

Structured Continuous Objective-Based Assessment of Resident's Performance at Point of Care (SCOPA)

Mohammed Hijazi

The assessment of the clinical performance of physicians-in-training is an important task. The critical care rotation is a mandatory rotation for most residency training programs and is designed to ensure the graduation of trainees who are able to initiate lifesaving management during medical emergencies. Ensuring that each resident fulfills the objectives of the rotation is of paramount importance. Unfortunately, the current assessment methods are subjective and suffer from many threats to validity and reliability that make the assessment inaccurate. In this review, the current assessment method is analyzed, and causes for inaccuracy are identified. A new model for assessment that is continuous, structured, objective-based and at the point of care (SCOPA) is proposed based on the best available assessment methods. Such a model might be useful for the assessment of trainee's performance in critical care as well as non-critical care rotations.

*From the Department of Medicine,
King Faisal Specialist Hospital
& Research Centre
Riyadh, Saudi Arabia*

*Correspondence:
Mohammed Hijazi, MD, FCCP,
Department of Medicine, MBC 46
King Faisal Specialist Hospital
& Research Centre
Riyadh, 11211
Saudi Arabia
Fax: 966-1-442-7499
hijazim@hotmail.com*

*Accepted for publication
November 2004*

Ann Saudi Med. 2005;25(3):193-197

Despite the fact that assessing the clinical performance of physicians-in-training is an important professional and public matter, most current training programs use subjective methods that yield suboptimal assessment.¹⁻⁵ The assessment of trainee performance during the critical care rotation is no exception. Inaccurate assessment of trainees is likely to have a negative impact on trainees, training programs, professional standards, patient safety, and public trust.

Current Situation

Successful completion of an adult critical care medicine rotation is a requirement for medical, surgical, emergency medicine, anesthesia, neurology, and neurosurgery residents. The goal of the rotation is to make sure that all trainees (the first line responders to emergencies) are able to recognize, assess and initiate first-line life-saving management of common life-threatening inpatient and outpatient conditions. Ensuring that all trainees successfully achieve this goal is of paramount importance for patient safety and optimal management during emergencies. Successful completion requires a minimum of an average rating using the standardized clinical performance assessment forms that are common for other specialty rotations. The forms are completed by attending physicians on the last day of the rotation based on the consensus opinion of all the available attending physicians. The assessment form covers knowledge, clinical skills, operative and interventional skills, and personality and ethics.

Residents are rated on a 5-point scale (unacceptable =1, below average =2, average =3, above average =4, outstanding =5) for each item and the average score is calculated and used for the final assessment.

The above methods for assessment are used by most other rotations and residency training programs.

Shortcomings of the Current Situation

Assessment of clinical performance requires evidence of validity to be interpreted in a meaningful way.⁶ The current assessment method is suboptimal. It lacks reliability, validity and evidence supporting its results, and the inferences made based on it cannot be generalized to all clinical performances of trainees. Contributing factors are subjectivity, undersampling, timing of the assessment, assessment form vagueness, and rater-related factors. Each of these factors is described in detail in the following paragraphs.

The assessment is completely subjective and does not reflect the objectives of the rotation. All the items in the assessment form are general and lack alignment with the training objectives. Some objectives are probably not assessed, thus missing some of the dimensions of clinical competence, resulting in a construct underrepresentation (CU) threat to validity.⁷ For example, one of the objectives of the rotation is to differentiate the types of shock and initiate first line management. There is nothing in the assessment to reflect fulfilling this objective, which means that some trainees will not possess the minimum acceptable performance in such a common life-threatening emergency. Unfortunately, no research specific to the ICU setting addresses this issue. All the dimensions of clinical performance are often not assessed; for example, data collection skills are frequently overlooked in the ICU, despite its importance in reaching a correct diagnosis, as shown by Bordage.⁸

Current assessment of performance is done retrospectively, which makes it likely to miss the truth. Many attending physicians state clearly that they do not recall the exact performance of resident A, or they confuse resident A with resident B. At most, raters can recall one or two aspects of trainee performance and miss all other dimensions of performance, causing both a CU and construct irrelevant variance (CIV) threat to validity.⁷

The 5-point scale is not well defined, making it open to different interpretation by raters, adding to the variability of the rating, causing a CIV threat to validity (above average for rater A might be the average for rater B).⁷

The assessment is variable, differing from one attending physician to the other, causing a CIV threat to validity (rater leniency versus severity, central

tendency, halo effect, and variable rater assessment skills).⁷ The poor quality of raters observation skills has been shown to result in missing 67% of the errors committed by residents assessed using a videotape.⁹

Raters are biased because they work closely with the trainees and see them working hard and doing a lot of on-calls. Moreover, there is a lot of competition between ICUs to attract residents for training. This makes raters give an above average or outstanding assessment to most trainees (inflation of the assessment), even if the rating fails to reflect accurate clinical performance (CIV threat to validity).⁷

There is reluctance to give a below average evaluation by most raters (CIV threat to validity).⁷ For example, not a single resident got a below average assessment at the end of the ICU rotation at my institution during the last 10 years, despite the fact that some of them are below average in the rater's judgments (Mum effect).²

In the busy work environment of the ICU, assessment of the trainee's clinical performance is not given enough priority and time by attending physicians. It has been shown that the lack of observation of trainees by faculty is common, and might be a major problem in making an accurate assessment of trainees performance.³

Despite the fact that residents spend most of their time with nurses, respiratory therapists, classmates, fellows, patients and their families, the evaluations are done by the attending physicians with little contribution from other health care workers.

Impact of the Current Situation

The objective of all training programs is to graduate trainees that are able to perform well in real life (professional competence). Programs assess and make inferences about trainee performance during their training and assume that their performance in all similar situations in real life will be the same (generalization).² By making an accurate assessment of trainee performance during training, we hope (an assumption) that such accurate inferences about performance under observed conditions will generalize to all similar situations in real life when they are not observed. Thus, inaccurate assessment impairs training and affects patient safety. It prevents the identification of poorly performing residents that did not fulfill the basic objectives of the rotation and blinds the program directors to poorly performing rotations that do not help residents to achieve the objectives of the rotation and hence prevents implementing corrective measures.

In summary, the current assessment method is subjective, unreliable, lacks validity evidence, and inferences cannot be generalized to trainee clinical performance in real life. This is likely to have a negative impact on the quality of training, the quality of residents, patient safety, and professional standards.

Proposed Solution

The objective of SCOPA is to improve the reliability and validity of the inferences made based on the assessment of the trainee's clinical performance during critical care rotation by:

- Making the assessment structured and standardized, using well defined assessment forms and scales that are encounter-specific and cover all the dimensions of clinical competence to improve content and response validity evidence.^{2,7} The use of structured assessment forms has been shown to improve the accuracy of raters in rating resident performance and decreasing rater variability.^{9,10}
- Allowing for continuous assessment by sampling multiple clinical performances by multiple observers throughout the rotation to capture most of the resident's typical performance. It has been shown that clinical performance is case specific, dependant on experience, training, interest, and personality factors.² Sampling multiple performances using multiple raters will minimize the effect of case specificity, rater variability, leading to improved reproducibility, validity and generalizability.^{2,7,11}
- Assessing the intended outcomes of the rotation (objective-based) to overcome the general subjective nature of the current assessment and to minimize missing some of the objective parameters (decreasing CU threat to validity).⁷ Making the assessment more focused is likely to help raters and improve accuracy.⁵
- Performing the assessment at the time of the performance of interest (at point of care) to avoid problems with recall and the making of blind observations. In addition to minimizing CIV threat to validity, it will provide immediate feedback to trainees.²

The objective of SCOPA is to observe what physicians do in day-to-day practice at the point of care using multiple assessment tools to make more accurate inferences about their clinical performance, which is the apex of the Miller triangle (Figure 1) of clinical competence.¹² Observations will be structured and objective-based to ensure completeness

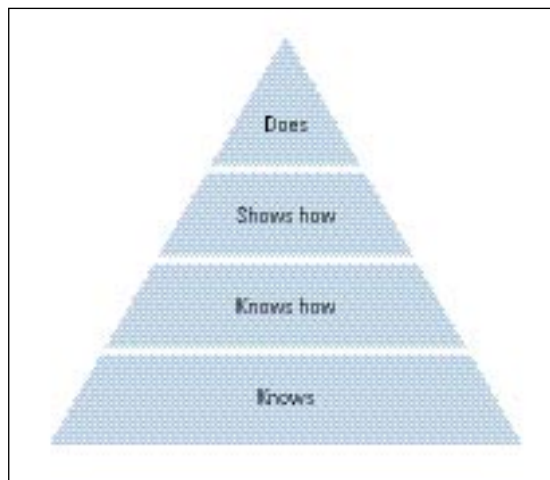


Figure 1. Miller's Pyramid of Clinical Competence.

and decrease variability. SCOPA entails the observation of three areas of clinical performance at the point of care using different tools: clinical skills, procedural skills, and attitude. Each will contribute equally to the trainee assessment ($1/3$ each).

Clinical Skills

Each patient encounter is an opportunity to observe the trainees clinical skills. The clinical skills that can be assessed during the encounter are history taking, physical examination, communication, decision-making, diagnostic skills, management skills, performance under emergency situations, reports, and records.

An accurate assessment instrument and process is an essential quality of accurate clinical performance rating.² To ensure that ratings are the result of the observed performance of interest and nothing else, a new assessment instrument (form) will be developed by the program director in consultation with the ICU staff. The form will be problem specific based on the objective of the rotation. For example, during the ICU rotation, trainees are supposed to deal with patients with shock, respiratory failure, altered level of consciousness, drug overdose, and metabolic disturbances. A problem-specific assessment form will be created for each of these encounters in a structured way (similar to the forms used in an objective structured clinical examination) that covers all the clinical skills areas stated above. Each skill will be rated based on a well-defined 5-point scale. The average rating for all skills will reflect trainee performance during the encounter.

The assessment is to be done by attending staff and fellows at the time of real patient encounter in the ICU. Multiple observations by multiple observers

are encouraged. It will help to assess the agreement between raters, which is an essential quality of accurate clinical performance rating.² Each resident will be handed a pocket size assessment booklet that includes all the assessment forms at the start of the rotation. The resident will be responsible for completing the form on a daily basis at the time of each encounter. In the future, electronic assessment forms will be created to facilitate the process (the current assessment forms are available electronically online). By using an accurate assessment instrument and process, doing continuous assessment of performance and achieving acceptable agreement between raters, one can assume that the inferences from the assessments can be generalized to the universe of trainees clinical performances at all times. The number of encounters needed to produce a reliable assessment in the ICU setting is not known. A pilot study will help to estimate the number of encounters needed.

An average rating in all assessments is the minimum acceptable performance for successful completion of the rotation. During the first half of the rotation, the assessment will be utilized (utility) for feedback to the trainees to identify areas for improvement (formative), while in the last half of the rotation, assessment will be used to make decisions about the successful completion of the rotation. It will contribute to one third of the total rotation rating. The whole process transforms what is done routinely at a subconscious level on a daily basis while working with trainees to a conscious process that is well structured and documented.

This assessment method of clinical skills is similar to the mini-clinical evaluation examination (mini-CEX) of clinical skills which has been shown to have validity evidence.¹³ Moreover, it improves the reliability of the resident's clinical performance assessment because of the multiple encounters and multiple raters involved and it provides better feedback and training for residents.¹³⁻¹⁵ The main differences between the SCOPA and mini-CEX are the specificity of the assessment form to the encounter and the critical care environment; the effect on reliability and validity is to be determined.

Procedural Skills

Each procedure performed by the trainee is an opportunity to observe trainee procedural skills and provide both formative and summative feedback. While there are currently no valid methods for assessment of procedural skills in the ICU setting, direct structured observation of trainee performance

while doing procedures is more likely to yield a reliable and valid assessment as compared to the current use of logbooks and subjective retrospective assessment of skills at the end of the rotation.

Direct observation of practical skills (DOPS) is a method of assessment developed specifically for assessing practical skills by the Royal College of Physicians in the United Kingdom.¹ It requires an observer to directly assess the trainee while doing the procedure. Similar methods of assessment will be used during the ICU rotation. Assessment forms for central venous catheters and arterial line insertion, airway maintenance, endotracheal intubation, bag-valve-mask ventilation, and thoracocentesis will be developed by the program director in consultation with the ICU staff. The trainee will be handed the forms at the start of the rotation and will be responsible to hand the forms to the observer at the time of the procedure. Critical care fellows and attendings will do observations. The observations during the first half of the rotation will be used for formative assessment while the last half of the rotation will be used for summative assessment. The number of observations required to produce a reliable assessment will be addressed by a pilot study.

Attitude

In a multisource feedback (360° assessment) assessment, all health workers (nurses, respiratory therapists, peers, fellows, attendings) will be responsible for completing a structured assessment form that addresses punctuality, attitude, communication, respect, teamwork, reliability, cooperation, enthusiasm, participation in scientific activity, and curiosity of the trainees.¹⁶ For each trainee, one assessment form will be given to each health care team supervisor at the start of the rotation and on a monthly basis thereafter. The ratings average will be calculated and used as the final rating at the end of the rotation. All raters will be anonymous and the assessment will be both formative (copy to the trainee on a monthly basis) and summative. The training program secretary will be responsible for distributing and collecting the forms. Reliability and validity data will need to be assessed. No similar use for such a method has been reported in the ICU setting.

Summary

The current assessment of resident clinical performance falls short of being acceptable. A more structured and objective assessment that samples multiple real patient encounters by skilled observers, such as

SCOPA, is needed. Despite being a major change from the current situation, SCOPA provides an objective assessment of what trainees do in real life (performance), which is the apex of Miller's pyramid and the ultimate objective of assessment.¹² What SCOPA does is provide a structure to such observations, linking it to the rotation objective, and completing the loop by using it for feedback and assessment. It is a

way to reorganize what we are doing now. It brings assessment to life (the bedside) rather than the conference room, written exam paper, or the artificial situation of simulation. The real challenge is to make assessment a hospital and staff priority, assign it more time, and to make it a part of daily routine. This requires faculty development and support from the hospital administration, stakeholders, and the public.

References

1. Wilkinson J, Benjamin A, Wade W. Assessing the performance of doctors in training. *BMJ*. Sep 20 2003;327(7416):s91-2.
2. Williams RG, Klamen DA, McGaghie WC. Cognitive, Social, and Environmental Source of Bias in Clinical Performance Ratings. *Teach Learn Med*. 2003;15(4):273-92.
3. Holmboe ES. Faculty and the Observation of Trainees' Clinical Skills: Problems and Opportunities. *Acad Med*. January 1, 2004 2004;79(1):16-22.
4. Gray JD. Global rating scales in residency education. *Acad Med*. Jan 1996;71(1 Suppl):S55-63.
5. Crossley J, Humphris G, Jolly B. Assessing health professionals. *Med Educ*. Sep 2002;36(9):800-4.
6. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ*. Sep 2003;37(9):830-7.
7. Downing SM, Haladyna TM. Validity Threats: Overcoming Interferences with Proposed Interpretations of Assessment Data. *Med Educ*. In press.
8. Bordage G. Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Acad Med*. Oct 1999;74(10 Suppl):S138-43.
9. Herbers JE, Jr, Noel GL, Cooper GS, Harvey J, Pangaro LN, Weaver MJ. How accurate are faculty evaluations of clinical competence? *J Gen Intern Med*. May-Jun 1989;4(3):202-8.
10. Noel GL, Herbers JE, Jr, Caplow MP, Cooper GS, Pangaro LN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents? *Ann Intern Med*. 1992;117:757-65.
11. Turnbull J, MacFadyen J, Van Barneveld C, Norman G. Clinical work sampling A new approach to the problem of in-training evaluation. *J Gen Intern Med*. Aug 2000;15(8):556-61.
12. Miller G. The assessment of clinical skills/ competence/performance. *Acad Med*. 1990;65:563-7.
13. Holmboe ES, Huot S, Chung J, Norcini J, Hawkins RE. Construct Validity of the MiniClinical Evaluation Exercise (MiniCEX). *Acad Med*. August 1, 2003 2003;78(8):826-30.
14. Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. *Ann Intern Med*. Mar 18 2003;138(6):476-81.
15. Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med*. Nov 15 1995;123(10):795-9.
16. Lockyer J. Multisource feedback in the assessment of physician competencies. *J Contin Educ Health Prof*. Winter 2003;23(1):4-12.
17. Kelly M, Cantillon P. What the educators are saying. *BMJ*. December 13, 2003 2003;327(7428):1393.