Chapter 6
Assessment of Clinical Education

Anthony Luder

Clinical studies which constitute at least half of the time devoted to training doctors, are the most important part of an MD course. In most countries, the demonstration of the ability to incorporate acquired clinical knowledge and skills in the care of patients constitutes the final step in the completion of a medical degree, entitling the graduate to practice as a doctor. The successful study of clinical medicine requires the mastery of a complex mix of theoretical, practical, and humanistic skills. These span the gamut of Miller's knowledge and behavior pyramid—knowing, knowing how, showing how and doing [1]. The progressive assessment of the clinical student through all these levels is required in order to ensure that doctors have the personal and professional skills and attributes upon which modern medicine depends. This is a formidable task [2].

Clinical medicine is taught and learned in a wide variety of inpatient and ambulatory locations, ranging from hospital inpatient and emergency departments, institutes, clinics, and operating rooms to community centers and doctors' offices. None of these places have teaching or learner appraisal as their priority; rather their raison d'être is the care and treatment of patients. The time a student or resident may spend in any one environment may be as short as a few days or as long as many months. Longitudinal clerkships provide the opportunity for students to place patient cohorts, rather than specific disciplines to be the organizational center of their clinical experience. Longitudinal clerkships of 3 to 12 months have been successfully implemented for selected students in many medical schools with academic success equal to that of classic discipline-based block rotations. Most clerkships are compulsory, but elective and/or selective clinical experiences, which are also important for broadening students' education, have an important role to play.

A. Luder (✉)
Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel

Department of Pediatrics, Ziv Medical Center, Safed, Israel
e-mail: luder.a@ziv.health.gov.il

Although the essence of clinical teaching is based on the ages-old apprentice model, much effort is given to the strengthening of scientific knowledge and evidence-based reasoning both for clinical practice and also for research and development. Most students perform their studies in small groups or as individuals. They meet a wide variety of health professionals who inevitably have a broad range of abilities as teachers. As well as learning medicine, the successful student also has to master working within the healthcare system, a skill which some learn faster than others. The student is required to make a rapid transition (the "white coat" ritual) between the safe and sterile academic classroom of the initial years to the real life maelstrom of actual people with illnesses and suffering.

The transition to clinical medicine and the complexities of patient care place inevitable stresses on the students. Theoretical universal knowledge gained in the preclinical stages must be applied to complicated disease presentations in a great broad diversity of individuals often under emotional and temporal pressure. There are also major challenges in communication, empathy, ethics, professionalism, and endurance. These are the backdrops against which students must be evaluated in robust and dependable ways.

Whatever methods are used to appraise students, they must be judged against the following criteria [3]:

- Transparency. Staff and students must clearly know in advance what will be tested, when, where, and how.
- Pedagogic robustness. Testing must be as objective and reliable as possible, in the sense that if the same student were to be tested twice in the same circumstances, the same result would be expected.
- Appropriateness. The testing method should test the aspect required and be relevant. It should be of as high quality as possible. Written examinations should preferably be all new, be original and be carefully written, prepared, edited, and vetted by experienced examiners with time devoted and protected to this endeavor. In practice this high bar is not always achievable.
- Fair. Administration should be strictly supervised in quiet, clean, comfortable, and ordered surroundings.

The best way to achieve all these goals is to use a variety of methodologies and at different times during the clinical curriculum [2]. Prerotation testing primarily examines base knowledge learned predominately in the preclinical curriculum; in-rotation evaluations test performance; and completion appraisal should test both these as well as integrated competencies such as clinical reasoning, clinical judgment, and therapeutic interventions. This will provide the best chance of fair and accurate judgment of a student as he learns and develops over time. The following scheme is a general plan that can be adapted as necessary according to place, time, and specialty.

Prerotation Testing

It is often desirable that a clinical placement begins with the students having a basic understanding of the clinical specialty before the first day. The syllabus should set out clearly what is expected. The students may complete self-instruction by reading selected parts of textbooks or online modules as well as participating in lectures or small group activities such as projects, seminars, or tutorials. Increasingly students have simulated and real-patient encounter experience even during the preclinical period. The implementation of personal portfolios may be useful in many faculties to enable staff and students to keep records and track their experiences, achievements, and learning materials [4]. These are especially important for documentation during longitudinal clerkships.

> **Portfolios [4]**
> Portfolios are personal records of data, experience, and achievement kept and updated by the student. They used to be printed but now they are generally electronic. Among the items recorded can be the following:
>
> - Basic course information: syllabus, curriculum, timetables, diaries, campus maps, contact information, lesson plans, exam practices.
> - Course materials including reference lists, articles, book sections, data and information sheets, presentations, video and sound clips, pictures, illustrations.
> - Patient and procedure records, clinical log books.
> - Progress charts: subjects mastered or requiring further study, problems requiring solution, marks or grades achieved, priority lists for study and action, questions for faculty.
> - Discussion or chat spaces for interaction with other students and faculty, links to Email and social media.
> - Reflection and personal thought.
> - Feedback.
>
> Portfolios are essentially private records but the student and staff can define certain parts for sharing with staff or even public view. In some medical faculties, portfolios can be awarded quality grades for completeness and detail. The portfolio can permit teaching staff to make additions such as entering feedback or marks.

In many of the longer traditional clinical rotations, students may have a formative or summative assessment before the placement. This encourages knowledge retention and revision, which will optimize the learning experience during the rotation itself. Written examinations using a form of multiple-choice format are popular for this purpose since they are efficient and robust [5].

Multiple-Choice Question (MCQ) Examinations [5]

Writing Multiple Choice Questions requires patience and time. Questions should focus on core subjects and not peripheral or trivial themes. The level of knowledge or judgement tested should be appropriate for the experience and level of the learner and "copy-pasting" questions from higher-level exams is not appropriate. In general, all questions should be new, but in practice not more than 20% of the questions should have appeared on exams within the past 3 years. The following are some guidelines for successful MCQ's:

- The question root should be as clear and concise as possible, usually not more than 2–3 lines. Only relevant information should appear.
- There should be a mix of difficulty, so that about a third of the questions are hard, a third medium and a third easier.
- The examination should be analysed for psychometric validity and robustness after completion and invalid questions deleted.
- Single best answer (SBA) questions are now most favored since they test clinical knowledge, not the ability to interpret grammar or think inversely, which are skills not being tested by the exam.
- The answer options should not introduce new elements not included in the root.
- The answer options should all be approximately the same in length and complexity.
- It is bad practice to write answers in the negative ("all the following are right except") as this can confuse the examinee.
- Consecutive questions should not depend on the answer of the previous question, since failure in one question will necessarily lead to failure in the next.
- Questions should be plain with one best answer and not include options such as "all of the above are correct" or "none of the above".
- It is important that incorrect answer options are as plausible as possible but not deliberately misleading.
- Questions should not be based on matching a list of options with a list of answers.
- The number of answer options may be 4 or 5 but should be consistent.
- Allow about 90 s for each question and sufficient time (30 s) to transfer and check responses to computer cards, if these are separate.
- Short tests can include 20–30 questions; final examinations may include between 150 and 200 questions.
- In the examiner's copy the correct response should be clearly marked (by bolding or adding a symbol) and a page reference to the course text material appended for later reference, appeal, and quality control.

In-Training Evaluation

This stage is critical since the student can be helped to improve and optimize his experience in real time, while he is still in the rotation and there is time to correct faults and difficulties. It is the opportunity for the assessment of performance in practice or, in Miller's taxonomy, the stage of "doing" [1]. Many ways of evaluation can and should be exploited in order to maximize coverage and minimize bias:

1. Individual and group tutor meetings. The student will have regular tutor encounter assessments during which progress and obstacles can be openly discussed and solutions for problems identified. A summary meeting will take place at the completion of the placement.
2. Student seminars. These provide an opportunity for each student to learn a topic in detail and gain experience in presentation to colleagues, who gain from his efforts accordingly. These should be based on actual patient cases and experiences. The attending tutor will assign a grade based on quality of delivery and information presented.
3. Patient clerkings. Each student will be given the task of periodically submitting full written reports of clerked patients, for example, weekly in a long rotation. The reports should be original work and include a full history, physical examination, problem list, investigation plan, list of suggested and differential diagnoses, clinical and prognostic discussion, treatment, and follow-up plan. These submissions should be carefully read by the tutor or clinical instructor and returned with annotations, corrections, and comments. They may also be used for oral case discussions during which each item of the patient write-up is awarded a mark.
4. Participation in routine activities. The clinical staff will form a cumulative impression of the student's motivation, behavior, appearance, participation, knowledge, skills, insight, and judgment. Examples of activities include rounds, meetings, presentations, clinics, and procedures. The student should be observed at patient encounters whenever possible. The so-called Mini Clinical Evaluation Exercise (mini-CEX) and observed procedure (OP) are structured forms of these evaluations during which a staff member provides formal feedback after an actual clinical task [6]. As many staff members as possible should contribute to this process in order to ensure as nonsubjective a result as possible.

Mini-CEX
The clinical task is pre-prepared and constructed. This might be the examination of a body system, the taking of a specific history or the performance of a task like taking a blood sample.
Observed Procedure
The student is observed directly or indirectly during a clinical experience in real-time with actual patients, such as clerking or counseling a patient.

5. Task list. It is important for medical teaching staff and students to have a clear alignment of expectations before the rotation begins. An excellent way to follow progress is the use of a clinical logbook [7]. The student will be presented with a comprehensive list of medical (and where appropriate nursing or other para-medical) activities and procedures in which they will be expected to witness and participate. This is not designed to be a spectator activity, but the student is expected to know the indication, purpose, dilemmas, and any possible complications of each medical activity and procedure. The responsible staff member will sign off on each satisfactorily completed task. A post-graduate variation of the log book is known as multi-source feedback (MSF). Other faculties might use an electronic 'log book'.

Clinical Log Book [7]

Clinical log books list for the student and staff a record of procedures and tasks successfully completed during the attachment. The student has a staff member sign off on each item completed. These tasks form part of the requirement for completing the attachment. Examples of these tasks vary greatly between different departments but some general ones could include:

- Submission of written patient intakes for assessment, including history, physical examination, problem list, suggested investigations, diagnosis and differential diagnosis, clinical discussion, management, and follow-up.
- A list of on-calls performed.
- Medical procedures observed or performed and understood: births, operations, resuscitations, blood taking, insertion of tubes, catheters, and drains, LP, bone-marrow or biopsy, insertion of venous or arterial lines, point-of-care measurements or imaging.
- Nursing procedures observed or performed and understood: measuring temperature, pulse, oximetry, blood pressure, height and weight; feeding non-independent patients; changing dressings; evaluation of pain; administration of drugs and fluids by different routes.
- Seminars delivered or clinical case-conferences managed.

Table 6.1 summarizes how the competencies are assessed using these multiple methods.

Completion Appraisal

A student learns multiple competences during his clinical studies. His ability to perform the multilayered tasks demanded in medicine is evaluated at the completion appraisal, although there may be intermediate tests as well. In the past, most of the

Table 6.1 Table of competencies and evaluation methods for in-training students

Competency	Tutor interaction	Task logbook	Student case presentations	Student seminars	Participation in routine clinical activities	Written patient clerking notes
Progress and obstacles	✓	✓				✓
Exposure to procedures		✓			✓	
Clinical knowledge			✓	✓	✓	✓
Clinical judgment			✓		✓	✓
Patient integration			✓			✓
Study goals			✓	✓		
Professionalism	✓				✓	

burden of clinical evaluation has been placed on this aspect of testing. Unfortunately, there are still too many examples of inadequate pre- and in-rotation performance evaluation which fail to identify a failing student or provide encouragement to the snail or the stumbler. The result might be an unexpected relative or complete failure at the final testing hurdles, a dispiriting and occasionally devastating experience for the student. For the faculty, the failure might be considered a huge waste of resources. It is clearly incumbent on the teaching institution to provide a system which is based on the methodological criteria listed above.

Four main systems of evaluation are in use in different parts of the world: written, oral, practical, and simulation. In an ideal system, use will be made of all of these options as they each have strengths and weaknesses.

- Written tests are used particularly to test knowledge, clinical thinking, and problem-solving. They can be of different forms, including multiple-choice, multiple-answer, short, and long essay forms. Single-best-answer (SBA) types of multiple-choice tests have proven to be very popular since they lend themselves easily to computerization for rapid marking and psychometric validation analysis, which makes them relatively light on resources. Written examinations can be broadly based and include graphic and pictorial items and online material. On the other hand, creating high-quality tests is a complex and highly demanding process (MCQ tests discussed above).
- Oral tests can take the form of virtual case discussions over 10–20 min during which the student is taken through a structured staged scenario by one or more examiners, and is awarded marks for the correct completion of each stage. This is a demonstration of "knows" and "knows how." Such cases can make full use of graphics, sound, and video clips and any other enrichment that modern technology can provide.

Oral Tests ("Viva Voce")

One or two examiners test the ability of the student to understand a clinical scenario and develop a clinical discussion. The student is given a short one or two-line summary of the case ("a 50-year-old man with a productive cough" or "a 7-year-old girl with abdominal pain"). The student may make notes and is required to ask pertinent questions related to the case and then the case develops with information supplied by the examiner including physical findings, photographs, illustrations, and investigations, according to the requests of the student. The student summarises the problems and suggest diagnoses and management. At the end, some questions may be asked by the examiner regarding the disease process, its treatment, and follow-up. The examiners have a list of ready-made questions they must ask and required answers to evaluate, so that standardization is maximized. A quiet comfortable room is essential and the examiners should give their entire attention to the process.

- In practical tests, the student is expected to complete a defined task under observation ("show how"), which may be direct or remote via a camera or one-way window. Examples include taking a history, performing a partial or full physical examination, or perhaps demonstrating a procedure such as testing a urine sample or performing an ECG. It should be borne in mind that the presence of the observer often affects the performance. However, in the objective structured long examination record (OSLER), a student is typically left unobserved with a real standard or simulated patient for up to an hour and then has a feedback interview with the examiner for another 20–30 min. If observed, the student should be evaluated both on the technical aspects of the task (e.g. by the use of a checklist) and his behavior, language, confidence, structure, fluency, and deportment during the session. In cases in which an actor, standard, or real patient is involved, they should also provide their own feedback using standard rubrics [8].

OSLER [8]

OSLER examinations require planning to ensure that prepared and cooperative patients are available. This can be a logistical challenge for large numbers of students and therefore OSLER examinations are not always appropriate. Patients should be thoroughly briefed and if possible "standard" patients with fixed known physical signs used. These patients can also provide feedback. The fields to be assessed include presentation, physical examination, investigation, appropriate management, and clinical acumen. The student can be left unobserved for 20–30 min to complete the history and examination, since the act of observing can be a confounder for some students. In well-planned and executed examinations, OSLER have been shown to have good validity and reliability.

- Simulation has become increasingly important as an educational and evaluation methodology [9]. It offers a wide range of possibilities using models, manikins, and an almost unlimited range of computerised, virtual, and online technologies. Simulation allows excellent appraisement of "knows how" and "shows how". It has become increasingly used in postgraduate assessment especially in the emergency, trauma, and surgical specialties. Simulation examinations are best carried out in specialized education center which can provide a high-quality observation environment, logistic and timing control, and evaluation tools.
- The Objective Structured Clinical Examination is a system designed to provide exposure to all these examination techniques by the employment of numerous stations. At each station, the student may be tested by a different examiner, using a different method [10]. The greater the number of stations and examiners, the less possibility exists for interference by subjective, random, and extraneous factors, and, therefore, the overall process is more stable and the result more reliable. OSCEs are used both for formative and summative assessments. OSCEs are demanding of time and resources, but their advantages have made them universal since their introduction in the 1970s, and they are now considered to be the most dependable and valid way of performing clinical evaluation in undergraduate and postgraduate settings.

Objective Structured Clinical Examination (OSCE) [10]
- A typical OSCE would include 10–24 stations with each station lasting from 5 to 15 min.
- Each station should examine a specific and focused skill or item, and include itemized history taking using simulated patients, physical examination technique with actors or mannikins or models, clinical thinking, interpretation of images, physiological and laboratory investigations, and report preparation.
- Evaluation can be performed using check lists for itemized skills and rubrics for style and fluency. Both examiners and actors/patients can provide assessments.
- Stations can be duplicated or combined when longer tasks are involved.
- If feedback is provided after the station by an examiner or using a standard "school solution", time should be allotted accordingly. Feedback can also be given after the OSCE. Feedback should be given within a reasonable time and involve details of the student's performance at each station.

- The overall mark is not necessarily an arithmetic average of all the stations; sometimes very poor performances at a number of tasks may fail a student even if in other competences he has succeeded and gained an overall pass grade. Whatever system is adopted, it should be clear to all before the examination.
- Rigorous pre-planning, staff and examiner briefing and real-time logistic control are vital for a successful OSCE. There should be clear announcements 2 min and 30 s before the end of each station and strict adherence to timetables. Typically, there would be a 2-minute space between stations for preparation and thought. Students should be instructed to use the toilet before the examination and that no eating, drinking or smoking is permitted during the examination. The smartphones and computers of students and staff should be switched off or removed. It is common to provide light refreshments if the examination lasts for more than 2 h.

Table 6.2 summarizes the pros and cons of completion appraisal methods.

Table 6.2 Table of completion appraisal systems—pros and cons

System	Pro	Con
Written exams including MCQ	• Test knowledge • Simple to administer • Cheap	• Require validation • Don't test competency
Oral voce viva	• Tests clinical thinking and problem-solving • Can use images and graphics • Resembles real life	• Labor intensive • Difficult to standardize
Practical observation/ OSLER	• Tests actual skills • Real or actor patients can be used	• Time-consuming • Subjective • Logistically complex
Simulation	• Safe—no harm can be done • Tests "show how" • Broad range of technologies available	• Expensive • Require special equipment and setup
OSCE	• Maximizes validity • Multiple stations with a variety of tasks	• Time- and resource-demanding • Labor-intensive

Summary

In this chapter the challenges involved in assessing clinical medical students at different stages of their training have been described. Miller's taxonomy provides a clear framework of progress in the process of learning medicine, and each stage requires tailored methods and techniques of evaluation. Detailed descriptions of these as well as their relative advantages and drawbacks have been described. Common to all systems of assessment is the need for reliability, validity, and academic dependability. It is hoped that the reader has gained a clear idea of the options available and some practical assistance in implementing them.

Take-Home Messages
- Assessment of clinical students should be appropriate for stage and purpose, continuous, transparent, robust, and fair.
- Appraisal techniques include written, oral, observational, practical, and simulation methodologies.
- Portfolios and logbooks are useful tools for continuously following academic and professional progress.
- Different methods may be used at different times, places, and stages of training.
- A combination of evaluation techniques is most likely to provide the highest quality evaluation of medical learners.

References

1. Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine (Supplement), 65*, S63–SS7.
2. Harden, R., & Laidlaw, J. (2016). *Essential skills for a medical teacher* (2nd ed.). Amsterdam: Elsevier. ISBN: 9780702069581.
3. Epstein, R. M. (2007). Assessment in medical education. *The New England Journal of Medicine, 356*, 387–396.
4. Carraccio, C., & Englander, R. (2004). Evaluating competence using a portfolio: A literature review and web-based application to the ACGME competencies. *Teaching and Learning in Medicine, 16*, 381–387.
5. Case, S., & Swanson, D. (2000). *Constructing written test questions for the basic and clinical sciences* (3rd ed.). Philadelphia: National Board of Medical Examiners.
6. Kogan, J. R., Bellini, L. M., & Shea, J. A. (2002). Implementation of the mini-CEX to evaluate medical students' clinical skills. *Academic Medicine, 77*(11), 1156–1157.
7. Patil, N. G., & Lee, P. (2002). Interactive logbooks for medical students: Are they useful? *Medical Education, 36*(7), 672–677.

8. Gleeson, F. (1997). AMEE Medical Education Guide No. 9. Assessment of clinical competence using the Objective Structured Long Examination Record (OSLER). *Medical Teacher, 19*, 7–14.

9. So, H. Y., Chen, P. P., Wong, G. K. C., & Chan, T. T. N. (2019). Simulation in medical education. *The Journal of the Royal College of Physicians of Edinburgh, 49*, 52–57.

10. Pugh, D., Touchie, C., Humphrey-Murto, S., & Wood, T. J. (2015). The OSCE Progress Test – Measuring clinical skill development over residency training. *Medical Teacher, 38*(2), 1–6.