dauphinee, Jacques, Klass, Winslade.
Acquisition of data: Tamblyn, Dauphinee, Wenghofer, Jacques, Klass, Smee, Girard, Bartman.
Analysis and interpretation of data: Tamblyn, Abrahamowicz, Dauphinee, Wenghofer, Jacques, Klass, Blackmore, Girard, Du Berger, Bartman, Hanley, Buckeridge.
Drafting of the manuscript: Tamblyn, Abrahamowicz, Dauphinee, Smee, Blackmore, Girard, Bartman, Hanley.
Critical revision of the manuscript for important intellectual content: Tamblyn, Dauphinee, Wenghofer, Jacques, Klass, Winslade, Du Berger, Hanley, Buckeridge.
Statistical analysis: Tamblyn, Abrahamowicz, Girard, Du Berger, Bartman, Hanley, Buckeridge.
Obtained funding: Tamblyn, Dauphinee.
Administrative, technical, or material support: Tamblyn, Dauphinee, Wenghofer, Jacques, Klass, Smee, Blackmore, Winslade.
Study supervision: Tamblyn, Abrahamowicz, Dauphinee, Jacques, Klass.
Financial Disclosures: None reported.
Funding/Support: The Medical Council of Canada (MCC) and the Canadian Institutes of Health Research (CIHR) provided funding for the study. The CIHR provided operating funds for the study and fellowship support for Dr Winslade.
Role of the Sponsors: The CIHR had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. The MCC, the College of Physicians and Surgeons of Ontario, and the College of Physicians of Quebec, as organizations, were not involved in the approval of this article. Coauthors from the Medical Council of Canada (Dr Blackmore and Mss Smee and Bartman), the College of Physicians and Surgeons in Ontario (Drs Klass and Wenghofer), and College of Physicians of Quebec (Dr Jacques) provided oversight for the retrieval, linkage, and anonymization of data from their respective institutions, as well as input on the design and conduct of the study; collection, management, and interpretation of the data; and preparation, review, and approval of the manuscript Mss Girard and Du Berger performed the statistical analysis.

 ©2007 American Medical Association. All rights reserved.

Downloaded from www.jama.com at Taipei Medical University on January 3, 2010

REFERENCES

1. Levinson W, Roger D, Mullooly J, Dull V, Frankel R. Physician-patient communication: the relationship with malpractice claims among primary care physicians and surgeons. *JAMA*. 1997;277(7):553-559.

2. Franks P, Fiscella K, Shields CG, et al. Are patients' ratings of their physicians related to health outcomes? *Ann Fam Med*. 2005;3(3):229-234.

3. Hickson GB, Clayton EW, Githens PB, Sloan FA. Factors that prompted families to file medical malpractice claims following perinatal injuries. *JAMA*. 1992; 267(10):1359-1363.

4. Beckman HB, Markakis KM, Suchman AL, Frankel RM. The doctor-patient relationship and malpractice: lessons from plaintiff depositions. *Arch Intern Med*. 1994;154(12):1365-1370.

5. Kohatsu ND, Gould D, Ross LK, Fox PJ. Characteristics associated with physician discipline: a case-control study. *Arch Intern Med*. 2004;164(6):653-658.

6. Stelfox HT, Gandhi TK, Orav EJ, Gustafson ML. The relation of patient satisfaction with complaints against physicians and malpractice lawsuits. *Am J Med*. 2005; 118(10):1126-1133.

7. Baldwin LM, Larson EH, Hart LG, Greer T, Lloyd M, Rosenblatt RA. Characteristics of physicians with obstetric malpractice claims experience. *Obstet Gynecol*. 1991;78(6):1050-1054.

8. Bovbjerg RR, Petronis KR. The relationship between physicians' malpractice claims history and later claims: does the past predict the future? *JAMA*. 1994; 272(18):1421-1426.

9. Sloan FA, Mergenhagen PM, Burfield WB, Bovbjerg RR, Hassan M. Medical malpractice experience of physicians: predictable or haphazard? *JAMA*. 1989; 262(23):3291-3297.

10. Hickson GB, Federspiel CF, Pichert JW, Miller CS, Gauld-Jaeger J, Bost P. Patient complaints and malpractice risk. *JAMA*. 2002;287(22):2951-2957.

11. Hickson GB, Pichert JW, Federspiel CF, Clayton EW. Development of an Early Identification and Response Model of Malpractice Prevention. *Law and Contemporary Problems*. 1997;60(1):7-29. Medical Malpractice: External Influences and Controls, part 1.

12. Vincent C, Young M, Phillips A. Why do people sue doctors? a study of patients and relatives taking legal action. *Lancet*. 1994;343(8913):1609-1613.

13. Hickson GB, Clayton EW, Entman SS, et al. Obstetricians' prior malpractice experience and patients' satisfaction with care. *JAMA*. 1994;272(20):1583-1587.

14. Moore PJ, Adler NE, Robertson PA. Medical malpractice: the effect of doctor-patient relations on medical patient perceptions and malpractice intentions. *West J Med*. 2000;173(4):244-250.

15. Papadakis MA, Teherani A, Banach MA, et al. Disciplinary action by medical boards and prior behavior in medical school. *N Engl J Med*. 2005;353(25):2673-2682.

16. Fiscella K, Meldrum S, Franks P, et al. Patient trust: is it related to patient-centered behavior of primary care physicians? *Med Care*. 2004;42(11):1049-1055.

17. Stillman P, Swanson D, Regan MB, et al. Assessment of clinical skills of residents utilizing standardized patients. *Ann Intern Med*. 1991;114(5):393-401.

18. AAMC. Report III: Contemporary Issues in Medicine: Communication in Medicine. http://www.aamc.org/meded/msop/. Washington, DC: Association of American Medical Colleges; 2006:1-29. Medical School Objectives Project. Accessed August 10, 2007.

19. Reznick RK, Blackmore D, Dauphinee WD, Rothman AI, Smee S. Large-scale high-stakes testing with an OSCE: report from the Medical Council of Canada. *Acad Med*. 1996;71(1)(suppl):S19-S21.

20. Whelan G. Educational Commission for Foreign Medical Graduates: lessons learned in a high-stakes, high-volume medical performance examination. *Med Teach*. 2000;22(3):293-296.

21. Papadakis MA. The Step 2 clinical-skills examination. *N Engl J Med*. 2004;350(17):1703-1705.

22. USMLE. Harris Interactive Poll, Results from: United States Medical Licensing Examination. http://www.usmle.org/news/Step2CSNews/Harris.asp. Accessed August 10, 2007.

23. Hoppe RB. *Performance Assessment: Why Should Medical Schools Care?* Henry R, Mavis B, Mullan P, Pryor O, Solomon D, eds. East Lansing: Office of Medical Education Research and Development, Michigan State University; 2002.

24. Bogdonoff MD. The clinical-skills examination. *N Engl J Med*. 2004;351(5):507-509.

25. Musco S. The clinical-skills examination. *N Engl J Med*. 2004;351(5):507-509.

26. Henry SG. Playing doctor. *JAMA*. 2005;294 (17):2138-2140.

27. Hallock JA, Melnick DE, Thompson JN. The step 2 clinical skills examination. *JAMA*. 2006;295(10): 1123-1124.

28. Federation of State Medical Boards. Elements of a modern state medical board. http://www.fsmb.org/pdf/GRPOL_Elements_Modern_Medical_Board.pdf. 2006. Accessed August 10, 2007.

29. Federation of State Medical Boards. A guide to the essentials of a modern medical practice act–tenth edition. http://www.fsmb.org/pdf/2003_grpol_modern_medical_practice_act.pdf. 2003. Accessed August 10, 2007.

30. Dauphinee WD. Self regulation must be made to work. *BMJ*. 2005;330(7504):1385-1387.

31. Federation of Medical Regulatory Authorities of Canada. http://www.fmrac.ca. 2007. Accessibility verified July 16, 2007.

32. Federation of State Medical Boards. http://www.fsmb.org. 2007. Accessibility verified July 16, 2007.

33. Mandin H, Dauphinee WD. Conceptual guidelines for developing and maintaining curriculum and examination objectives: the experience of the Medical Council of Canada. *Acad Med*. 2000;75(10):1031-1037.

34. Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decision-making skills. *Acad Med*. 1995;70(3):194-201.

35. Scherpbier AJJA, van der Vleuten C, Rethans J-J, van der Steeg AFW, eds. *Optimizing the Input of Physician Examiners in Setting Standards for a Large Scale OSCE: Experience With Part II of the Qualifying Examination of the Medical Council of Canada.* Dordrecht, the Netherlands: Kluwer Academic Publishers; 1997.

36. Osborne JW. Effect sizes and disattentuation of correlation and regression coefficients: lessons. http://pareonline.net/getvn.asp?v=8&n=11 8[11]. 2003. Practical Assessment, Research & Evaluation. Accessed August 10, 2007.

37. Hastie TJ, Tibshirani RJ. *Generalized Additive Models.* London, England: Chapman & Hall; 1990.

38. Hanley JA. A heuristic approach to the formulas for population attributable fraction. *J Epidemiol Community Health*. 2001;55(7):508-514.

39. Signorini D. Sample size for Poisson regression. *Biometrika*. 1991;78(2):446-450.

40. Edelstein RA, Reid HM, Usatine R, Wilkes MS. A comparative study of measures to evaluate medical students' performance. *Acad Med*. 2000;75(8):825-833.

41. Roberts RG, Friedsam D, Beasley JW, Helstad C, Moberg DP. The state of quality reports: comparing states by their rankings. *WMJ*. 2006;105(8):60-66.

42. Federation of State Medical Boards. State medical board disciplinary reports: 2005; trends in physician regulations–2006. http://www.fsmb.org/pdf/PUB_FSMB_Trends_in_Physician_Regulation_2006.pdf. April 2006. Accessibility verified July 16, 2007.

43. Halperin EC. Grievances against physicians: 11 years' experience of a medical society grievance committee. *West J Med*. 2000;173(4):235-238.

44. Lumsden MA, Bore M, Millar K, Jack R, Powis D. Assessment of personal qualities in relation to admission to medical school. *Med Educ*. 2005;39(3):258-265.

45. Yedidia MJ, Gillespie CC, Kachur E, et al. Effect of communications training on medical student performance. *JAMA*. 2003;290(9):1157-1165.

46. Ben-David MF, Klass DJ, Boulet J, et al. The performance of foreign medical graduates on the National Board of Medical Examiners (NBME) standardized patient examination prototype: a collaborative study of the NBME and the Educational Commission for Foreign Medical Graduates (ECFMG). *Med Educ*. 1999;33(6):439-446.

47. Page G, Bordage G. The Medical Council of Canada's key features project: a more valid written examination of clinical decision-making skills. *Acad Med*. 1995;70(2):104-110.

48. Cleary TA, Linn RL, Walster GW. Effect of reliability and validity on power of statistical tests. *Sociol Methodol*. 1970;2:130-138.

©2007 American Medical Association. All rights reserved.

Downloaded from www.jama.com at Taipei Medical University on January 3, 2010

High-stakes OSCE International Workshop 高標準 OSCE 國際工作坊(*II*)

時間:99 年 1 月 9 日(星期六)~10 日(星期日），8:30-17:00

組別	序號	姓名	職位	E-mail	學校/醫院
1	1	許聖德	創傷科主治醫師		三軍總醫院
1	2	蕭正文	大腸直腸外科主治醫師		三軍總醫院
1	3	黃天祐	胃腸科主治醫師		三軍總醫院
1	4	陳思州	急診部主治醫師		三軍總醫院
1	5	張耀田	主治醫師	yutein@mail2000.com.tw	台中榮民總醫院
1	6	陳卷書	主治醫師	chencs@mail.vghtc.gov.tw	台中榮民總醫院
1	7	陳雅芳	總醫師	saggi@mail.vghtc.gov.tw	台中榮民總醫院
1	8	王建得	主治醫師	jiaander@mail.vghtc.gov.tw	台中榮民總醫院
1	9	王振宇	主治醫師	cheuyu@mail.vghtc.gov.tw	台中榮民總醫院
1	10	陳志中	急診部主治醫師	e12150@vghks.gov.tw	高雄榮民總醫院
1	11	劉文忠	整形外科主治醫師	dliu@vghks.gov.tw	高雄榮民總醫院
1	12	劉斯顥	神經外科主治醫師	shliu@vghks.gov.tw	高雄榮民總醫院
1	13	葉玫伶	護理部副護理長	mlyeh@vghks.gov.tw	高雄榮民總醫院
1	14	葉同成	一般內科主治醫師	tcyeh@vghks.gov.tw	高雄榮民總醫院
1	15	謝昌勳	內科學科助理教授	10324@yahoo.com.tw	國防醫學院
1	16	黃清峰	小兒學科助理教授		國防醫學院
1	17	黃加璋	主治醫師	extarboy25@gmail.com	臺北榮民總醫院
1	18	詹哲彰	主治醫師	ccchan@vghtpe.gov.tw	臺北榮民總醫院
1	19	陳盈如	主治醫師	yichen7@vghtpe.gov.tw	臺北榮民總醫院
1	20	徐德福	主治醫師	tfhsu@vghtpe.gov.tw	臺北榮民總醫院
1	21	何橈通	主治醫師	ltho@vghtpe.gov.tw	臺北榮民總醫院
1	22	許維倫	主治醫師	wlhsu2@vghtpe.gov.tw	臺北榮民總醫院
2	1	李宜恭	急診科主任	A122514@tzuchi.com.tw	大林慈濟分院
2	2	陳品汎	新陳代謝科主任	dm926840@tzuchi.com.tw	大林慈濟分院
2	3	蔡坤維	老人醫學科主任	cktsai@ tzuchi.com.tw	大林慈濟分院
2	4	陳國瑚	醫師	cherish730721@hotmail.com	台北慈濟分院
2	5	楊慶輝	醫師	cherish730721@hotmail.com	台北慈濟分院
2	6	王采芷	國北師護理系副教授	tsaejyy@ntcn.edu.tw	台灣專科護理師學會
2	7	童恒新	國北師護理系助理教授	henghsin@ntcn.edu.tw	台灣專科護理師學會
2	8	蔡秀鸞	理事長	sltsay@ntcn.edu.tw	台灣專科護理師學會
2	9	余博芮	台大護理系講師	pojuiyu@ntu.edu.tw	台灣專科護理師學會
2	10	王錦榮	主治醫師	wang2260@cgmh.org.tw	林口長庚紀念醫院
2	11	崔克宏	科主任	t2130@cgmh.org.tw	林口長庚紀念醫院
2	12	張寓智	主治醫師	yjc0601@cgmh.org.tw	林口長庚紀念醫院
2	13	林宗男	主治醫師	t12244@cgmh.org.tw	林口長庚紀念醫院
2	14	林建志	主治醫師	lin0227@cgmh.org.tw	林口長庚紀念醫院
2	15	唐瑞平	科主任	rptang@adm.cgmh.org.tw	林口長庚紀念醫院
2	16	陳宗鷹	教學部主任	chenyting@tzuchi.com.tw	花蓮慈濟醫院
2	17	謝明蓁	臨床技能中心主持人	jeany.jeany@msa.hinet.net	花蓮慈濟醫院

High-stakes OSCE International Workshop 高標準 OSCE 國際工作坊*(II)*

時間:99 年 1 月 9 日(星期六)~10 日(星期日），8:30-17:00

組別	序號	姓名	職位	E-mail	學校/醫院
2	18	黃君后	內科加護病房護理師	arita0401@yahoo.com.tw	花蓮慈濟醫院
2	19	劉岱瑋	醫學系組長	dwliu3800@mail.tcu.edu.tw	慈濟大學
2	20	陳立光	醫學院院長	clk@mail.tcu.edu.tw	慈濟大學
2	21	謝美玲	護理系教師	shaer@mail.tcu.eud.tw	慈濟大學
2	22	朱紹盈	教師發展中心組長	shaoyin01@mail.tcu.eud.tw	慈濟大學
2	23	林玉娟	護理系教師	witch@mail.tcu.edu.tw	慈濟大學
3	1	高潘福	醫學系核子醫學科主任	pfkao@yahoo.com.tw	中山醫學大學
3	2	王子源	新陳代謝科主治醫師	yuan.w16@msa.hinet.net	中國醫藥大學
3	3	方信元	外科學科助理教授	fanghy@mail.cmuh.org.tw	中國醫藥大學
3	4	周致丞	急診部主治醫師	fremen@emome.net	中國醫藥大學
3	5	簡維宏	主任	dtpc25@gmail.com	亞東紀念醫院
3	6	陳芸	主任	ychen@mail.femh.org.tw	亞東紀念醫院
3	7	林姵君	課長	rachel@mail.femh.org.tw	亞東紀念醫院
3	8	陳嘉君	總醫師	b8703100@tmo.edu.tw	亞東紀念醫院
3	9	鍾明敏	總醫師	Stargoose1977@yahoo.com.tw	亞東紀念醫院
3	10	王祖鑑	主治醫師	Jonath.wang@mac.com	耕莘醫院
3	11	鍾耀文	主治醫師	yaw781123@yahoo.com.tw	耕莘醫院
3	12	劉樹泉	主任	lan1009@gmail.com	耕莘醫院
3	13	洪焜隆	副院長	jane@cgh.org.tw	國泰綜合醫院
3	14	劉榮森	教學部部主任	jungsen@cgh.org.tw	國泰綜合醫院
3	15	陳德芳	外科部部主任	Jac7@cgh.org.tw	國泰綜合醫院
3	16	莊美芳	NP 護長	fang525100@cgh.org.tw	國泰綜合醫院
3	17	簡志誠	院長	phh@cgh.org.tw	國泰綜合醫院
3	18	蔡欣恬	教研室副主任	hsinaka@cgh.org.tw	國泰綜合醫院
3	19	謝燿州	副研究員	R001805@ms.skh.org.tw	新光醫院
3	20	李賢發	課長	R000277@ms.skh.org.tw	新光醫院
3	21	陳玟霖	護理長	R003190@ms.skh.org.tw	新光醫院
3	22	陳宏毅	藥師	T004895@ms.skh.org.tw	新光醫院
3	23	連恆煇	醫師	hhlhhl@cgh.org.tw	輔仁大學
4	1	蔡詩力	主治醫師	bientsai@gmail.com	台大醫院教學部
4	2	池岸軒	住院醫師	jediboy@gmail.com	台大醫院教學部
4	3	許麗齡	老師	chengyu@mail.tmch.org.tw	國立臺北護理學院
4	4	顏如娟	主治醫師	m701061@gmail.com	臺北醫學大學
4	5	謝銘勳	醫學系主任	shiemin@tmu.edu.tw	臺北醫學大學
4	6	王正信	助理教授	Wangjs6@tmu.edu.tw	臺北醫學大學
4	7	邱艷芬	院長	yfchao@tmu.edu.tw	臺北醫學大學
4	8	王佳慧	兼任講師	wangch@tmu.edu.tw	臺北醫學大學
4	9	張念中	副院長	ncchang@tmu.edu.tw	臺北醫學大學
4	10	黃薇珍	助理	osce2008@tmu.edu.tw	臺北醫學大學

High-stakes OSCE International Workshop 高標準 OSCE 國際工作坊*(II)*

時間:99 年 1 月 9 日(星期六)~10 日(星期日)，8:30-17:00

組別	序號	姓名	職位	E-mail	學校/醫院
4	11	王聖儒	小兒科主治醫師	97318@wanfang.gov.tw	臺北醫學大學・市立萬芳醫院
4	12	蘇鴻文	婦產科主治醫師	90144@wanfang.gov.tw	臺北醫學大學・市立萬芳醫院
4	13	林俊佃	內科部主治醫師	Jdlin1971@yahoo.com.tw	臺北醫學大學・市立萬芳醫院
4	14	王定宇	一般外科主治醫師	wdy2@wanfang.gov.tw	臺北醫學大學・市立萬芳醫院
4	15	蕭炳昆	一般外科主治醫師	pkhsiao@wanfang.gov.tw	臺北醫學大學・市立萬芳醫院
4	16	何為斌	骨科主任	weipinho@wanfang.gov.tw	臺北醫學大學・市立萬芳醫院
4	17	劉偉翰	一般內科主治醫師	h2an20@gmail.com	臺北醫學大學・市立萬芳醫院
4	18	施俊明	部主任	cmshih53@tmu.edu.tw	臺北醫學大學附設醫院
4	19	黃群耀	主治醫師	cyhuang@tmu.edu.tw	臺北醫學大學附設醫院
4	20	鄭仰志	主治醫師	tmuh0000@tmu.edu.tw	臺北醫學大學附設醫院
4	21	許嘉麟	復健科醫師	pkpkpk100@gmail.com	臺北醫學大學附設醫院
4	22	區慶建	主治醫師	hk.au@msa.hinet.net	臺北醫學大學附設醫院
5	1	王明淑	技能中心專責護理師	wms9801@ms1.mmh.org.tw	馬偕紀念醫院
5	2	李君儀	醫師 4358	jyl5891@ms2.mmh.org.tw	馬偕紀念醫院
5	3	許希賢	技能中心主任	hsu5936@ms3.hinet.net	馬偕紀念醫院
5	4	林珍如	醫師 4443	alfielin@yahoo.com.tw	馬偕紀念醫院
5	5	陳兆弘	醫師 4277	chchen@ms1.mmh.org.tw	馬偕紀念醫院
5	6	彭純芝	醫師 4566	a4566@ms2.mmh.org.tw	馬偕紀念醫院
5	7	蔡樹榮	醫師 4089	tree@ms1.mmh.org.tw	馬偕紀念醫院
5	8	簡定國	醫師 5874	anderson@ms1.mmh.org.tw	馬偕紀念醫院
5	9	蔡維德	急診主治醫師	cktsai920@hotmail.com	馬偕紀念醫院
5	10	蕭偉仁	急診主治醫師	aa9701@yahoo.com.tw	馬偕紀念醫院
5	11	林慶忠	OSCE 組長 4063	unny.lin56@gmail.com	馬偕紀念醫院
5	12	蔡哲嘉	學士後醫學系副系主任	jerchia@kmu.edu.tw	高雄醫學大學
5	13	張高評	手術技能中心主任	kapich@kmu.edu.tw	高雄醫學大學附設中和紀念醫院
5	14	張智豪	整形外科主治醫師	igor8301023@yahoo.com.tw	高雄醫學大學附設中和紀念醫院
5	15	林彥克	外傷科主治醫師	yenko@ms16.hinet.net	高雄醫學大學附設中和紀念醫院
5	16	李坤泰	臨床技能中心主任	kuntai.lee@msa.hinet.net	高雄醫學大學附設中和紀念醫院
5	17	林季緯	家醫科主任	ed104283@edah.org.tw	義大醫院
5	18	羅錦河	內科部長	ed104676@edah.org.tw	義大醫院
5	19	梁正隆	一般醫學外科主任	ed100183@edah.org.tw	義大醫院
5	20	蘇智銘	主治醫師	08261@shh.org.tw	臺北醫學大學・署立雙和醫院
5	21	林聖傑	主治醫師	09261@shh.org.tw	臺北醫學大學・署立雙和醫院
5	22	蘇勇誠	主治醫師	chenny.su@gmail.com	臺北醫學大學・署立雙和醫院
5	23	林裕峯	學術副院長	linyf@shh.org.tw	臺北醫學大學・署立雙和醫院
6	1	王憲变	臨床技能中心主任	why8@ms61.hinet.net	台南永康奇美醫院
6	2	簡志強	腎臟內科主治醫師	ccchien58@yahoo.com.tw	台南永康奇美醫院
6	3	朱宗信	秘書長	tschu@ntu.edu.tw	台灣醫學教育學會
6	4	王明誠	主治醫師	wangmc@mail.ncku.edu.tw	成大醫院

High-stakes OSCE International Workshop 高標準 OSCE 國際工作坊*(II)*

時間:99 年 1 月 9 日(星期六)~10 日(星期日)，8:30-17:00

組別	序號	姓名	職位	E-mail	學校/醫院
6	5	賴宏瑛	護理師	h2360902@yahoo.com.tw	成大醫院
6	6	蔡新中	外科主任		行政院衛生署台中醫院
6	7	沈宜靜	內科 vs		行政院衛生署台中醫院
6	8	廖妙淯	家醫科 vs		行政院衛生署台中醫院
6	9	賴其萬	醫學教育講座教授	cwlai@kfsyscc.org	和信致癌中心醫院
6	10	簡竹君	主治醫師	drchien@adm.cgmh.org.tw	林口長庚紀念醫院
6	11	高聖博	標準化病人中心主持人	kaoshengpo@gmail.com	花蓮慈濟醫院
6	12	陳德全	主治醫師	puppy@cgmh.org.tw	高雄長庚紀念醫院
6	13	伍碧琦	內科護理長	rita@ cgmh.org.tw	高雄長庚紀念醫院
6	14	刁茂盟	兒科主治醫師	tmm@ cgmh.org.tw	高雄長庚紀念醫院
6	15	藍國忠	婦產科主治醫師	blue@cgmh.org.tw	高雄長庚紀念醫院
6	16	林寶源	整形科主治醫師	paoyuan@cgmh.org.tw	高雄長庚紀念醫院
6	17	孫灼均	一般外科主治醫師	Cksun6985@ cgmh.org.tw	高雄長庚紀念醫院
6	18	歐育哲	婦產科主治醫師	Ou4727@ cgmh.org.tw	高雄長庚紀念醫院
6	19	劉克明	醫學系解剖學科教授	kemili@cc.kmu.edu.tw	高雄醫學大學
6	20	張家昌	胃腸肝膽科主治醫師	g15539@cgmh.org.tw	基隆長庚紀念醫院
6	21	葉集孝	心臟外科主任	hsiuhui@adm.cgmh.org.tw	基隆長庚紀念醫院
6	22	黃金洲	主治醫師	huangchinchou@gmail.com	臺北榮民總醫院