

# Workplace-based assessment

M. H. Davis, G. G. Ponnampereuma, D. Wall

## Introduction

Miller (1990) took forward thinking about assessment in the healthcare professions by identifying four levels of assessment; knows; knows how; shows how; and does. Miller's work focused attention on a largely unassessed level, what the doctor does in real life settings. This has come to be called performance-based assessment. Rethans et al (1991) emphasised the importance of performance-based assessment by showing that scores awarded to general practitioners by simulated patients in an examination setting were significantly higher than the scores awarded for the same tasks in a real-life setting. Hence the need to assess what healthcare professionals do in practice. Traditional examinations cannot assess what the candidate 'does' in real-life settings and a battery of new assessment tools has been produced to assess performance: work-based assessment. Most work-based assessment tools are based on checklists and rating scales.

## Checklists

Checklists contain a list of items or statements about the trainee's behaviour or performance in an identified situation. The person checking the items directly observes the trainee during a procedure, task or performance of a skill. The checklist can accommodate sequential or non-sequential items.

Checklists:

- provide a convenient method of recording trainee performance, i.e. not time consuming
- are objective
- can be standardised as the observers mark all the trainees using the same checklist
- are ideally suited for tasks that can be broken down into several steps by task analysis and where the performance of these steps is an 'all or none' process that needs no grading, e.g. asking the patient's name.

Checklists, however, do not give information about how well a task has been performed, and so are limited to assessing procedures that do not require estimation of the quality of performance.

## Rating scales

Rating scales can be either global measures of overall performance or several rating scales measuring different aspects of performance. If suitably anchored with behavioural descriptors, rating scales, or rubrics can provide objective information which also indicates the standard of performance.

“In contrast with the unstructured descriptions of behaviour gathered in anecdotal records, rating scales provide a systematic procedure for reporting observers' judgements”

Gronlund & Linn 1990

Rating scales are particularly useful for assessing areas that are difficult to test by conventional methods, such as personal attributes, attitudes, generic competencies and professional attributes such as reliability, trustworthiness and time keeping.

## Advantages

Rating scales:

- can be validated to suit a particular group of candidates
- are flexible and can be used in different settings; e.g. the hospital ward, outpatient department, community, etc.
- have high validity, if they are designed to assess important learning outcomes of the course

“The specific learning outcomes specify the characteristics to be observed and the rating scale provides a convenient method of recording our judgements”

Gronlund & Linn 1990



- are standardised. The same rating scales are used to assess all the trainees
- can be used to indicate whether or not the trainee has reached the required standard, i.e. the required standard can be indicated on the rating scale
- provide ratings of trainee behaviour under usual practice circumstances. Thus, they have high authenticity
- can be used to quantify the trainees' overall ability in a given performance by calculating the average of the trainee ratings
- can be used to provide feedback to both the trainer and trainee, particularly if the scale is anchored on observable, clearly defined behavioural descriptors
- can be used to profile the progress of the trainee over time. This will assist early identification of poor performers in order to initiate remedial action
- are unobtrusive
- can be employed at minimal cost
- can assess sustained performance over a period of time rather than the snap shot obtained during an examination: rating scales are actual and on-going.

### Disadvantages

Disadvantages of rating scales include:

- problems in the use of rating scales
- subjectivity
- low reliability
- potential for adverse influence on the relationship between tutor and student.

### Problems in the use of rating scales

Guilford in *Psychometric methods* (1974) lists six common errors made by raters when using rating scales:

- *Error of leniency.* Difficulties may arise if the rater likes the candidate. Some raters may give the candidate a higher rating than is appropriate, while others may overcompensate and give a lower rating.

☛ Raters should be made aware of this potential problem and trained to separate the candidate's performance from their feelings about the candidate

- *Error of central tendency.* Many raters tend to ignore the extremes of the scales and concentrate on the rating points near the centre of the scale.

☛ Instruct raters to make full use of the scale, and anchor points on the scale using descriptors

- *Halo effect.* The appearance of the student may influence the rater's judgement.

☛ Use several raters' opinions to minimise the positive and negative influences of attributes that are not being assessed

- *Logical error.* Where competencies appear to be logically related, e.g. empathy and verbal communication, the raters tend to give similar ratings.

☛ The provision of clear descriptors distinguishing the competencies will help to avoid this tendency

- *Proximity errors.* Raters tend to give similar ratings to traits, which are placed close together especially if they are not disparate traits, e.g. reflective ability and professionalism.

☛ Place similar traits some distance apart on the rating form

- *Contrast error.* If raters think that they are exceptionally good at certain skills, they tend to rate candidates lower than most other raters.

☛ Point out this tendency during rater training sessions

### Subjectivity

If there is only one rater, even when well-designed rating scales are employed, there is a possibility that the assessment may be subjective and open to the influences of hawks and doves.

### Low reliability

The reliability of the rating scale may be low when only one tutor's view is represented and when the contact time between student and tutor has been short.

### Adverse influence on relationship between tutor and student

A potential risk is that the observation of the student necessary for continuous assessment may affect the relationship between the student and the tutor, who is oftentimes the rater. Emphasis placed on a consistently high level of performance during the observation period in order to obtain the highest point on the scale may be unrealistic and students may be discouraged.

### Improving the quality of rating scales

Several measures can be employed to improve the quality of rating scales.

*Provide feedback to raters on the quality of ratings.*  
The quality of rating scales can be variable and



depends on the effort the rater expends in ensuring high standards and in following best practice when completing the rating scale. It is important to maintain pressure on raters to ensure that the rating scales are taken seriously and that adequate attention is paid to their completion. Those who are consistently hawks or doves should be informed of their stance.

*Provide training in the use of the rating scales.*

Training in the use of rating scales improves reliability. Training should be directed at ensuring that raters understand the scale, are given practice in their use, understand the common errors listed above and can defend their views particularly to their superiors. When a consensus view is needed, there may be a tendency for the views of the head of department to dominate. Raters need to be trained to defend their views on students and for heads of department to accept consensus opinions.

*Use several raters.* Increasing the number of raters may accommodate the views of several tutors and other members of the healthcare team working with the student. The use of several raters is likely to improve both objectivity and reliability.

*Recognise the limitations.* Use of a rating scale is not a good way of testing knowledge.

*Anchor the rating scales.* Where rating scales are employed it is important to anchor the scales: that is, to define points on the scale concretely to compensate for doves and hawks and to improve reliability.

*Compensate for a central tendency.* Raters may be unwilling to rate students at either end of the rating scale. This may effectively reduce a five-point scale to one of three points, which may be insufficiently discriminating. This problem can be overcome by:

- anchoring the scale with clearly defined descriptors
- the use of seven or more points in the scale, so that raters still have the choice of a number of points; in practice it is wise to limit the number of points to below 10 as beyond this anchoring is difficult and the range of choice becomes confusing
- varying the number of points on the scale so that raters cannot run down a central point on all scales
- using even numbers of points so that raters must choose one or other side of the midline, assuming the midline is the indicator of acceptable performance.

*Ensure adequate contact time between rater and student.*

In practice, it is not likely to be worthwhile employing rating scales where the clinical attachment is under 4 weeks. Anecdotal evidence suggests that daily contact for a minimum of 6 weeks is necessary.

## Tools for assessing performance in the workplace


### Mini-clinical evaluation exercise (Mini-CEX)

The American Board of Internal Medicine (ABIM) developed the mini-CEX to assess short, specific tasks within a patient encounter (e.g. history taking, examination of the cardiovascular system). The mini-CEX replaced the CEX (Clinical evaluation exercise), which assessed an entire patient encounter (e.g. both history taking and physical examination). The main reason for this change was to enhance sampling of the assessment material, i.e. patients with different conditions and to increase the number of examiners or raters. Originally developed for formative assessment, it has also been used for summative assessment. Although initially developed for postgraduate assessment, it has also been successfully adopted in undergraduate assessment. A patient encounter in the undergraduate version, however, is reported to take much longer (30–45 minutes), than the original, 15–20-minute, postgraduate mini-CEX.

The success of the mini-CEX has led to the development of a similar tool to assess professionalism in the workplace, called the P-MEX (Professionalism mini evaluation exercise) (Cruess et al 2006).

In the mini-CEX, the supervisor directly observes the trainee performance during a patient encounter. The trainee will be rated on interviewing skills, physical examination, humanistic qualities, clinical judgement, counselling, organisation and efficacy, and overall competence, using a 9-point rating scale, where 1–3 is unsatisfactory, 4–6 is satisfactory, and 7–9 is superior. Immediately after the encounter, there is a feedback session lasting 5–10 minutes.

Holmboe et al (2004) found that mini-CEX ratings for videoed patient encounters could differentiate different levels of candidate ability. Between 12 and 14 patient encounters are sufficient to achieve a reliability of 0.8 (Norcini et al 1995, Holmboe et al 2003). If, however, results with a variability of one rating point or less on the rating scale with 95% confidence are acceptable, approximately four encounters are sufficient. In this case, additional encounters may only be needed for borderline candidates (Norcini & Burch 2007).

 Examiner training is crucial for the successful implementation of the mini-CEX

### Multisource feedback (360° assessment)

Multisource feedback (MSF) collects evidence about the candidate from multiple sources, e.g. senior colleagues, junior colleagues, peers, nurses, other healthcare workers, and patients. Some MSF assessments



also include self-assessment. Since various tools are used to collect evidence from the above sources, MSF is a collection of assessment tools such as rating scales and checklists.

This method can be used for both summative and formative assessment. Though more widely used in postgraduate (Whitehouse et al 2002) and continuing medical education settings (Sargeant et al 2005), it has been used also in undergraduate medical education (Norcini & Burch 2007).

An advantage of MSF is that it does not require the pre-identification of material (clinical cases) and time for assessment (Norcini & Burch, 2007). In other performance assessments such as DOPS and Mini-CEX, the pre-identification of a specific patient encounter and time for assessment makes them less authentic than MSF.

The assessment of foundation year doctors in the UK is an example of MSF in practice. The trainee doctors nominate eight assessors, including senior consultants, junior specialists, nurses and other health professionals. These assessors use a structured questionnaire to assess the doctor's ability in five General Medical Council outcomes; i.e. good clinical care, maintaining good medical practice, teaching and training, appraising and assessing, relationship with patients, and working with colleagues. The trainees also assess themselves on the same five outcomes. The completed assessment forms are collated by a central processing unit and individual feedback is prepared for each trainee. The results are provided graphically with each trainee rating compared with the corresponding national average. All qualitative feedback is also provided verbatim, but anonymised. The trainees go through the results with their supervisor and arrive at a future action plan. This process is repeated twice a year (Norcini & Burch 2007).

Examiners for each competence should be selected judiciously. Technical competencies such as clinical ability are best assessed by senior colleagues, whereas nontechnical competencies such as communication skills can be assessed by all the nominated raters

There is convincing evidence supporting the validity of MSF. Reliability is dependent on the assessment instrument (Swick et al 2006) and the assessor category. There are studies that required as many as 23 medical students or as few as five nurses to produce a reliability of over 0.8 (Massagli & Carline 2007).

#### Directly observed procedural skills

Directly observed procedural skills (DOPS) is designed for the assessment of clinical procedures or investigations such as lumbar puncture. DOPS is used in the UK in the foundation years and for registrar

training and assessment; for example in the medical specialties. The assessment utilises a nine-point rating scale somewhat similar to the mini-CEX. In some situations, for example foundation year 1, the nine-point scale is collapsed to a six-point scale that identifies, with regard to foundation year completion, whether the trainee is: below expectations; borderline; meets expectations; average or above expectations.

DOPS is designed to assess:

- understanding of indications for the procedure/investigation
- appropriate use of sedation or analgesia
- technical ability
- consideration for the patient
- management of complications
- interpretation of results.

Additional specific assessment forms are available for certain procedures such as cardiac catheterisation. In studies, the Joint Royal Colleges of Physicians Training Board (JRCPTB) found that they obtained reliable results with three observers each observing two procedures per trainee (<http://www.jrcptb.org.uk/Specialty/Documents/DOPS%20Guidelines%20Renal.pdf>). For the foundation years, four DOPS per year are recommended, but the frequency of assessments may vary by deanery. The time taken for the assessment depends on the length of the procedure and feedback to the trainee requires approximately a further 5 minutes.

#### Mini-PBAs

The surgical specialties in the UK use a type of DPOS, which they have renamed mini procedure based assessments (Mini-PBAs), for the assessment of minor surgical procedures such as the debridement of a wound or the closed management of a fracture.

#### Procedure-based assessment (PBA)

For the management of more major surgical procedures, however, the procedure-based assessment (PBA) is being piloted. Within orthopaedic surgery, examples of procedures for which specific PBAs are available include hemiarthroscopy for intracapsular fracture of neck of femur; operative fixation of Weber B fracture of ankle; and intramedullary nailing for femoral or tibial shaft fractures. The agreement of both trainer and trainee is necessary to trigger a PBA and it is the trainer's responsibility to provide the level of supervision appropriate for the individual trainee. The PBA rating form includes six aspects:

1. Consent
2. Pre-operative planning
3. Pre-operative preparation
4. Exposure and closure



5. Intra-operative technique
6. Post operative management.

The trainer rates the trainee on each of the six aspects as either:

- N – not observed or not appropriate for the trainee level of training and ability
- U – unsatisfactory or needs improvement
- S – satisfactory.

The PBA form contains areas for written comments and trainers are required to identify what needs to be done to improve when a U rating is given. There is an additional global rating for each procedure that the trainer is required to give:

- Level 0 – Insufficient evidence observed to support a judgement
- Level 1 – Unable to perform the entire procedure under supervision;
- Level 2 – Able to perform the procedure under supervision;
- Level 3 – Does not usually require supervision but may need help occasionally;
- Level 4 – Competent to perform the procedure unsupervised and can deal with complications.

The validity and reliability of PBAs are being studied at the time of writing.

### Case-based discussion

In the USA chart stimulated recall (CSR) is used to assess clinical decision making and the application and use of medical knowledge in real cases managed by the trainee. The format is usually a 2-hour, standardised oral exam with each case taking 5–10 minutes to assess. Reported reliabilities vary from 0.65 to 0.88 in the USA (<http://www.acgme.org/Outcome/assess/Toolbox.pdf>). The method has been adopted in England, Wales and Northern Ireland for foundation years' assessment, where it is called case-based discussion (CbD). Here, the trainee selects two sets of patient case notes, in which the trainee has made notes, and provides the case notes to a trainer. The trainer then questions the trainee on one of the cases and rates the trainee on:

- Medical record keeping
- Clinical assessment
- Investigation and referrals
- Treatment
- Follow-up and future planning
- Professionalism
- Overall clinical judgement.

A six-point rating scale is used that identifies, with regard to completion of training, the trainee as below expectations; borderline; meeting expectations; or

above expectations. There is also an opportunity to check an 'unable to comment' box where a specific competence has not been utilised in the case. The discussion is expected to last no more than 20 minutes including 5 minutes for feedback.

### Significant event analysis

The significant event analysis (SEA) is a story of things that have gone well and things that have gone badly. This story can be used as a means of improving patient care and safety and potentially minimising healthcare risk. The basic framework is as follows:

1. What happened?
2. Why did it happen?
3. What have you learned?
4. What have you changed?

The SEA requires that when things go wrong, we do not adopt a blame culture, but a learning culture and look carefully at what has happened in a formalised and structured way

Bowie et al (2003) described a simple four step model of SEA, used in general practice in the West of Scotland (Table 45.1).

### Audit

Audit in medical practice may be defined as the improvement in the quality of care through a series of steps including standard setting, peer review, implementation of change and re-evaluation. Audit is about

**Table 45.1**

Four-step model for significant event analysis

1. What happened?	Has personal impact Important to individual or organisation Causes reflection
2. Why did it happen?	Clear reasons sought
3. Was insight demonstrated?	Aware of previous suboptimal care Decision making process altered Awareness of risk demonstrated Level of personal responsibility linked to circumstances
4. Was change implemented?	Yes – describes implementation of relevant change No – risk of similar significant event unlikely No – unable to influence change but suggestions for change given

Bowie et al 2003



comparing one's work against a standard, trying to improve one's performance and then comparing again to see if improvements have in fact been made. Audit is now a professional responsibility of a doctor in the UK. Structured formats have been developed to guide the preparation and presentation of audits, and for the marking of audit projects. Completion of an audit is now part of the English foundation programme assessment (for doctors in the first and second year after qualification in the UK).

#### The five criteria audit schedule

Each of five criteria needs to be present to pass the assessment. The five criteria are:

1. Reasons for choice of audit – Should be clearly defined and reflected in the title. Should include potential for change.
2. Criteria chosen – Should be relevant to the subject of the audit. Should be justifiable, e.g. literature.
3. Preparation and planning – Should show appropriate teamwork and methodology in carrying out the audit. If standards are set they should be appropriate and justified.
4. Interpretation of data – Should use relevant data to allow appropriate conclusions to be drawn.
5. Detailed proposals for change – Should show explicit details of proposed changes.

This schedule is useful where the doctor is only in a post for a short time (such as the 4-month duration posts in the foundation programme) and where it is not always possible to re-examine practice after changes have been put in place. However, the audit is not complete and does not allow demonstration of a complete audit cycle. Thus, a further schedule was developed to allow demonstration of a full audit cycle, the eight criteria audit schedule.

#### The eight criteria audit

All eight criteria need to be present in order for a pass to be awarded. The criteria are similar to the five criteria audit above, but include the second data collection and conclusions about the lessons learned.

The eight criteria audit schedule is:

1. Reasons for choice of audit – potential for change and relevant to practice
2. Criteria chosen – relevant to audit subject and justifiable, e.g. current literature
3. Standards set – targets towards a standard with a suitable timescale
4. Preparation and planning – evidence of teamwork and adequate discussion where appropriate
5. Data collection 1 – results compare against the standard

6. Changes to be evaluated – actual example described
7. Data collection 2 – comparison with data collection 1 and the standard
8. Conclusions – summary of the main lessons learned.

Both of these audit assessment tools were developed in the west of Scotland (Lough & Murray 2001). Both are simple and well used tools, which enable the marking of an audit project. With regard to reliability, investigators in the West Midlands found that six, eight criteria audits were necessary to give a generalisability coefficient of 0.8 and 14 were necessary for a generalisability coefficient of 0.9.

#### Patient satisfaction surveys

The patient satisfaction survey is increasingly viewed as a key component in evaluating healthcare quality. In the UK since 2002 every hospital trust has been required to survey a sample of patients and report the results to the Healthcare Commission. In general practice, as part of the 2004 UK general practice contract, an annual patient satisfaction survey is a part of the Quality and Outcomes Framework.

“Patients may not be able to assess all aspects of a doctor's performance, but they do prefer doctors with excellent communication skills and sound up to date technical skills”

Coulter 2006

In general practice, the *General Practice Assessment Questionnaire (GPAQ)* is widely used and may be downloaded (<http://www.gpaq.info/>). It measures various aspects of care in the practice, including access to care, and not just the doctor–patient relationship. At least 50 completed questionnaires per doctor are required to obtain reliable results.

Another questionnaire is the Patient Satisfaction Questionnaire (PSQ) which may be downloaded from the Royal College of General Practitioners' website (<http://www.rcgp.org.uk>). It is now a requirement for the new MRCGP examination that candidates carry out a patient satisfaction questionnaire with at least 40 completed questionnaires being needed. Patients are asked to rate the doctor in ten areas using a seven-point scale.

The 10 areas are:

- Making you feel at ease
- Letting you tell your story
- Really listening
- Being interested in you as a whole person
- Fully understanding your concerns
- Showing care and compassion
- Being positive



- Explaining things clearly
- Helping you take control
- Making a plan of action with you.

The seven-point scale is: poor to fair; fair; fair to good; good; very good; excellent; and outstanding. There is also a global assessment.

### Video consultation assessment

The consultation between doctor and patient is a fundamental and crucially important part of the practice of medicine. Analysis of the doctor-patient consultation depends on the development and use of a model of the consultation and what is going on within it. There have been several models. These include those of Balint (1957), Pendleton and colleagues (1984), Neighbour (1987) and more recently the Cambridge Calgary model of the consultation (Silverman et al 1998), which built on Pendleton's work. At present many medical and dental schools use the Cambridge Calgary model to teach and assess their students' consultation skills.

Video consultation assessment is a very incisive and revealing method of study of the consultation. There are several safeguards which should be mandatory when using this method. Signed informed consent from the patient is essential before the patient comes into the consulting room, and the form should be again signed by the patient afterwards. Specific forms are available for this purpose, which follow General Medical Council guidelines. The explanation on the consent form should explain that the video recording will only be of the doctor and the patient talking together and no examinations of the patient will be filmed. If an examination is to take place the camera will be switched off. Also the camera will be switched off immediately if the patient requests it at any time during the consultation.

There are several assessment frameworks with which to look at the elements of the doctor-patient consultation. The Consultation Observation Tool (COT) has been developed for the new MRCP examination and includes five main elements and a global overall assessment.

The five elements are:

1. Discover the reasons for the patient's attendance
2. Define the clinical problem
3. Explain the problem to the patient
4. Address the patient's problem
5. Make effective use of the consultation.

Each element is rated on a four item scale: insufficient evidence; needs further development; competent; excellent.

For examination six video consultation assessments are necessary in the first year, six in the second year, and 12 in the third year. However, in most cases the method is used regularly much more frequently than this, in order to teach and improve communication skills.

Knowing how to give feedback constructively is crucial to make the method work in practice. Pendleton's rules are used to give constructive feedback and were specifically developed for the purpose

### Pendleton's rules

Learner goes first and carries out the activity (the consultation)

Questions on facts of points *for clarification only* are allowed at this stage

Learner says *only* what (s)he thought was done well  
Teacher then says *only* what (s)he thought was done well

Learner then says what (s)he thought could have been done better – on reflection

Teacher then says what (s)he thought could have been done better and in discussion gives constructive advice on how to improve.

### Summary

Checklists and rating scales have been in use for many years, but recent developments in the form of workplace-based assessment are exciting innovations that provide potential to combine learning with assessment. Experience with the use of workplace-based assessments is building, and while some of the assessments described in this chapter have stood the test of time; others may not. What is already clear, however, is that training the trainers in the use of workplace-based assessments is essential for their successful implementation. The instruments described in this chapter assess different aspects of performance. Which instruments to employ and how they are combined into an assessment system is dependent on the outcomes to be assessed. A profile of the trainee's strengths and weaknesses in terms of the learning outcomes can be constructed using a combination of workplace-based assessments. The potential to provide feedback to the trainee is a feature of all workplace-based assessments facilitating the progression of the trainee over time to reach the required standard; employing assessment to support learning.

### References

- Balint M 1957 The Doctor, his patient and the illness. Pitman Medical, London
- Bowie P, McKay J, Lough M 2003 Peer assessment of significant event analyses: being a trainer confers an advantage. *Education for Primary Care* 14:338-344
- Coulter A 2006 Can patients assess the quality of health care? *British Medical Journal* 333:1-2



- Cruess R, McIlroy J H, Cruess S et al 2006 The professionalism Mini-Evaluation Exercise: a preliminary investigation. *Academic Medicine* 81(10):S74-S78
- Gronlund N E, Linn R L 1990 Measurement and evaluation in teaching, 6th edn. Macmillan, New York
- Guilford J P 1974 Psychometric methods, 2nd edn. McGraw-Hill, New York
- Holmboe E S, Hawkins R E, Huot S J 2004 Effects of direct observation of medical residents' clinical competence training: a randomised control trial. *Annals of Internal Medicine* 140(11):874-881
- Holmboe E S, Huot S, Chung J, Norcini J J, Hawkins R E 2003 Construct validity of the minClinical Evaluation Exercise (MiniCEX). *Academic Medicine* 78:826-830
- Lough J R M, Murray T S 2001 Audit and summative assessment: a completed audit cycle. *Medical Education* 35:375-363
- Massagli T L, Carline J D 2007 Reliability of a 360-degree evaluation to assess resident competence. *American Journal of Physical Medicine and Rehabilitation* 86(10):845-852
- Miller G E 1990 The assessment of clinical skills/competence/performance. *Academic Medicine* 65(9):s63-s67
- Neighbour R 1987 *The Inner Consultation*. Kluwer Academic, Lancaster
- Norcini J J, Blank L L, Arnold G K, Kimball H R 1995 The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Annals of Internal Medicine* 123:795-799
- Norcini J, Burch V 2007 Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher* 29:855-871
- Pendleton D, Schofield T, Tate P, Havelock P 1984 *The consultation: an approach to teaching and learning*. Oxford University Press, Oxford
- Rethans J J, Sturmans F, Drop R et al 1991 Does competence of general practitioners predict their performance? Comparison between examination setting and actual practice. *BMJ* 303:1377-1380
- Sargeant J, Mann K, Ferrier S 2005 Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Medical Education* 39(5):497-504
- Silverman J D, Kurtz S M, Draper J 1998 *Skills for communication with patients*. Radcliffe Medical Press, Abingdon
- Swick S, Hall S, Beresin E 2006 Assessing ACGME competencies in psychiatry training programmes. *Academic Psychiatry* 30:330-350
- Whitehouse A, Waltzman M, Wall D 2002 Pilot study of 360 degree assessment of personal skills to inform record of in-training assessments for senior house officers. *Hospital Medicine* 63:172-175