

# ASSESSMENT IN TRAINING: PURPOSE AND RECENT CHANGES

*Dr Nigel de Kare-Silver<sup>1</sup>*  
Associate Director

*Paul Booton BSc FRCP*  
*MRCGP<sup>2</sup>*

1 London Deanery for  
Postgraduate Medical and Dental  
Education  
Stewart House  
32 Russell Square  
London WC1B 5DN

GP Principal  
Gladstone Medical Centre  
5 Dollis Hill Lane  
London NW2 6JH

2 Deputy Head of School of  
Public Health (Teaching)  
School of Public Health  
Imperial College,  
London  
W6 8RP

**Correspondence to:**  
**Dr Nigel de Kare-Silver**  
Associate Director  
London Deanery for Postgraduate  
Medical and Dental Education  
Stewart House  
32 Russell Square  
London WC1B 5DN

## ABSTRACT

The activity of assessing undergraduate students or qualified doctors on training programmes must be fit for purpose. The assessment method must be able to demonstrate that the person, deemed as qualified by that assessment, can indeed perform the tasks expected of the role he is now allowed and consequently expected to undertake as a result of his success. There have been recent changes in assessment processes applied to doctors on general practice specialty training programmes. Their rationale and their detail is described. The ePortfolio as a recording resource and as a tool to assist the process of determining competence is described.

## KEY MESSAGES

New methods of assessment in specialty training have radically altered the pathway junior doctors have to undertake to qualify as General Practitioners. The new method involved competency based assessment of fitness to perform as a practicing GP.

## WHY THIS MATTERS TO ME

NS is involved in specialty medical education and is an Associate Director at the London Deanery with responsibility to lead on assessment. In the summer of 2007 a new MRCGP award was introduced necessitating the undertaking of assessments throughout the whole of their 3 year training programme by doctors on the scheme. It is my belief from personal conversations with non-training colleagues that this new process is poorly understood and that there is even frequent lack of awareness of any change at this level. Some of the techniques used in the new MRCGP process are likely to appear in revalidation: multi source feedback and patient satisfaction questionnaires. Some of the tools used, an electronic portfolio, appear to be on the horizon for revalidation.

It is very likely that increasing numbers of GPs in London will be used in training including GPs who are not currently trainers and GPs who do not currently work in general practice. This is likely as a result of the change in structure of the general practice training programme with an increase in the balance of time spent in practice placements from the current 1 year to at least 18 months, compounded by a likely extension of training from the current three years to five with a consequence of even further time in practice. The capacity of general practice trainers and premises is already stretched and placements may well be in practices with no experience of training, with supervision working on a hub and spoke type model.

PB is heavily involved in undergraduate medical assessment and has academic interests in this area. Properly designed assessments drive learning, offer feedback and accurately identify those who have reached a quality threshold. Poorly designed assessments may offer false reassurance, fail the competent and potentially put the public at risk. It follows that investment in quality assessment represents time and resources well spent.

## ASSESSMENT

Qualifying Assessments must be fit for purpose. The assessment method must be able to demonstrate that the person, deemed as qualified by that assessment, can indeed perform the tasks expected of the role he is now allowed and consequently expected to undertake as a result of his success. Someone, licensed by an assessment system, to work independently, as a fully qualified GP, as a result of passing through an assessment process, must be fit, able and competent to undertake that role. If not, then the assessment system has let down both the doctor and the people whose health problems he serves<sup>1</sup>.

The new MRCGP qualification is now a mandatory exit requirement for those wishing to practice as General practitioners in any capacity. It therefore has a crucial role in setting a quality threshold on doctors entering the discipline and establishing a benchmark for quality across the specialty.

This paper reviews the methods used in the new MRCGP assessment and discusses these in the context of learning an assessment theory.

Trainees in general practice currently spend 3 years in a mixture of hospital and primary care training posts during which they undertake a variety of assessments. On qualification they are entitled to practice as General Practitioner in the UK.

## RECENT CHANGE

Work Place Based Assessment, was introduced by the RCGP for all doctors in training for general practice in September 2007<sup>2</sup>. This change was brought about as a move to improve the education of doctors in training. Competency based curriculums have been introduced and in parallel with this assessment processes that could truly determine whether a trainee, undertaking their

specialty programme, was indeed progressing towards becoming competent to practice<sup>3</sup>.

Work Place Based Assessments which take a variety of forms but are just part of the assessments that the trainee needs to undertake. They are also obliged to pass a knowledge based examination (a multiple choice question paper), called AKT (Applied Knowledge Test) and a role play (OSCE format) examination called CSA (Clinical Skills Assessment).

Assessments and learning are recorded on an electronic, web based portfolio. This portfolio is used to record their work place based assessments, examination passes and other, perhaps lesser intensity, but nevertheless key criteria of their requirements such as certification in Cardio Pulmonary Resuscitation or Out of Hours experience and competence.

This change in assessment methods, introduced by the RCGP, is part of a change in assessment processes across all medical specialty training programmes in the UK. This change has been introduced as part of the shift in emphasis of assessments, directed, led and approved by the Post Graduate Medical Education Training Board (PMETB)<sup>4</sup>.

It is important that assessments for training should meet the key criteria of Validity and Reliability.

## RELIABILITY AND VALIDITY OF ASSESSMENTS

Reliable assessments are assessments that are reproducible and repeatable. Thus a perfectly reliable assessment would always give the same student at the same point in their learning the same mark. This would provide a flat, transparent and equitable process for all candidates. This clearly is both theoretical and aspirational: a student cannot sit the same exam twice! Even if we had the perfect exam, student performance varies for non- academic reasons. Moreover, the more we attempt to control the variables in an exam to make them identical, the more we divorce it from the realities of clinical practice leading to the tension that in general reliable assessments have low validity and valid assessments low reliability. Furthermore in the real world of patient care, patients, people, are predictable in one sense: they are inherently variable.

Reliability is measured statistically, usually using Cronbach's alpha. Using this approach, the most

inherently reliable assessments are computer marked written questions – such as multiple choice, single best answer and extended matching questions of which the latter 2 are currently regarded as the most technically satisfactory. Clinical examinations which are inherently far more variable have been standardised using simulated, patients rigid scenarios and checklist marksheets most commonly using the Objective structured clinical examination (OSCE) or Clinical Skills Assessment (CSA) originally developed by Ronnie Harden in Dundee<sup>5,6</sup>. These have been used widely across many educational fields for several decades and in medical education have become an accepted norm, with newer variants such as the RCP PACES exam conforming to the same principles.

However there are several problems and limitations with such assessments. Factual knowledge may be assessed across a wide range of topics through multiple choice type questions but the integration of knowledge skills and attitudes into real patient encounters is much more difficult, if not impossible, to assess in a traditional examination manner. Exams can structure questions and exercises to gain insight as to whether the candidate can apply their knowledge, but only in the theoretical setting. Attempts to look at the candidates' ability to practice medicine with patients, or indeed on patients is limited. OSCE's attempt this but in a reductionist way, looking at small parcels of history taking or clinical management skills which cannot measure the candidate's ability to manage a whole complex patient encounter and which the process demands is measured and marked identically, thus making it difficult to simulate real clinical situations where there may be a number of equally valid approaches to any encounter. This reductionist approach means that much is lost in terms of seeing whether the doctor can apply those skills measured to the reality of managing patients. This limits the ability of the examiner to be confident either any examination technique on its own, or indeed all together, are indeed fit for purpose.

Reliable assessments are furthermore limited in other ways.

### **“HIGH STAKES” ASSESSMENTS**

Examinations, where the outcome is critical to the progress and future aspirations of the student such as traditional exams, say 'A' Levels or GCSE's termed 'high stakes' assessments. The use of a

gambling term in this context is particularly unfortunate! In this context it is critically important that the exam should accurately differentiate those who genuinely deserve to pass from those who don't, thus reliability has historically taken precedence over validity. Paradoxically therefore the exam tests those things we can measure reliably as proxies of the things we really need to know about the candidate.

Many factors impact on the reliability of an assessment. Amongst these are:

- Candidates can have good luck: each of the few topics they are familiar with arises in the exam paper on the day.
- Different examiners may mark to different standards and individually may have “good” and “bad” days
- Patients inevitably vary, but actors too will vary in any scenario in the challenge they present to the student
- Poorly written questions throw up paradoxical results
- With the factor of recency, an examiner who is pleased with the concluding answers to viva questions can become forgetful and forgiving of earlier errors.
- OSCE's are difficult to organise and can be partially predicted by the candidate with only a limited amount of change from one year to the next. Candidates may work very hard preparing for likely topics and gain a result which does not reflect their actual ability.

All professionals, indeed all school children, are aware of the weak classroom member who surprises and does well in important exams and the strong classroom colleague who consistently fails or underperforms.

Thus in an insufficiently reliable examination there is the real risk that the very weak student or the very weak doctor, can perform very well as an examination candidate yet return to work with their weaknesses undiminished, now in possession of the document of achievement that permits their career to progress. Conversely, the good doctor, respected and admired by their colleagues and peers, can fail to pass. Candidates work hard to get through these exams. Candidates will work hard to learn say the hexose-mono-phosphate shunt, or lists of reactions of ATP and become

skilled at these to a high level, if these are made the hurdles they must cross before progressing to the next stage of their career. This begs the question: if candidates work hard to improve their performance for examinations, why not examine them on those aspects and areas most critical and relevant to the work they are expected to perform? This strategy is part of the rationale for workplace based assessments.

## TRANSFERANCE

Several authors,<sup>7,8,9,10</sup> discuss the need to be able to readily adapt taught skills from the classroom to the practice of patient care. This is called “transference” and it is also known as “context dependant learning” This is a weakness in areas of medical education: that doctors know what to do and have practiced and rehearsed what is required in theoretical and simulation exercises, even shadowed others, but when faced with the reality of a situation where they are required to take the lead, they are unable to transfer the skills from preparation and learning into practice.

This is a particular problem in examinations: A candidate who has worked for the exam may triumph in this setting but be unable to perform in real clinical settings, a candidate who is competent in the surgery may struggle to demonstrate that competence in the exam

## HIGH VALIDITY ASSESSMENTS

Validity embraces the concept that the assessment corresponds to what we are expecting of candidates in the real world. It has the difficulty that it is not measurable by a simple statistic in the way that reliability can make some claims to be. The most “scientific” approach to validity is predictive validity where the exam is judged by the later performance of candidates in the real world which itself is both difficult and problematic. Generally we rely on “face validity”, ie does it look right, and “construct validity”, the extent to which the question measures the objectives set. These are at best qualitative judgements usually carried out by the group setting the examination as a consensus on the quality and relevance of the question.

Thus high validity assessments attempt to assess that the doctor can perform and do the job he is supposed to do, the task he has been and is being trained to perform in real clinical situations or those that carefully simulate them. This term, performance, is vitally important. The term demands that an assessment must be fit for purpose in adjudging that the person assessed, the

trainee GP, can actually do the work they are expected to do and not merely pass examinations. As we have seen, too much emphasis historically has been placed on seeing whether students and trainees can pass exams, and insufficient emphasis on whether they can perform in the role expected of them as a general medical practitioner.

To this end, in creating the (n)MRCGP examination, the college created “Work Place Based Assessments” (WPBA). The trend towards WPBAs is seen across postgraduate medical education in foundation training and specialty training. In these assessments trainees are assessed in their working environment, as opposed to the classroom or examination hall. They are based on the domains of critical areas of practice defined by the college They are thus assessed on the work they are expected to be perform in their role as General practitioners.

There are several different types of WPBA used in the college’s assessments.

- DOPs Directly Observed Procedures
- Mini-CEX Mini Clinical Evaluation
- CBD Case Based Discussion
- COT Clinical Observation Tool
- MSF Multi-source feedback
- PSQ Patient Satisfaction Questionnaire
- CSR Clinical Supervisor Report
- ESR Educational Supervisor Report

DOPS:- There are a total of 19 procedures training GPs must demonstrate competence in. These procedures are those common to the expectations of the normal working practice of the general practitioner and include regularly performed tasks such as both male and female genital examination.

Mini-CEX:- This is an observed assessment of the trainee undertaking a history and examination on a patient. This is used during hospital attachments; in the general practice setting it is replaced by the COT.

CBD:- A presentation by the trainee of a case to his trainer or consultant, looking at the history, examination, diagnosis, management plan, likely course and prognosis.

COT:- It is an observed consultation taking place in general practice according to the carefully defined criteria and standards of consultation expected of a GP performing in practice.

MSF:- A global assessment of performance by colleagues with whom the trainee works, both administrative and clinical, professionalism and attitudes.

PSQ:-A questionnaire format to be completed by patients, an assessment of what patients think of the trainee as a doctor, their doctor.

Trainees are required to complete these assessments throughout the course of their training programmes. They are required to complete a minimum number of each assessment type defined in blocks of six months. In order to complete these assessments they must be observed by the person carrying out the assessment and furthermore, observed in the work place.

CSR:- A report towards the end of each post by the clinical supervisor – the departmental consultant

ESR:- A report by the educational supervisor on progress towards becoming a general practitioner and proceeding towards being competent as an independent GP.

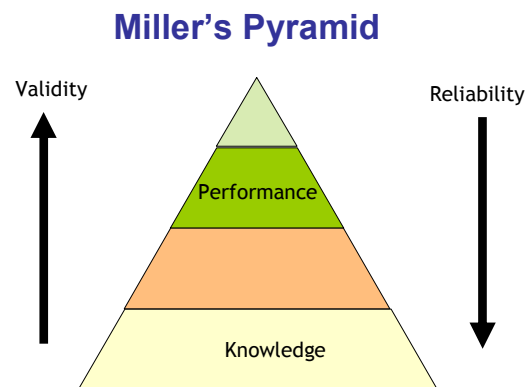
### TRIANGULATION

This term, borrowed from social science research, is a key feature of the ePortfolio and Work Place Based Assessment. It is based on the concept that taking different views of the trainee by different people and from different perspectives is more likely to establish the true level of achievement of the trainee. In order to be effective as a valid assessment tool, and to give reassurance of the quality of the assessment and to improve reliability assessments are thus carried out by many different people over the course of the trainee’s training programme. Many different types of assessment are undertaken and they are undertaken in many different work places over the course of a three year rotation. Thus the system provides a holistic picture of the trainee subject to assessment, not a snapshot of their ability to raise their bar and pass a two hour examination only to let their work standards slip immediately afterwards. Indeed, because of the multiplicity of assessments, as well as being valid, the assessment process is theoretically capable of adequate reliability, trainees being exposed to so many different assessment formats in so many different places and by so many different people, that the personal biases of individual assessors are minimized.

Thus in summary, reliable assessments offer a level

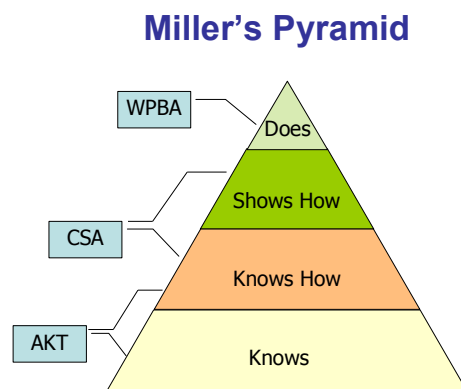
playing field to all candidates but do not necessarily assess whether the candidate can perform the job he is expected to perform. Valid assessments acknowledge the variability in patient care and test performance in the work place, but are not necessarily consistent or reproducible.

Miller<sup>11</sup> sets out assessment in terms of a pyramid and applies this pyramid in two ways. As Reliability improves, Validity lessens and vice versa. Knowledge is at the bottom of the pyramid but also the widest section of it.



Knowledge is vital for doctors who need a broad overview of clinically related science in order to work and function. This can be assessed reliably but knowledge alone, although necessary, is not sufficient.

He then applies his pyramid to the examination and assessment formats.



Again, knowledge is at the base of the pyramid and we can see here it is assessed through knowledge based examinations. The AKT, an MCQ format examination, tries to test the skills in the

application of knowledge by the trainee but sits low in the pyramid. The CSA examination, the role play exam, sits higher. However, the exercises are in obviously artificial settings designed to fit the ten minute window allowed by the examination format. This lessens their validity.

Work place based assessments, testing how a trainee performs in his role of work as a doctor, as a GP sits comfortably at the top of the pyramid.

## **SUPERVISION**

### Clinical Supervisor

Throughout their time in secondary care posts the trainees have a clinical supervisor. This is normally the consultant in charge of their attachment within their clinical department. He is asked to complete a report on their progress towards the end of their posts and determine whether the trainee has progressed within the post as would be expected of someone of their career grade and within that specialty.

### Educational Supervisor

All this in isolation, with the most meticulous recording, would be of limited value. Throughout their training programme, the trainees have an Educational Supervisor, most often a GP trainer, with the responsibility to look at all the individual assessments being completed and ensure the trainee's progress is on track and towards target. The Educational Supervisor completes a review on a six monthly basis commenting on overall progress, curriculum coverage and progress towards each of the competencies. The Educational Supervisor's role is to monitor the training and learning, give direction where there are weaknesses and to challenge the learning that has been documented. One of the Educational Supervisor's roles is to give feedback to the Programme Directors (Course Organisers) both on the progress of the trainee and on any weaknesses in the posts they are undertaking.

The clinical and educational supervisors' roles are importantly formative just as much as they are summative.

## **RECORDING**

All assessments are recorded on the ePortfolio. The assessments are recorded as they are prepared and are saved onto the web based service.

In addition to assessments, learning is recorded as the trainee proceeds through their programme. Learning is recorded on a module called the 'learning log'. Trainees are encouraged to take a reflective approach to their learning, being asked not only to document what they have learnt but also how this learning fits in with their current practice and may change the way they practice. During the course of their programme trainees have to demonstrate, through their learning, that they are working towards covering the general practice curriculum, and must be able to demonstrate complete coverage by the end of their programme.

Exam modules, are recorded as they are taken onto the ePortfolio, with both the pass or fail status and the actual scores. Mandatory requirements such as satisfactory CPR and AED certification likewise appears here as does confirmation of OOH experience.

## **REVIEW**

On an annual basis the central deanery undertakes a review of every single general practice specialty trainee's portfolio. In London there are currently just over 1000 trainees to review. The process is called ARCP - Annual Review of Competency Progression. It is at this stage that trainees are adjudged as being fit to progress, be it to the next year of their training or in due course for Certification of Completion of Training, or otherwise.

It is important to emphasise that all trainees are reviewed and there is no attempt at sampling a few or selecting those deemed most likely to be weak. The ePortfolios are reviewed by teams of deanery educators consisting of an Associate Director, a local scheme teaching Programme Director and a trainer.

Trainees deemed by the review panel as being weak and potentially needing extra support including extensions to training are called for an interview. The interview panel is enhanced by the presence of a lay member to ensure fairness, equity of process and appropriate decision making.

## **CURRENT USE**

The ePortfolio is universally used as a tool to monitor all London GP trainees. Like any tool used in education or in intervention in medicine it has strengths and weaknesses.

### Strong areas

Trainees are responsible for documenting evidence of their teaching and learning and for the documentation of certain specific exercises.

Without this evidence their training they cannot progress and in London the consistent and rigorous standards applied has led to a culture of near universal completion.

As educators and trainees are becoming increasingly familiar with these processes they are being used more rigorously as assessment tools. There is an increasingly critical use of the assessments to identify strong and weak areas of performance by the trainee and the educational supervisors and clinical departmental assessors to direct where the trainee should focus their learning.

### Weak Areas

There remain both educators and trainees who clearly regard the entire matter as a box-ticking exercise and ignore its intrinsic educational value. There are educators who are scared to use strong clear language and give plain messages about weaknesses their trainees exhibit, even when the circumstances demand this. There appears to be a hope amongst some educational supervisors, that those who undertake the review will detect a trainee's weaknesses through subliminal scripts, wording and messages, even where this has been masked in ambivalent or complimentary language. There is thus a hope that the decision an individual is not performing to either expected or required standards will be taken by someone else in the educational system rather than the educational supervisor's own name being against a negative decision.

As in any assessment process there are individuals who by nature push things to the limit in terms of deadlines and superficiality of engagement. ePF is no different in this respect and again lines have to be drawn where behaviour is deemed unacceptable. In many parts subjective and most difficult to rectify.

Calibration of assessors is emerging as a major problem. By its nature, assessment of a trainee's performance is not objective. Though reduced to a series of criteria against which the assessor must

make his judgement, within each criteria the decision on achievement of a satisfactory level of performance remains the personal view of the assessor. One thus depends on the experience of the assessor to decide whether performance is satisfactory. Descriptions and definitions do exist and are available but are not always precise in their guidance.

Clinical supervisors, the departmental consultants responsible for the trainee during a hospital or other secondary care attachment, have a variable attitude to the exercise. Although required to use the portfolio and participate in the assessments, the London Deanery GP department has limited influence of the way they handle their documentation. The deanery can only attempt to facilitate their positive use of the ePortfolio by inviting them to training events and ask local teaching Programme Directors to do their best to invite consultants to attend training sessions.

### **.CONCLUSION**

Work place based assessments are being use as a major component within the (n)MRCG examination. They are being used in conjunction with the AKT and CSA examinations which themselves represent a shift away from traditional classroom based remote exercises towards respectively applied knowledge and consultation role play. The work place based assessments are thus more valid tools to assess the trainee in performance of the role expected of a practicing GP than traditional examinations. The work place based assessments, because they are carried out in the work place, are subject to variability in as much as neither no two patients will be alike nor indeed two assessors. Hence they are less reliable. This loss of reliability is partially compensated by the requirement that multiple assessments need to be carried out in a variety of settings by different assessors over the three year programme. This feature is known as "triangulation". As a result of assessments in the three areas: CSA, AKT and WPBA, the trainee can satisfactorily demonstrate he is competent in each of the competency areas defined by the RCGP.

The trainee is obliged to have all assessments entered onto his ePortfolio. This can then be picked up by any of the trainee's educators and inspected and is mandatorily reviewed regularly by

the deanery in a process known as ARCP (Annual Review of Competency Progression).

A culture of positive use of the ePortfolio, for the mutual and continuing benefit of both trainee and supervisor, needs to be pervade the entire process.

Medicine 2008; 15 (11): 1079-1087

- 11 Miller GE (1990) The assessment of clinical skills, competence/performance. Acad Med; 1990 (9): 565.569

---

## REFERENCES

- 1 Crossley J, Humphries G, Jolly B. Assessing Health Professionals. Medical Education. 2002 Sep; 36 (9): 800 – 804
- 2 Wilkinson JR, Crossley JG, Wragg A, Mills P, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. Medical Education 2008; 42(4): 364-73
- 3 Mayor S (2002). UK royal colleges publish competency based curriculums. BMJ 2002; 325: 1378
- 4 Preparing doctors for the future: about PMETB. Postgraduate Medical Education and Training Board. [http://www.pmetb.org.uk/fileadmin/user/Communications/Publications/Preparing\\_Doctors\\_for\\_the\\_future\\_About\\_PMETB.pdf](http://www.pmetb.org.uk/fileadmin/user/Communications/Publications/Preparing_Doctors_for_the_future_About_PMETB.pdf)
- 5 Schuwirth LWT, van der Vleuten CPM (2003). ABC of learning and teaching in medicine. Written Assessment. BMJ 2003; 326: 643-645
- 6 Smee S. ABC of learning and teaching in medicine. Skill Based Assessment. BMJ 2003; 326: 703-706
- 7 Laksov BK, Lonka K, Josephson A. How do medical teachers address the problem of transfer? Adv Health Sci Educ Theory Pract. 2008; 13(3): 345-60
- 8 Agnew SP, Dumanian GA. A simple test: please draw the incision. It is not as easy as we may think. J Hand Surg. Am 2008;33 (10): 1894-8
- 9 Wittert GA, Nelson AJ. Medical Education: revolution, devolution and evolution in curriculum philosophy and design. Med J Aust 2009 6; 191 (1): 35-7
- 10 Lammers RL et al. Teaching and assessing Procedural skills using simulation; Metrics and Methodology. Academic Emergency