

Assessing Students – Clinical Competence Versus Performance

John Ruedy

The recent elaboration of the range of physician competencies upon which the quality of health care is dependent has fostered the development of a variety of methods of assessing medical student competencies and performance. Such assessments are essential in providing feedback to students to guide their learning and to faculty on the success of the curriculum in achieving competency outcomes. In addition they provide evidence that students have achieved minimum requirements for progressing. Well-designed Observed Structured Clinical Examinations (OSCEs), Mini-Clinical Examinations (Mini-CEXs) and some forms of Multi-Source Feedback (MSF) can meet acceptable standards of validity and reliability and are feasible. Competency assessments are limited in predicting how a student will actually act in the work situation particularly in humanistic skills. More emphasis needs to be placed on student performance, in such competencies as communication and professionalism, in a variety of settings by a number of observers.

IeJSME 2007: 1 (1): 15-21

Key words: assessment, evaluation, clinical competence, performance, multi-source feedback

The assessment of the competence and performance of medical students is becoming of increasing importance. This has been prompted by the explicit recognition over the last decade of the importance of the range of physician competencies upon which medical care is dependent. This has spawned the development of competency-based medical educational programs – both postgraduate and undergraduate. For example, the Royal College of Physicians and Surgeons of Canada has identified that postgraduate students must attain competence in seven domains: medical expert, communicator, collaborator, manager, health advocate, scholar, and professional (the so-called CanMEDS roles).¹ A similar set of educational outcomes for undergraduates has been outlined by the Institute for International Medical Education.² The competency-based curriculum of the International

Medical University is based on a similar array of outcomes (Table 1)

In addition, physician regulatory bodies in the United Kingdom, Australasia and Canada have increased their vigilance of the continued competency of physicians in practice and have introduced performance assessments to address these concerns.³⁻⁶

In concert with these initiatives the assessment of undergraduate and postgraduate medical students has expanded beyond the traditional focus upon the evaluation of knowledge to the assessment of specific competencies and there has been renewed interest in the assessment of how students act and interact in the learning environment and in patient care. Moreover, there has been increased recognition of the importance of formative assessment to assist students in understanding their abilities and in guiding their continued learning.⁷

The ultimate assessment of student and physician performance in the provision of health care is an evaluation of the health outcomes of individual patients or communities of patients. This has been facilitated by the development of electronic patient records and the establishment of acceptable standards or benchmarks. Such outcomes are not applicable to students since their contribution to health care delivery is difficult to separate from that of their supervisors. Student assessment is therefore dependent upon the evaluation of component skills upon which outcomes are dependent including knowledge, knowing how to undertake tasks, demonstrating the performance of such tasks as well as observations on how the students actually undertake tasks in the relevant learning environments (Figure 1). In the current language of medical educators showing that they can perform is referred to as competence and demonstrating that they actually do this in practice is referred to as performance.⁸

Student assessment provides an opportunity to provide feedback to faculty on whether curriculum

Professor Emeritus, Dalhousie University, Halifax, Nova Scotia, CANADA

Address for Correspondence:

John Ruedy, 1210 Cromwell Road, Halifax, NS, B3H 4L2, CANADA
Email: johnruedy@aol.com, jruedy@Dal.Ca

outcomes are being achieved, to evaluate whether students are meeting minimum requirements to proceed (summative assessment) and to provide feedback to individual students to motivate them for continued improvement (formative assessment).⁴

Many novel methods of competency assessment have been developed over the last decade following the pre-eminent introduction of the Observed Structured Clinical Examination (OSCE) by Hardin over 25 years ago. There has been less development of performance assessment particularly at the undergraduate level (Table 3).

In determining which methods to use several characteristics need to be addressed. (Table 4). First, it is important that the method provides accurate information on what is being assessed. Content validity can be inferred when the evaluation emphasizes material that is considered important by experts, requires appropriate skills and is a broad sample of the defined material. A measure is said to have concurrent validity if the results are similar to those of other measurements of the same skills. The evaluation is said to have predictive validity if there is a known association with another outcome. A measure is defined as reliable if the assessment approach provides consistent results regardless of when it is used and who uses it. A few assessment methods stand out as having validity and reliability- for others there is little evidence or they have been shown not to meet these standards.

Competency assessment

Among methods to assess competency OSCEs have had the closest scrutiny and if well-conducted have content validity and reliability. For example, Sloan, in a 38-station OSCE for 56 surgical residents at three levels of training demonstrated validity by the fact that experienced students performed at higher levels: senior residents performed better than junior residents who, in turn, performed better than interns.⁹ Reliability assessed by coefficient alpha was high (0.91). Hodges has cautioned that it is incorrect to say that OSCEs have validity pointing out that they are an accurate

evaluation of competence only in the context in which the examination is taken.¹⁰ This is, of course, true of all other methods of competency assessment. The OSCE may not measure competencies equally well. Volkan analyzed the psychometric structure of a comprehensive OSCE taken by 169 Harvard medical students at the end of the third year of medical school. Two factors fit well with the model- information gathering and reasoning and information exchange, whereas case management did not.¹¹

Probert compared the performance of 30 final year students in an OSCE and a long case oral examination with faculty in-training assessments.¹² Whereas OSCE results had a positive association with faculty assessments, oral examination results had an inverse association. Tamblyn has studied the relationship between the scores of 614 family physicians on licensing examinations (predominantly OSCEs) and their performance in their first 18 months of practice. The scores were significant predictors of a number of measures of quality of care including consultation rates, mammography screening and appropriate prescribing.¹³ Thus there is some evidence that OSCE assessment of competencies has predictive value of actual performance.

Assessment of the clinical competency of a student by oral examination or by observing the student perform a complete history and physical examination is logistically difficult. In addition these forms of competency assessment are psychometrically unsound because of the inter-observer and inter-patient variability with the limit to the number of encounters that are practically achievable.

The second competency assessment method that has been shown to have validity and reliability is the Mini-Clinical Examination (Mini-CEX) introduced by the American Board of Internal Medicine for resident assessment. This consists of a single faculty member observing the resident taking a focused history and doing a limited physical examination and discussing the findings. The encounters are limited in time to

approximately 20 minutes making it feasible to have multiple observations in a variety of settings by a variety of faculty. If structured the Mini-CEX has been shown to be valid and reliable.¹⁴ Holmboe assessed the ability of 38 faculty to rate 348 taped performances of standardized residents upon scripted Mini-CEX scenarios. Faculty were able to discriminate among poor, marginal and high performers in history taking, physical examination and counseling skills.¹⁵ Durning compared the Mini-CEX results of 23 residents with faculty in-training evaluations. There were strong correlations between assessments of history taking, physical examination, judgement, humanistic skills and overall clinical competence.

Performance assessment

A major limitation of competency-based assessments is that they are undertaken in an artificial environment with the student aware that an assessment is being made. For example, students can readily learn by rote the checklist items included in the assessment of communication skills and physical examination and their assessment may have little bearing on what they will do in practice. Although there are many opportunities for a variety of observers to directly observe the actual performance of the student in the learning environment in interactions with other students and faculty as well as with patients, peers and co-workers in the clinical setting, the logistics of ensuring that such observations are valid and reliable and in documenting the observations have been frustrating.

Most of the efforts in performance assessment have been on faculty undertaking ward or practice based assessments- generally referred to in North America as in-training assessment. The ward or clinic should be opportune environments in which to assess a student. They allow for multiple observations over periods of time in a variety of clinical settings. In practice, the potential of this type of assessment has not yet been realized with faculty rarely observing a student completing a history or physical examination and seldom recording observations and providing feedback to the student in a timely manner. In addition,

advantage is rarely taken of multiple observers. Most assessors are not specially trained nor do they systematically receive feedback on the quality of their assessments. The quality of faculty assessments of students has been repeatedly questioned.¹⁶⁻²⁰ Faculty responsible for students often have a dual relationship with them that can compromise critical assessment and feedback. The team relationships between faculty and student in providing patient care can interfere with forthright assessments. These and other factors raise serious issues as to the validity and reliability of most in-training assessment processes.

A variety of factors have been identified as important in a successful In-training assessment system (Table 5). Unfortunately, although considerable effort is often expended in the design of the rating form, much less is expended in ensuring that assessments are systematically integrated into the responsibilities of faculty.

There is little doubt that an important impediment to maintaining an effective in-training evaluation process is logistical. Ways have to be found to imbed such an activity systematically into the work schedule of faculty. Turnbull has reported on “clinical work sampling”.²¹ Responsible physicians were asked to complete an assessment of the student related to the admission of each patient as well as an overall performance assessment at the time of each patient’s discharge. This systematic approach to assessment was found to be feasible, reliable and valid. Hemmer’s report emphasizes the importance of convening assessors to review assessments as a group.²² There was a higher rate of detection of students needing improvement or being unacceptable in regard to professionalism as a result of such meetings compared with the reports contained in checklists or written comments. The issues of validity and reliability of in-training assessment systems have recently been reviewed by others.²³⁻²⁵

The results of a review of 23 postgraduate programs in a Canadian medical school are shown in Table 6. In-training assessment has been used as an important assessment process in Canadian programs for many

decades and it is thought that these observations are characteristic of the status of such assessment processes in postgraduate programs in Canada. Although most rating forms reflect the goals and objectives of the clinical posting and feedback to residents occurs regularly few programs use evaluators other than physicians and few programs provide physician training in this form of assessment.

In the assessment of humanistic characteristics too little advantage has been taken of the variety of individuals who interact with students in the learning environment. Patients, peers and other health professionals can provide invaluable input. McCleod et al have studied the correlations among faculty, patient and self-evaluations.²⁶ Comparisons of the ratings of 33 residents, 13 faculty and 792 patients showed a high correlation between patient and faculty ratings of students on such qualities as respect, integrity, compassion (overall correlation coefficient 0.52). Resident self-ratings correlated less well with patient ratings (overall correlation coefficient 0.24).

Little advantage has been taken of tutor, self and peer assessment in contemporary educational programs that are emphasizing student-student and student-tutor interactions in a variety of small group sessions. Inappropriate behaviours of students in their clinical years, in postgraduate training and by physicians in practice are rarely a surprise to those who have known the individual in difficulty. There should be an increased emphasis on early recognition of students with compromised humanistic skills through multi-source feedback.

The validity and reliability of performance evaluation are dependent of the number of observations and the number and variety of observers. Multisource feedback is increasingly finding a place in physician assessment and its characteristics have been reviewed.^{27,28} The acceptance of this type of performance assessment by physicians should be a stimulus to its wider use in undergraduate and postgraduate educational programs.

Summary

Competency and performance assessments of students are requirements in contemporary undergraduate and postgraduate medical education programs. Validity, reliability and feasibility have been best established for a well executed OSCE or set of Mini-CEXs. Competency assessment methods are compromised by not necessarily predicting how a student will perform in the work-place particularly in relationship to humanistic skills. More emphasis needs to be placed on performance assessment of the student through multi-source feedback in a variety of situations in which the student is interacting with peers, faculty and patients.

REFERENCES

1. Neufeld JR, Maudsley RJ, Pickering RJ, Turnbull JM, Weston WW, Brown MJ. Educating future physicians for Ontario. *Acad Med* 1998; 7:1133-48.
2. 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.
3. Miller SH. American Board of Medical Specialties and repositioning for excellence in lifelong learning: maintenance of certification. *J of Cont Educ in Health Prof* 2005;25(3):151-156.
4. ACGME Outcome Project. General Competencies. <<http://www.acgme.org/oooutcome/comp/compFull.asp>> Accessed 12/12/06. Accreditation Council for Graduate Medical Education, Chicago, Illinois, 2001.
5. General Medical Council, Good Medical Practice. London: GMC, 2001.
6. Finucane P, Bourgeois-Law G, Ineson S, Kaigas TA. A comparison of performance assessment programs for medical practitioners in Canada, Australia, New Zealand and the United Kingdom. *Acad Med* 2003;778(8) : 837-843.
7. Schuwirth LWT, van der Vleuten C. Changing education, changing assessment, changing research? *Med Educ.*2004; 38: 805-812.
8. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002; 287(2): 226-235.
9. Sloan DA, Donnelly MB, Schwartz RW and Strodel WE. The Objective Structured Clinical Examination. The New Gold Standard for Evaluating Postgraduate Clinical Performance. *Annals of Surgery* 222(6); 1995: pp 735-742.
10. Hodges B. Validity and the OSCE. *Medical Teacher* 25(3); 2003: pp 350-354.
11. Volkan K, Simon SR, Baker H and Todres ID. Psychometric Structure of a Comprehensive Structured Clinical Examination: A Factor Analytic Approach. *Advances in Health Sciences Education* 9; 2004: pp 83-92.
12. Probert CS, Cahill DJ, McCann GL and Ben-Shlomo Y. Traditional finals and OSCEs in predicting consultant and self-reported clinical skills of PRHOs: a pilot study. *Medical Education* 37; 2003: pp 597-602

13. Tamblyn R, Abrahamowicz M, Brailovsky C et al. Association Between Licensing Examination Scores and Resource Use and Quality of Care in Primary Care Practice. *JAMA* 280; 1998:pp 989-996.
14. Norcini JJ, Blank LL, Duffy FD and Fortna GS. The Mini-CEX: A Method for Assessing Clinical Skills. *Annals of Internal Medicine* 138(6);2003:pp 476-481.
15. Holmboe ES, Hout S, Chung J, Norcini J and Hawkins RE. Construct Validity of the MiniClinical Evaluation Exercise (MiniCEX). *Academic Medicine* 78; 2003: pp 8226-830.
16. Holmboe ES. Faculty and the Observation of Trainees' Clinical Skills: Problems and Opportunities. *Academic Medicine* 79; 2004: pp 16-22.
17. Noel GL, Herbers JE, Cooper GS, Pangaro LN. How well do internal medicine faculty members evaluate the clinical skills of residents. *Ann Int Med* 117; 1992:pp 757-765.
18. Herbers JE, Moel GI, Cooper GS, Harvey J Pangaro LN and Weaver MJ. How accurate are faculty evaluations of clinical competence? *J Gen Int Med* 4; 1989: pp 202-208.
19. Elliott DL, and Hickam DH. Evaluation of physical examination skills. Reliability of faculty observers and patient instructors. *JAMA* 258; 1987: pp 3405-3408.
20. Kalet A, Earp JA and Kowlowitz V. How well do faculty evaluate the interviewing skills of medical students? *J Gen Int Med* 9; 1993 :pp 179-184.
21. Turnbull J, MacFadyen J, van Barneveld C, and Norman G. Clinical Work Sampling. A New Approach to the Problem of In-Training Evaluation. *Journal of General Internal Medicine* 15; 2000: pp 556-561.
22. Hemmer PA, Hawkins R, Jackson JL and Pangaro LN. Assessing How Well Three Evaluation Methods Detect Deficiencies in Medical Students' Professionalism in Two Settings of an Internal Medicine Clerkship *Academic Medicine* 75(2); 2000: pp 167-173.
23. Arnold L. Assessing Professional Behaviour: Yesterday, Today, and Tomorrow. *Academic Medicine* 77; 2002: pp 502-515.
24. Holmboe Es, Hawkins RE and Huot SJ . Effects of Training in Direct Observation of Medical Residents' Clinical Competence. A Randomized Trial. *Annals of Internal Medicine* 140(11); 2004: pp874-881.
25. Newble DI .Assessing clinical competence at the undergraduate level. *Medical Education* 26; 1992: pp 504-511.
26. McCleod PJ, Tamblyn R, Benaroya S. Faculty Ratings of Resident Humanism Predict Patient Satisfaction Ratings in Ambulatory Medical Clinic. *J Gen Int Med* 9; 1994: pp 321-326.
27. Lockyer J. Multisource Feedback in the Assessment of Physician Competencies.
 - a. *The Journal of Continuing Medical Education in the Health Professions* 23; 2003: pp 4-12.
28. Sargeant J. Multi-Source Feedback for Physician Learning and Change. Ph D Thesis, University of Maastricht: ISBN 0-9781130-0-4, 2006.

Table 1. Roles and competencies of the contemporary physician and medical student as identified by the Royal College of Physicians and Surgeons of Canada, the Institute for International Medical Education and the International Medical University

| Royal College of Physicians and Surgeons of Canada ⁽¹⁾ | Institute for International Medical Education ⁽²⁾ | International Medical University |
|---|--|---|
| Medical expert | Scientific knowledge | Application of basic science tin the practice of medicine |
| | Clinical skills | Clinical skills |
| Communicator | Communication skills | Communication skills |
| Professional | Professionalism | Professionalism, ethics and personal development |
| Scholar | Critical thinking and research | Critical thinking and research |
| | Information management | Self-directed learning and information management |
| Health advocate | Population health | Disease prevention and health promotion |
| Collaborator | | Family and community issues in health care |
| Manager | | |

Table 2. Contemporary competency assessment methods

| |
|---|
| Chart Stimulated Recall Oral Examination(CSR) Objective Structured Clinical Examination (OSCE) Observed long cases Procedure, Operative, or Case Logs Portfolios Patient management problems Simulations and Models Standardized Oral Examination Standardized Patient Examination MiniClinical Evaluation Exercise (Mini-CEX) Written Examination (MCQ, Short Answer) Chart stimulated recall |
|---|

Table 3. Contemporary performance assessment methods

| |
|---|
| Ward and practice-based assessment Checklists Global ratings Multiple Source Feedback (MSF) Clinical Work Sampling Chart audit/ patient review |
|---|

Table 4. Desirable qualities of assessment processes

| | |
|--------------------------|--|
| 1. VALIDITY | <i>MEASURES WHAT IT IS SUPPOSED TO MEASURE</i> |
| 2. RELIABILITY | <i>YIELDS CONSISTENT RESULTS</i> |
| 3. PREDICTIVE (VALIDITY) | <i>PREDICTS AN ASSOCIATION WITH ANOTHER CHARACTERISTIC</i> |
| 4. PRACTICALITY | |

Table 5. Key quality markers of successful In-training assessment systems

| |
|---|
| 1. Checklists not global rating 2. Multiple observations 3. Multiple evaluators 4. Multiple methods 5. Timeliness of recording 6. Evaluator feedback sessions 7. Evaluator training |
|---|

Table 6. Reviewers opinions on the quality of in-training assessment systems in 23 postgraduate programs in a Canadian medical school in 2004

| Characteristic | No or not acceptable Number | Acceptable or excellent Number | Percent |
|--|-----------------------------|--------------------------------|---------|
| Forms reflect goals and objectives | 6 | 17 | 74% |
| Forms consist of checklists of tasks | 12 | 11 | 48% |
| Forms completed more often than at end of rotation | 14 | 9 | 39% |
| Formal feedback occurs regularly | 5 | 18 | 78% |
| Physician evaluators receive training | 19 | 4 | 17% |
| A variety of other health professionals evaluate | 15 | 8 | 35% |
| There is peer evaluation | 14 | 9 | 39% |
| There is structured self evaluation | 18 | 5 | 22% |
| Patients provide recorded evaluations | 19 | 4 | 17% |

Figure 1. Component skills that determine health outcomes

