Background:
In 1999, the ACGME endorsed six general competencies for residents as part of an ongoing attempt to emphasize educational outcomes as part of residency program assessment and accreditation. (1) Although the focus of these reforms has been on post-graduate training, the competencies certainly are applicable to medical students who aspire to become excellent clinicians, and can help guide the development of assessment tools to measure clinical and professional performance of students in the emergency department.

The six general competencies are: patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and systems-based practice. Though not emergency medicine-specific, these competencies are relevant to the practice of all clinical medicine and their standards can be modified and adjusted to the emergency setting, the level of training of the student, and the pre-determined standards of acceptable performance determined by the course director.

Similarly, Epstein and Hundert, defined professional competence as the successful integration of the following skills: the skill to acquire and use knowledge to solve problems, the skill to synthesize biomedical and psycho-social data in clinical reasoning, the skill of effective communication, and the emotional awareness necessary to apply all of these skills during patient care. (2) They also outlined a framework for assessment, which incorporates Miller’s progressive levels of competence: knows (factual recall), knows how (solves problems, describes procedures), shows how (demonstrates skills), and does (applies in real practice). (3)

This chapter addresses the evaluation of students’ clinical and professional performance (knows how, shows how, and does), with an eye to the ACGME competencies and the ACGME Toolbox of Assessment Methods©. (4) It outlines and reviews various assessment tools which may be used for this purpose. Their application in the setting of an emergency medicine rotation is described and discussed.

Overview:
Until the 1960s, assessment of medical trainees’ clinical skills relied on the results of oral and written examinations, supplemented by resident and faculty ratings of students’ clinical abilities. Faculty ratings have historically been based on observation, direct or Socratic questioning, and case discussions. Unfortunately, these evaluations were often limited, based on casual, random observation by untrained observers, and were subject to inter-patient variability in clinical problem-solving requirements. In the last half of the 20th century computer simulations, evaluations of data interpretation (i.e. ECGs, x-rays), and case write-ups were added to the evaluator’s toolbox of student assessment methods. In 1999, Kassebaum published a survey of 97 U.S. medical schools’ evaluation techniques. (5) The authors found that faculty and resident ratings still contribute 50 - 70% of a student’s grade in core clinical rotations. However, the majority of these ratings were based on a composite of random, limited presentations and partially observed physical examinations.

The best assessment instruments should yield valid, accurate data which is consistent, and therefore, reliable in reproduction. The assessment method should also be feasible within the bounds of time, training, and equipment costs, and applicable to the specific clinical circumstances being measured. For example, although an assessment tool may work in some clinical settings, it may not apply in the emergency department. Finally, it should provide useful information on the student’s abilities as compared to the competencies being measured. (6)

The assessment should be consistent with the rotation’s curriculum goals and objectives. The criteria for “excellent”, “satisfactory”, and “fail” should be based on pre-determined and specified standards, which are available to the student. Finally, clinical competence is multi-dimensional. No single assessment tool is usually able to evaluate all of the competencies of interest to the clerkship director. Therefore, multiple assessment tools are often necessary to completely evaluate the student’s clinical and professional performance. (6)
(I) Evaluation forms for distribution to faculty and/or residents:

Global rating scales are judgments of general abilities and are considered “potentially applicable methods” of evaluation by the ACGME for the evaluation of the development and implementation of patient care plans, medical procedures, and teamwork. Composite evaluations of observed student performance are still the most common means of assessing clinical skills. The clerkship director cannot be present on every shift with every student and often relies on the feedback of fellow faculty and senior residents, if available. This composite method of evaluation is inexpensive and a minimal burden on time and equipment resources. However, due to their very nature, evaluation forms may be fraught with other problems.

Unfortunately, global rating scales are subject to the problems of inter-observer reliability, incomplete observation of student performance, or ratings by untrained observers (busy residents). Reliability and validity may also be limited by a lack of discrimination between various competence domains, the “halo effect”, and rater bias. Lamantia demonstrated such inter-observer variability in faculty evaluation of emergency medicine resident skills during videotaped standardized patient encounters. The authors found that raters’ variability in scores was similar whether they used general assessment forms or specific, objective performance-based criterion checklists. Despite these findings, faculty evaluations of students may be improved by explicitly stating the objective criteria for grading on evaluation sheets.

For example, Go published a competency-based approach to clinical evaluations of students in the emergency department at University of Missouri-Kansas City School of Medicine. The author developed a five-point rating scale to evaluate students’ clinical skills in ten competency areas. This performance matrix was reviewed by the student and attending physician at the end of each shift, then entered into a central database for final cumulative evaluation. This method allowed for instant formative feedback on the students’ performance and improved physicians’ compliance with the evaluation system to nearly 100%. Similarly, Bandiera published the results of a ten-sub-domain global assessment form with a five-point rating per sub-domain. Unfortunately, they found that the more complex the rating matrix, the less discrimination there was between the sub-domains being assessed.

Based on a review of the above literature, the author makes the following practical suggestions for the successful use of evaluation forms by faculty and residents:

- Evaluation matrices should be specific, but simple; one form should not attempt to evaluate too many domains of competence.
- Specific examples should be given of what level of skill or behavior is considered “fail”, “satisfactory”, “honors”, etc.
- Faculty and residents should be trained in the use of an evaluation matrix card and familiarized with the specific criteria for evaluation at various levels of expected competence.
- The evaluation form should be simple and easy to use if students will distribute it to faculty and residents at the end of each shift.
- If evaluations are distributed by e-mail, hard-copy, or web-based, attaching digital photos of students to the forms will help remind faculty and residents of who they are evaluating.

(II) Direct observation in the clinical setting:

Direct observation of an entire student-patient encounter, especially when compared to a predetermined checklist of performance-based criteria for competence, may be the “most desirable” method of evaluating interviewing skills and proficiency with medical procedures, per the ACGME. This evaluation method is similar to the “long case” used in the United Kingdom in evaluating clinical clerks. Direct observation of the student’s history-taking, physical exam and procedural skills can provide a wealth of information on the student’s competence in that one clinical encounter. When followed by Socratic questioning or a brief oral examination on the clinical topic, the long case can provide additional information on the competencies of investigatory thinking, application of basic science knowledge in the clinical setting, and the use of evidence-based medicine to solve clinical problems.

Unfortunately, the evaluation of students using non-standardized patients in the clinical setting may not provide reliable or valid information on the student’s overall competence. Inter-patient variability in clinical problems and a patient’s ability to be interviewed are only two potential confounders to this method of evaluation. Wass compared the use of non-standardized patients with standardized patients in an OSCE for the evaluation of history-taking skills. The authors found that if history-taking long cases
with non-standardized patients were observed, it would take approximately 3.5 hours and 10 patient encounters to produce a reliable test of history taking.

Although it could provide excellent formative feedback to a student to have an entire patient encounter observed and critiqued during each shift, the use of long cases, or non-standardized emergency department patient encounters to evaluate students would require many hours of evaluator time and may not be acceptable to patients. As a formal method of evaluation, direct observation may be unfeasible on many busy emergency medicine rotations. It may be generally impractical unless there is a dedicated teaching faculty or teaching resident present in the department whose sole responsibility is the observation and critique of students’ clinical performance. However, if time allowed and specific criteria for evaluation were developed, several scheduled observations might provide additional data to evaluate the above-mentioned competencies.

(III) Simulated patient encounters:

(i) According to the ACGME Toolbox, the OSCE (observed structured clinical examination) is an excellent and most desirable method for evaluating interviewing and patient counseling skills, physical examination techniques, interpersonal and communication skills, and professionalism. It also provides information on the student’s decision-making and practice-based learning. Most OSCEs employ multiple assessment stations with standardized patient encounters, and data interpretation exercises. Students rotate through each station and are observed and evaluated using checklists of pre-determined performance-based criteria.

Valid and reliable OSCEs usually require at least seven, and usually ten to 20 stations of variable length. Inter-observer reliability is improved if multiple evaluators are present at each station. OSCEs require significant time, equipment resources, and finances. They may not be feasible on a frequent (monthly) basis for the routine evaluation of students. Team OSCEs have also been described which consist of teams of five students rotating through five different clinical stations. (11) During each station, each student performs a different task and is evaluated. Although the authors found this method of evaluation more feasible than standard OSCEs, there was some variability in any individual student’s experience at any one station.

(ii) Standardized patients (SPs) are trained persons who simulate a medical condition in a standardized way in order to evaluate the clinical skills of trainees. SPs are often used in OSCEs and according to the ACGME, are the most desirable way to evaluate clinicians’ caring behaviors, communication, physical exam and listening skills, and professionalism. SPs are already part of the Canadian medical licensure process and the Educational Commission for Foreign Medical Graduates certification process. (12) SP examinations are under development for the U.S. Medical Licensing Examination and are used in comprehensive examinations at many medical schools in the U.S. SP examinations have also been used to evaluate emergency medicine residents’ patient assessment, chart documentation, and overall patient care. (13) Of note, the NBME has announced an intention to make a clinical skills examination a requirement for Part II of the USMLE in 2004.

Groups of SPs are often associated with medical schools. However, the cost of training and hiring SPs for routine student evaluation may make this method of assessment unfeasible. An alternative might be allied health professional students (PA, EMT, NP) who could also be hired and trained to simulate standardized clinical complaints at a lesser cost. Case scenarios for SPs should incorporate the specific competencies being evaluated, most often history and physical exam skills, and be assessed through specific checklists of pre-determined skills to be performed during the SP encounter.

(iii) Simulators and models are gaining popularity in the teaching and evaluation of clinical skills. High-fidelity simulators assess performance in situations, which closely resemble reality and imitate clinical problems. Procedural skills such as intubation, tube thoracostomy, and venous cannulation can be learned, performed and evaluated in a “safe” setting, which puts no patient at risk. (14) In fact, the ACGME lists simulators and models as one of the most desirable methods for evaluating procedural proficiency. Simulations also provide information on the student’s ability to develop and implement a clinical management plan, apply knowledge of basic sciences to clinical problems, and use analytical thinking. High-fidelity simulators have been used to evaluate students’ core knowledge and anesthesia skills. (15) More recently, Murray published the use of a simulation of acute trauma with hypotension to evaluate the clinical skills of 3rd and 4th year medical students, and PGY-1 emergency medicine residents.
(16) Once the case had been approved, students and residents were rated using both a checklist of weighted “essential behaviors” and a holistic rubric intended to assess thought processes, action, and integration in the resuscitation and care of the patient. The authors found that the simulation was a reliable and valid measure of clinical performance.

Although simulation centers are gaining acceptance and being built at some medical schools, not all clerkship directors have access to such technology. A cheaper alternative is a paper case simulation, or patient management problem (PMP), similar to the ABEM oral examination. Paper cases will not allow observation of procedural skills. However, if cases are designed appropriately, they can provide information on students’ analytical thinking, application of basic sciences to clinical problems, and use of evidence-based medicine to manage case scenarios.

(IV) Evaluation of procedural skills:

The evaluation of proper and safe technique when performing procedures is best evaluated by direct observation of clinical work measured against a checklist of essential actions during a procedure, or through performance on simulators or models. Global rating scales often ask about procedural competence, but are subject to the potential problems listed above. Surveys of nurses and patients themselves may give some information on the student’s technique, but these perspectives may not answer the desired questions of individual steps involved in a more complex competency, or technique itself. Students may keep self-reported logbooks of procedures performed, but these are also not sufficient to evaluate the technique of the student.

Procedures can be performed in non-clinical environments and observed for appropriate activities and skills. For example, if available, students may practice and be evaluated on central line placement, cut downs for venous access, intubation, and tube thoracostomy placement in an anatomy lab. Alternatively, they may be evaluated on appropriate splinting technique in a splinting lab in which they fashion and apply various splints to each other.

(V) Other ACGME toolbox ideas, which might be applicable to the clinical evaluation of student performance:

(i) Chart stimulated recall (CSR) oral examinations are standardized oral examinations based on the patient cases, which the student has seen during the clerkship. The student presents copies of his/her charts and a sample is selected for a standardized examination of the clinical problem. This method of evaluation is not financially burdensome and allows for a most desirable evaluation of clinical decision-making and the ability to formulate and carry out a management plan. CSRs also assess the student’s investigatory and analytical thinking skills and ability to use evidence from scientific studies to solve clinical problems. In addition, chart reviews can become sessions in evaluation charting skills themselves.

(ii) Case write-ups and presentations of clinical cases seen by the student may also show his or her ability to cogently present and analyze a clinical problem, as well as critically assess and synthesize current literature on a topic.

(iii) 360º evaluation instruments allow “multiple people in a person’s sphere of influence” to provide feedback on such behaviors as teamwork or communication skills. These ratings evaluations could be completed by emergency nurses, allied health personnel, patients, and peers, and provide additional data to that of faculty and resident evaluations. A number of evaluations are needed to correlate with other methods of evaluation, and may provide better formative feedback for the student, rather than a summative assessment of clinical performance.

(iv) Portfolios are in use in the U.K. and Canada and are gaining acceptance as methods of documenting some clinical competencies in graduate medical education. Student portfolios might include a combination of case write-ups, literature reviews, procedure logs, and formative feedback forms obtained from faculty and residents over the course of one clinical rotation.

Summary:
- The clinical competencies to be assessed should be determined.
- These competencies should be made explicit and understandable.
- Multiple evaluation tools should be considered in order to assess various aspects of clinical competence and performance.
• Students should be aware of the expectations for clinical performance and the assessment tools being used to evaluate them.
• Faculty and residents should be trained in the use of chosen assessment tools.
• Faculty and residents should be trained in how to be specific observers of clinical performance.
• The more observations of student performance, the more reliable the data.
• The more observers of student performance, the more reliable the data.
• Using various means of assessment successfully requires creativity.

References:
1. The ACGME Outcome project: http://www.acgme.org/outcome
4. Toolbox of Assessment Methods© 2000 ACGME and ABMS. A product of the joint initiative of the ACGME Outcome Project and the Accreditation Council for Graduate Medical Education (ACGME), and the American Board of Medical Specialties (ABMS). Version 1.1, September 2000.
6. Lynch DC, Swing SR: Key considerations for selecting assessment instruments and implementing assessment systems. The ACGME Outcome project: http://www.acgme.org/outcome