

Assessment

Dr Martin Rhodes

MA (Lond) MBBS FRCGP

GP and Honorary Senior Lecturer, Dept of Primary Health Care and General Practice
Faculty of Medicine, Imperial College, London

This paper was first written in 2003 as part of a project led by the London Deanery to provide a web-based learning resource to support the educational development for clinical teachers. It was revised by Judy McKimm in 2007 with the introduction of the Deanery's new web-based learning package for clinical teachers. Each of the papers provides a summary and background reading on a core topic in clinical education.

Aims

This paper:

- Provide an introduction and overview of the major assessment methods and underlying educational principles used in healthcare education and training
- Raise awareness of issues concerning assessment, feedback and appraisal in the clinical context
- Enable you to apply knowledge and principles to everyday practice

Content

- Introduction
- Dimensions of assessment:
 - Purpose
 - What is assessed
 - What the assessment is measured against
 - Quality in assessment (validity and reliability)
- Assessment instruments:
 - Tests of knowledge (MCG, EMQ, Short answer/MEQ, essay, computerised assessments)
 - Tests of skills (OSCE, long case, communication skills, in-course competency assessment)
 - Other assessment instruments (teachers assessment, viva voce, case analysis/case based orals, self assessment, portfolios, logbooks and rating scales, projects, patient questionnaires)
- Conclusion
- References

Introduction

This paper states the principles of good assessment practice and discusses some important issues relating them to the assessment of medical students and postgraduates. It then critically examines some common assessment instruments and describing their uses and their strengths and weaknesses.

Dimensions of Assessment

As with many things (especially medicine!), once you understand the terminology, you are half way there. Below is a list of terms commonly used in assessment which you should be familiar with.

First of all **assessment** itself:-

Assessment:- the use of a wide range of methods to determine the attainment of a learner or trainee

Evaluation:- A determination of how a particular educational input, course or programme (or indeed assessment) has performed

It is important to note that this distinction is not always used the same way in the literature especially from the USA.

Next **different types of assessment**:-

1. PURPOSE

Although assessment may be part of an evaluative process of an institution or course, or have other purposes, teachers use assessment for either summative or formative processes.

Summative assessment. To determine how well a learner or trainee has performed at a fixed point in a course or programme.

In healthcare and medicine, as elsewhere this may well result in **accreditation** or **licensing**, for example allowing a healthcare professional to practice, or **revalidation** (e.g. allowing a doctor to continue to practice in a speciality). When this occurs the assessment is termed '**high stakes**' (see box) as so much depends on it. Some of these tests are **minimum competence** (see box) which aim to discriminate between someone who is fit to continue a course of study or enter into specialised training or practice against someone who isn't. It is not only in the interests of fairness and transparency is it important that such a test should have clear criteria against which candidates are judged. All adults but especially professionals need to be aware of their own strengths and weaknesses, their thinking and consequently their learning needs (**metacognitive skills** or "knowing about knowing") and therefore need clear criteria against which they can judge themselves. Unfortunately this has been all too

rare in medical examinations which could learn from the example of National Vocational Qualifications (NVQs) (*Jessup, 1991*) and of those in other health professions.

Summative assessments are reported in the form of grades, marks or pass or fail scores.

Formative assessment. Takes place during the course or programme and feeds back into the learning and teaching process with an aim of helping the learner learn (and for the teacher to evaluate how to help the learner learn). For formative assessment to work well, learners need to feel 'safe' to expose deficiencies, while in summative assessment the learner tries not to show deficiencies. This limits the use of a summative assessment process for formative use. The nature and timing of the **feedback** is important for this to happen. The feedback needs to take place when the learner is ready for it and must consist of executable advice. Pendleton et al. (1984) has given guidelines on how this can be done, see below.

RULES OF GIVING FEEDBACK (AFTER PENDLETON ET AL)

1. Check the learner wants and is ready for feedback
2. Let the learner give comments / background to the material that is being assessed
3. The learner states what was done well
4. The observer(s) state what was done well
5. The learner states what could be improved
6. The observer(s) state **HOW** it could be improved
7. An action plan for improvement is made

Giving grades alone is not particularly useful in formative assessment without additional feedback as it does not tell the learner how to improve. Team working and learning is becoming increasingly important in health care and **peer assessment** can be very powerful when done well. This is the basis of the annual **appraisal** system for doctors working in the UK's National Health Service. **Appraisal** systems should allow formative, developmental assessment to be given in a systematic way.

A demonstration of metacognition is in self-assessment when a learner measures his or her performance against previous performance. This is called **ipsitive performance**.

Problems Of High Stakes Minimum Competency Assessment

'You get what you assess, you don't get what you don't assess, and you build assessments towards what you want educators to teach'

In minimum competency testing with a few 'key' criteria with competency at a basic level you only get learning at a superficial level in those criteria. The wider complex agenda involving deep learning is ignored

In high stakes testing:

- It is hard to underestimate the effect of the test on the learning of the individual and the teaching
- Teachers teach to the test (rather than the underlying content and constructs)
- Past exams define the curriculum which therefore falls into the control of the examination board
- Society treats the test as a major goal rather than a fallible indicator

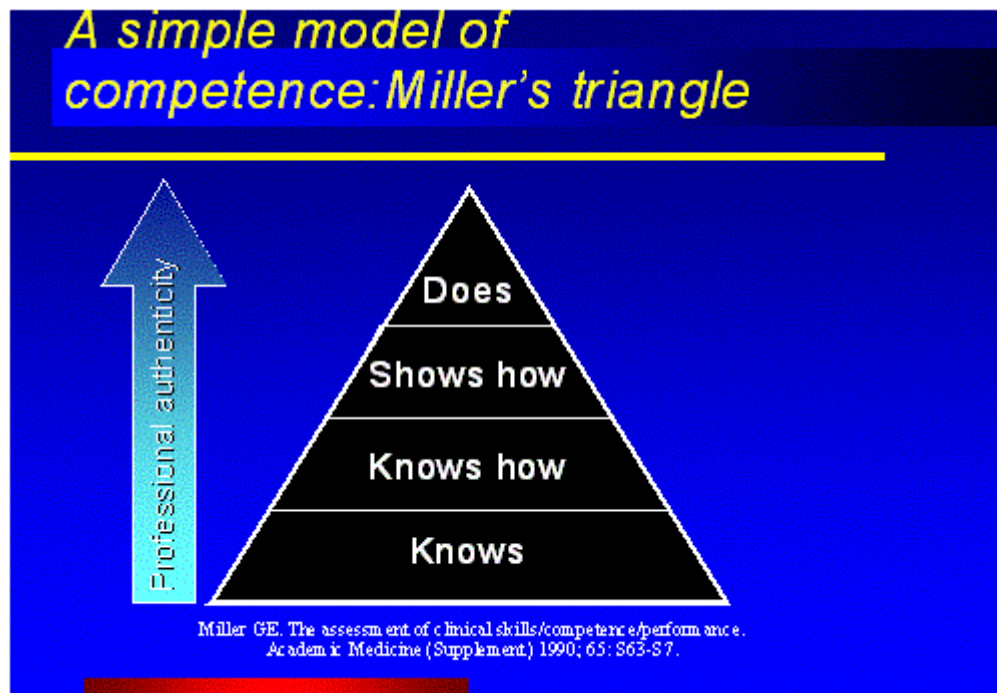
See also *Assessment terminology: a glossary of useful terms* from the Building Tool Room at <http://www.newhorizons.org> (follow links to teaching and learning strategies) and the Higher Education Subject Centre websites, for example **Medicine, Dentistry and Veterinary Medicine** at www.medev.ac.uk contains some downloadable materials on assessment, just follow the links as these are regularly updated and added to. The Subject Centres provide useful teaching and learning resources, hold conferences and other events and provide funding for educational projects.

2. WHAT IS ASSESSED

When formulating an assessment programme the starting point is to define precisely what is being assessed as this will determine which instrument to use.

Competency is the ability to show that the learner has the knowledge and skills in order to move forward to the next stage of training or to be able to do the job that is being tested for. All healthcare learners and juniors have tests of competence included in the formal assessment process.

Assessment of **Performance** determines if the doctor, team or institution is actually applying the competencies in the care of patients.



It is the highest level of assessment of professional authenticity and therefore has the most validity. In practice there is a gradation from basic competency to performance as shown in the diagram (*Resnick and Resnick, 1992; Madaus, 1988*). Performance assessment looks at the whole range of work of an individual or a team. Examples are found within the Quality Team Development (QTD) programme of the Royal College of General Practitioners (RCGP), the Performance procedures of the General Medical Council (GMC), National Clinical Assessment Authority (NCAA) and the Commission for Health Service Improvement (CHI).

3. WHAT THE ASSESSMENT IS MEASURED AGAINST

Assessments were traditionally referenced against all candidates being assessed at the same time to provide a rank order. This is called **normative (or norm-referenced) referenced assessment** and can be useful for example when there are limited places on a programme and thus a need to **discriminate** to find the best learners. However, there are problems using normative assessment (see box) especially when applied to adult professionals.

Disadvantages Of Normative Assessment

- There is low intrinsic motivation to learn the content of the programme
- Increase anxiety and insecurity among candidates
- Undermines effort
- Learning is less efficient for poorer learners (who may never have a chance of passing)
- Competition can lead to poorer social and professional relationship with peers (important where peer working has advantages)
- Covers up misunderstanding (the candidate may get the right answer for the wrong reasons)

The alternative is to measure a candidate's ability against a set of criteria that define the educational outcome that the test assesses. This is called **criterion-referenced assessment**. It has the advantage that:-

- In theory all candidates can pass an assessment
- Candidates can determine what skills and knowledge they still have to acquire before they can pass, or when they fail where they need remedial help.
- In its purest form (such as NVQs) where candidates have to achieve all the criteria to pass, employers and others can be confident that the employee has the knowledge and skills written as criteria.

There is a move towards criterion referenced assessment and there are processes that try and make instruments which are often used for norm-referenced assessment (such as Multiple Choice Questions) more criterion referenced. Some, using experts (such as the Angoff and Ebel technique) have been used, but ultimately unless there are specified criteria, candidates cannot make judgements about why they have reached a particular score.

4. Quality In Assessment

"We may isolate ...characteristics that a good examination should possess and which thus merit research attention. The foremost quality of a good examination is **validity**: a valid examination does what it is designed to do. Second comes **reliability**: a reliable examination consistently and accurately measures whatever it is measuring. Thirdly, a good examination should have a **beneficial rather than harmful backwash effect on the curriculum and teaching**" (*Nuttal and Willmot 1972 p.12*)

A. Validity:

For our purposes, validity can be considered of three types,

- **Face validity** - does an examination or test look valid and does it make sense as a reasonable way to assess the subject?
- **Content validity** - does the assessment cover the appropriate and necessary content of what is to be assessed? This is usually an exercise that 'experts' do. It is important that all content of the curriculum is assessed (or it won't be learnt) and there is always a danger that skills and attitudes that are hard to assess will be missed. This will affect the **utility** of the examination – what use the results of the mean to those who take the examination and what assumptions their teachers/employers can make about the successful candidate. Of course, the content is often so large that it cannot all be tested in an examination and it is good assessment practice to **sample** widely when this happens.
- **Construct validity** - Is the assessment an adequate measure of the construct: that is, the underlying skill being assessed. An example might be a examination of medical knowledge written in Russian. For an English doctor, the primary construct being tested is knowledge of Russian rather than medicine.

B. Reliability

'Reliability is concerned with the accuracy with which the test measures the skill or attainment it is designed to measure. The underlying reliability questions are: would an assessment produce the same or similar score on two occasions or if measured by two assessors?' (*Gipps, 1994, p67*). Reliability can be thought of as 'reproducibility'.

In a normative referenced examination this would mean that the candidates would be put into the same rank order if the examination were repeated. This is helped if the examination produces a high range of scores (and standard deviation) to allow **discrimination**. However, repeating the examination is not feasible and would bring in its own bias. Reliability can be determined using a number of methods but is mostly done using computerised statistical analysis by comparing the performance of all items in the test with all others. This is known as alpha coefficient, a coefficient of internal consistency (an alpha coefficient greater than 0.8 is thought excellent).

This is not applicable where there is criterion referencing. Here the results are compared between different assessors examining the same material. This is the inter-rater coefficient. High reliability in a criterion referenced assessment implies that criteria should be clear and unambiguous and assessors should be trained and subject to a quality assurance programme. As in most things in life, the more effort you put into an assessment, the more you will get out.

Another way that ensures a reliable result is to test a construct in different ways and **triangulate** the result as you might in a research project. Triangulation (as you may expect) uses three different methods or contexts. So communication skills might be tested by using simulated patients, watching real consultations and by an examination (e.g. OSCE, objective structured clinical examination) or alternatively by assessments carried out over time in different contexts and by different people.

It should be noted that in a minimum competence examination where pass marks of 95% of the candidates pass and each item has a high pass mark, both alpha coefficient and inter-rater reliability will be automatically high. In such cases especially where the material is non-standardised (such as assessment of communication skills with real patients) another statistical analysis, Cohen's kappa is used.

There is often a **tension between validity and reliability** in an assessment. Medical and healthcare practice uses a set of complex constructs integrated to help patients in a holistic way and therefore assessments (especially of performance) are complex if they are to be valid. This can put a strain on the reliability as it is difficult to write unambiguous criteria. Although educators place the primacy of validity over reliability, there has been a tendency to emphasise reliability by test developers probably because there is a numerical value given to reliability not given to validity. Although both are important, there is a legitimate view that that for criterion referenced assessment, reliability is "the baggage we bring from normative assessment" (*Jessup, 1991*). Burke and Jessup (1990) further elaborate. Two dissonant judgements would suggest unreliability. They suggest that when this occurs there should be a check of the performance against the external criteria rather than each other with a view to maximising validity.

Lastly it should be stated that the reliability figures relate to a particular examination with a particular set of learners and examiners rather than an examination *per se*.

If all this seems complicated, it is! There is a lot of literature on assessment in medicine, healthcare and in general education. A useful website is the Deliberations site which contains articles and references at:

<http://www.lgu.ac.uk/deliberations/assessment>

The **effect of testing** on learning and teaching has previously been referred to, but there are some other factors that affect quality of an assessment. For example examinations in the health professions are usually assessed by practising professionals. They have clinical commitments and have to squeeze teaching and

assessment in between clinical work. A superb examination with high validity and reliability will inevitably be complex and time consuming and may not be **feasible** or **affordable**.

Self assessment activity

Think of an examination or assessment process with which you have been involved and try to complete the following checklist

1.	PURPOSE What is the purpose of the examination/ assessment process? Is this made clear?	
	If summative is it a. High Stakes	
	b.minimum competence	
	If formative describe what and how feedback is given.	
2.	Is the examination NORMATIVE or CRITERION REFERENCED ?	
	If criterion referenced where are the criteria specified?	
	If norm referenced, has any technique been used to make it more criterion referenced?	
3.	Is the examination measuring COMPETENCY or PERFORMANCE ?	
4.	QUALITY	
	How have the exam developers ensured the examination is valid?	
	Does the examination measure the construct it is meant to measure and does it sample widely and sufficiently?	
	Have any results been published on the reliability of the process? What techniques have been used to ensure this?	
	What is the process of examiner selection and training?	
	What quality assurance processes are in place?	

Assessment Instruments

This section critically examines some of the instruments used in medical and other healthcare assessments and examinations.

TESTS OF KNOWLEDGE

Multiple Choice Questions (MCQs)

These are tests of knowledge that have one great advantage- they can be marked by computer. This means that there is no examiner time needed in marking (although good questions can be very difficult to write) and each item can be easily analysed for its discrimination and internal reliability. The answers are usually written in pencil on a card that can be read by an optical mark reader and computer. In medicine it used to be a tradition that the candidate received one mark for a correct answer and one mark deducted for an incorrect one, this created gender bias (in favour of men) and most examinations do not now have negative marking.

The title is called the stem and each question is called an item with each item being marked as true or false, for example:

SICKLE CELL DISEASE IS ASSOCIATED WITH

1. priapism.
2. impaired fertility in women.
3. an increased incidence of stroke.
4. gall stones in the majority of patients.
5. an enlarged spleen after the first decade of life.

There are considerable problems with using negatively marked MCQs as the only knowledge test because they:

- Can be gender biased (towards males)
- Are inevitably normative
- Cannot easily test 'grey' areas of knowledge
- Test superficially unless very experienced writers are setting the questions
- Do not reflect how adults, learn, retain and recall knowledge

Extended Matching Questions (EMQs)

Also allow computerised marking with its advantages but their format obviates some of the disadvantages of MCQs (see box). They are replacing MCQs in many examinations. Each question has a theme, a set of options, a lead-in statement and a number of items. The items used are usually patient vignettes or scenarios, this reflects a more problem solving approach. As one learner told his teacher 'you actually have to see patients to pass this exam.'

Example of EMQ used in undergraduate medical examination

DIAGNOSIS OF BREATHLESSNESS ON EXERTION (Theme)	
ACUTE MYELOID LEUKAEMIA	BRONCHOPNEUMONIA
CHRONIC LYMPHATIC LEUKAEMIA	CHRONIC OBSTRUCTIVE PULMONARY DISEASE
CONGESTIVE CARDIAC FAILURE	IRON DEFICIENCY ANAEMIA
PULMONARY EMBOLUS	PULMONARY FIBROSIS
PULMONARY TUBERCULOSIS	SARCOIDOSIS

Match the description of the patient with the most likely diagnosis. Each option may be used once, more than once or not at all. (Lead in statement).

Items

1. A 70 year old retired boilermaker. Gives a five-year history of exertional dyspnoea, and a dry cough. The patient is a non-smoker. On examination, fine crackles are heard at the base of both lung fields.
2. A 25 year old refugee from Somalia. She has a productive cough for last three months with haemoptysis and night sweats. Chest X-ray shows a mottled appearance in both by fields.
3. 13 year old boy who has been increasingly tired over last month. On examination he looks pale, has a large bruise over his right thigh and a firm palpable liver and spleen.
4. A 60 year old publican who smokes 20 cigarettes a day. He has a 10 year history of having a 'smokers morning cough' in which he expectorates clear sputum. This is worse in the winter when it sometimes turns green and he has to go to his general practitioner for antibiotics. On examination he has poor air entry over both lung fields and his Peak Expiratory Flow Rate is 210 litres/min.
5. An 80 year old lady with a two month history of exertional dyspnoea, but also wakes up short of breath at night. She has also noticed that during the day over the last month her ankles have been swelling.

Advantages of EMQs

- More problem solving and pattern recognition. Reflects how adults think
- Test application of knowledge in context
- Patient centered
- To prepare: 'hit the wards' and then books
- Wider standard deviation - more discriminatory

Short answer / modified essay questions (MEQs)

These are questions that take between 10-15 minutes to answer, and often have 10 to 15 questions in a paper. In the box there are some examples taken from the Membership of the Royal College of General Practitioners (MRCGP) and from 'finals' at Imperial College School of Medicine.

Examples Of Short Answer / Modified Essay Questions (Meqs)

MRCGP

1. A patient Mr. Burton is a single man of 32 who is a travelling salesman. He attends for follow up after a chest infection a week ago for which you prescribed antibiotics. He says he feels "90% better" but on examination you find a tachycardia of 104/minute. You find no abnormal physical signs.

What likely areas of information should you now explore with Mr. Burton?

2. Alan Fraser, aged 27, is the next patient to see you. He runs an office furniture business with one other partner who departed on holiday the previous day. Mr Fraser breaks down in the surgery in tears, saying that he can no longer cope.

How would you decide what to do?

Imperial Finals

1. A woman of 70 presents with back pain and weakness in both legs developing over three to four weeks.

Discuss the differential diagnosis and your initial management.

2. A mother brings her two years and five months old son to see you, her GP, because she is concerned that he says very little. What information are you going to elicit from her that will help you to evaluate this situation. What specialist clinics are you going to refer the boy to?

Advantages and disadvantages of short answer/MEQs

Advantages

- Samples quite widely therefore improving the reliability of the overall result
- Able to examine at a deeper level than MCQs (improved validity):
- Diagnostic logic
- Critical thinking
- Problem solving
- Potential for giving constructive feedback

Disadvantages

- Needs care to ensure sampling is wide enough to cover a broad curriculum (there is potential for bias using examiners 'hobby horses')
- Takes clinicians time to formulate the marking scheme and to mark
- Needs examiner training and quality assurance programmes to ensure reliability
- The criteria are not transparent to candidates

Full Essay questions

These have the advantage that they can examine in depth the knowledge about one topic and can be structured to examine the candidate's analytical skills, critical thinking, application of knowledge and problem-solving abilities. When used formatively within a course they can be used to measure how far the learner has understood the constructs being taught and help the learners with their learning. In addition, learners/trainees and qualified staff need to communicate in writing clearly and concisely as part of their work and this can also be tested by an essay. However, writing may be the primary construct being tested when an essay is used summatively (a learner who writes well, legibly and coherently may do better than one who doesn't but has greater depth of knowledge and understanding). Essays can also be useful to draw together a number of themes or knowledge from different parts of the course and in particular to enable the learner to demonstrate understanding of 'softer', more conceptual issues such as ethics, law, sociology and psychology. They can enable a learner to explore issues in depth and allow for integration of different course areas.

The main problem with having a few long essays as the only form of summative assessment is that there is only narrow sampling of the syllabus that is being tested. Marking criteria must be clearly drawn up and indications given where marks can be gained or lost. There is more reliance on the experience and judgement of the examiners as learners have more leeway to stray away from expected answers and in some topics there are no right or wrong answers, we might be looking to mark the way that an argument is constructed or awareness of comparative methods or health systems for example. Essays are therefore prone to subjectivity and marker bias and are thus less reliable as assessment methods than other more objective methods.

Computerised assessments

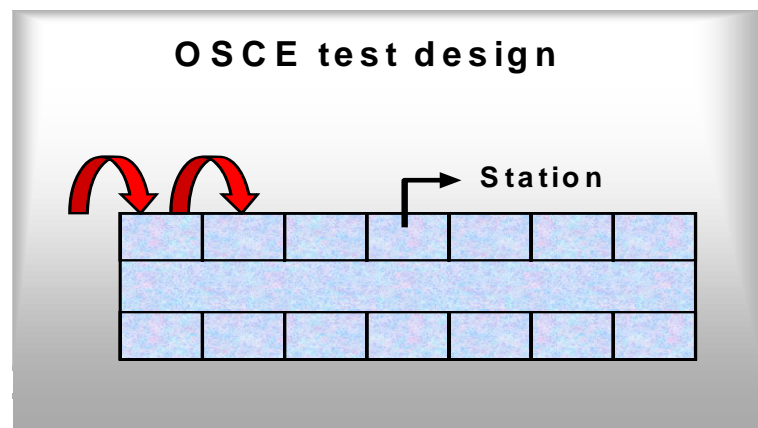
Technology allows learners to be tested in an innovative way. It has the potential to allow scenarios to develop along different routes and allows the candidate to weigh the importance of information given, ask for appropriate and additional information the candidate thinks appropriate, make decisions and see the result of those decisions. It can take account of the learners own self-assessment and can be standardised to allow learners at the same level to be tested. It makes them a valuable learning resource allowing learners to learn at their own pace without much input from teachers.

These sort of assessments take resources to set up and develop which limit their use for summative assessment. In addition, there is the risk of contamination (learners getting to know the questions before the examination).

An example of a good computerised assessment is the Phased Evaluation Programme from the Royal College of General Practitioners.

Tests Of Skills

1. Objective Structured Clinical Examination (OSCE).



In this examination, the learners travel through a series of stations where they are asked to perform a specific task (see box). At each station there is a task with instructions for the learner and some material (e.g. a mannequin, a simulated or real patient a laboratory test result) and an examiner. The learner spends a fixed amount of time at each station (usually five or ten minutes) before moving onto the next station (see diagram - after Van der Leuten C).

Advantages of OSCEs	Disadvantages of OSCEs
<ul style="list-style-type: none"> • Lots of learners can be examined at once • Wide sampling can be achieved • Potential for constructive feedback to learner (rarely done well – usually just a grade) • Psychometrically reliable when well done • Particularly useful for novices and testing basic skills 	<ul style="list-style-type: none"> • Take lots of resources over a short time (e.g. an out-patient clinic might have to be closed for a day to examine a cohort of learners) • Many examiners needed • Simulated patients need to be trained and be consistent between candidates • Needs to be slick to work well • Doesn't look at the whole patient:- difficult to examine complex tasks and expertise.

A variation on an OSCE is a **simulated surgery** where the examined doctor stays in a consulting room and simulated patients attend at fixed intervals. This can test much more complex issues than the simple OSCE, but needs well trained, usually professional actors.

There has been a lot of work in the Netherlands (cited by Norcini, 2002) on testing doctors with simulated patients turning up at their place of work. This allows a good test of actual performance as opposed to competency.

Examples of OSCE stations:

history taking e.g.

- mental state
- drug history

other oral communication e.g.

- giving information
- writing a prescription
- writing a referral letter

examination e.g.

- cardiovascular
- respiratory

psychomotor e.g.

- venepuncture
- cervical smear
- suturing (often on mannequins)

problem solving e.g.

- data interpretation

writing e.g.

- a prescription
- a referral letter

Example of an OSCE station for a GP registrar

CONTRACEPTION

Instruction for patient

You are a 17 year old typist. You have been building up courage to come and see the doctor and have decided to open with "I'd like to go on the pill". You have been having sex with your 18 year old boyfriend for three months. At first he used a condom which neither of you like and lately he has not been using anything and he says he is being careful (you are not quite sure what this means).

You smoke 10 cigarettes a day (your boyfriend also smokes) and have regular periods and no past medical or gynaecological problems. You had an injection against German Measles at school.

You do not know a lot about sex but know it can lead to babies and also to a nasty disease called AIDS especially if you do not use a condom. You do not know very much about your partner's previous sexual history but suspect he has had other partners.

Your knowledge of contraception is virtually nil as you were away when they had that lesson at school but you are afraid to show your ignorance to your boyfriend or girlfriends by asking. You do know that some of your friends have boasted about the pill and it seems simple and safe enough.

You are a little scared that the doctor will tell your mother who does not know you are having sex and you think she would be shocked if she knew.

Marking schedule

HISTORY (10 marks)

Sexual
Contraceptive
Relationship
Medical
Family
Contraceptive
Gynaecological
Rubella
Smears
Smoking

EXPLANATION (10 marks)

Other Methods of Contraception
How pill works and disadvantages of pill
How to take the pill
What to do if you forget
When it will not work e.g. antibiotics
and D & V, first two weeks
How pill works and disadvantages of pill
Starting on 1st day
Leaflet to build up information
FP 1001
Advice about smoking
Smear test
When to seek medical help?

EXPLORING (10marks)

Any questions?
Patient's ideas and concerns Informing other medical attendants

EXAMINATION (10MARKS) OTHER (10 MARKS)

BP	Non-judgmental
Breasts	Clear use of language understandable to a simple 17 year old.
Pelvic	Supportive attitude about seeking
Urine	Contraceptive advice.
Weight/Height	Appropriate use of time.
Rubella	

The long case

The long case is a long established tradition in final undergraduate medical examinations in the UK. The candidate takes a medical history and performs a physical examination on a selected patient taking half to three quarters of an hour. The findings are presented to the examiners, who question the candidate about the case. The patient is often brought to the examination especially but the candidate is not observed. While this is good in a formative assessment, it has many problems as a summative instrument and this has been articulated by many authors with suggestions how to improve this instrument (eg. Norcini, 2002) by having a more structured and criterion based approach.

Advantages of the Long Case:

- Real patients presenting a complete and realistic clinical challenge.
- Relatively easy to arrange.

Disadvantages of the Long Case:

- Shown to have poor reliability due to:-problems of case specificity, differences between examiners and the aspects of competency that are assessed.

Communication skills

Communication skills were traditionally not examined summatively in medicine although other health professions placed great emphasis on these skills. Since the early 1990s, communication skills have been emphasised as important by the General Medical Council and all specialities. They are therefore examined in membership examinations for the Royal Colleges and in undergraduate medical examinations. General Practice and psychiatry led the way and there are a number of good validated models used formatively and summatively such as in the RCGP examination <http://www.rcgp.org.uk/>. There are also some poorer summative instruments. Direct observation or the use of video is used the latter having the advantage that it can potentially increase reliability although there can be technical problems and ethical considerations (e.g. patient confidentiality).

In-Course competency assessment

One of the biggest problems of summative assessment is the logistics. Many schools of nursing and medicine have to examine 300 or more learners at a time. While this causes considerable problems for the written examination, it makes a comprehensive examination of skills almost impossible. One way around this is to examine these skills during the learner's clinical attachment. In course assessment allows the assessment of skills that do not fit into the (short) OSCE format, because they take longer than the time that can be feasibly allowed, or because they must be assessed in the context in which they are carried out. This also has the advantage of being able to assess the learners on a number of occasions and in different contexts and therefore gain a much more rounded and more accurate assessment of the learners' performance.

In-course competency assessment**Challenges:**

- Potential for poor reliability
- Teachers reluctant to fail their learners
- Possibility for bias related to learner behaviour
- Mixing of teaching and assessment functions
- Needs excellent record keeping and sharing of information about learner
- Teachers' time

Advantages:

- Validity: as the 'show how' level
- Patient availability
- Ownership and participation by teachers
- Encourages learners onto the wards
- Allow constructive feedback
- Is not time constrained
- Encourages meta cognition

Prerequisites:

- Involvement and ownership by teachers
- Acceptance by learners
- All assessors need to be trained
- Quality assurance measures in place
- Time available for the assessment
- Quality assurance measures in place
- Workable criteria developed

The example shown below is one used for General Practice in the second clinical year of undergraduate training at Imperial College School of Medicine.

Imperial College School Of Medicine

Department of Primary Health Care and General Practice:- competency assessment for CORE 1

- The learner should achieve all the sub-tasks over a series of consultations although not necessarily in any single consultation. This can be done at one time at a pre-booked surgery or at different times (e.g. the last patient of the day).
- A second doctor may be used to observe and this should help to ensure reliability
- The teacher needs to observe the history taking and examination.
- Learner should be given this form at the beginning of the placement to ensure they understand the criteria they are being judged against.
- This is currently a formative assessment exercise and learner must get constructive feedback to help them improve.
- Please can you give feedback on how it worked for you on the accompanying forms.
- Use one set of forms for each of the patients seen and a another form for a summary(please mark this 'summary')

PATIENT IDENTIFICATION

Core Task	Sub-Task	
A. Takes a patient-centred history	A1. Attempt to elicit the patient's ideas and concerns about the presenting symptoms or illness	Y/N
	A2. Explores the patient's presenting problem	Y/N
	A2. Determines the patients expectations of the consultation	Y/N
	A3. Evaluates the effect of the illness on the patient's functioning and within the context of their environment (e.g. work, home family and carers)	Y/N
	A4. Summarises	Y/N
B. Examination:- competent in	B1. An examination appropriate for the clinical condition	Y/N
C. Defines the clinical problem (s) (to be assessed by questioning afterwards).	C1. Makes a credible diagnosis using the information elicited	Y/N
	C2. Formulates a three dimensional diagnosis	Y/N
	C2i. Physical diagnosis	Y/N
	C2ii. Psychological	Y/N
	C2iii. Social	Y/N
	C3. Determines a management plan using a range of resources available in primary care	Y/N

During the placement the learner should show competency in the following examinations (please tick)

D1. Cardio-vascular examination (including blood pressure)	
D2. Respiratory examination	
D3. Examination of the GI system	

OTHER ASSESSMENT INSTRUMENTS

Teachers' assessment

Formative assessment is an essential part of the learning process and all teachers need to have appropriate skills in order to give accurate, timely and constructive feedback. Teachers are expected to help prepare learners for examinations so need to be aware and gain experience in the examinations their learners take (being an examiner is a good way of doing this).

However, teachers are increasingly called upon to make summative assessment judgements on their learners. There are some inherent difficulties in doing this for example teachers form close relationships with their learners and might feel loath to fail them. More importantly, formative assessment needs the learner to feel safe to expose deficiencies, so having a summative assessment role for a teacher creates a tension. Lastly, teachers may find it difficult to make summative judgements

especially concerning knowledge and skills. However, they are best placed and feel more confident when asked to make judgements concerning professional attitudes and behaviours, at least at the postgraduate level (Rhodes, 1998). These would include, timekeeping, ability to work as a member of a team and integrity.

In order to be able to make summative judgements about knowledge and skills on a learner, teachers need training and clear criteria against which to make those judgements (see in-course competency assessment above)

The **Viva Voce examination** is another time honoured 'rite of passage' for medical students in the UK and other summative examinations. The candidate spends a quarter to half an hour usually with two examiners who explore relevant issues. Vivas are often used as the defining instrument determining whether a candidate passes or fails when borderline.

The hypothesis is that by actually seeing the candidate and asking questions, the examiners can determine whether the learner can 'join the club'. In theory the examiners can respond and probe to the candidate's question and vivas can quickly determine depth of knowledge understanding. In reality vivas are rarely performed well for reasons shown in the box, although when effort is put into selection and training of examiners and determining what the viva is meant to examine, they can be useful (Wakeford, 1999).

POTENTIAL PROBLEMS OF VIVA VOCE EXAMINATION

- Tend to examine knowledge better tested by other instruments
- Material is unstructured with examiners spending time on their 'hobby horses' which may have little validity
- Usually too short to allow adequate sampling of contents
- Examiners can be biased by other aspects than the construct being examined (the diffident pregnant female may have unfair advantage with the male examiners)
- Poor inter-rater coefficients of reliability between pairs of examiners

Case analysis / case based orals

This is another type of oral examination performed on medical graduates centred on the care the doctor has actually given. The medical records form the basis of the discussion which allows the exploration of themes, the reasoning behind the care the patient has given as well as an examination of the record-keeping itself.

The main strength of such an examination is that it centres on the doctor's performance giving it high validity as an instrument. It also needs little in the way of resources. The case-based oral can be used formatively or summatively and is used in the General Medical Council and National Clinical Assessment Authority procedures of performance. It suffers many of the potential problems of other viva voce examinations as explained in section 2.

Jennet (1995) has suggested standardised questions that might be used in a case based oral (or Chart Stimulated Recall as it is called in North America) and these are shown in the box below.

Diagnosis

- What specific features of the presentation led to your decision that the diagnosis was.....?
- Were there any other conditions you ruled out in terms of (Symptom)? Describe how you ruled these conditions out.

Investigations

- What approach did you choose with regard to tests and investigations for the presenting problem?
- Were there any tests or investigations that you thought about and deferred or ruled out?

Management

- I note that you prescribed (Drug) What were your thoughts regarding this choice of medication?
- Were there other medications that you considered and ruled out? What factors influenced your decision.

Patient factors

- We know that patient characteristics have some influence on decision-making. What was special about this patient that had some influence on your choices?

Practice

- Is there anything special about your practice that influenced your management of this patient? (Examples might be availability of investigations or access to other health professionals)

Follow-up

- What factors influenced your choice of follow-up time?

Self assessments, portfolios, logbooks and rating scales

The ability to self-assess should be an aim of an educational programme for professionals. To do this learners need to know:

- What they need to know
- An ability and time to reflect on what they know and how to move their learning forward (for a fuller discussion see Schon, 1993).

An aid to what they need to know might be a log book which states the core competencies or skills which the learners or trainees need to acquire. They or their teacher can tick these off when they have experienced the patient with the problem, demonstrated competence at a clinical skill or procedure or that they understand a topic (See below for example).

Portfolios can also be used as a developmental learning log or tool. Portfolios usually comprise printed or electronic materials which can be in different formats and include e.g. copies of a diary of a professional's learning or practice containing pieces of evidence such as articles, handouts from a lecture or a video of the professional consulting or carrying out procedures. However, the vital part of the portfolio is a piece of reflection (reflective commentaries) on how the evidence is

important in demonstrating learning and development. The development and discussion about a portfolio can be aided by a mentor (who might be senior or a peer) who aids the reflective process. Portfolios form the basis of the revalidation and appraisal system recently introduced in postgraduate medicine.

For a report describing how portfolios are used in undergraduate and postgraduate medicine and other examples see the Medicine, Dentistry and Veterinary Science (MEDEV) Subject Centre website at <http://www.medev.ac.uk>

Another method of self-assessment which can also be used by teachers is a **rating scale**. A rating scale consists of two statements of observable behaviour in a learner separated by a scale. The two statements are effectively opposites, the first representing what is really an ideal behaviour by the professional or learner while the other represents reprehensible behaviour. The evaluator places a mark on the scale to represent where he or she feels the doctor has reached at a given time. A rating scale scores highly in terms of validity as a method of assessment but poorly on terms of fairness and discrimination. Indeed rating scales have been criticised because judgements are highly subjective and contaminated by the attitudes of the evaluator. However, they are very useful for self-assessment.

An example for general practitioners is the Manchester Rating Scale which is used for general practitioners. One of the statements from this scale is shown below.

Example From A Log Book For GP Registrars

GYNAECOLOGY

CONSULTANT: NOTES

TOPIC	Date	Date
	Early in job 0 - 4	Late in job 0 - 4
HISTORY		
Take a full gynaecology history		
Take a sexual history		
EXAMINATION		
Abdominal examination		
Bimanual examination		
Speculum examination		
Cervical smear		
Swabs		
ACUTE ILLNESSES & EMERGENCIES		
Ectopic pregnancy		

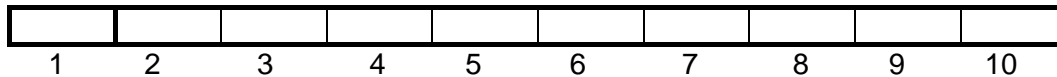
TOPIC	Date	Date
	Early in job 0 - 4	Late in job 0 - 4
Spontaneous abortion		
Ovarian cyst		
PID		
Acute STD		
Herpes		
CHRONIC CONDITIONS		
Dysmenorrhoea		
Amenorrhoea		
Menorrhagia		
Pruritus vulvae		
Incontinence		
Uterine tumours - fibroids - malignancies		
Prolapse		
Subfertility		
Ovarian masses		
Postmenopausal bleeding		
Cervical neoplasia		
Dyspareunia		
Endometriosis		
Infertility		
Vaginal discharge		
SELF LIMITING		
Pre-menstrual syndrome		
Cervical erosion		
Physiological discharge		
PSYCHOSOCIAL		
Contraception		

TOPIC	Date	Date
	Early in job 0 - 4	Late in job 0 - 4
Therapeutic abortion		
Sexual abuse/Rape		
Psycho-sexual problems		
STD		
HIV/AIDS		
DEVELOPMENTAL		
The menarche/normal sexual development		
Abnormal sexual development		
The menopause		
PREVENTIVE		
Screening for genital neoplasia		
Osteoporosis/HRT		
COMMUNICATION		
Counselling for termination		
The HIV positive patient		
Explaining practical procedures		
TEAMWORK		
The Family Planning Clinic		
Contact tracing		
Social Worker		
INDICATIONS & COUNSELLING for		
Hysterectomy		
Laparoscopy		
Hysteroscopy		

ONE OF THE CRITERIA FROM THE MANCHESTER RATING SCALE

SCALE 2. History taking: Special skills

- 1 The doctor's history taking tends to follow a rigid approach. Sometimes psychological and social factors are not effectively pursued and psychiatric symptoms are not thoroughly investigated.
- 10 The doctor's history taking is flexible. Psychological and social factors are followed up when appropriate and psychiatric symptoms are thoroughly investigated.



Projects

The main purpose of doing a project is to allow the learner to engage with the content and constructs of the subject, thus helping learning. However, as has been stated before it is only likely to motivate the learner if it is assessed. There are many examples of short and long term projects in undergraduate medicine, for example following a patient or family over time; clinical audit; health care management; care of a specific condition (chronic or acute) or the work and purpose of a clinical specialty or client/patient group.

Projects should for the most part reflect the way health professionals work, namely in a multidisciplinary and inter-disciplinary way. Using the ideas from discussions with peers will enhance a project and give it more construct validity. However, care must then be taken to separate the contribution of the individual learner for assessment purposes and this may be difficult when the project arises from multiple discussions and a consensus between peers.

Patient questionnaires

One obvious way of finding out about a doctor's performance is to ask the patients. However, there are a lot of problems associated with sampling and confidentiality. Most patients will usually give good references to their doctors. Several validated patient questionnaires have been produced for General Practitioners. One is given as part of the Quality Team Development (QTD) programme of the Royal College of General Practitioners ([hypertext link to RCGP website](#)) and another is the Improving Practice Questionnaire (IPQ).

In Conclusion

There is no such thing as the perfect assessment instrument. Each has advantages and disadvantages and combining them to form a programme that fits the particular learning programme is usual practice. The overarching aim of any individual assessment (e.g. examination, competence based test, observation) or assessment programme is to ensure that the right assessment method is matched to the type of learning and in particular that the assessment actually assesses the specific learning outcomes as defined in the programme. Using different forms of assessment, both summative and formative, at different times and in different contexts is essential to ensure that a learner's or trainee's learning is assessed fairly and accurately and that feedback on learning and performance is made a central part of the learning process. Individuals have different learning styles and so some people will be better at some

forms of assessment than others. Traditions in some professions or specialties tend to favour certain types of assessment methods and as teachers we must aim to consider alternatives from other areas of education as they can often have much to offer.

References and bibliography

Burke J. and Jessup G. (1990) Assessment in NVQs. In Horton ed. *Assessment Debates* Open University, Hodder & Stoughton.

Department of General Practice, University of Manchester (1988). *Rating scales for Vocational Training in General Practice*. Occasional Paper 40. RCGP

Gipps C (1994) *Beyond testing: Towards a Theory of Educational Assessment* London: Falmer.

Jennet P. (1995) Chart stimulated recall: a technique to assess clinical competence and performance. *Education for General Practice*, 6, 30-34.

Jessup G. (1991) *Outcomes: NVQs and the emerging models of Education & Training*. The Falmer Press, London.

Madaus G. (1988) The influence of testing on the curriculum, in Tanner (Ed) *Critical Issues in Curriculum*, 87th Yearbook of NSSE Part 1, Chicago, IL, University of Chicago Press.

McCormack J et al. (1981). *Teaching general Practice*, Kluwer Medical, London.

Miller GE. (1990) The assessment of clinical skills/competence/performance. *Academic Medicine* (Supplement); 65: S63-S7.

Neighbour R. (1989). *The Inner Consultation* London MTP Press

Norcini J (2002) The death of the long case? *British Medical Journal*. 324:408-409.

Nuttall D. and Willmot A. (1972) *British Examinations:- Techniques of analysis* NFER

Pendleton D. et al (1984) *The consultation-An approach to teaching and learning*. Oxford. University Press

Pietroni P. (1988) *The Consultation-The art of caring for the patient* London Heinemann

Resnick LB and Resnick DP (1992) Assessing the thinking curriculum: New tools for educational reform, in Gifford B. and O'Connor M. (Eds) *Changing Assessments: Alternative views of Aptitude, Achievement and Instruction*, London, Kluwer Academic Publishers.

Rethans JJE et al. 1991 A method for introducing standardized (simulated) patients into general practice consultations *British Journal of General Practice*, 41, 94-96

Rhodes M. and Wolf A. (1997) The summative assessment package: a closer look., *Education for General Practice* **8** 1-7.

Rhodes M. (1998) Evaluating the trainer's report in the summative assessment of general practice registrars: a qualitative approach. *Education for General Practice* **9** 40-45.

Schon D. (1983) *The reflective practitioner: How professionals think in action*, , Basic Books New York.

Stott CP Davies RH (1979) The exceptional potential in each primary care consultation. *Journal of the Royal College of General Practitioners*. 29:201-5

Van der Leuten C. (1999) The Lord Cohen Lecture. *Defining and assessing the competencies required by doctors*. <http://www.educ.unimaas.nl/Lord-Cohen-lecture.htm>

Wakeford R *et al* (1995) Improving oral examinations: selecting, training and monitoring examiners for the MRCGP. *British Medical Journal* Vol. 311 pp931-935.